Trends in the History of Science

Laurent Mazliak
Glenn Shafer
Editors

The Splendors and Miseries of Martingales

Their History from the Casino to Mathematics

Birkhäuser

Birkhäuser

Trends in the History of Science

Trends in the History of Science is a series devoted to the publication of volumes arising from workshops and conferences in all areas of current research in the history of science, primarily with a focus on the history of mathematics, physics, and their applications. Its aim is to make current developments available to the community as rapidly as possible without compromising quality, and to archive those developments for reference purposes. Proposals for volumes can be submitted using the online book project submission form at our website www.birkhauser-science.com.

Laurent Mazliak · Glenn Shafer
Editors

The Splendors and Miseries of Martingales

Their History from the Casino to Mathematics

Birkhäuser

Editors
Laurent Mazliak
Sorbonne Université
LPSM
Paris, France

Glenn Shafer
Rutgers Business School
Newark, NJ, USA

ISSN 2297-2951 ISSN 2297-296X (electronic)
Trends in the History of Science
ISBN 978-3-031-05987-2 ISBN 978-3-031-05988-9 (eBook)
https://doi.org/10.1007/978-3-031-05988-9

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This book is published under the imprint Birkhäuser, www.birkhauser-science.com by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Introduction

$$\int_{\{\nu=k\}} X_k \, d\mathbf{P} \leq \int_{\{\nu=k\}} X_n \, d\mathbf{P},$$

$$a\mathbf{P}(\nu \leq n) \leq \int_{\{\nu \leq n\}} X_\nu \, d\mathbf{P} = \sum_{k=0}^{n} \int_{\{\nu=k\}} X_k \, d\mathbf{P} \leq$$
$$\leq \sum_{k=0}^{n} \int_{\{\nu=k\}} X_n \, d\mathbf{P} = \int_{\{\nu \leq n\}} X_n \, d\mathbf{P} \leq \mathbf{E}(X_n).$$

The mathematical concept of a martingale appears today as one of the essential tools of modern probability theory. Formalized only at the end of the 1930s, even though we can now see it in the pioneering seventeenth and eighteenth centuries work of Blaise Pascal and Abraham De Moivre, the concept gives the discipline an efficient way to obtain a myriad of fundamental results through the relatively elementary verification of a property directly based on the notion of conditional expectation. As the name from the world of gambling indicates, martingales came into mathematics in the 1930s as strategies for betting. The central mathematical preoccupation of the mathematician who first promoted the name, Jean Ville, was with the asymptotic properties of the evolution of the player's capital. But the extension of the concept to continuous time brought into view the martingale property of certain random processes, especially the two most important ones, Brownian motion and the Poisson process. With the help of this property, a new type of integral was defined, the stochastic integral, and beyond that a brilliantly original differential calculus, whose results have not ceased to grow in importance since the middle of the twentieth century. When we also bring into view the connections between discrete- or continuous-time Markov processes and various properties of martingales, we see that by the end of the last century martingales had invaded all aspects of probability theory and its applications.

The literature on the history of probability has not kept pace with this growing importance of martingales. The special issue of the *Electronic Journal for History of Probability and Statistics* that we devoted to the history of martingales in 2009 was, to our knowledge, the first attempt to gather texts and documents that traced with some precision the genesis and the trajectory of the concept through the history of the mathematics of randomness in the twentieth century. The present work can be seen as a considerably enriched second edition of this issue of the electronic journal. This seemed to us a necessary initiative for two reasons. The first is the cessation of the publication of the e-journal, which unfortunately did not survive the death of its co-founder Marc Barbut in 2011: the fragility inherent to the perennial availability of a discontinued online journal made us think that it would be judicious to guarantee this availability through a book published both in paper and online. The second reason is more profound: in the more than 10 years that have passed since 2009, new research has extended, corrected, and completed many of the texts that we presented in 2009, and newly discovered documents and newly emerging insights have added important elements to the puzzle that a historiographic construction always constitutes. As a result, most of the texts presented in this book are either entirely new or significantly enhanced.

The book has four principal parts, ordered more or less chronologically. A fifth part presents annotated transcriptions of archival documents that enrich the historical account.

The first part of the book, entitled *In the beginning*, considers some aspects of "martingales before martingales". In a text full of verve and literary erudition, Roger Mansuy traces the genealogy of the name. He tells us that a martingale can be a part of a horse's harness, a part of a sailboat's rigging, a man's coat, or even a courtesan. Lexicographers have advanced various hypotheses for the word's origin. Mansuy begins with its use in mathematics and works back to the picturesque city of Martigues, on the French Mediterranean coast. Along the way one encounters a king's breeches, fencers, a sailor's dance, a prophetess, and an imprisoned Javert in Victor Hugo's *Les Misérables.*

The second chapter, by Glenn Shafer, focuses on martingales in games of chance. Shafer asks what made a betting strategy a martingale and what made martingales so deceptively alluring to casino gamblers. The decade after the French revolution marks the beginning of a written record that casts some light on these questions. By the beginning of the nineteenth century, as Shafer shows, any betting strategy could be called a martingale. The casino history has more than an antiquarian interest, for we still live with the seduction of martingales, inside the casino and outside—whether at the horse track, in high finance, or at home as day traders and internet gamblers.

The third chapter is a profound study by Bernard and Marie-France Bru, which tries to locate the magic of martingales in a paradox that appeared very early in the history of probability: a fair game can become unfair at infinity. This paradox was fully understood and mastered only in the 1940s by Émile Borel, who resolved it using the theory of denumerable probability he had introduced in 1909. Borel's reflections on the paradoxes of infinite play were stimulated by a debate beginning around 1910 with the biologist and uncompromising determinist Félix Le Dantec. Despite Le Dantec's fierce contention that probabilities are of no use in science, Borel learned something profound and far-reaching about probability's applications from him: they generally depend on equating a small or zero probability with impossibility. He struggled for decades to reconcile this insight with his denumerable probabilities.

The fourth and final chapter of the first part, by Salah Eid, focuses on the exchanges between the Danish analyst Børge Jessen and the French probabilist Paul Lévy, who arrived from their different starting points at convergent insights that became central to martingale theory. Their main results have come to be known as Jessen's theorem and Lévy's lemma. Jessen saw his theorem as an extension of the Fubini-Lebesgue theorem of 1907–1920. Lévy saw his lemma as an extension of Borel's strong law of large numbers of 1909. In letters between the two authors, each wanted to see the other's result as a trivial consequence of their own. Jessen sought a level of abstraction that proved unattainable, but his interaction with Lévy can be seen as the origin of a now-standard version of the martingale convergence theorem.

The second part is entitled *Ville, Lévy, and Doob*, as it focuses on the three principal protagonists in the emergence of the mathematical concept of a martingale: Jean Ville, Paul Lévy, and Joseph Doob. Though he may be considered chronologically the second in line, we have placed Jean Ville as the hero of the first chapter of this part, because he was certainly the one who baptized the concept with the name "martingale". In this chapter, Glenn Shafer enlarges the question, asking what led Ville to think about probability in terms of betting games and explaining how betting games allowed him to link Borel's denumerable probabilities with Richard von Mises's concept of probability as frequency in a collective. Shafer sees Ville as excavating the martingales already hidden in probability theory, foreshadowing their role not only in Lévy's and Doob's measure theory but also in two complementary theories—algorithmic complexity theory and game-theoretic probability.

As already noted, Ville's martingales were preceded by Lévy's. But Lévy was focused on extending the law of large numbers and other theorems about sequences of independent random variables to dependent random variables. In the second chapter of the second part, Laurent Mazliak explains how Lévy showed that this extension was possible when each random variable has mean of zero given the preceding ones. Under this condition, the sequence of cumulative sums is a *martingale* as Jean Ville would define the word, but Lévy never focused on this sequence of cumulative sums as a mathematical object. In this respect, his was not a theory of martingales. Moreover, he never showed much interest in the properties of martingales studied by Ville and Doob. In fact, as Mazliak documents, Lévy had a troubled relationship with Ville and generally disdained his mathematical work.

The third mathematician of the founding triad, Joseph Leo Doob, put stochastic processes into Kolmogorov's framework, systematically and with brilliant success. In the 1940s, he developed a theory of martingales that eclipsed what Lévy and Ville had contributed to the concept. In the third chapter of the second part, by Bernard Locker, provides penetrating insights into Doob's fundamental contribution from the vantage point of his lecture on applications of the theory of martingales at the international colloquium on probability organized in Lyon in 1948 by Maurice Fréchet. It was here, at Lyon, that Doob first used the word "martingale" systematically in his own work, that this work was first presented in France, and that the mathematical world first saw how easily the concept of a martingale yields the strong law of large numbers and the almost sure consistency of Bayesian estimation.

The third part entitled *Modern probability*, recounts how martingales came to play so central a role in this modern mathematical theory. Its three chapters are provided by witnesses who had an important role in the evolution of the field after the Second World War. The first chapter is by Paul-André Meyer, who was the architect of the general theory of processes in the 1960s and 1970s. This theory forms the basis of all subsequent studies using continuous time processes more general than Brownian motion or Poisson processes. The chapter is a translation of a text written by Meyer on the eve of the year 2000, which traces his perception of the evolution of the theory over half a century, an evolution which, of course, involves much more than the concept of martingales. Meyer emphasizes the founding role of Doob's *Stochastic Processes*, published in 1953, which presented tools and topics that fueled probabilistic research for the rest of the century: filtrations, stopping times, martingales, Markov processes, diffusions, Itô's stochastic integral, and stochastic differential equations. The period from 1950 to 1965 was dominated by the study of Markov processes and their connections with potential theory and martingales. In the period from 1965 to 1980, martingales became more prominent, along with the stochastic integral, excursions, the general theory of processes, and stochastic mechanics. The review extends into the 1980s, discussing the Malliavin calculus and noncommutative probability theory.

The second chapter of the third part, provided by another important actor of the period, Shinzo Watanabe, focuses on the contributions of the vigorous and productive Japanese school of probability. Though Japanese scholars did not contribute directly to martingale theory before 1960, many of their contributions after 1960 were based on the stochastic calculus that Kiyosi Itô first introduced in 1942. Itô's collaboration with Henry McKean on the pathwise construction of diffusions attracted wide interest from students in Japan. Subsequent Japanese contributions in the 1960s included adaptations of results on Markov processes to martingales, such as Itô and Watanabe's multiplicative analog of the Doob-Meyer decomposition, which involved the introduction of local martingales, contributions to stochastic integration for square-integrable martingales and semi-martingales, and contributions to the representation of martingales. Japanese contributions after 1970 included Itô's reformulation of the stochastic calculus in terms of stochastic differentials, Itô's circle operation, the Itô-Tanaka formula, and the Fukushima decomposition.

The third chapter of the third part is an autobiographical account by Klaus Krickeberg. As a university student at Humboldt University in the difficult conditions right after the Second World War, he was attracted to mathematics by the brilliant teaching of the famous analyst Erhard Schmidt. After moving from Berlin to Würzburg in 1953, he became acquainted with Doob's work on martingales and Dieudonné's counterexample to Doob's martingale convergence theorem. This led to his work on the role of Vitali-type conditions in the convergence. After obtaining his *Habilitation* in Würzburg, he spent a year in Doob's group at the University of Illinois. During this period, he proved that every $\mathbf{L}^1$-bounded increasing semimartingale with a directed index set converges stochastically. His further work on martingales at Würzburg from 1957 to 1964 continued to emphasize Vitali conditions.

The fourth part, entitled *Modern applications*, reviews a few of the applications of martingale theory. Its first chapter, by Laurent Bienvenu, Glenn Shafer and Alexander Shen, recounts the role played by the concept of a martingale in the algorithmic understanding of randomness. In the 1930s, Jean Ville used martingales to improve Richard von Mises's and Abraham Wald's concept of an infinite random sequence, or collective. After the development of algorithmic randomness by Andrei Kolmogorov, Ray Solomonoff, Gregory Chaitin, and Per Martin-Löf in the 1960s, Claus-Peter Schnorr developed Ville's concept in this new context. Along with Schnorr, Leonid Levin was a key figure in the development in the 1970s. While Schnorr worked with algorithmic martingales and supermartingales, Levin worked with the closely related concept of a semi-measure. In order to characterize the randomness of an infinite sequence in terms of the complexity of its prefixes, they introduced new ways of measuring complexity: monotone complexity (Schnorr and Levin) and prefix complexity (Levin and Chaitin).

The second chapter in the fourth part provided by Tze Leung Lai, describes how martingales came into his world of mathematical statistics, first in sequential tests and confidence intervals, then in time series, stochastic approximation, sequential design of experiments, and stochastic optimization. Lai sketches the trajectories of many other statisticians that he met along the way, emphasizing the roles of Harold Hotelling, Abraham Wald, and Herbert Robbins in their creation of the environment for the study of martingales at Columbia University and then his own subsequent work at Stanford University. At Stanford, he came to see stochastic optimization as a unifying theme for the use of martingales in statistical modeling.

Another applied field where martingales have made a great contribution is survival analysis. In their chapter, Odd O. Aalen, Per K. Andersen, Ørnulf Borgan, Richard D. Gill, and Niels Keiding trace the development of martingales in survival analysis from the mid-1970s to the early 1990s. This development was initiated by Aalen's Berkeley Ph.D. thesis in 1975, progressed in the late 1970s and early 1980s through work on the estimation of Markov transition probabilities, nonparametric tests, and Cox's regression model, and was consolidated in the early 1990s with the publication of the monographs by Fleming and Harrington and by Andersen, Borgan, Gill, and Keiding. The authors see this development as an unusually fast technology transfer of pure mathematical concepts, primarily from French probability, into a practical biostatistical methodology. It was possible because the martingale ideas inherent in the deep understanding of temporal development intrinsic to the French theory of processes were already quite close to the surface in survival analysis.

The last chapter of the fourth part by Tyrone Duncan, describes the role of martingales in stochastic control. Martingales provide a natural way to avoid some major mathematical difficulties that arise when Hamilton--Jacobi--Bellman partial differential equations are used to solve optimization problems. The brief history of stochastic control given by Duncan commences with work during the Second World War in the United States and proceeds through the 1950s and 1960s. The focus is on the solution of problems in potential theory arising in stochastic control and related stochastic filtering problems. Duncan's text is complemented by an appendix by Laurent Mazliak, who discusses how martingales play a central role in a more general probabilistic approach to stochastic control that does not involve the HJB partial differential equation. El Karoui and several of her collaborators showed in the 1980s how the concept of a martingale leads to the formulation of a powerful optimality criterion for the control of very general processes and the existence in distribution of optimal controls.

A fifth part presents some archival material that casts light on the history of probability theory in the twentieth century and documents some of the conclusions of earlier chapters. Many of the documents are letters, transcribed with commentary. Bernard Bru and Salah Eid comment on letters exchanged by Børge Jessen and Paul Lévy from September 1934 to August 1935, concerning the relation between what we now call Jessen's theorem and Lévy's lemma. As explained in Eid's earlier chapter, each author contended, with some justice, that the other's results could be derived from their own, and we now think of both as versions of the martingale convergence theorem. The letters have survived in the Jessen archives at the University of Copenhagen. An initial letter from Jessen is missing, but we have Lévy's five letters and drafts of two of Jessen's letters. Bru and Eid also include a 1947 letter from Harald Bohr and Jessen to Lévy, which does not contribute to the mathematical discussion but shows the continued relationship between the parties.

Bru and Eid also present ten letters dealing mainly with two erroneous theorems published by Joseph Doob in 1938, concerning probability measures on abstract spaces. The first concerned the construction of measures on infinite-dimensional product spaces from their finite-dimensional margins (the Daniell-Kolmogorov construction). The second concerned the existence of regular conditional probabilities. Ten years after the theorems were published and used by Doob's students, counterexamples were independently discovered by Jean Dieudonné and by Erik Sparre Andersen in collaboration with Børge Jessen. The correspondence begins when Jessen writes to Doob about the counterexample he and Sparre Andersen had discovered. It reveals that Doob had not encountered Jessen's theorem until then, and it suggests that Doob's return to martingales, which he had left aside after his initial work on them in 1940, was inspired by his learning about Jessen's work on the topic.

In 1984, Pierre Crépel contacted Jean Ville, who had retired from the University of Paris in 1978, to ask him about the sources of his thinking about martingales.

The next document is an English translation of their correspondence and a narrative based on Crépel's notes from a face-to-face interview. Ville recounts not only his work on martingales in the 1930s but also his perceptions of mathematical teaching and research in France during the period; his experience with Maurice Fréchet and Émile Borel in Paris and with Karl Menger in Vienna; and his own subsequent mathematical career.

Next Laurent Mazliak presents a series of recently discovered letters from Paul Lévy to Maurice Fréchet, which complement the large collection of such letters published by Marc Barbut, Bernard Locker, and Laurent Mazliak in 2014. The newly discovered letters cover many topics, but in particular they extend what we know about Lévy's vision of martingales and his negative view of Jean Ville.

Finally, Laurent Bienvenu, Glenn Shafer, and Alexander Shen provide translations of material written by Andrei Kolmogorov and Leonid Levin. This includes a 1939 letter from Kolmogorov to Maurice Fréchet, in which Kolmogorov explains his views on the connection between probability theory and its applications. It also includes documentation of the early history of Kolmogorov complexity in the USSR: abstracts of talks by Kolmogorov and letters from Levin to Kolmogorov.

This book is surely not the last word about the history of martingales in probability theory. We have scarcely touched on the concept's ever-greater role in the past 40 years, and we have not pretended to foretell what is to come. We have surely overlooked many fascinating aspects of probability's martingales, and the history our authors have told may well be retold, with yet more detail, context, and insight. But all authors have provided a cornucopia of insight, sufficient to explain how and why martingales have become so fundamental. We heartily thank them. We salute the memory of those among them no longer with us, Marie-France Bru, Niels Keiding, Bernard Locker, and Paul-André Meyer.

July 2022

Laurent Mazliak
Glenn Shafer

Contents

Part I In the Beginning

The Origin and Multiple Meanings of *Martingale*

Roger Mansuy

Abstract

A martingale can be a mathematical object, a gambling strategy in a casino, part of a horse's harness, part of a sailboat's rigging, a man's coat, or even a courtesan. The wide variety of objects described by the word is both intriguing and challenging. Lexicographers have advanced various hypotheses for the word's origin. Here we begin with its use in mathematics and work back to the picturesque city of Martigues, on the French Mediterranean coast. Along the way we will encounter a king's breeches, fencers, a sailor's dance, a prophetess, and an imprisoned Javert in Victor Hugo's *Les Misérables*.

Keywords

Martingale · Joseph Doob · Jean Ville

1 Introduction

For mathematicians, a martingale is a kind of stochastic process that was first studied in the mid-20th century. But if you search for *martingale* in the database of all the texts digitized by the Bibliothèque Nationale de France, you find several thousand books, the oldest dating from the 16th century. If you look more closely at the list of results obtained, you will see that in addition to books of mathematics, there are collections of tips for gamblers, texts about lottery regulation, volumes on military uniforms, horse riding manuals, technical booklets for the sailing navy, and even literary works.

This jumble is easily explained by the polysemy of the word; while mathematicians agree fairly well on what it means for a stochastic process to be a martingale,

R. Mansuy (✉)
Lycée Saint-Louis, 44 Boulevard Saint Michel, 75006 Paris, France
e-mail: roger.mansuy@ac-paris.fr

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*,
Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_1

it is difficult to pull together all the other uses of the word, the relationships between the different meanings often begin confusing. This is illustrated by an anecdote about Joseph L. Doob told by Laurie Snell in [40, p. 122]:

> One day Doob received a package from his former student, the well-known mathematician, Paul Halmos. In the package was a handsome leather strap with brass rings. After several inquiries, Doob was told that he had received a martingale.

In fact, Halmos's humorous gift was designed for a horse, not for a mathematician.

The connection between equestrian and mathematical vocabulary was extended when John Hammersley began to generalize the notion of martingale and named the stochastic processes he was studying harnesses. He justified his choice of the word in [21, p. 92]:

> The idea behind the terminology is the following. In gaming, a martingale is a fair gambling system, and this is probably the immediate source of the stochastic sense of "martingale." But in turn, the gaming term seems to have its origin in the equestrian sense of the word "martingale". In that sense, a martingale is a strap that prevents a horse from throwing up his head.

As we will see, Hammersley was mistaken. The etymologies of the different meanings of *martingale* for gamblers and equestrians are not so directly connected.

Many dictionaries propose the Spanish word *almartaga*, of Arabic origin, as the etymology for *martingale*. This word is still in use in Spanish, where it refers to a bridle used by the rider to dismount rather than to a strap that attaches to the saddle girth. However, the Spanish lexicographer Joan Corominas [10] refutes this appealing hypothesis; in fact, as the word *almartaga* is only found in Castilian, it must have originated on Iberian soil, where the use of *martingale* to designate the strap has never been seen. Beyond this first inconsistency, Corominas also provides more sophisticated arguments concerning the use of Arabic suffixes in the creation of Castilian words, which tend to show the impossibility of an etymological tie between *almartaga* and *martingale*.

We must therefore abandon this track and start again in our backwards quest from our current mathematical usage.

2 From Probability Back to Gambling

Twentieth-century pioneers in the study of the stochastic processes we now call martingales included Sergei Bernstein, Paul Lévy, Jean Ville, Émile Borel, and the aforementioned Doob. Their contributions are discussed in other chapters of the present volume.

The historical record shows that mathematicians learned the word *martingale* from gamblers, for whom a martingale was a betting strategy. In the 18th century, gamblers talked about *the* martingale: double your bet repeatedly. But by the 20th century many betting strategies were called martingales. Mathematicians had long contended that such strategies are futile; according to Bernard Bru, the conclusion that no strategy can assure a win in an unfavorable game goes back at least to the ancient Greek philosopher Xenophon of Athens. In the first years of the 19th century, mathematicians were already using *martingale* when making this point; the futility of the doubling martingale was pointed out in books by the mathematicians Condorcet (posthumously in 1805 [7]) and Sylvestre-François Lacroix (1816, [24]). A century later, we find the mathematician Louis Bachelier explaining that

> aucune combinaison, martingale ascendante ou descendante, progression arithmétique ou géométrique ou quelconque[1]

was just as futile as the basic doubling martingale [3, Chap. 10].

The futility of a betting strategy took on new significance in the early 20th-century in the work of Richard von Mises [26]. For von Mises, the futility of a strategy for selecting trials of a repeated event on which to bet was a fundamental aspect of the meaning of probability. But von Mises did not use *martingale* in this connection, and gamblers' martingales were always more complicated than his strategies. Instead of merely choosing which trials to bet on, and always betting the same amount on the same outcome when they do bet, gamblers' martingales varied the amount bet and often the outcome they bet on. Jean Ville's insight was that von Mises's viewpoint was therefore incomplete. All betting strategies, at least all those that do not risk more than some amount fixed at the outset, should be considered.

Ville's fundamental result was that for any event of probability zero in an infinite sequence of trials, there is a betting strategy that begins with finite capital, does not risk more than this initial capital, and turns it into infinite capital if the event of probability zero happens. When he first stated this result, in a note in the *Comptes rendus* in 1936 [43], Ville called any such betting strategy a *martingale*. This may have been the first time the word was used with this generality. Many decades later, in a letter to Pierre Crépel that is translated into English in Crépel's chapter in the present volume, Ville recalled that he had learned the word from a certain Monsieur Parcot, a gambler who was a relative of his wife. He may have also encountered it in writings on games of chance by mathematical colleagues.

In his thesis, published in 1939 [42], Ville took a further step that proved very influential. The strategies are in a one-to-one correspondence with the sequences of random variables representing the player's capital—i.e. the capital processes. So Ville left aside notation for the strategies and worked with these sequences of random variables, which, as he said, "sufficiently define" the martingales.

Ville wrote in French. It was his American contemporary Joseph L. Doob who brought the new mathematical meaning of *martingale* into English. Already a leading authority on probability theory among American mathematicians, Doob reviewed Ville's book for the *Bulletin of the American Mathematical Society* in 1940 [15]. He

[1] English: any combination, any ascending or descending martingale, any progression, whether arithmetic, geometric, or whatever.

later adopted Ville's use of *martingale* in his own work, especially in his influential 1953 book on stochastic processes [16].[2]

3 Are Martingales Foolish?

The next step seems more subtle: where does the term used by gamblers come from? The entry for martingale in the dictionary of the Académie Française was introduced in 1762, in the fourth edition [1]: "Jouer à la martingale, c'est jouer toujours tout ce que l'on a perdu".[3] This definition is repeated by many books on gambling, as in this passage by James Smyll in 1820 [39, pp. 64–65]:

> Plusieurs personnes jouent en Martingale, c'est-à-dire qu'ayant perdu *un* écu elles en mettent *deux* pour le coup suivant, et si elles perdent encore, elles en mettent *quatre* toujours en doublant, 8, 16, 32, 64, 128, 256 etc. de sorte qu'après avoir perdu huit coups de suite et conséquemment 255 écus et mettant encore 256 écus pour le neuvième coup, en gagnant ce dernier coup elles ne regagneroient que ses pertes des huit coups précédens et un écu en sus.
>
> D'autres jouent la Martingale en augmentant leurs mises de manière à ce que chaque coup leur fasse gagner un écu, ou masse s'ils ne sautent pas.[4]

Such strategies risk ruining unlucky players, and many authors warn against this danger. In 1801 [29, p. 118], Parisot writes:

> Avant d'entreprendre une martingale, la prudence commande de s'assurer d'abord si l'on est en état de la soutenir pendant le nombre de tirages présumés nécessaires pour la faire réussir.[5]

In his memoirs [8], Giacomo Casanova recalls martingaling at Venise's Casino in 1754. He was lucky enough that he never lost six times in a row:

[2] See the interview [41]. Here Doob also explains why he preferred "lower semimartingale" to "supermartingale" in 1953: "the name "supermartingale" was spoiled for me by the fact that every evening the exploits of "Superman" were played on the radio by one of my children."

[3] English: To play the martingale is to always bet all one has lost.

[4] English: Many people play Martingale, i.e. having lost *one* ecu they put in *two* on the next round, and if they lose again, they put in *four* always doubling, 8, 16, 32, 64, 128, 256 etc., so that after losing eight rounds in a row and consequently 255 ecu and still putting in 256 ecu for the ninth round, by winning this last round they will only regain their losses from the previous eight rounds and one ecu in addition. Others play the Martingale by increasing their bets so fast that they net an additional ecu for each round if they do not go bankrupt.

[5] English: Before embarking on a martingale, it is prudent to first make sure that one is in a position to support it for the number of draws presumed necessary to make it a success.

> J'ai pris tout l'or que j'ai trouvé, et pourtant avec la force qu'en terme de jeu on dit à la martingale, j'ai gagné trois et quatre fois par jour pendant tout le reste de carnaval. Je n'ai jamais perdu la sixième carte. Si je l'avais perdue je n'aurais plus eu de fond qui consistait en deux milles sequins.[6]

In 1760 [14], Denis Diderot tells his correspondent and lover Sophie Volland that the aging baron of Dieskau "a fait la martingale avec nous".[7] Going further back, the abbé Prévost's dictionary [30], beginning in 1750, provides a definition of *martingale* limited to the game of Faro: the gambler doubles his stake at each loss "pour se retirer avec un gain sûr, supposé qu'il gagne une fois".[8] These references take us back to the 18th century but provide no clue as to how the word was adopted by gamblers.

One hypothesis, tenuous on its face, is that the word comes from the Provençal expression *jouga a la martegalo*, which means "to play in an absurd and incomprehensible way" [27]. One can easily understand that the strategy of doubling the stake might have seemed absurd for players who lived before the Age of Enlightenment, who adamantly believed that bad luck was a sign of fate.[9] Few French sources support this hypothesis, but more light comes from across the Channel: Randle Cotgrave's 1611 French-English dictionary [11] gives *à la martingale* the meaning "absurdly, foolishly, untowardly, grossely, rudely, in the homeliest manner" and even quotes a use of the expression "philosopher à la martingale". This expression appears already in the 16th century [9, p. 52].

This hypothesis of a Provençal origin is supported by the fact that many other words used by gamblers were borrowed from this language; for example, the game of cards called Baccara(t) takes its name from a Provençal expression meaning "going bankrupt". Having found some substance in this lead, we need to follow it further.

4 An Excursion Around Martigues

Having reached a new stage in this quest, it is now necessary to understand the origin of the expression *jouga a la martegalo*. Further examination of Frédéric Mistral's Provençal dictionary [27] indicates that the word *martegalo* also refers to the residents of Martigues, to whom is attributed a certain "gaping", "naïveté", "banter". "Le Martigue" then designated the pond of Berre, which had given its name to the city created on April 21, 1581, by the fusion of three boroughs (Ferrières, Jonquières and L'isle) bordering the opening to the Gulf of Fos. The isolated situation of this area "a valu à ses habitants une réputation de naïveté proverbiale".[10]

[6] English: I took all the gold I could find, and by the trick called *à la martingale* in the language of gaming, I won three or four times a day throughout the rest of the carnival. I never lost the sixth card. Had I lost it, I would have had nothing left of the two thousand sequins with which I began.
[7] English: played the martingale with us.
[8] English: in order to quit with a sure profit, provided that he wins once.
[9] Pascal's ideas on probability, especially his famous wager on the existence of God, had not yet spread much.
[10] English: brought to its residents a reputation for proverbial naivety.

For completeness, it must be said that the source of the place name is a matter of debate. The authoritative explanation [32] is *Stagnum Marticum*, the pond of stones, but other more or less far-fetched theses coexist. The place name might derive from that of an ancient city, a priestess, a Roman general…

Step by step, Martigues is emerging as the destination of our quest beginning with probability and gambling. Before skipping to the equestrian meaning of *martingale*, let us explore several meanings more or less directly associated with the region of Martigues.

The martingale garment worn by François Rabelais's Panurge is a pair of breeches with an opening at the back (in Rabelais's words, "un pont-levis de cul pour plus aisément fienter" [31]).[11] Jacob Le Duchat, a French scholar of the 18th century, adds these comments in his reference edition of the works of Rabelais [17, p. 82][12]:

> Cette forme de culottes qui étoit encore en usage du tems de Rabelais prit son nom des Martégaux, peuple du Martégue en Provence qui l'avoient inventée …On a dit Martingale pour Martégale, comme Portingal, qui dans nos vieux livres est le nom du Portugal.

The currency, between the 16th and the 18th centuries, of references to "des habits à l'espagnole, à l'italienne et particulièrement à la napolitaine, à la flamande, à la martingale" [25, entry *grégues*][13] suggests that *martingale* was used in this case to describe people by referring to where they lived.

Martingale breeches seem to have been worn quite widely; Brantôme [5] confirms, a century later, that François I, the king of France from 1515 until his death in 1547, wore them:

> Ce brave chevalier avoit une complexion en luy, que, toutes les fois qu'il vouloit venir au combat, il falloit qu'il allast à ces affaires [...] et pour ce portoit ordinairement des chausses à la martingale.[14]

La Curne's Old-French Dictionary [34] states that the fashion originated in Martigues, but that these pants

> étoient encore à la mode environ l'an 1579, entre les mignons de la cour, qui les faisoient servir à tout autre usage que celui pour lequel on les avoit inventées.[15]

Martingale remains in use as the name of a half-belt in the back, which can still be seen on some jackets and also in fencers' outfits. By analogy, it also refers to a strip

[11] English: a drawbridge on the ass that makes excretion easier.

[12] English: This form of breeches, which was still in use at the time of Rabelais, took its name from the Martégaux, the people of Martégue in Provence who invented it …We said Martingale for Martégale, like Portingal, which in our old books is the name of Portugal.

[13] English: clothes in Spanish, Italian (particularly) Neapolitan, Flemish, Martingale styles.

[14] English: This brave knight had to have a bowel movement every time he wanted to begin fighting and therefore usually wore martingale pants.

[15] English: were still in fashion around 1579 among the minions of the court, who used them for quite different purposes than the one for which they were created.

of cloth or leather attaching the handle of the weapon to a fencer's forearm or wrist. In [6], Georges Breittmayer explains that in a duel respecting the code of honour, "les extrémités de la martingale ne doivent pas pendre ni former aucune boucle où l'épée adverse puisse s'engager".[16]

Martingale is also associated, in a more anecdotal way, with a sailors' dance still taught by some folkloric dance associations.[17] Darluc makes a clear link between the temperament of sailors and this typical dance [12, p. 402]:

> La franchise, la candeur, font en général le caractère des matelots ; ils ne savent point dissimuler leurs sentimens, ni mettre un frein à leurs passions ; ils sont jaloux de leur liberté & extrêmement vifs, ce qui les fait paroître brusques & impatiens : ils aiment la danse; on nomme *martingale*, celle qu'ils exécutent ici avec le plus d'action.[18]

It seems that this dance consists mainly in repeatedly stamping the ground roughly with the heel. It has remained little known, even though the report on the voyage of the 14-year old Charles IX (King of France from 1560 until his death in 1574) and his court to Brignoles on October 26, 1564 [4] states that

> les habitants s'étudièrent à lui donner du plaisir par la gentillesse des danses de la contrée [...] danses que l'on appelle la volte et la martingale[19]

At the entry "Martegues ou Martigues", Louis Moreri [28] explains, "Le Sieur Soleri parle de l'enjoüement et des dances des habitans du Martigue, d'où est venu le Proverbe, *Danser à la martingale*."[20]

The famous novelist Alexandre Dumas, in his picturesque trip through Provence [18], reports a related usage that has not been corroborated: "chez le peuple provençal, pour dire bien danser, on dit: danser à la martingale".[21]

Still more picturesquely, a prophetess from Provence named Claude Scotte called herself Martingale (with a capital M, or sometimes with the misspelling Martingalle) in her correspondence between 1617 and 1628. Her letters are full of Provençal quatrains, of visions (groups of angels and holy apparitions), and of doubtful predictions. An heir to the throne, the future Louis XIV, will arrive only in 1638, more than ten

[16] English: The ends of the martingale must not hang down or form any loops where the opposing sword can be caught.

[17] Mentioned by Randle Cotgrave [11]. Its location is revealed by its earliest citation, in the work of a Provençal jurist Antonius Arena [2].

[18] English: Frankness and candor are generally what make the character of sailors; they do not know how to hide their feelings, nor how to put a brake on their passions; they are jealous of their freedom and extremely lively, which makes them appear abrupt and impatient: they love dance; they give the name *martingale* to the one they perform here with the most action.

[19] English: the citizens tried to please him through the gentleness of the dances of the area [...], dances called the "volte" and the "martingale".

[20] English: Sieur Soleri talks about the cheerfulness and dances of the inhabitants of Martigue, where the proverb came from: dancing à la martingale.

[21] English: The Provençaux, to say "dance well", say "dance à la martingale".

years after her prophecy. The stronghold of La Rochelle, demolished by Richelieu in 1628, will be conquered many years after the prediction by the unfailing Martingale. These letters are always punctuated by requests for a pension or donations and by anecdotes showing la Martingale's self-proclaimed pious and devout life. Indeed, there are a great number of petitions such as: "votre majesté aura égard aux services que Martingale a rendus à la France" [36].[22]

Finally, the word *martegalo* also refers to a sailboat and to a rope attached below the bow spit to help hold down the flying jib above. It is not surprising that the famous sailors called "martégaux" gave their name to these objects. Numerous documents, including [12] and [37], attest to the boldness and talent of the martégaux for net fishing, which they practiced as far away as the south of Italy and Andalusia. But this particular rope raises further questions.

5 Back to Harnesses

Adjusting for the scale, the sailor's martingale resembles so closely the horseman's running martingale that it can be mistaken for it. Aside from the collar that attaches it to the horse's neck, the running martingale is a strap that runs from the bottom of the saddle girth through the front legs and then bifurcates to attach to each of the reins by means of rings; see Fig. 1. It keeps the horse from throwing its head while the rider has at least one hand free to play polo or use a weapon, etc.

The various dictionaries previously cited almost all mention martingales for horses, without ever supplying a convincing etymology. The oldest citation seems to the one in John Florio's 1598 Italian-English Dictionary [19]. Starting from there, any philologist, no matter how scholarly, is reduced—for lack of additional information—to conjecture: does the name stem from the analogy with the Mediterranean sailors' rope? Is it a fortuitous similarity of sound (originating from a lexical association) in Provençal between a local expression and an ancient word from another Mediterranean language? Could Walter Skeat possibly have been right when he claimed in his etymological dictionary [38], without evidence, that the dolphin striker on a boat is called a martingale because of its resemblance to the horse's martingale? Presently it is impossible to answer these questions categorically (Fig. 2).

On the other hand, we can clearly see the origin of the name *martingale* for the yoke imposed on Javert in Victor Hugo's famous novel *Les Misérables* [22]:

> Pour plus de sûreté, au moyen d'une corde fixée au cou, on ajouta au système de ligatures qui lui rendaient toute évasion impossible cette espèce de lien, appelé dans les prisons martingale, qui part de la nuque, se bifurque sur l'estomac, et vient rejoindre les mains après avoir passé entre les jambes.[23]

[22] English: your Majesty will take into consideration the services that Martingale rendered to France.
[23] English: For greater security, by means of a rope fixed to his neck, they added to the system of bonds that rendered all escape impossible the type of ligature called a *martingale* in the prisons,

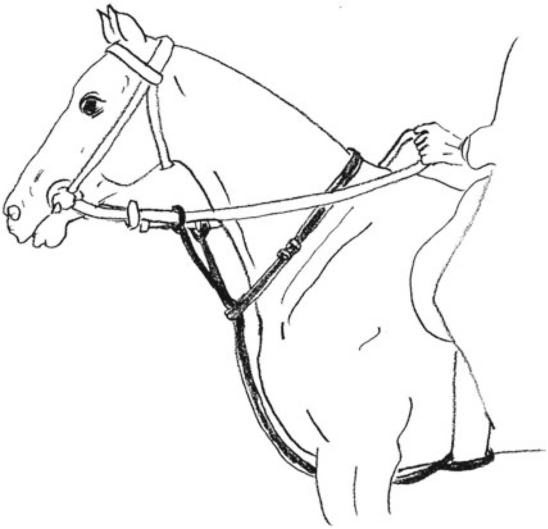


Fig. 1 There are several types of martingales for horses. The *running martingale*, shown here, is probably the type Halmos gave to Doob. Its two rings connect to the reins held by the rider, preventing the horse from throwing its head. Another type, the *standing martingale*, connects to the horse's harness without bifurcating *Drawing: courtesy of Nell Painter*

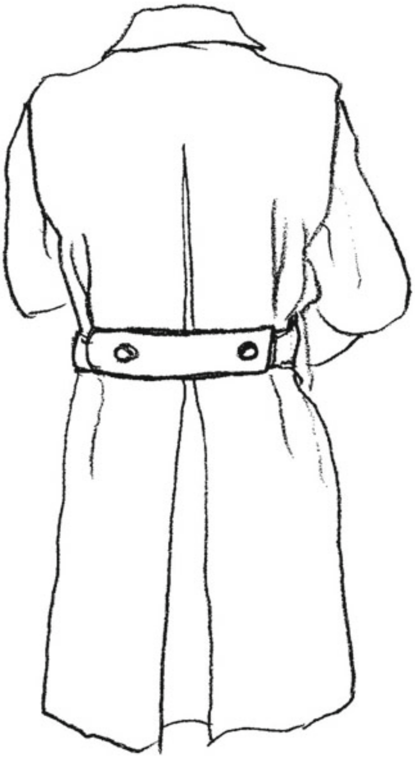


Fig. 2 Men's martingale coats, showing a belt in the back, are still popular. It is believed that this garment and its name derive from 16th-century breeches that closed in the back *Drawing: courtesy of Nell Painter*

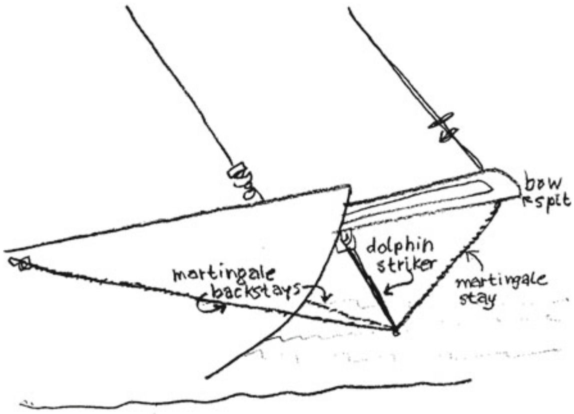


Fig. 3 The martingale sketched here in red helps the bow spit withstand the upward pull of the jib sail. A spar, here given the self-explanatory name *dolphin striker* but sometimes itself called a martingale, is stabilized by the martingale's two back stays while it holds the martingale's front stay at an angle that allows it to hold the bow spit down *Drawing: courtesy of Nell Painter*

Indeed, there can be little doubt: this meaning of *martingale* derives from the horse gear even if it sounds like another entirely different use of the word. We can be sure that it is not merely an invention by the great poet, because we find corroboration in the testimony of a former prisoner [23] (Fig. 3).

6 The Ultimate Treachery of Martingales

Before concluding this lexicographic journey, we must still examine one last long forgotten meaning. According to slang dictionaries dating back to the 18th century [13,33], *martingale* was used in the vernacular langage to refer to "putain, coureuse, courtisane, femme de mauvaise vie, d'une conduite déréglée."[24] This very particular meaning accounts for the word's appearance in burlesque plays like Paul Scarron's *Virgile travesti* (in the fourth book of this parody of Virgil's Aeneid) [35]:

> Vous êtes un homme bien sage
> Eh quoi pour vos folles amours
> Voudriez vous bien passer vos jours
> À faire le Sardanapale
> Et servir une martingale?

which starting from the back of the neck, divides over the stomach, and is fastened to the hands after passing between the legs.

[24] English: harlot, courtesan, street-walker, woman of low virtue, of unregulated conduct.

Cependant le fils de Cypris,
Suivant sa vieille martingale,
Aborda la rive infernale.[25]

or Jean-Louis Fougeret de Montbron's *La Henriade, travestie en vers burlesques* [20]:

Qu'aux piés de quelque martingale,
Ainsi qu'Hercule à ceux d'Omphale
Le pleutre fasse le calin,
Et file du chanvre ou du lin.[26]

Yet at this stage, nothing can be said about the origin of this meaning of the word. Here, perhaps, is the ultimate treachery of martingales!

Acknowledgements An earlier version of this article appeared in French as "Histoire de martingales" in *Mathématiques & Sciences Humaines/Mathematical Social Sciences*, 43th year, no. 169, 2005(1), pp. 105–113. The author is indebted to Ronald Sverdlove, who translated the French article into English, and to the artist Nell Painter, who provided the sketches.

References

1. Académie Française: Dictionnaire de l'Académie Françoise, 4 edn. Veuve Brunet, Paris (1762)
2. Arena, A.: Provincialis de bargardissima villa desoleriis ad suos compagnones studiantes qui sunt de persona friantes bassas dansas de novo. Nourry, Lyon (1528)
3. Bachelier, L.: Le jeu, la chance et le hasard. Flammarion, Paris (1914)
4. Bouche, H.: La chorographie ou Description de Provence, vol. 2. Aix (1664). Cote BNF FOL-LK2-1390 (2)
5. de Bourdeilles (dit Brantôme), P.: Grands capitaines français. Paris (1655)
6. Breittmayer, G.: Code de l'honneur et du duel. Devambez, Paris (1918). Cote BNF 8-R-28857
7. Condorcet, Jean-Antoine-Nicolas de: Élémens du calcul des probabilités, et son application aux jeux de hasard, à la loterie, et aux jugemens des hommes (1805)
8. Casanova, G.: Histoire de ma vie, manuscrit, livre III 282v (1789–1798). Cote BNF NAF 28604 (3)
9. de Cholières, N.: Les après disnées du seigneur de Cholières (1587). Cote BNF 8-Y2-52759
10. Corominas, J.: Diccionario critico etimologico, vol. III. Gredos, Madrid (1980)
11. Cotgrave, R.: A dictionarie of the French and English tongues. Islip, London (1611)
12. Darluc, M.: Histoire naturelle de la Provence, vol. 1. Niel, Avignon (1782)
13. d'Hautel, C.L.: Dictionnaire du bas-langage ou des manières de parler usitées parmi le peuple. Schoell (1808)
14. Diderot, D.: Letter to Sophie Volland, November 6, 1760

[25] English:You are a very wise man So why for your foolish loves Would you want to spend your days Behaving as Sardanapalus And serving a martingale?
Yet Cypris's son, Following his old martingale, Arrived at the infernal shore.

[26] English: At the feet of any martingale, Like Hercule at Omphale's May the coward cuddle, And spin hemp and flax.

15. Doob, J.: Review: Jean Ville, Étude critique de la notion de collectif. Bull. Amer. Math. Soc. **45**(11), 824 (1939)
16. Doob, J.: Stochastic processes. Wiley, New York (1953)
17. Duchat, J.L.: Œuvres de Maître François Rabelais, publiées sous le titre de : Faits et dits du géant Gargantua et de son fils Pantagruel, avec la Prognostication pantagruéline, l'épître du Limosin, la Crême philosophale, deux épîtres à deux vieilles de moeurs et d'humeurs différentes et des remarques historiques et critiques de M. Le Duchat, vol. II (1732)
18. Dumas, A.: Voyage pittoresque en Provence (1853)
19. Florio, J.: A worlde of wordes. Hatfield, London (1598). An enlarged version appeared in 1618 under the title "Queen Anna's new world of words".
20. Fougeret de Montbron, J.L.: La Henriade, travestie en vers burlesques (1751)
21. Hammersley, J.M.: Harnesses. In: Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability (Berkeley, 1965/66), vol. III: Physical Sciences, pp. 89–117. Univ. California Press, Berkeley (1967)
22. Hugo, V.: Les Misérables. Pagnerre, Paris (1862). V, I, VI
23. Joigneaux, P.: Prisons de Paris, par un ancien détenu. Self-published (1841)
24. Lacroix, S.F.: Traité élémentaire du calcul des probabilités. Courcier, Paris (1816). Second edition 1822
25. Ménage, G.: Dictionnaire étymologique de la langue française. Briasson, Paris (1750)
26. von Mises, R.: Wahrscheinlichkeit, Statistik und Wahrheit. Springer, Wien (1928)
27. Mistral, F.: Lou Tresor dòu Felibrige ou dictionnaire Provençal-français. Petit, Raphaèle-les-Arles (1979)
28. Moreri, L.: Le grand dictionnaire historique, ou Le mélange curieux de l'histoire sacrée et profane, vol. 2(1). Lyon (1683). Cote BNF G-1045
29. Parisot, S.A.: L'art de conjecturer à la loterie, ou Analyse et solution de toutes les questions les plus curieuses et les plus difficiles sur ce jeu, avec des tables de combinaisons et de probabilités, et diverses manières de jouer, toutes fondées sur le calcul et confirmées par l'expérience. Paris (1801). Cote BNF V-20006
30. Prévost, A.A.F.: Manuel lexique ou dictionnaire portatif des mots François. Didot, Paris (1750)
31. Rabelais, F.: Gargantua. Juste, Paris (1534)
32. Rostaing, C.: Dictionnaire étymologique des noms de lieux de France. Larousse, Paris (1963)
33. Roux, P.J.L.: Dictionnaire comique, satyrique, critique, burlesque, libre et proverbial. Lyon (1735). Cote BNF X-14022
34. de La Curne de Sainte-Palaye, J.B.: Dictionnaire historique de l'ancien langage François, vol. 7. Favre, Niort (1875)
35. Scarron, P.: Le Virgile travesti (c. 1648–1653)
36. Scotte, C.: Factum au Roy et à Nosseigneurs de son conseil, pour Claude Scotte, dite Martingalle, prophétesse provençale, aux fins d'avoir récompense de Votre Majesté pour les signalés services qu'elle et son feu mari ont rendus à Votredite Majesté et au public (1628). Cote BNF: 8-LN27-13672
37. de Séguiran, H.: Procès verbal du président du parlement de Provence et surintendant général de la navigation et du commerce (1633). Archive du service historique de la marine au château de Vincennes: 5H25888A
38. Skeat, W.W.: An Etymological Dictionary of the English Language. Clarendon Press (1910)
39. Smyll, J.: Tactique des jeux de hasards: recherches sur les meilleures manières d'y jouer et de jouer avec assurance de gain. Hinrichs, Leipzig (1820). Cote BNF V-52897
40. Snell, J.L.: Gambling, probability and martingales. Math. Intelligencer **4**(3), 118–124 (1982)
41. Snell, J.L.: A conversation with Joe Doob. Statist. Sci. **12**(4), 301–311 (1997)
42. Ville, J.: Étude critique de la notion de collectif. Gauthier-Villars, Paris (1939)
43. Ville, J.A.: Sur la notion de collectif. Comptes rendus **203**, 26–27 (1936)

Martingales at the Casino

Glenn Shafer

Abstract

For more than two centuries, various betting strategies have been called *martingales*. What makes a betting strategy a martingale? Why have martingales been so popular among gamblers and casinos? The decade after the French revolution marks the beginning of a written record that casts some light on these questions. This record centers on Trente et Quarante, the game that drove the profits of the gaming spas of northwestern Europe and the casinos of Paris before the invention of Roulette in its modern form. Players in Trente et Quarante made successive even-money bets on red or black. *Martingaling* initially meant doubling one's bet to recover from a loss. By the beginning of the 19th century, any betting strategy could be called a martingale. A review of this history will help us understand how seductive martingales can be. We still live with the perils of this seduction, inside and outside the casino.

1 Prelude

From the French poet and philosopher Antoinette Du Ligier de la Garde Deshoulières (1638–1694), Verse XIV of her *Réflexions Diverses* [42, p. 97]:

> Les plaisirs sont amers d'abord qu'on en abuse:
> Il est bon de joûër un peu,
> Mais il faut seulement que le jeu nous amuse.
> Un joûëur d'un commun aveu,
> N'a rien d'humain que l'apparence;
> Et d'ailleurs il n'est pas si facile qu'on pense
> D'être fort honnête homme & de joûër gros jeu.
> Le desir de gagner qui nuit & jour occupe

G. Shafer (✉)
Rutgers Business School, 1 Washington Place, Newark, NJ 07102, USA
e-mail: gshafer@business.rutgers.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*,
Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_2

Est un dangereux éguillon.
Souvent, quoyque l'esprit, quoyque le coeur soit bon,
On commence par estre dupe,
On finit par estre fripon.

In translation, sadly without the rhyme and 12/8 m[1]:

Pleasures are bitter as soon as they are abused:
It is good to gamble a little,
But only if the game is merely to amuse.
A gambler, by common assent,
Is human only in appearance;
And besides it is not as easy as one thinks
To be a truly honest man and play a big game.
The desire to win, which takes hold day and night
Is a dangerous poison.
Often, no matter that the mind, the heart be good,
You begin by being a dupe,
You end by being a rogue.

2 Introduction

Double or quits, double or nothing, *quitte ou double*, *doppelt oder nichts*. When you lose a bet, bet twice as much. By winning, you will recover your loss and also make the gain you had first hoped for. This idea for turning loss into gain is surely as ancient as gambling itself (Fig. 1).

Repeated doubling following repeated losses can hardly be less ancient. As the losses mount, the redoubling becomes a strategy of desperation, and the gambler is ruined. The earliest known literary enactment of this spectacle is the 13th century fabliau of St. Peter and the Minstrel, which explains why there are no minstrels in hell. There once was a minstrel in hell, we are told. Satan had left this minstrel in charge when he and his devils went to hunt more souls. St. Peter, spying an opportunity, engaged the minstrel in a game of *hazart*, a three-die ancestor of craps: St. Peter's gold against the minstrel's souls. The minstrel gambled away all the souls. Satan, enraged when he returned, expelled the minstrel from hell.[2]

We do not know when *martingale* was first used to name the strategy of repeated doubling. But as Roger Mansuy notes in his chapter in the present volume, the abbé Prévost recorded this meaning for the word in his 1750 French dictionary, saying

[1] Unless otherwise noted, translations from the French are by the author of this chapter.
[2] Noomen and van den Boogaard provided the definitive edition of the fabliau in modern French in 1983 [66, pp. 129–159]. Sheldon's translation of the old French into modern English, published in 1912 [84], is readily available on-line.

Fig. 1 Jules Arnous de Rivière, celebrated French chess master and organizer of chess tournaments at Monte Carlo, wrote his books on games of chance under the pseudonym Martin Gall. Here, on the cover of the book he published on Roulette and Trente et Quarante in 1882 [46], the beautiful croupier becomes a goddess of chance at the seaside, standing on the roulette wheel and raking in the gamblers' gold with cupidity's help. *Source* gallica.bnf.fr/Bibliothèque nationale de France

that a player is following a martingale when he repeatedly doubles in order to be sure of ending up with a gain, supposing that he wins at least once.[3]

By 1800, there were many martingales at the casino. In Armand Croizette's *Le Masque tombé*, performed at the Théâtre Molière on January 10, 1801, the rogue

[3] [74, vol. 2, p. 490].

Melval tells his intended dupe Dorsange that he has an infallible martingale, better than the martingale of Spa, the Darquin, the Foudroyante, and all other martingales past, present, and future.[4] If martingaling meant playing wildly like someone from Martigues, as Mansuy argues, the martingalers' play was not to be rigidly prescribed.

Croizette's mockery and right thinkers' sermons were of little avail; there were ever more martingales, martingalers, and sellers of martingales in the 19th century. How did the rogues and dupes convince themselves and others that the martingales would succeed?

3 The Casino

Following the model of Bath and Tunbridge Wells in England, a number of towns in principalities bordering on France became cities of pleasure in the second half of the 18th century, building dance halls and gaming rooms around spas that claimed beneficial mineral waters. The archetype was the town of Spa in the Pays de Liège, now part of Belgium. Spa's heyday was launched in 1762, when the prince-bishop of the principality of Liège granted the town's magistrate a monopoly on public gaming.

Development of the gaming spas (*Spielbäder*, they were called in German), was irresistible for the rulers of the principalities, who shared in the profits from the monopolies they granted. But as competition grew, the spas became capitalist enterprises, which attracted entrepreneurs and both noble and bourgeois investors, and which developed a business model for casinos that spread throughout Europe.[5]

Emblematic of the professionalization of the gaming spas was the entrepreneur Pierre Nicolas Huyn (1753–1843). A native of Lorraine and a successful businessman in Koblenz, Huyn had an interest in the casinos in the free imperial city of Aachen (Aix-la-Chapelle in French), which hosted gaming year-round, even in the winter when Spa's baths were closed. In 1786, he acquired the monopoly on gaming at Ems. Eventually attracting additional investors, he built a permanent casino open year-round in Ems in 1824, helping make it one of the most exclusive of the gaming spas. In an 1862 broadside against gambling, Charles de Birague reported that Ems had become the rendez-vous of the aristocracy of birth and finance, where those who gambled made a show of not caring about their money.[6]

---

[4] [34, Scène XI, pp. 25–26].

[5] See Russell Barnhart [10] for Bath and Tunbridge Wells, Manfred Zollinger [98, especially pp. 229–256] for the Spielbäder in the Holy Roman Empire.

[6] Thierry Depaulis [40, p. 18], [39, pp. 49–53], Charles de Birague [19, pp. 28–29]. Concerning gaming in the winter at Aix-la-Chapelle [86, p. 29].

3.1 Trente et Quarante

In 1788 Huyn issued a pamphlet on games of chance.[7] In French, of course: the language of the casinos and of the European elite in general. The upright casino owner did not tell potential customers how easily they could become rich. Instead, he explained the house's advantage in each game he described, urging those who played to recognize that they were paying for their pleasure.

Thirteen of Huyn's 51 pages are devoted to Trente et Quarante, which he correctly describes as the least disadvantageous of casino games to the gambler, and the most honest. It was not a new game; Italians had been playing it in the 16th century, and Molière had mentioned it in *The Miser*, first performed in 1668. It is not an innocent game, for many fortunes have been lost playing it. But the house's modest advantage is transparent, cheating is nearly impossible, and card counting is futile. Six decks of cards are shuffled together—312 cards in all. Each player may then bet on red or on black, putting the money they bet in corresponding cells on a *tapis vert* (green tablecloth) like the one shown in Fig. 2. The bets having been made, the dealer deals two rows of cards face up, one for black, one for red. In each case, he stops when the total of the points on the faces—an ace counts as 1 and each face card as 10—is at least 31. The greatest possible total is 40; thus the name Trente et Quarante (Thirty and Forty). Other names for the game were Rouge et Noir (Red and Black) and Trente-un (Thirty-one). The round is won by the color whose total is closest to 31; the bettors on that color receive payoffs matching the amounts they had bet. If the two colors are tied, the bets are off; but if the tie is 31 to 31, the house takes half of everything the players have bet. Then the process is repeated, until the deck is exhausted. In the long run the house takes about 1.1% of the money bet.[8]

When Trente et Quarante was played between friends or in the salons of the nobility in the early 18th century, there was no house and no advantage for the house; onlookers could make side bets on how the numbers came out, but the basic game was between the player who cut the cards and the player who dealt them. One legend had it that a certain Hazon or Azon introduced the house's advantage at Spa, first claiming half the stakes for any tie and then retreating to making the claim only for the 31–31 tie.[9]

[7] The cover page discloses only the title, the author, and the year 1788. The book could have been printed in France or smuggled into France, especially as the royal administration of the book trade was already crumbling [51, Chap. 1].

[8] In calculating the fraction taken by the house, we can include or exclude rounds resulting in a tie at 32 or more. Excluding these rounds, Huyn estimated the house's take to be $1\frac{1861}{6589}\%$, or about 1.3%. Calculations by Poisson in 1825 [22,72] and even more precise recent calculations show it is a little less. According to Ethier [43, p. 628], it is 0.011 998 000 when ties are excluded and 0.010 945 685 when ties are included.

[9] The early 18th century game, as it was differently played in France and in Florence, is described in [61]. In 1798 G. N. Bertrand wrote that Hazon invented the banker's advantage [15, p. 3], and this name was repeated in 1799 by de Gaigne [45, p. 28]. *Spéculateur*, writing in 1809 and echoed by Grégoire in 1863, placed the invention in Spa and named the inventor Azon [87, pp. 19–20], [47, p. 34]. A compilation of tales about Paris gambling dens written for the purpose of blackmail and

Fig. 2 Frontispiece in Debrett's *Faro et le Rouge et Noir* [5], showing a *tapis vert* for Trente et Quarante. The *tailleur* (dealer) stands at T; the *croupier*, who settles the bets, stands at C. Players put their money on red or black. Instead of giving up half their bet on a 31–31 tie, a player may gamble it on the next round by pushing it inward to the next compartment. If the next round is again a 31–31 tie, the player may again postpone by pushing their money even further inward. *Source* Google Books

There is no good reason for the two rows of cards to be called red and black. But French playing cards had been half red and half black since the 16th century, making red versus black the natural 50–50 bet. The 50–50 odds of Trente et Quarante were infinitely more reliable than having a rogue deal a single card, flip a coin, or spin an 18th century roulette wheel with its unnumbered but alternately red and black pockets.

In his discussion of Trente et Quarante, Huyn comments on martingales, large and small. A large martingale doubles many times if it keeps losing. A small martingale might double only once or twice. In either case, the player might resume betting with a standard amount after finally winning or giving up the doubling. Huyn insists that neither a martingale nor any other manner of playing can diminish the house's advantage:

> In the long run all events equalize, and the house, having more chances in its favor, absolutely must win. If a player has the good luck to make a big win, he will lose it again in small

---

smuggled into France in 1781 gave Hazon as the name of a card-sharp who ran the public games at the house of the Venetian ambassador in Paris [1, p. 23]. Azon was the name of an early 13th century scholar who argued that magistrates had the power to enforce their judicial decisions.

> chunks, just as he will lose, in one large chunk, what he has won with a martingale, because no matter how large the martingale is, it will fail in proportion to what it can gain.[10]

In 1803, Huyn's pamphlet was reprinted, supposedly in Amsterdam, to be sold in bookstores there and in the Hague, Leiden, and Brussels. But its contents gained a wider audience sooner than that, through the industry of the press magnate and encyclopedist Charles-Joseph Panckcouke. As part of his huge *Encylopédie méthodique*, Panckcouke published three volumes on mathematics, concluding the third volume with a separately paginated and bound *Dictionnaire des jeux* (*Dictionary of Games*) of over 300 pages, which appeared in 1792 [67]. Its entry for Trente et Quarante is essentially a copy of Huyn's 15 pages on the game, except that every sentence is slightly paraphrased. No one was looking over Panckcouke's shoulder in 1792, the royal censor having collapsed.[11]

The following year, 1793, the English compiler John Debrett published a short book entitled *Faro and Rouge et Noir*. Faro (Pharaon in French) had long been played in England. Rouge et Noir (Trente et Quarante) was novel in England, or so the author thought, for the reader is told all about its French vocabulary and exactly how the French played it.[12] We also learn how different players try to win the game:

> Some think the only way to win is, to follow the runs, that is, when a colour has won twice or thrice, to take money on it, imagining it is in luck, and going to win many times. Others oppose it, and stake their money on the colour, which lost last. Some wait until a colour has lost several times before they play, and then, if perchance it loses, follow it till it wins; conceiving, because it is sixty-three to one, that one colour does not lose six times following, and one hundred and twenty-seven to one against seven successive events, that one event has some relation to another, and that when a colour has lost six times, the odds are in favour of its winning the seventh: others double their stakes until they win, which is called the martingal game. This is a desperate mode, and unless a punter has a very large capital, and punts at a very rich bank, will probably ruin him. …

This is the earliest instance of *martingal* in its gaming sense that I have found in English. The spelling is notable; it mimics the pronunciation of *martingale* in French. We see this spelling in some later English sources; Charles Babbage used it in 1820, and it was still listed as an alternative to *martingale* in the 1914 edition of Whitney's dictionary.[13]

Debrett's report on the stratagems of players in Trente et Quarante is capped by an explanation of why they cannot succeed:

> Whether one event at play has any relation to the next in succession is a problem, the solution of which is of the last importance to every person who regulates his play on that principle. When a novice considers the great apparent disparity of the odds against a given number

[10] [52, p. 34].
[11] [51, Ch. 1].
[12] Who was the author? Debrett may have paid someone recently arrived from France to draft the book, but because he did not name an author, I will attribute the book's contents to him.
[13] Babbage [8], Whitney [96, Vol. 6, p. 3643].

> of successive events, and one more, and the paucity of many successive similar events, he is induced to infer, that after having lost six times, because the odds were twice as many to one that he did not lose seven, that therefore the chance is now in his favour: but if he unprejudicedly considers the circumstance of the cards being so indiscriminately shuffled, that the cards dealt bear no relation to, and cannot govern those in the pack, he must conclude that it is an even chance that either colour wins next.
>
> Let a person pursue at an even game of chance whatever honest mode of play he can devise, although he may win or lose for a week or a month, a series will always approach equality.

Don't expect to beat the house. We can hardly fault this advice, wisdom born of long experience. But are the arguments given by Huyn and Debrett fully coherent? Everything equalizes in the long run, Huyn says. A series will always approach equality, Debrett says. Yet these very assertions seem to open a path to beating the house. When a series is too far out of balance, bet that it will return to balance. This thought has played a role in many a betting system.

In fact, the odds against a series of a given length being in or close to balance are greater when the length is greater. The anticipated balance may lie far in an imagined future. If you know probability theory, you can begin to explain this using De Moivre's theorem: if Y is the number of red wins in n rounds of Red and Black, the standard deviation of Y's distance from $n/2$ is $\sqrt{n}/2$. So the chance of a gambler who bets on red every time being out a given amount of money grows with $\sqrt{n}$. But this is esoteric theory. Even today, not one gambler in a thousand has a working understanding of De Moivre's theorem. How many did in 1800? Many studied De Moivre's *Doctrine of Chances*, which was all about calculating chances in games, and De Moivre had inserted his theorem in the book's second edition in 1738. But only a few mathematicians could have deduced the theorem's practical implications from De Moivre's statement of it. Even statisticians began to use the theorem widely only in the 1830s, after Fourier and others provided elementary expositions of Laplace's generalization.[14]

3.2 The Business Model

Huyn's emphasis on the equalization of events and on the smallness of the house's advantage obscures the importance of another factor: the persistence of the gambler. This persistence puts a gambler with limited resources in great danger and steadily increases the house's gain even when the gambler has great resources.

In theory, the numbers of reds and blacks will repeatedly equalize, and exactly so, even after great departures from equality in one direction or another. But the wins and losses of a persistent gambler will not keep equalizing, because he or she will run out of money before the house does. If a gambler repeatedly bets the same amount, and the house can risk 1000 times as much as the gambler, then the gambler will run

[14] For De Moivre's theorem as he presented it, see [89, Chap. 4] and De Moivre's own words on p. 239 of the second edition or p. 247 of the third edition of [37]. Fourier [44].

out of money first with probability 99.9%, even if the house has no advantage on each bet.

The certainty of the gambler's ruin was demonstrated mathematically in 1802 by André-Marie Ampère, a mathematician now remembered for his work on electromagnetism. As Ampère observed, many authors had documented how the passion for gambling leads to ruin, but gamblers had paid little attention to this empirical evidence, because they were accustomed to seeing only randomness in the events that lead to ruin. Being a devout Christian, Ampère hoped they would pay more attention to his demonstration that ruin is a mathematical consequence of the laws of chance when a gambler plays persistently. They did not. Even mathematicians paid little attention to Ampère's result at the time. But it was popularized by Joseph Bertrand nearly a century later, and it is now seen as mathematically pathbreaking—the first formulation and serious use of the idea of *almost sure* or probability-one convergence. Ampère showed that when the house is infinitely rich and play can continue indefinitely (a mathematical idealization, to be sure), the persistent gambler is ruined with probability one even if the house has no advantage.[15]

A real casino has only finite capital. So it follows from Ampére's theorem that it needs an advantage in order to avoid its own certain ruin if play continues indefinitely and ever more gamblers bring new money to the table. The 1.1% advantage in Trente et Quarante, as small as it may appear, may do very nicely, because it takes 1.1% of each bet, not merely 1.1% of the money the gambler takes out of his pocket to put on the table. A dealer might deal 200 rounds in a day.[16] On average, a gambler who makes a one-unit bet on each round will need to take only about 12 units out of their pocket but will lose about 18% of it (1.1% of 200 units, or about 2.2 units). This percentage goes up steadily if the gambler persists for many days. An extreme example: a gambler who bets one unit 200 times a day for 180 days takes about 430 units out their pocket on average and loses on average 396 units—more than 90%.[17]

Some of casino's take being devoted to lavishly entertaining its patrons, the house also profits when they move a little faster along the road to ruin. They do this when they make ever larger bets, such as those suggested by martingales. We will not be surprised when we see the house advertising martingales.

[15] Ampère [3]; Bertrand [16, Chap. VI]. At first, Ampère thought that he had shown that the gambler would be ruined with probability one even if he had an advantage over the house, but Laplace showed him that this was an error; see [23, Vol. 1, §3].

[16] One 312-card deck provides 25 to 32 rounds [43, p. 634]. In the gaming spas of the 1790s, a single séance would last a couple of hours and use two to four decks, and there would be two or three séances a day [86, pp. 58–59].

[17] These calculations assume that the gambler has enough capital that they are not risking ruin by persisting for the 200 rounds or the 200×180 rounds.

3.3 The Paris Casinos

The fully legal and stable environment of the gaming spas made them one source of fashion and authority for late 18th-century France, where gambling was always popular but never fully legal. Later authors tell us that Aix-la-Chapelle and Spa had been the cradle of Trente et Quarante, and that the game had been the rage in France since the 1780s.[18] The French revolution only strengthened France's interaction with the spas, first as the French military repeatedly occupied them and then as their principalities were incorporated into the French republic.

The French royal court was another source of fashion, for the monarchy encouraged gambling among the nobility as a way of keeping them out of rebellious mischief. The monarchy also authorized lotteries; a national lottery was established in 1776. But opposition in the *parlements* and by other opinion leaders continued to prevent official legalization of public gaming in France. By the end of the *ancien régime*, the police had to regulate gambling on their own, licensing some venues, ignoring those where nobles and officials were prominent participants, and trying to suppress others. Criticism of this toleration of gambling became a way of expressing opposition to the regime.

Indoor tennis courts, or *jeux de paume*, provided one venue for gambling by the less privileged classes. Tennis had been very popular during the turbulent 17th century, when leaders of men needed to exhibit their physical prowess. In more peaceful times, as tennis's popularity declined, the *jeux de paume* began to offer billiards, public spectacles, and betting.

The revolution was no more successful in suppressing gambling than the monarchy had been, and by the end of the 1790s, Paris had acquired casinos of the same elegance as those of the gaming spas. In 1806, when his empire included the gaming spas, Napoleon put them and the Paris casinos under a common licensing regime. Unlicensed gambling dens persisted, of course; they were commonly called *tripots*, this being an ancient name for the *jeux de paume*. Not being as well capitalized as the casinos, the tripots relied more often on prostitution or rougher methods.[19]

4 Gamblers' Fallacies

Alongside the martingale, we learn about the paroli from Huyn and Debrett. Here the player doubles not after losing, but after winning. Parolis are an integral part of Faro, and players used them in Trente et Quarante as well. The word *paroli* came into French from Neapolitan Italian; American gamblers turned it into *parlay*.

Opposite fallacies are at work here; a player who has won by betting on black thinks black is hot and plays a paroli, while a player who has lost by betting on red

---

[18] G. N. Bertrand wrote in 1798 that the game had been in fashion for twenty years [15, p. vi]. *Spéculateur* wrote in 1809 that Aix-la-Chapelle and Spa were the game's *berceaux* [87, p. 3].

[19] [12, 13, 48, 55].

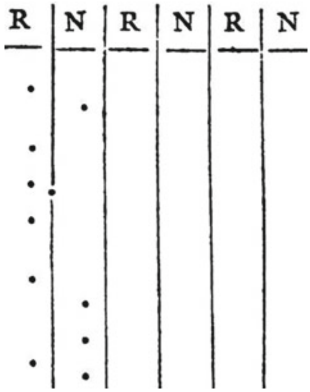


Fig. 3 Fragment of a slip with pinpricks showing successive outcomes in Trente et Quarante, from Debrett's *Faro and Rouge et Noir* [5, p. 24]. Each pinprick represents a win for red (R for *rouge*), a win for black (N for *noir*), or a 31 to 31 tie (on the line). *Source* Google Books

thinks red is still overdue and plays a martingale. Debrett tells us that players do both. Contrary to the way *gambler's fallacy* is used in today's scholarly literature, gamblers have many fallacies.[20]

Debrett explained that the casinos provided their customers with slips of paper printed with columns, as in Fig. 3. You can track successive wins by red (R) and black (N) by punching successive holes in the columns with a thick pin, punching the hole on the line for a 31 to 31 tie. Thus equipped, you can take preceding outcomes into account however you like. By the early 19th century, gamblers had assembled punched slips from all the public casinos (though de Birague professed doubt that the pin was known at Ems), so they could be used to devise and test betting systems. Eventually such assemblages were published in gambling manuals along with advice on systems. One author included 25,000 pinpricks in a book in 1809; another included about 50,000 in 1863.[21]

There were plenty of betting systems and plenty of entrepreneurs eager to sell them. The avalanche of printed material produced by the disappearance of the censor and the proliferation of printing presses in revolutionary France included many works marketed to novice gamblers. Some invoked magic and astrology rather than arithmetic.[22] No doubt many have disappeared. But many have survived. The Edict of Montpelier, promulgated by Francis I in 1537, mandated that all books printed in

[20] For a list of gamblers' fallacies published in 1812, see [29, pp. 65–67]. Martin Gall argued in 1882 that the paroli fallacy came first, the martingale fallacy then following when gamblers were disappointed by their parolis [46, p. 241]. The earliest use of the term *gambler's fallacy* I have seen is in Adam Leroy Jones's *Logic* in 1909; Jones reports that the name is given to the fallacy that "because a coin has come up heads several times in succession, it is therefore more likely to come up tails the next time" [53, p. 223].

[21] Debrett [5, pp. 73–74]; de Birague [19, p. 30]; 25,000 pinpricks in [87, pp. 227–288]; 50,000 in [47, pp. 148–260].

[22] See [6], published in 1801, to learn how to win in the lottery by interpreting your dreams and the mysteries of numbers, etc.

France be deposited in the Royal Library, and many manuals for gamblers are still to be found at the Bibliothèque nationale de France.[23]

A look at a few of these books will help us understand the 19th century's betting systems and martingales. We will organize our exploration around a handful of authors who wrote near the beginning of the century: two moralists, a rogue, a mathematician, and a gambler. The moralists, G. N. Bertrand and Alexandre-Toussaint de Gaigne, manage to tell us how to play martingales while ostensibly telling us not to. The rogue, identifying himself only by the initials H.E.B.C., sells a system that can be seen as transparently fraudulent by anyone who knows a bit of probability theory. The mathematician, Sébastien Antoine Parisot, is earnest and even competent, but can confuse confusion with insight. The gambler, Jacques-Joseph Boreux a.k.a. James Smyll, is equally earnest and more colorful: thwarted cavalryman, inventive craftsman, skilled draftsman, designer, and wide-ranging intellectual.

4.1 Two Moralists

G. N. Bertrand and Alexandre-Toussaint de Gaigne together show us the world of Trente et Quarante at the very end of the 18th century. We know nothing about Bertrand beyond his name on his book, *Trente-un dévoilé*. Self-published in France in 1798, it was a collection of poems, anecdotes and sermons about the evils of the game. De Gaigne, on the other hand, was a well known figure: former officer in the infantry, royal censor, author of a military dictionary, an 18-volume collection of poems, and other works. In his *Mon histoire au Trente-un*, published in London in 1799, a cheerful female narrator recounts losing her husband's money in every popular way at Trente et Quarante in a Paris casino.[24]

In these two books we see already a variety of betting systems that were marketed throughout the 19th century and into the 20th. In particular, we see two popular and representative systems; Bertrand describes the d'Alembert, as it is now called, and de Gaigne describes *le tiers and le tout*.[25]

- The d'Alembert is a toned-down martingale. You start by making a one-unit bet. Every time you win with a one-unit bet, you make a one-unit bet again. Every time you win with a larger bet, you make your next bet one unit smaller. Every time you lose, you make your next bet one unit larger. If you stop when the number of losses, counting from the first loss, is equal to the number of wins since that first loss, then your net gain will equal that number of wins; see the example in Table 1. Here the fly in the ointment is the possibility that you will run out of money before the number of wins catches up with the number of losses.

[23] Depaulis has cataloged books on the rules of card games published before 1850 [38].
[24] [15,45]. De Gaigne was also known as Alexis-Toussaint de Gagne.
[25] [45, Chap. XI].

- *Le tiers et le tout* (the third and the whole) is a toned-down paroli. You begin by putting three units on the table. Then you bet one unit, which is one third of what you have on the table. Whenever you win a bet, you leave your bet and your winnings all on the table and bet one-third of the total. Whenever you lose a bet, you bet what you still have on the table, and if you lose it also, you take another three units out of your pocket and start over. You continue until you have won six times, say, in a row. The six straight wins produce a gain of $3 \times (4/3)^6 - 3 \approx 14$ units. In the meantime, you have lost 3 units every time you lost twice in row. So you hope that this happened only four times or fewer.

On the surface, both books tells us emphatically and repeatedly not to gamble. Bertrand preaches directly. De Gaigne's narrator mocks her fellow gamblers and herself for their foolishness, occasionally admonishing the reader not to follow her example; the best *marche*, she tells us, is the *marche rétrograde*: march right back out of the casino. Yet the reader senses the attraction both authors feel to the game and to the promise offered by their betting systems.[26]

Bertrand's mixed feelings about the d'Alembert are summed up in this sentence:

> …today this *sublime discovery* is considered the *nec plus ultra* of Trente-un; and I attest that it is a crime to be its apostle, just as it is a crime to be the apostle of any system.[27]

The d'Alembert may be the most popular betting system of all time.[28] Could it be Delval's martingale of Spa? We will learn more about its seductiveness from Jacques-Joseph Boreux.

Bertrand gave the d'Alembert an awkward name: *la progression croissante en perte, et décrosissante en gain* (progression increasing in loss, decreasing in gain). Later authors simplified this to *la montante et descendante* (the increasing and decreasing). The legend that d'Alembert had invented the system was already known to Bertrand, who properly called it a calumny. But in 1882 the influential Martin Gall stated the legend as fact. In 1892, Norwood Young reported that the system was known at Monte Carlo simply as *the d'Alembert*, and this name is still used today.[29]

Like Bertrand, de Gaigne's protagonist tells us about many ways of playing. She knows twenty or thirty martingales. But she concludes her narration with *le tiers et le tout*, showing the same enthusiasm for it as she has shown for her earlier errors, and not yet being disabused. In the mould of past and future sellers of systems, she suggests that this may be the one to bankrupt the casinos.

[26] [51, Chap. 6].
[27] [15, p. 63].
[28] There are many variations, of course. In 1821 [8], Charles Babbage considered the generalization in which the amount by which the bet is increased or decreased is a constant perhaps different from 1. Consistent with his focus on computation, Babbage showed that an appropriate choice of notation allows an easy calculation of the capital after p wins and q losses [9]. This received critical notice from the young Antoine Augustin Cournot [32, vol 1, pp. 304–307].
[29] [46, pp. 243–244], [97, p. 454], [43, pp. 289–292].

The 19th century brought a plethora of betting systems that Bertrand and de Gaigne had not yet seen. But these two authors had already painted the general picture. "Martingale" had a double meaning. On the one hand, martingaling means doubling your bet; on the other hand, any system is a martingale. As Bertrand wrote,

> It means betting all you have lost, the quickest way to bankrupt yourself. It means imposing on yourself any rule whatever that makes the outcome depend on chance.[30]

Moreover, these authors' favorite betting systems shared a structure that persisted. Their essential element was a rule for determining the amount to bet next based only on previous wins and loses, ignoring whether the previous bets had been on red or on black. Some later authors called this the *massage*—the determination of the *masses* to bet. The *massage* determined the mathematical properties of the system. The gambler added to it an *attaque* or *marche*, which gave guidance each time on whether to place the bet on red or on black and when to make it—immediately or perhaps after some color or sequence becomes hot or overdue.[31]

4.2 The Blatant Rogue

In 1803, a *Manuel des jeux de hasard* appeared in three Paris bookstores, its author identified only by his initials, H.E.B.C. It described the usual casino games, including Trente et Quarante and Roulette, as well as the national lottery, which had been discontinued by the revolutionary government only from 1793 to 1797. The subtitle also promised to give readers systems, based on calculation and repeated observation, for playing these games to their advantage.[32]

Any book that purports to tell the reader how to win at the casino (or how to win in the stock market, for that matter) faces a simple question: Why are you telling us instead of getting rich yourself? H.E.B.C. answers that he writes to save his fellow citizens from gambling's perils. The first third of the book is devoted to decrying the evils of gambling; he wants the first consul (Napoleon) to severely repress the casinos. Some people will never manage to resist the lure of gambling, but perhaps his advice will save a few from utter ruin. This advice included maxims that sellers of martingales echoed throughout the 19th century: limit your martingale, settle for small gains, risk a little to gain a lot, be prudent, keep cool, persevere.

H.E.B.C.'s Roulette is the modern game, not the older game described by Huyn in 1788. The older Roulette was merely a faster and often less honest game of Red and Black. The new Roulette was a Parisian innovation of the late 1790s: the red and black pockets of the wheel were numbered, and the game was married with Biribi, a game that had been played with numbers drawn from a bag. Surely this had been

---

[30] [15, p. 41].

[31] Martin Gall explained the distinction between *attaque* and *massage* in 1882 [46, p. 260].

[32] [49]. The book bears no date, but it was advertised as a new book in early 1803 in several Paris newspapers, beginning on January 19 in the biweekly *Journal Typographique et Bibliographique*.

tried before, but the Parisians perfected it.[33] The new Roulette was spectacularly successful, quickly spreading across Europe. It dazzled by its speed, the opportunity to bet while the wheel is still in motion, and the variety of possible bets. The casinos loved it too. A gambler might make 600 bets in a day.[34] The house's advantage being 1/19 (two zeros out of 38 pockets), 600 simple bets on red or black require the gambler to take about 41 units out his pocket on average, of which the house keeps 32 on average. Trente et Quarante remained popular because of its more reasonable advantage for the house, but now it too was made a little less monotonous; in addition to betting on red or black, players could bet whether the first card dealt by the dealer would match the winning color.

For Trente et Quarante, H.E.B.C. offered two betting systems: small martingales and small parolis. His small-martingale system goes this way:

- Wait until you see two reds or two blacks in a row. Then bet one unit on the opposite color.
- If you win, stop. If you lose, then double your bet—i.e., bet two units on the same color. No matter how this second bet comes out, stop.
- Resume betting when you have again seen two reds or two blacks in a row.
- Quit when you have gained 4 units.

H.B.E.C. points out, with enough words to distract, that each time you start betting, you lose only if the outcome goes against you twice in a row. The probability of this is only $1/4$. So you win three times as often as you lose. He neglects to mention that each loss is three units—one on the first bet and two on the second—whereas each gain is only one unit.

The small-paroli system goes this way:

- Wait until series (several reds or several blacks in a row) are more frequent than alternations.
- Then bet on the color that has just won. If you win, play a paroli—i.e., bet both the money you first bet and the money you won on the same color winning again. If you win again, perhaps risk another paroli.
- Quit as soon as you have won a few units this way.

H.E.B.C. considered his small-paroli system riskier than his small-martingale system, because of the choices that the gambler must make. But perhaps it was appropriate for the bolder gambler.

---

33 The earliest surviving detailed description of the new Roulette was given by de Gaigne's narrator in 1799 [45, p. 70]; see also [11,39,40]. The oft repeated conjecture that Blaise Pascal invented Roulette has no basis in fact; it stems no doubt from Pascal's having studied, under the pseudonym Amos Dettonville, a mathematical curve that he called *la roulette*, the curve traced out by a point on a wheel as it rolls on a flat pavement [30].

34 [20, p. 180].

After explaining these systems and additional systems for Roulette and Pass-dix, H.E.B.C. acknowledges that individual gamblers often do not have deep enough pockets to make them work. So he proposes the formation of a counter-bank, which would finance such gamblers and put the casinos out of business once and for all.

H.E.B.C.'s facile fraud reminds us that most gamblers are not mathematicians. Sound arithmetic is not needed to sell them books. Many of the betting systems sold to those who frequented 19th century casinos, whether in books, in pamphlets, or merely on slips of paper sold at the entrance, must have been equally flimsy.

Here is a much later example: a tourist guide for casino goers published in 1875 under the name Guide Sextius. France outlawed its casinos in 1836; Britain in 1845, and Belgium and Germany in 1872. Guide Sextius told the French how to visit the closest remaining casinos, in Monaco, Switzerland, and Spain. It described the venues and the two principal games, Roulette and Trente et Quarante. Its last 15 pages patiently explained, with tables and arithmetic and only one hidden deliberate error, why a player who repeats the same simple bet for 10 days will come away with large winnings. Did the casinos subsidize the Guide Sextius? We do not know, but in 1892, Norwood Young informed his fellow Britons that the Monte Carlo casinos published betting systems to help novices get started.[35]

4.3 The Failed Mathematician

Bernard Bru has called Sébastien Antoine Parisot the great theoretician of the maturity of chances. We know little about his life. On the best evidence we have, he was born in 1761, held a minor administrative post, and died in 1812.[36] His two books, which appeared in 1801 and 1810, were not written for gamblers or would-be gamblers. They were books of mathematics, meant to win him a post as a mathematician, perhaps, he hints, at the Collège de France.

The title of Parisot's first book, *L'art de conjecturer à la loterie*, evoked Jacob Bernoulli's celebrated *Ars conjectandi*. The book studied combinations and permutations and applied his results to the French lottery, where 5 numbers are drawn at random from the numbers from 1 to 90. Then the book went off the rails. Parisot fancied that the probability of drawing a particular number increases the longer it fails to appear, and he undertook to calculate this probability. It exceeds $1/2$, he concluded, on the 13th draw. After that, one might advantageously play a martingale.[37]

Parisot considered martingales that aim only to recover money lost and martingales that aim for a net gain. For bets at even odds, the two types go this way:

[35] Guide Sextius [81], Young [97].

[36] These guesses, provided by Bernard Bru, are based on official records reconstructed after a fire and on genealogical sources.

[37] The probability of getting the number at least once in 13 drawings is $1 - (85/90)^{13} \approx 52\%$ [68, p. 101]. For the history of the French lottery, see [90].

- To recover our loss, we bet everything we have lost so far.[38] So long as we are losing, this stratagem produces the sequence of bets 1, 1, 2, 4, 8, etc.
- To recover our loss and also win what we had initially sought, we double each time the previous bet.[39] This produces the sequence 1, 2, 4, 8, etc.

Parisot generalized this picture, allowing the odds to be other than even (he was out to beat the lottery) and allowing the desired gain to be different from the initial bet.[40]

The 1801 book gained little notice, but Parisot stubbornly returned to work, producing in 1810 his *Traité du calcul conjectural*. He still used only combinatorics, but now he offered applications to commerce, insurance, and even metaphysics. And he extended his erroneous calculations on lotteries to casino games, explaining in both cases how to find "the number of drawings one should allow to pass between two appearances of the same outcome in order to attack it with advantage and at its highest point of maturity".[41] He wrote boldly, not hesitating to attribute errors to Fermat, Bernoulli, and Montmort, and he submitted his work to the Académie des Sciences. Jean-Baptiste Biot gave a verbal report on the book to the Académie on December 26, 1809. The report could not have been positive. Parisot had no future as a *savant*, and we are left wondering about how his disappointment played out and about the circumstances of his death so soon afterward.[42]

Biot having seen Parisot's second book, we can be confident that his colleague Laplace, a quick and omnivorous reader, was also aware of it. In his lecture on probability at the *École Normale* in 1795, Laplace had mentioned people's errors, illusions, and contradictions when estimating probabilities. They believe erroneously in streaks of good and bad luck, and at the same time they think that a number is more likely to appear in the lottery if it has been absent for a while. They similarly think that heads is more likely to appear in a sequence of coin flips if it has been absent for a while, contrary to Laplace's Bayesian theory, which discerns in the absence of heads a possible bias towards tails. In 1814, when Laplace first expanded this lecture into his *Essai philosophique* on probability, he added a comment on the futility of martingales (without using the word *martingale*) and wrote that he had personally confirmed that the anomalies in the outcomes of the lottery were within limits that could be expected from randomness.[43]

---

[38] This is the basic definition of *martingale* given by the abbé Prévost in 1750 and by the dictionary of the Académie Française in 1762 and 1798 (4th and 5th editions). See Mansuy's chapter in this volume.

[39] This is the basic definition of *martingale* given by the dictionary of the Académie Française in 1835 (6th edition).

[40] [69, pp. 105ff].

[41] [69, p. 309].

[42] A few scholars, including Adolphe Quetelet and the German jurist Johann Heinrich Bender [14, pp. 30–32], cited Parisot in the 19th century, but they seem to have done so carelessly, not taking the time or lacking the training to read his books.

[43] Laplace in 1795 [59, pp. 159–160], Laplace in 1814 [58, pp. 78–79].

Laplace's example notwithstanding, 19th century mathematicians did not waste much time refuting martingalers. The Marquis de Condorcet and Sylvestre-François Lacroix did mention martingales, by name, and their futility—Condorcet before his death in 1795 and Lacroix in 1816. Condorcet's comment was about martingales in even-money bets in the casino; Lacroix was writing about the lottery. But they did not mention systems such as the d'Alembert and *le tiers et le tout*, and they did not pause over the fallacy of the maturity of chances.[44]

After Parisot, we do not find other authors trying to gain standing as mathematicians by developing a general theory of the maturity of chances. Some sellers of betting systems made arguments like Parisot's but without Parisot's fatal clarity and generality. Most relied instead on claims that their systems had been validated by experience. Some claimed that experience showed theory to be inaccurate. Others claimed their methods were supported by theory but emphasized the validation by experience. Even Parisot had believed that observation could provide more reliable limits on maturity than those given by his mathematics.[45]

4.3.1 Emphasizing experience

A milestone had already been marked by the publication in 1809 of an ambitious guide to Trente et Quarante by an author who signed himself only *un Spéculateur*. This is the book with 25,000 pinpricks. It labeled itself a second edition. Its lengthy title, which I will abbreviate to *Essai sur le trente-un*, promised to give the reader the means to play with advantage. We do not find it in the Bibliothèque nationale de France, but it appears to have been popular. It attracted a two-page review in the *Journal Général de la Littérature de France* and a listing in 1811 in the *Edinburgh Review*. François Corbaux junior complained that it was practically the gambler's catechism.[46]

We can see why *Essai sur le trente-un* would be popular. It was sober, seemingly competent, and not dogmatic. *Spéculateur* seemed to be a master of probability theory. He tabulated the chances for short sequences of reds and blacks (*figures*) and calculated chances for Trente et Quarante based on approximations that he declared fit for purpose. He refuted common errors, including Parisot's (though he did not name him). He dismissed H.E.B.C. as an ignoramus and de Gaigne's novel as an amusement (also not calling them by name). He was not really sure that any system would work, and he joked that no system can be bad—otherwise you could make money for sure by doing the opposite. But he had recorded the outcomes of about 1,560,000 rounds of play, and seeing that this data did not completely agree with his own probability theory, he had concluded that it is possible to beat the house in a number of different ways. He found that *le tiers et le tout* worked better than

[44] Condorcet posthumously in 1805 [26, pp. 119–120], Lacroix [57, p. 110].

[45] Parisot [69, pp. 311–312].

[46] *Essai sur le trente-un* [87]. *Journal Général* in 1809 [7, pp. 252–253]. The *Edinburgh Review* listing, in its quarterly list of new publications from February to May 1811 in volume 18, p. 172, gives a price of 10 shillings. Corbaux [29, p. 83].

the d'Alembert, which he called *la montante and descendante* while acknowledging the attribution to d'Alembert and the argument from equilibrium. But he seemed to think that any *massage* would work with the right *attaque*. Rather than applying the *massage* to red and black, apply it to some more complicated figure that has not been happening often enough, say two reds in a row or two blacks in a row. And diversify: work with a confederate who bets on a different figure; when the plan says you are to bet on red and your confederate on black, or vice versa, only the one intending to make the larger bet really bets, after deducting from his planned bet the smaller planned opposite bet. Thus you are less likely to run out of money and less likely to lose a lot to the house on 31–31 ties.

François Corbaux, who had described himself as a trader (*négociant*) in an earlier book on exchange rates, published his *Essais métaphysiques et mathématiques sur le hasard* in 1812. As the title suggests, he drew metaphysical conclusions from observing Trente et Quarante:

> The more one repeats these observations, the more occasion one has to perceive that chance is an instrument in nature's hand; obeying a really existing force unique to it, this instrument produces diverse predetermined results according to immutable laws, from which chance can never depart more than momentarily, then soon returning to limits prescribed for it, giving way to a force of reaction equal to the force of the action that produced its deviations.

But he also criticized both Parisot and *Spéculateur* for their fallacies. He contended that if there really exists a betting system that gives the gambler an advantage, it must be based on the fact that probability theory makes only approximate predictions. You would have to take advantage of the random deviations in some way, and this would require new ideas and more observation. He presented his book as the first volume of a larger work, which might eventually tell the reader how to bet to win. He was ready to print the second volume, along with a sequence of 131,072 consecutive pinpricks from Trente et Quarante, just as soon as enough readers subscribed to pay the costs of printing. We have no evidence that this happened.[47]

Another Parisian, J. B. Chamois, sold both books and elaborate charts for tracking the outcomes of roulette spins and bets on them. In 1817, Chamois published the first edition of his *L'art du bien-jouer à la Roulette*. He gave tables for the three degrees of maturity Parisot had invented—*mineure*, *moyenne*, and *majeure*.[48] When a particular roulette number, say, has not appeared for the number of spins given by its minor maturity, it might be time to bet on it. But the player must also attend to the earlier record, for if the number had been appearing too often previously, then

[47] Corbaux [29]: the quotation is on p. 174, the plea for subscriptions on p. viii, the explanation of the need for observations on pp. 74–75, the footnotes on Parisot on pp. 54–55 and 72. In the first he praises Parisot; in the second he notices and disagrees with Parisot's argument for the maturity of chances. His book on exchange rates had appeared in 1802 [28]. He later established himself, at least temporarily, as a gentleman in England, where he promoted a scheme for insuring middle-class newly-weds against the hazard of having too many children to support [27].

[48] Compare, for example, Chamois's table for the three degrees of maturity for Roulette [24, 4th edition, p. 40] with Parisot's table for Roulette [69, p. 226].

its absence might only be a compensation, and then perhaps the player should wait for the major maturity. Once the player has begun his martingale, how long should he persist? We get tables for this as well. In contrast with Corbaux, Chamois knew his market. He gave his readers clear instructions, while leaving room for them to exercise their sense of how the game is going. He published at least three subsequent editions of his book, in 1818, 1823, and 1828. In the 1823 edition, he introduced a geometric version of the d'Alembert: you double your bet when you lose, halve it when you win. He called this the *martingale graduelle*, perhaps echoing Jacques-Joseph Boreux's name for the d'Alembert, *martingale graduée* (see Sect. 4.4 below).

Chamois reports that he had joined the army at the outset of the French revolution, and that he had begun his study of games of chance watching idle soldiers gamble. Over the twenty succeeding years, he had observed 100,000 rounds of play. This experience confirmed, he wrote, limits on maturity given by calculation:

> The limits that we have fixed …concord perfectly with calculation; but experience has proven to us more particularly that play based on the points of maturity we have determined are, without contradiction, superior to any sort of combination, and that if it is possible that an extreme delay should occur, the resulting loss would not be large enough to risk the player's purse …[49]

He presented no calculations.

Chamois's book probably remained in print, at least in pirated form, for many decades after the 1820s. In 1866, Van Kornicker's bookstore in Antwerp announced the 18th edition of a book with a lengthy title very similar to those Chamois had used.[50] Many others also sought to find the limits of maturity in ever more extensive observation. In 1863, Grégoire recalled the pioneering work of *Spéculateur* but argued that it had been based on too few observations. With yet more observations, he found periods for the different figures, the more complicated ones taking longer to return to equilibrium. Thus he could propose cleverer ways to diversify, putting together *massages* with differing periods that balanced each other just right. At about

[49] [24, 4th edition, p. 37].

[50] The announcement appeared in the *Feuilleton du Journal Général de l'Imprimerie et de la Librarie*, 55^{e} Année, 2^{e} Série, N^{o} 32, 11 Août 1866, p. 696. The title was *L'art de bien jouer à la Roulette, Indiquant la maturité, la limite des chances et les règles pour les attaquer sur-le-champ et avec succès, avec des tables de mises et des nouvelles cartes à marquer, suivi d'un aperçu sur la meilleure manière de jouer le Trente-un*. It was said to have been written by a former employee of 113, a Palais-Royal casino notorious under the second empire. Chamois varied his full title from edition to edition, but for the first edition it was *L'art du bien-jouer à la Roulette, ou principes et analyse des chances, de leurs périodes, retards et retours; indiquant la maturité et la limite de chacune en particulier; accompagnées des instructions et de règles pour les attaquer sur-le-champ et avec succès, suivi de tables de mises appropriées à chaque nature de chance, terminées par de nouvelles cartes à marquer, au moyen desquelles on peut, d'un coup-d'oiel, connaître et juger la situation de jeu sur toutes les faces de combinaisons*.

the same time, J. Jouet de Lanciduais was making his own exhausting observations in his search for the *equilibre corrélatif absolu*.[51]

4.3.2 Parisot's Fallacy, Broadly Defined

Parisot calculated the small probability that a single number, color, or figure will be absent for some lengthy period and then imagined that this prolonged absence still has the same small probability when the period is almost over and the absence has been verified so far. This error (or fallacy or sophistry, if you please) can equally well be applied to any event that involves many rounds of a game and has a small probability at the outset. Authors after Parisot found many such events.[52]

One author, who published in 1829 and signed himself only *un Amateur*, used Parisot's fallacy to design *attaques* for a raft of popular *massages*. Suppose, for example, that you are willing to play a large martingale to recover a loss, doubling up to 11 times, usually the most possible within the minimum and maximum bets allowed by a casino. How should you attack? *Amateur* suggests recording the reds and blacks for 19 rounds, then waiting until the first 8 of these rounds is repeated exactly. Once you see this, you play the martingale against the next 11 being repeated exactly as well. You will surely win, as the odds against a particular sequence of 19 being repeated exactly are astronomical.[53]

Parisot's and *Amateur*'s prescriptions could be cumbersome. There was a lot to track. In 1862, a former Paris croupier proposed a much simpler technology. Take some slips of pinpricks from previous rounds, and use them to define the *attaque* for a martingale that doubles up to 9 times. On each round, bet against the color that won on the corresponding round in the sequence of pinpricks. You will win unless the new sequence of outcomes exactly replicates the sequence represented by the pinpricks, and this has a very small probability. The author had worked at Frascati, Paris's most elegant casino during the years before 1836, when France's legal casinos were shut down. He recommended the popular German gaming spas of the 1860s: Homburg, Wiesbaden, Baden, Nauheim, and Ems. His book was entitled *La Californie germanique*: if lose your fortune foolishly betting in the stock market, don't go to California to pan for gold; find your gold in Germany.[54]

4.3.3 Maturity Debunked and Compensation Contested

In the latter half of the 19th century, *maturity of chances* became a term of criticism and mockery. The mockery began in French. Charles de Birague may have

---

[51] [47,54].

[52] As did authors and gamblers long before Parisot. We saw Parisot's fallacy being debated in Debrett's *Faro and Rouge et Noir*. We could also cite the St. Petersburg paradox and d'Alembert's doubts, topics with literatures too vast to tackle here. It is sometimes said that d'Alembert committed the gambler's fallacy, but it is fairer to say that he questioned the sufficiency of a priori arguments for counting chances [23, Vol. 2, p. 337].

[53] [2, esp. pp. 19–20].

[54] [4].

launched it in 1862, but it was performed most effectively by the famous French magician Jean-Eugène Robert-Houdin in 1863, in his *Les Tricheries des grecs dévoilées*. Robert-Houdin drew a memorable portrait of the gambler Raymond, who first appears explaining the maturity of chances and proclaiming all the clichés of the martingaler, then reappears some years later destitute. In 1870, in the second volume of his history of gambling, Andrew Steinmetz retold Robert-Houdin's account of Raymond in English. Richard Proctor, prolific and famed on both sides of the Atlantic as an expositor of science, subsequently mocked the maturity of chances in an 1872 magazine article and then in a book on gambling. Robert-Houdin's book was translated into English in 1882 as *Card Sharping Exposed*.[55]

The phrase *maturité des chances* became less common in French after Robert-Houdin. But the debate between martingalers and their critics continued, using terms such as *compensation* and *retour à l'équilibre* (return to equilibrium). *Compensation* had been one of *Spéculateur*'s favorite words. Chamois had insisted that compensation and equilibrium are not futile theories.[56] Sometimes the critics seemed to agree; as we have seen, Huyn and de Birague had argued that systems are futile precisely because the gambler's wins and loses will seek equilibrium. Lacroix and Antoine Augustin Cournot, mathematicians who were among the critics, themselves used *compensation* in discussions of the law of large numbers, and sometimes their unguarded statements could sound like support for the martingalers' arguments. In his 1816 textbook, for example, Lacroix stated that players' gains compensate their losses:

> …This compensation can never be exact, but multiple trials tend incessantly to produce it…. People that constantly play social games with each other, when their abilities are about equal and the terms of the bets are equal, see their losses and gains come close to each other in the long run…[57]

In 1882, the celebrated French chess champion Jules Arnous de Rivière published a comprehensive but playful treatise on games of chance under the pseudonym Martin Gall. Martin Gall was no one's fool. He had all the advice from professional mathematicians he might want, and he included a good deal of probability theory and calculation in his 338 pages. He emphasized how long it sometimes takes chances to equalize, packaging this insight as an aphorism: equilibrium is a conjecture; deviation is a certainty. He mocked the system-builders of his day, including Grégoire, Jouet de Lanciduais, and many an anonymous author. But then he playfully turned his insight about compensation on its head. If red and black are certain to be out of balance, then let's bet on their being out of balance. This led to more calculations. Like de Gaigne and Corbaux before him, he teased the reader, explaining why every proposal for beating the house is faulty but holding out hope, to the very end of the book, that something else might work. All these authors were out to sell books, after

---

[55] Robert-Houdin [78, Chap. VI], Steinmetz [88, vol. II, pp. 253–259], Proctor [75,76].
[56] [24, 4th edition, p. i].
[57] [57, §65, p. 105].

all, and if you want to sell a book about casino games you had best hold out some hope that the reader will learn how to win from it. Even de Birague had slipped the phrase *vrai système* (true system) into the title of his book and given it color with a tongue-in-cheek suggestion, at the very end, that you might wait to bet until some figure was far, far out of equilibrium.

The debate continued into the 20th century. In 1929, the British politician and gentleman gambler Sir Philip Richardson explained the futility of taking advantage of the return to equilibrium using simple arithmetic. In 1936, Marcel Boll was still refuting betting systems based on compensation. In 1939, Émile Borel was still writing about the vast fluctuations possible for martingales and *l'illusion du retour à l'équilibre*, in the middle of the half-century-long debate with Félix Le Dantec that is chronicled by Marie-France Bru and Bernard Bru in their chapter in this volume.[58]

4.3.4 Perpetual Motion Machine?

In 1843, the mathematician and philosopher Antoine Augustin Cournot compared betting systems to perpetual motion machines. Just as the laws of physics tell us that such a machine is impossible, the laws of probability tell us that beating the house is impossible. If a game is unequal, Cournot explained, no method of playing can destroy the unequal conditions of the opposing players.[59] The comparison with physics may elevate probability theory, but it also reminds us that it is a theory. A theory's predictions can be wrong. There will always be believers in perpetual motion machines. There will always be gamblers confident that they can beat the house.

The sense in which probability theory predicts the failure of betting systems can be explained using inequalities published by Andrei Markov in 1900 and by Jean Ville in 1939. Suppose a gambler engages in a game of pure chance in which the house has no advantage. He begins with a certain capital and cannot risk more. He follows a betting system that tells how he will bet on each round, depending only on previous outcomes, and when he will stop, so that his final capital depends only on chance. Then his final capital is a nonnegative random variable with expected value equal to his initial capital. Markov's inequality says that the probability is at most one-half that the final capital is twice or more the initial capital, at most one in ten that it is ten times as large, at most one in a hundred that it is one hundred times as large, and so on. Ville's inequality strengthens this statement: the probability is less than one in a hundred that the amount the gambler has in hand in the course of play will ever be more than one hundred times his initial capital, no matter how long he plays.[60]

But the gambler already knew that the way to multiply your money by a lot is to bet on an event of small probability. So where is the prediction? To make a given small probability into a prediction, we need to adopt the principle that an event of small probability selected in advance will not happen. Cournot and Markov and Ville

[58] Richardson [77], Boll [20], Borel [21, §26].
[59] [31, §62].
[60] [63], [94].

adopted this principle, but they might just as well have adopted directly the principle that a betting system, chosen in advance and followed, will not multiply your money a lot, saving themselves much mathematics and calculation.[61]

Perhaps this is why the debunking of betting systems, as in the books by de Birague in 1862, Martin Gall in 1882, and Hiram Maxim in 1904, can feel so repetitive and empty. After they tell us that probability, in its role as an empirical theory, predicts that a betting system will not succeed, what more do these authors really have to say?[62]

4.4 The Many-Talented Gambler

In 1820, Hinrichs's bookstore in Leipzig commissioned the printing of two volumes in French, a book on gambling entitled *Tactique des jeux*, and its *Atlas*, containing 40 tables illustrating the author's calculations and 16 color plates depicting casinos and games, including the plate shown in Fig. 4. The books described the gaming spas of the 1790s, their ambiance, and their games. The author, signing himself James Smyll, engineer, had gambled in these privileged casinos, he tells us, and had studied the games mathematically. He had known Pierre Nicolas Huyn. He respected Huyn's probity as everyone did, but he did not agree with Huyn that the house's advantage could never be overcome. The casino is like an army that has overwhelming advantages in men and arms but is under orders to stand in place; it can be defeated by nimble guerrillas who attack repeatedly, always biding their time until many probabilities work to their advantage, then withdrawing quickly. Such military analogies pervade the book, and the author bolsters his military credentials by citing Jacob Mauvilon and boasting of his friendship with Jean Noël Bouchotte at Aix-la-Chapelle shortly before Bouchotte became minister of war for the French Republic.[63]

James Smyll was a pseudonym for Jacques-Joseph Boreux, born in 1755 into a family that quarried black marble in the small town of Dinant, now in Belgium just north of France. Like Spa, Dinant was then part of the principality of Liège, under desultory Austrian hegemony within the Holy Roman Empire. Aspiring to a military career, Boreux managed to enlist at the age of 16 and served for a year in a particularly elegant cavalry regiment, distinguished by its African tambour corps. But his father persuaded him to return to the family business. While retaining a fascination with the military, he worked with his father from 1774 until 1792. Together they supervised the quarrying and cutting of marble for churches, mausoleums, and public buildings, providing workmen with designs and models, and supervising the transport of the marble and the construction of alters and other marble works in locales as distant as

[61] [31,63,94]. Cournot had already developed his philosophy of probability in the period from 1829 to 1834 [32, vol 2, pp. 717–721]. Shafer and Vovk use the futility of betting systems as a foundation for mathematical probability [83].
[62] [19,46,64].
[63] [86, pp. 3, 11, 19].

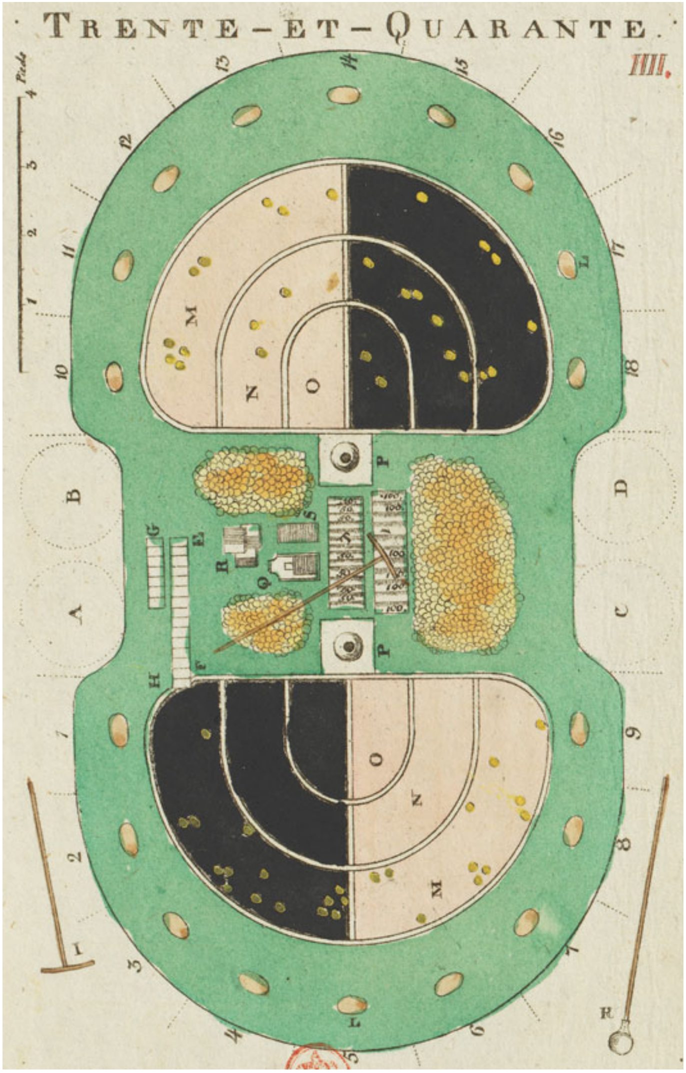


Fig. 4 Boreux's depiction of a table for Trente et Quarante at Spa or Aix-la-Chapelle in the 1790s [85, Plate IIII]. The dealer stands at A. Croupiers stand at B, C, and D. Eighteen players can sit at the table; others stand and use tools like the one shown at K to move their money. Piles and rolls of coins are shown; this is the money the house is risking on the séance. Occasionally this money would be exhausted, enhancing the illusion that a player could break the bank. The casino had plenty of money in reserve for the next séance. We see discarded cards at R, fresh decks at S, and chandeliers P and P. Cards just dealt are in the rows E–F and G–H. *Source* Bibliothèque nationale de France

Ghent to the west and Aix-la-Chapelle and Cologne to the east. Boreux also found time to study painting, design, sculpture, architecture, and even carpentry, skills he used to design marble ornamentations and to construct clay and wooden models to be sculpted in marble. In his spare time, he sought out books that illustrated military fortifications.

Boreux's mother kept the books for the family business. In the early 1780s, when his parents declined, his two sisters took over the books, his younger brother having left home to become a priest. The business and the family's unity suffered. The sisters resented Boreux's heated study and his spending on art supplies; he faulted their spending on clothes and questioned their probity in managing the family's property. After their father's death in 1790, the tension exploded into a dispute over inheritance, and Boreux, always a bachelor, left the family home and business. By 1797 he had left the Pays de Liège, and by 1799 he was established in Leipzig, where he apparently lived and worked for 20 years.

We know these details only from Boreux's own testimony in two hand-written documents giving his side of the family dispute, which he wrote in 1823 in an effort to extract money that he felt his siblings owed him and that he needed to pay off old debts. One was a narrative he threatened to publish; the other was a letter to his brother. The documents were preserved by descendants of the brother's illegitimate son and published in 2011 by the historian Jean-Louis Van Belle. Van Belle was primarily interested in what the documents tell about the exploitation of marble, but he also followed up on clues Boreux provided about his life after the marble business, which had collapsed in the 1790s. The principality of Liège had its own revolution in 1789, triggered by a struggle over the monopoly of gaming at Spa. It then became a theater of the military struggle between France and its neighbors, until the French Republic annexed it in 1795. It would be a long time before marble was again wanted for churches.

We know few details of Boreux's life in Leipzig, but Van Belle's research has revealed that he established a collaboration with Christian Gotthold Eschenbach, a prominent professor of chemistry and medicine, and that he left behind a series of publications, beginning in 1799 and continuing at least until 1820. Some were in German and some in French, France being ascendant by 1800 and even Saxony being under Napoleon's sway from 1806 to 1813. Some of Boreux's publications described inventions (exploding cannon balls, heating and cooking stoves, a vented toilet, even an elaborate mouse trap), some were maps, some were reflections on a variety of topics. James Smyll's book on games of chance seems to be the last of the series.[64]

Gambling is not mentioned in the documents Van Belle published, but the tale they tell coheres with James Smyll's account of himself, leaving us every reason to think that both the documents and the book were telling Jacques-Joseph Boreux's truth. Liège and Aix-la-Chapelle were among nearby destinations for the family's marble, and Spa was not far afield; Boreux would have known these towns in the

[64] Van Belle [92].

1780s. In the letter to his brother, he disputes the suggestion that he had already left the Pays de Lìege in 1793, reminding his brother that he had rented a room in Aix-la-Chapelle only because it was impossible to stay long in an inn there, just as in Spa, Wiesbaden, and Pyrmont, etc. These were all gaming spas, and it must have been common knowledge between the brothers that gambling was part of Boreux's attempt to build a new livelihood. Boreux also mentions that he was buying up quarries, a move that might be described, given the condition of the business, as doubling down or martingaling.

When Boreux mentions, in *Tactique des jeux*, that he had gambled in the privileged gaming spas in the 1790s, he adds that he had never gambled in Leipzig, where the gambling dens were more numerous than streets. He knew how hard one must work to win and how easily one then spends the winnings on the many acquaintances made along the way. Did he gamble again when he was back in the Pays de Liège in the 1820s trying to settle his old debts? We do not know, but it appears that he did not manage to resume his life as author and illustrator in Leipzig, remaining instead in Dinant for an impoverished old age, dying in 1846 at the age of 91.

There are literary references in *Tactique des jeux* dating from only shortly before it appeared in 1820. But Boreux describes the games as he had known them in the 1790s. As Thierry Depaulis has pointed out, Boreux describes Trente et Quarante and Roulette essentially as Huyn had done in 1788; he only briefly describes the new Roulette, which he calls "Roulette-Biribi".[65]

4.4.1 The Subtle Seduction of the d'Alembert

The subtitle of Boreux's book promised to demonstrate mathematically, by theory and by practice, a method for winning. The method was based on the d'Alembert. Who had taught Boreux the method? Pierre Nicolas Huyn.

In Boreux's telling, Huyn had learned the method from a certain Monsieur P., who had called the d'Alembert the *martingale graduée* (graduated martingale). Show me a game where the different chances end up equalizing from time to time, said Monsieur P., and I will find, by means of the *martingale graduée*, a simple and easy way of playing guaranteed to win.[66]

Table 1 reproduces and elaborates the first example Boreux gave to illustrate the power of the d'Alembert. After 20 rounds of play, he has netted 4 units (the last entry in Column C). Of this he can expect to lose only about 1 unit to the house, the house's advantage on average being between about two per cent of the total of all the bets, which is 58 units in this example (the sum of column A).[67]

Will the d'Alembert usually produce this kind of success? In fact, it will. As G. N. Bertrand told us in 1798, it has the remarkable property that if you begin with a

[65] On p. 16, Boreux mentions the 8th edition of d'Yverdun's encyclopedia; on p. 171, he mentions a story from the fourth volume of Clauren's *Erzählungen*, which appeared in 1819 [25]. Depaulis [40, p. 18].
[66] [86, p. 83].
[67] For Boreux's discussion, see [86, pp. 84–92].

Table 1 Boreux's first example of the d'Alembert [86, p. 85], with details added. The player bets for 20 rounds. He takes money out of his pocket and puts it on the table as needed to cover his bets, not returning any net gain to his pocket until he stops betting.—Column A shows the amount bet on each round: one unit so long as he is winning but then two units in round 4 because of the loss on round 3, etc.—Column D shows the total the player has taken out of his pocket so far. This goes up only when what he already has on the table is not enough to cover his bet.—Column E shows how much the player has on the table at the end of each round: E = C + D

Round	A	B	C	D	E
	Amount bet	Win or lose	Cumulative net gain	Total out of pocket	On table at end of round
1	1	W	1	1	2
2	1	W	2	1	3
3	1	L	1	1	2
4	2	W	3	1	4
5	1	W	4	1	5
6	1	L	3	1	4
7	2	L	1	1	2
8	3	L	−2	2	0
9	4	L	−6	6	0
10	5	W	−1	11	10
11	4	L	−5	11	6
12	5	L	−10	11	1
13	6	W	−4	16	12
14	5	W	1	16	17
15	4	W	5	16	21
16	3	W	8	16	24
17	2	L	6	16	22
18	3	W	9	16	25
19	2	L	7	16	23
20	3	L	4	16	20

loss, then when the numbers of wins and losses equalize, your net gain is equal to the number of wins. We have examples in Table 1. The first loss is on round 3; it is immediately equalized by the win on round 4, and the net result is that the two rounds together increase the player's net gain by one unit. This is because you increased your bet after losing, so that the winning bet was larger than the losing bet. The same thing happens when the equalization takes longer. The twelve rounds from 9 through 20 begin with a loss and end up with 6 wins and 6 losses, producing an increase in Column C of 6, from –2 to 4.[68]

This martingale is very seductive, Boreux tells us, because most people who try it immediately have a winning streak. But it is also very dangerous, because sometimes

[68] For a formal proof, see [43, pp. 289–291].

Fig. 5 Boreux's use of pinpricks to track how far red is ahead of or behind black [85, Plate VI]. *Source* Bibliothèque nationale de France

wins and losses do not equalize, even approximately, before you run out of money. So before putting it into practice, you must find a way of playing, a *marche* or *attaque*, that will bring you back where you began from time to time in spite of the variation.[69] Boreux had found that *marche*, he tells us, for the games of Trente et Quarante, Roulette, and Biribi. His experience had shown that these games do produce, more or less often, the required equalization of colors. So you track which color is ahead with pinpricks, and always bet on the color that is behind, not betting when they are equal (Fig. 5).

Here is the consummation of Boreux's military analogy. You attack only when you have two probabilities in your favor—the probability that wins and losses will equalize and the probability that red and black will equalize.[70] Remember also that you are the guerrilla attacking a mighty army. Don't push your luck. When you have risked a little and made a nice winning, move on. Come back later to another table or another séance.

For those who know mathematical probability, the success of the d'Alembert seems easy to explain. The expected gain is zero, but this zero is composed of a high probability of a small gain and a small probability of a huge loss. The probability of the huge loss being so small, we may play a long time before it finds and bankrupts us. The method of attack, betting on a color that is behind, is irrelevant.

[69] [86, pp. 85–86].

[70] This idea is echoed in the essay "Nemesis", published by the Danish philosopher Johan Ludvig Heiberg in 1827: you bet simultaneously on the transition from bad to good luck and the transition from one color to the other [50, p. 111].

But there is a more subtle aspect of the d'Alembert's seduction, which Boreux may have been the first to note in print. Not only does it usually win. It usually yields a high return on investment. Risk a little to gain a lot! Discussing the example detailed in Table 1, Boreux boasted that he won 4 units after only 20 rounds even though he had only 21 units in his pocket. In fact, he only took 16 units (the last entry in column D) out of his pocket. Winning 4 while risking only 16 is a 25% profit on your capital.

This aspect of the d'Alembert's success is also not unusual. If we repeatedly play it for 200 rounds—about the number of rounds we could expect to play in a day of Trente et Quarante, and we have a deep pocket, we can expect to double the money we actually take out of our pocket about 90% of the time. We can even expect to multiply the money we take out of our pocket by ten about 25% of the time.[71]

This aspect of the d'Alembert and other popular 19th-century martingales surely helped sustain their popularity in the face of the unrelenting success of the casinos. The British politician Henry Labouchere, heir to a great fortune, reported in 1877 that he "invariably" paid the expenses of his trips to Homburg from what he won using his favorite martingale. He could easily have forgotten the time when he lost a small fortune, shrugging it off as the result of one too many glasses of wine or mere inattention. Monsieur Rogier, writing in 1896, asked why the martingaler of more modest means, after trying his system at home, with his own roulette, and always obtaining marvelous results, then finds disaster at the casino. Rogier's explanation, sustained in a 23-page pamphlet, was that casinos cheat. But perhaps the modest martingaler, simulating play with his own bottomless pocket of play money, fell prey to the same illusion as Mr. Labouchere.[72]

How could we disabuse Labouchere and his less wealthy imitators of their illusions? We can only remind them not to walk into the casino with a credit card. Before trying out a gambling strategy, fix the amount of money, say A, you are willing to risk. Put A in your pocket and evaluate your final capital B relative to A, not relative to how deep into your pocket you actually dig in the course of each play.

5 Betting Systems and Game Theory

By the beginning of the 20th century, the invention of betting systems was both a business and a pastime. After decades of watching and deploring gambling, the Nobel-prize winning Flemish poet Maurice Maeterlinck published these observations in 1920:

> I have no intention of reviewing all these systems, which are innumerable and of unequal value: the *paroli* pure and simple, that artless, violent doubled stake which leads straight to disaster; the D'Alembert and all its variants; the descending progressions; the differential methods; the *montant belge*; the *parolis intermittents*; the snowball; the *photographie*; the

[71] [33].

[72] Rogier [79]. Labouchere's comments [56, p. 195] are quoted by Ethier [43, p. 313] and others.

> staking of equal amounts on certain groups of figures, which is a Chinese puzzle demanding days of patient observation before it is attacked; and many others which I forget, from the most clear-cut to the most mysterious, which are sold at a high price, to credulous beginners, in sealed envelopes containing what is everybody's secret and with all, or nearly all, of which I have become acquainted, thanks to the kindness of an erudite player. A detailed account of those most frequently used will be found in D'Albigny's treatise, *Les Martingales modernes*, in Gaston Vessillier's *Théorie des systèmes géométriques*, in Hulmann's *Traité des jeux dits de hasard*, in Théo d'Alost's *Théorie scientifique nouvelle des jeux de la roulette, trente-et-quarante, etc.*, and, above all, in the *Revue de Monte Carlo*, which has given a system in every issue since the date of its foundation some fifteen years ago.[73]

D'Albigny's book, published in 1902, was typical of the books Maeterlinck cited. D'Albigny laid out 23 betting systems, each with its *attaque* and its *massage*. He was always definite about the *massage*, but often left the reader leeway in the choice of the *attaque*. In the text of the book, the word *martingale* is used only a couple of times, and then with its primitive meaning, where lost bets are doubled. Yet the title of the book seems to label all the systems, or at least their *massages*, as "martingales", confirming that the word still had the same double meaning in 1902 that G. N. Bertrand had noted in 1798.

It seems that when the name *martingale* was used broadly, it ususally referred only to the *massage*. You must further choose your *attaque*. For Smyll's Monsieur P. for example, the *martingale graduée* is not a complete betting system; he constructs a betting system from it when you show him a game where chances equalize from time to time.

Ambiguity about the exact meaning of *martingale* was natural in the 19th-century, when the study of games of chance was relatively informal. In the early 20th century, however, the theory of games formulated by Émile Borel and developed by John von Neumann brought a new level of formality into the picture[74] In Borel's and von Neumann's concept of a game, the players and their allowed moves are fully specified, and a strategy for a player is a precise mathematical object that specifies the player's every move. For every round in the game, it tells the player how to move, as a function of the information the player has at that point. Who are the players in the casino? We are accustomed to saying, with justice, that the gambler is playing against the house. Certainly the house is taking the gambler's money. But the house is passive in Trente et Quarante. There are only two players making moves: the gambler, who bets, and reality, who decides the outcome of each round: red or black. Thus a strategy for a gambler who receives no additional information in the course

---

[73] Maeterlinck [62, pp. 146–148], translation by Alexander Teixeira de Mattos; d'Albigny [35], Vessillier [93], d'Alost [36]. Whereas d'Albigny and d'Alost purport to show the player how to win, Vessillier undertakes to explain the futility of betting systems. I have not seen Hulmann's book. See also Charles Derennes [41], who listed by name over 500 systems published in the *Revue de Monte Carlo*, and the notorious Doctor Petiot's book [71]. The English-language bibliography on betting systems is much thinner; 19th century and early 20th century examples include [17, 18, 70]; see also the references in [43, pp. 311–316] and in the bibliographies [91, pp. 1155–1272] and [60].

[74] See the references following the notes by Fréchet, Borel, and von Neumann on pp. 95–127 of *Econometrica* 21(1), 1953. On the role of Borel in game theory, see [65].

of play is a rule that specifies, for each finite sequence of reds and blacks, the amount the gambler is to bet on the next round, and (if that amount is nonzero), whether the bet is on red or on black.[75]

In 1936, three French mathematicians published work on martingales with the formality of the new game theory: Marcel Boll, René de Possel, and Jean Ville. The three almost certainly knew each other and may have discussed martingales. Both Boll and de Possel commented on the countless martingales that gamblers had proposed; Boll devoted many pages to the misunderstanding of compensation, and de Possel mentioned the thousand and one martingales described in specialized journals. For all three authors, a martingale was a strategy for successive betting that began with a fixed amount of money and could not risk more (and therefore, we may add, produced a nonnegative capital process). They did not assume that the bets were fair, because most often the house has an advantage.[76]

Boll was an extraordinarily prolific author on science, mathematics, and philosophy. He was well known for his role in popularizing the philosophy of science of the Vienna logical positivists. His discussion of martingales appeared in a 382-page book on probability and games of chance that was replete with references to Borel. The book mentions de Possel and Ville, along with Borel, as being among the few genuine authorities on the mathematics of probability.[77]

De Possel was a young mathematician, then in the Bourbaki group. He mentioned martingales in a monograph on game theory based on lectures he had given at Nice. Although it was brief and elementary and considered only a few games, the monograph distinguished between games where the players see each others' moves (as in Trente et Quarante) and games where they do not (the focus of von Neumann's contribution).

Whereas Boll and de Possel discussed strategy and probability in particular games of chance, Ville's purpose was more theoretical: he was generalizing Richard von Mises's concept of a collective. Focusing on the case of independent trials of an event of probability p, he considered all strategies for betting at odds p to $1 - p$, where the bettor's information on each trial consists of the outcomes of the previous trials.

In the thesis Ville published three years later, in 1939, Ville gave a different twist to the notion of a martingale. In addition to developing the idea in his 1936 note, he also introduced martingales into a more idealized picture involving a sequence of random variables. Here we imagine that the gambler can bet on successive variables by buying payoffs at prices given by their conditional expected values, and the gambler's martingale is represented by his nonnegative capital process. The house's

[75] The strategy itself will specify when it will stop, after which the bets are all zero. It must also respect the house's limits (minimum and maximum) on individual bets and the limits imposed by the gambler's and the house's wealth.

[76] Boll's 1936 book [20], de Possel's lectures [73], Ville's mathematical note [95].

[77] For Boll's career, see [80]. Boll's mention of de Possel and Ville [20, p. 356].

advantage disappears; Ville calls the martingale a fair game. Thus were born the martingales of modern mathematical probability.[78]

We often think of the newly abstract form taken by mathematics in the early 20th century as an unalloyed advance. But did we lose something when we replaced gamblers' martingales and parolis with an abstract mathematical object drained of any explicit purpose? Are we now blind to parolis played by entrepreneurs or martingales played by financiers? Could we avoid some of the perils of hypothesis testing if we remembered the seduction of the d'Alembert?[79]

Acknowledgements This chapter was inspired by conversations with Bernard Bru and owes much to his advice. It has also benefited directly from conversations with and suggestions by Harry Crane, Stewart Ethier, Niels Keiding, Roger Mansuy, Laurent Mazliak, Josh Miller, Nell Painter, Steve Stigler, Jean-Louis Van Belle, and Volodya Vovk.

References

1. Agripinæ: Les joueurs et Mr. Dusaulx. Written for Jean-Claude Jacquet de la Douay for the purpose of blackmail, smuggled into France (1781)
2. Amateur (anonymous): Art de bien jouer trente-et-un. Bréauté, Paris (1829)
3. Ampère, A.M.: Considérations sur la théorie mathématique du jeu. Perisse, Lyon (1802)
4. Ancien croupier de Frascati (anonymous): La Californie germanique. Poulet-Malassis, Paris (1862)
5. Anonymous: Faro, and Rouge et Noir. John Debrett, London (1793)
6. Anonymous: Mystères de la loterie dévoilés en faveur des actionnaires de la Loterie nationale de France. Labrousse, Paris (1801)
7. Anonymous: Cinquième classe, oeuvre mélées. Journal Général de la Littérature de France **12**(8), 249–256 (1809)
8. Babbage, C.: An examination of some questions connected with games of chance. Transactions of the Royal Society of Edinburgh **9**, 153–177 (1821)
9. Babbage, C.: On the influence of sign in mathematical reasoning. Transactions of the Cambridge Philosophical Society **II**(II), 325–377 (1827)
10. Barnhart, R.T.: Gamblers of Yesteryear. GBC, Las Vegas (1983)
11. Barnhart, R.T.: The invention of Roulette. In: W.R. Eadington (ed.) Gambling Research: Proceedings of the Seventh International Conference on Gambling and Risk Taking, Vol. 1. Public Policy and Commercial Gaming Industries throughout the World, pp. 295–331. University of Nevada, Reno (1988)
12. Barnhart, R.T.: Gambling in revolutionary Paris — The Palais Royal: 1789–1838. In: W.R. Eadington, J.A. Cornelius (eds.) Gambling and Public Policy: International Perspectives, pp. 541–562. Institute for the Study of Gambling and Commercial Gaming, Reno (1991). An abridged version appeared in *Journal of Gambling Studies*, 8(2):155–161, 1992.
13. Belmas, É.: Jouer autrefois: Essai sur le jeu dans la France moderne (XVI^e–XVIII^e siècle). Champ Vallon, Seyssel (2006)
14. Bender, J.H.: Die Lotterie, eine juristische Abhandlung. Mohr, Heidelberg (1832)

[78] Ville's thesis: [94]. I discuss Ville's ideas in more detail in my other chapter in the present volume. In his chapter, Bernard Locker discusses Joseph Doob's adoption of Ville's word *martingale*, dropping the assumption of nonnegativity, in 1948.
[79] [82, 83]

15. Bertrand, G.N.: Trente-un dévoilé, ou la folie du jour, dédié à la jeunesse. Self-published, Paris (1798)
16. Bertrand, J.: Calcul des probabilités. Gauthier-Villars, Paris (1889)
17. Bethell, V.: Ten Days at Monte Carlo at the Bank's Expense. William Heinemann, London (1898)
18. Bethell, V.: Monte Carlo Anecdotes and Systems of Play. William Heinemann, London (1910)
19. de Birague, C.: La Roulette et le Trente-et-Quarante, ou Le vrai système des jeux de hasard. Chaumerot, Paris (1862)
20. Boll, M.: La chance & les jeux de hasard. Larousse, Paris (1936)
21. Borel, É.: Valeur pratique et philosophie des probabilités. Gauthier-Villars, Paris (1939)
22. Bru, B.: Poisson, the probability calculus, and public education. Electronic Journal for History of Probability and Statistics **1**(2) (2005)
23. Bru, M.F., Bru, B.: Les jeux de l'infini et du hasard. Presses universitaires de Franche-Comté, Besançon, France (2018). Two volumes
24. Chamois, J.B.: L'art du bien-jouer à la Roulette, first edn. Self-published, Paris (1817). Second 1818, third 1823, fourth 1828
25. Clauren, H.: Erzählungen, vol. 4. Hilscher, Dresden (1819)
26. Condorcet, M.d.: Élémens du calcul des probabilités, et son application aux jeux de hasard, à la loterie, et aux jugements des hommes. Royez, Paris (1805)
27. Corbaux, F.: On the natural and mathematical laws, concerning population, vitality and mortality. Self-published, London (1833)
28. Corbaux junior, F.: Dictionnaire des arbitrages simples. Self published, Paris (1802). Two volumes
29. Corbaux junior, F.: Essais métaphysiques et mathématiques sur le hasard. Self-published, Paris (1812)
30. Costabel, P.: Essai sur les secrets des *Traités de la Roulette*. Revue d'histoire des sciences **15**(3/4), 321–350 (1962)
31. Cournot, A.A.: Exposition de la théorie des chances et des probabilités. Hachette, Paris (1843)
32. Cournot, A.A.: Ecrits de jeunesse et piéces diverse. Presses universitaires de Franche-Comté (2010). Edited by Bernard Bru and Thierry Martin
33. Crane, H., Shafer, G.: Risk is random: The magic of the d'Alembert (2021). Working Paper 57, www.probabilityandfinance.com
34. Croizette, A.: Le Masque tombé, ou le Bal de l'Opéra, comédie en un acte, mêlée de vaudevilles, par les Cit. Armand, Chateauvieux, et Bonel. L'imprimerie d'Egron, Paris (1801)
35. d'Albigny, G.: Les martingales modernes. Mathurins, Paris (1902)
36. d'Alost, T.: Théorie scientifique nouvelle du jeu de Roulette, Trente et Quarante, etc. Self-published, Bruxelles (1910)
37. De Moivre, A.: The Doctrine of Chances: or, A Method of Calculating the Probabilities of Events in Play. Pearson, London (1718). Second edition 1738, third 1756
38. Depaulis, T.: Les loix de jeu: Bibliographie de la litérature technique des jeux de cartes en français avant 1800. Suivi d'un supplément couvrant les années 1800–1850. Cybalum mundi, Paris (1994)
39. Depaulis, T.: Bingo! A material history of modern gaming. In: Zollinger [99], chap. 2, pp. 36–56
40. Depaulis, T.: La roulette au XVIIIe siècle: Essai de généalogie (2019). Academia.edu
41. Derennes, C.: La Fortune et le jeu. Georges-Anquetil, Paris (1926)
42. Deshoulières, A.D.L.d.l.G.: Les Poésies de Madame Deshouliers. Édition Nouvelle. Augmentée d'un tiers. Wetstein, Amsterdam (1694)
43. Ethier, S.N.: The Doctrine of Chances: Probabilistic Aspects of Gambling. Springer (2010)
44. Fourier, J.: Mémoire sur les résultats moyens déduits d'un grand nombre d'observations. In: J. Fourier (ed.) Recherches statistiques sur la ville de Paris et le département de la Seine, pp. ix–xxxi. Imprimerie royale, Paris (1826)
45. de Gaigne, A.T.: Mon histoire au Trente-un et celle de tous ceux qui le jouent. Bell, London (1799). Published anonymously

46. Gall, M.: La Roulette et le Trente et Quarante. Delarue, Paris (1882). Martin Gall was a pseudonym for Jules Arnous de Rivière.
47. Grégoire, G.: Traité complet du Trente-Quarante, de ses rapports avec la Roulette et de l'assimilation de ces jeux avec les Échecs et les Dames. Passard, Paris (1863)
48. Harouel, J.L.: De François 1er au pari en ligne, histoire du jeu en France. Pouvoirs: Revue française d'études constitutionnelles et politiques (139), 5–14 (2011/4). Available in English on Cairn International
49. H.E.B.C.: Manuel des jeux de hasard. Barthelemy, Pollet, Bidault, Paris (1803)
50. Heiberg, J.L.: Nemesis: A popular philosophical investigation. In: J. Stewart (ed.) Heiberg's *Contingency Regarded from the Point of View of Logic* and Other Texts, pp. 103–125. Museum Tusculanum Press, Copenhagen (2008). Originally published in Danish in 1827; translated by Stewart
51. Hesse, C.: Publishing and Cultural Politics in Revolutionary Paris, 1789–1810. University of California Press, Berkeley (1991)
52. Huyn, P.N.: La théorie des juex de hasard ou analyse du Krabs, du Passe-dix, de la Roulette, du Trente & Quarante, du Pharaon, du Biribi & du Lotto. Publisher not given. Reprinted in Amsterdam in 1803 (1788)
53. Jones, A.L.: Logic, Inductive and Deductive: An Introduction to Scientific Method. Holt, New York (1909)
54. Jouet de Lanciduais, J.: Influence de l'esprit aléatoire sur l'économie politique et social: Trente-et-Quarante dévoilé. Dentu, Paris (1859)
55. Kavanaugh, T.M.: Dice, Cards, and Wheels: A Different History of French Culture. University of Pennsylvania Press, Philadelphia (2005)
56. Labouchere, H.: Entre nous. Truth **1**(7), 193–197 (1877)
57. Lacroix, S.F.: Traité élémentaire du calcul des probabilités. Courcier, Paris (1816). Second edition 1822
58. Laplace, P.S.d.: Essai philosophique sur les probabilités, first edn. Courcier, Paris (1814)
59. Laplace, P.S.d.: Œuvres complètes de Laplace, Volume XIV. Gauthier-Villars, Paris (1912)
60. Lohner, H.: Roulette, Baccara, Trente-et-Quarante. International Bibliography of Casino Gambling, Its Aspects and History. Books on Demand GmbH, Norstedt, Germany (2007)
61. M., D.: Calcul du jeu appellé par les François, le Trente-et-Quarante, et que l'on nomme à Florence, le trente-et-un. Bernard Paperini, Florence (1739)
62. Maeterlinck, M.: Of gambling. In: Mountain Paths, pp. 133–161. Methuen, London (1920). Book translated by Alexander Teixeira de Mattos from *Les Sentiers dan la Montagne*, Fasquelle, Paris, 1919
63. Markov, A.A.: Probability Calculus (in Russian). Imperial Academy, St. Petersburg (1900). The second edition, which appeared in 1908, was translated into German as *Wahrscheinlichkeitsrechnung*, Teubner, Leipzig, Germany, 1912.
64. Maxim, H.S.: Monte Carlo Facts and Fallacies. Grant Richards, London (1904)
65. Mazliak, L.: The games of Borel and chance. some comments on Borel's role in the theory of games. In: M. Voorneveld, al (eds.) One Hundred Years of Game Theory: A Nobel Symposium. Cambridge University Press (2022)
66. Noomen, W., Boogaard, N.v.d.: Nouveau Recueil Complet des Fabliaux, vol. 1. Van Gorcum, Assen, Pays-Bas (1983)
67. Panckoucke, C.J.: Dictionnaire des Jeux, Faisant suite au Tome III des Mathématiques. Encyclopédie Méthodique, Paris (1792)
68. Parisot, S.A.: L'art de conjecturer à la Loterie. Bidault, Paris (1801)
69. Parisot, S.A.: Traité du calcul conjectural, ou l'Art de raisonner sur les choses futures et inconnues. Bernard, Didot, Courcier, Béchet, Paris (1810)
70. Persius, C.: Rouge et Noir. The Academicians of 1823; or the Greeks of the Palais Royal, and the Clubs of St. James's. Lawler and Quick; Stephen Couchman, London (1823). Charles Persius was a pseudonym for Charles Dunne.
71. Petiot, M.: Le hasard vaincu. Fantaisium, Paris (2018). First typeset edition, by François Montmirel, of a book self-published in manuscript by the author in 1946

72. Poisson, S.D.: Mèmoire sur l'avantage du banquier au jeu de trente et quarante. Annales de Mathématiques Pures et Appliquées **16**(6), 173–208 (1825)
73. de Possel, R.: Sur la théorie mathématique des jeux de hasard et de réflexion. Hermann, Paris (1936)
74. Prévost, A.F.: Manuel lexique ou dictionnaire portatif des mots François. Didot, Paris (1750). 2 volumes
75. Proctor, R.A.: Gambling superstitions. Cornhill Magazine **25**, 704–717 (1872)
76. Proctor, R.A.: Chance and Luck: A Discussion of luck, coincidences, wagers, lotteries, and the fallacies of gambling; with notes on poker and martingales, second edn. Longmans, Green, London (1887)
77. Richardson, S.P.: Systems and Chances. Bell, London (1929)
78. Robert-Houdin, J.E.: Les tricheries des grecs dévoilées: L'art de gagner à tous les jeux. Hetzel, Paris (1863). Translated as *Card-Sharping Exposed*, Routledge, London, 1882
79. Rogier, M.: Les secrets de la Roulette et du Trente et Quarante, avec méthode et système infaillibles, sans capital proprement dit. Self-published, Paris (1896)
80. Schöttler, P.: Marcel Boll et l'introduction du Cercle de Vienne en France. In: C. Bonnet, E. Nemeth (eds.) Wissenschaft und Praxis, Veröffentlichungen des Instituts Wiener Kreis 22, pp. 203–221. Springer International Publishing Switzerland (2016)
81. Sextius, G.: Itinéraires et notices, Monaco, Saxon-Les-Bains, Fontarabie, Saint-Sébastien & Madrid. Explication de la Roulette et du Trente et Quarante avec un moyen pratique pour gagner. Dépot Central des Almanachs, Paris (1875)
82. Shafer, G.: Testing by betting: A strategy for statistical and scientific communication (with discussion). Journal of the Royal Statistical Society, Series A **184**(2), 401–478 (2021)
83. Shafer, G., Vovk, V.: Game-Theoretic Foundations for Probability and Finance. Wiley, Hoboken, New Jersey (2019)
84. Sheldon, E.S.: Saint Peter and the Minstrel. Macmillan, New York (1812)
85. Smyll, J.: Atlas de la tactique des jeux de hasards, consistant en 16 planches coloriées et 40 tableaux de calculs spéculatifs et démonstratifs. Hinrichs, Leipzig (1820)
86. Smyll, J.: Tactique des jeux de hasards. Hinrichs, Leipzig (1820). James Smyll was a pseudonym for Jacques-Joseph Boreux.
87. Spéculateur (anonymous): Essai sur le Trente-un, second edn. Vente and Petit, Paris (1809)
88. Steinmetz, A.: The Gaming Table: Its Votaries and Victims, in All Times and Countries, Especially in England and in France. Tinsley Brothers, London (1870). 2 volumes
89. Stigler, S.M.: The History of Statistics: The Measurement of Uncertainty before 1900. Harvard University Press, Cambridge, MA (1986)
90. Stigler, S.M.: Casanova, "Bonaparte", and the Loterie de France. Journal de la société française de statistique **144**(1–2), 5–34 (2003)
91. United States Government: Gambling in America: Staff and Consultant Papers, Model Statutes, Bibliography, Correspondence. Commission on the Review of the National Policy Toward Gambling, Washington (1976)
92. Van Belle, J.L.: Le projet de factum de Jacques-Joseph Boreux (1755–1846), maître marbrier dinantais, écrivain, inventeur. La Taille d'Aulme, Braine-le-Château, Belgium (2011)
93. Vessillier, G.: Théorie des systèmes géométriques (à masses égales) appliqués aux chances simples de la Roulette. Delarue, Paris (1909)
94. Ville, J.: Étude critique de la notion de collectif. Gauthier-Villars, Paris (1939)
95. Ville, J.A.: Sur la notion de collectif. Comptes rendus **203**, 26–27 (1936)
96. Whitney, W.D., Smith, B.E.: The Century Dictionary: An Encyclopedic Lexicon of the English Language. Century, New York (1914)
97. Young, N.: Systems of gambling. The National Review **18**, 449–460 (1892)
98. Zollinger, M.: Geschichte des Glücksspiels: Vom 17. Jahrhundert bis zum Zweiten Weltkrieg. Böhlau, Wien (1997)
99. Zollinger, M. (ed.): Random riches : gambling past & present. Routledge, London (2016)

Émile Borel's Denumerable Martingales, 1909–1949

Marie-France Bru and Bernard Bru

Abstract

At infinity, a fair game can become unfair. This paradox was first fully understood and mastered in the 1940s by Émile Borel, who resolved it using the theory of denumerable probability he had introduced in 1909 and the theory of martingales developed by his student Jean Ville in the 1930s. Borel's reflections on the paradoxes of infinite play were stimulated by a debate beginning around 1910 with Félix Le Dantec, who questioned the practical value of probability theory. Borel learned from Le Dantec that probability's applications generally depend on equating a small or zero probability with impossibility, and he struggled to reconcile this insight with his denumerable probabilities. By 1949, the struggle had led him to the optional stopping theorem for heads and tails.

Keywords

History of probability theory · Émile Borel · Laplace · Martingales · Optional stopping · St. Petersburg paradox

Marie-France Bru (1943–2012) taught at Université Paris Diderot. Bernard Bru retired from Université Paris Descartes. Glenn Shafer of Rutgers University (e-mail: gshafer@business.rutgers.edu) has prepared this translation of the original French version [12, Vol. 1, Chap. 8], including its quotations, and provided brief footnotes based on the original's extensive endnotes.

M.-F. Bru
Université Paris Diderot, Paris, France

B. Bru (✉)
Université Paris Descartes, Paris, France
e-mail: leslogesb@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*,
Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_3

1 Introduction

Fair games, when allowed to continue indefinitely, can be mathematically unfair. Consider, for example, a fair game of heads or tails, with a mathematical coin that has the same chance of landing heads as of landing tails. Let F be the initial capital of a player who bets one euro each game. He wins a euro when the coin lands heads, loses a euro when it lands tails. We know that the mathematical expectation for the player's capital after any number of rounds is equal to F, the expectation of winning or losing each toss being zero. Everyone readily agrees that this remains true if the player, taking account of his wins and losses as he goes along, is allowed to stop whenever he wants, at least if the game ends after finitely many tosses with probability one. This insight greatly simplifies certain complicated calculations, as very quickly noticed by Abraham De Moivre and Nicolas Bernoulli and others after them, Joseph Bertrand in particular.

Yet if the player decides to stop after he has doubled his capital, the capital he expects is $2F$, not F, and since Ampère at least, we have known that this doubling occurs in finite time with probability one.[1] There is an intuitive contradiction here, a contradiction that evidently belongs to the domain of Émile Borel's denumerable probabilities and must be resolved within this framework. Before giving Borel's answers, let us first examine a historical example of this paradox, which has not failed to plunge the most distinguished scholars into a certain perplexity.

2 Martingales of Fathers of Families

Beginning in the 1816 edition of his *Essai philosophique sur les probabilités*, Laplace considered the following "illusion":

> Some have tried to explain the number of male births being greater than the number of female births by the common desire of fathers to have a son who will continue their name. Thus, imagining an urn filled with equally and infinitely many white and black balls, and supposing that a large number of people draw balls from the urn, each planning to stop drawing when they have drawn a white ball, they have believed that this plan should result in more white than black balls being drawn. It is true that after all the drawing is done, the number of while balls will be at least as great as the number of people drawing them, and it is possible that no black balls will be drawn.[2]

The first question raised by this passage is the origin of the illusion it discusses. Who are these "some" who think that fathers' desires for male descendants can change the possibility for the birth of boys in the entire population, accounting for there always being more boys born than girls, in the nearly fixed proportion

[1] For Ampère's theorem, published in 1802 [1], see [12, Vol. 1, Chap. 3].

[2] [17], p. 193 of the 3rd ed. 1816; pp. 204–205 of the 4th ed. 1819; p. 164 of the 1986 scholarly reissue of the 5th ed. 1825.

long observed but unexplained? We will then need to discuss the illusion itself and its connections with the martingale paradox, following Laplace and his immediate successors.

The question of the identity of the "some" can be resolved, partly thanks to the masterful edition of Laplace's correspondence brought to fruition by Roger Hahn [16].[3] There we find a letter dated 28 December 1783 from Georges Louis Le Sage, an eminent Genevan scholar and author of a corpuscular theory of gravity recognized by Laplace for its ingenuity but not adopted by him. At the end of the letter, mostly concerned with many other topics, Le Sage writes:

> In your "Mémoire sur les probabilités" submitted on 19 July 1780, you take it, sir, as morally certain "that the difference observed between the births of boys and that of girls in Paris is due the greater possibility of a boy being born". I will hazard a conjecture about this "greater possibility", a conjecture easy to verify by checking the record of baptisms.
>
> Married men of every class almost all want more male than female children, and they strongly want to have at least one male child to preserve their name and support the family. This has consequences when economic considerations lead them to balance whether to continue surrendering themselves gently to the call of nature or to deprive themselves. They bring many considerations into this balance: the desire to have a male heir if they do not yet have one, or even the desire to have more sons if (already having some) their last child was a daughter (as they think there should be some rule of alternation in the sequence of births). And they will not completely stop until they have finally obtained what they desire or give up, lest they burden themselves with too many offspring. In a word: perhaps the observed difference is due entirely to the difference between the youngest children in each home, and this is only due to their mother's and father's gambles based on their preference for our sex.

Laplace's was not slow to respond. One 23 February 1784, he wrote to Le Sage that he did not share his "sentiment" on the topic:

> If you consider an urn containing an infinite number of white and black tickets in a given ratio and suppose that about twenty million tickets come out on each draw, no imaginable reasoning by the people who draw these tickets will have any influence on the ratio of white to black among the tickets that should come out.

A peremptory argument as if one were needed: it is impossible to change the ratio of white to black tickets, because it is impossible. In short, Laplace simply affirms the principle that says a betting system is impossible in the game of heads or tails. Yet it has been said that this principle, seen more or less as a necessary consequence of the "fairness" of the game, is far from being so obvious as it seems. In any case, it merits discussion.

Le Sage had the entirely Genevan courtesy to avoid remarking on the weakness of the Laplacian argument from authority. He responded right away, on 30 March 1784, that we was well aware of the "remark" that "our reasoning cannot influence the ratio of whites and blacks that are drawn from a given urn", but this had not prevented

[3] The correspondence discussed here is in Volume 1: Letter 67, beginning on p. 135, Letter 72, beginning on p. 146, and Letter 76, beginning on p. 151.

him from making his conjecture about male children. "Because this particular case had not seemed to be covered by the general remark." But because of the complete confidence that he had in Laplace, he had decided to examine the question again and convinced himself that at least in the particular case where the couples decide to stop procreating after the birth of the first son, this cannot change the average number of boys relative to that of girls. Le Sage lays out his reasoning in a way as simple as convincing, which will be rediscovered by many authors after him.

Nonetheless, the fact that the martingale of fathers who stop procreating after their first son does not manage to change the distribution of genders at birth in the whole population in no way proves that some other more sophisticated martingale could not do so. Besides, sometimes a martingale can change the distribution, at least in theory. This is the martingale paradox or illusion. If the number of white and black tickets in Laplace's urn of births are equal at the beginning of time, and fathers draw at their convenience as many tickets as they want, we can propose to them, for example, that they stop procreating the first time the two genders return to equilibrium (so that the number of their daughters is equal to that of their sons) and then the next birth is a son. We can be almost certain that this will happen in a finite amount time.[4] So if all the fathers adopt this martingale, they will all have an extra son, and in the whole population thus created the difference between the number B of boys and the number G of girls will become infinitely large with probability one.

But if a martingale that makes the difference of genders $B - G$ become infinitely large is theoretically possible, we can be assured at least that no birth strategy can succeed in changing the ratio B/G. If the numbers of white and black tickets in Laplace's urn are equal, no strategy for stopping procreation is able to guarantee that the ratio B/G in the whole population obtained from indefinite drawing from the urn could equal 2 or some other number different from 1 that one wants (say 1.05, the gender ratio in France). This would contradict Borel's law of large numbers, which fixes the ratio in question for any indefinitely prolonged sequence of births, and this with probability one [3]. So Laplace was right, but not for the obviously insufficient reasons he gave. Contrary to what he seemed to think, there do exist martingales that make a player a sure winner in a fair game, but within certain limits and only if we let the game continue indefinitely, which leaves plenty of time beforehand to bankrupt any player who risks it with finite capital. The "remark" on the impossibility of a betting system is thus more subtle and one must take a closer look. This is what Borel undertook beginning in 1910, in the framework of his theory of denumerable probabilities, and even more after the second world war. Let's take a look.

[4] Ampère's theorem, cited earlier, implies that there will always be another return to equilibrium with probability one.

3 Borel's Martingales

Borel's probabilistic work had multiple strands—a mathematical strand, denumerable probabilities in the image of and in continuity with the Borelian theory of functions, a scientific strand concerned to show that the probability calculus was one of the mathematical keys to modern science, and a pedagogical strand meant to educate coming generations in the probability calculus and its universal applications in the best possible way. Now it happens that these strands were soon to come into conflict, precisely because the theory of denumerable probabilities seemed to lead to irreducible practical contradictions. For an activist like Borel, it was necessary to resolve the conflict in some dazzling way, either to go around it or to hide it for the better good of humanity (and he would do both). We can easily imagine how daunting this task was. The beginning of the 20th century can look like a sort of apotheosis of scholarly rationalism where you could not get away with fine words if you did not want to be immediately contradicted and even destroyed by an even more thorough rationality. The probability calculus was not in favor in France, neither with the mathematicians, nor the physicists, nor even the statisticians—an over-the-top rejection in the land of Laplace. The smallest doubt, the most tiny apparent contradiction could compromise the whole enterprise of rehabilitation. The difficulty seemed widely known. At the end of his probability course, Poincaré wrote:

> The probability calculus presents a contradiction in the very words that name it, and if I were not afraid of repeating here a phrase too often repeated, I would say that it teaches us above all one thing: to know that we know nothing.[5]

Already burdened with multiple responsibilities and teaching assignments, Borel took it upon himself in 1907 to give a course on the probability calculus at the Sorbonne two years in a row and to publish it under a Euclidean title, *Éléments de la Théorie des probabilités* [2], just as his article on denumerable probabilities [3] was officially appearing, no doubt the one compensating the other. In appearance, the course has a classical form, but be not deceived. Here is a course of combat where nothing is left in the shadows. The principles of the calculus are "well established; the consequences that we will derive from it by purely logical reasoning are rigorously demonstrated".

The first chapter of *Éléments* deals with the game of heads or tails. Nothing very original there, other that in the section at the end of the chapter, Borel discusses "some paradoxes". Here again, all the treatises on the probability calculus did the same, at least since Laplace, but Borel's treatment has the singularity that he sets out to occupy the ground of the "commonsense reasoning" that "many minds, otherwise excellent" prefer to "logical reasoning", particularly in "questions of probability". Borel was targeting in particular one of his most brilliant classmates in the École normale supérieure, "whose mathematical education was very serious", Félix Le Dantec. In

---

[5] This sentence appears on p. 274 of the first edition, 1896 [26]. Perhaps because of Borel's influence, it is not in the second edition, which appeared in 1912.

1907 [18,19], Le Dantec had advanced negationist views on the probability calculus in Borel's journal, the *Revue du mois*: "The probability of an isolated event is an idea that makes no sense", and in the case of events repeated often enough, the "law of large numbers" is "very imprecise" and ultimately useless, as good sense can do very well without it, and commonsense reasoning can as well, inasmuch as this "law" is a purely mathematical "stratagem" having no relation to the (often immoderate) interpretations given it.

Unlike Félix Le Dantec, who excelled at it, Borel was fairly maladroit at polemics. But he had to respond, and he did so fairly haughtily, which must have greatly irritated his interlocutor:

> In my opinion, it is of the greatest scientific and social interest that fundamental principles be accepted without reservation by as many people as possible; so if a few (commonsense) arguments can lead to this result, it is worth devoting a few lines to it, even if they are unnecessary from an absolutely mathematical point of view.

So the task is to show that common sense (or *sentiment*, as Borel called it) is insufficient in questions of probability; calculation is needed. Mischance had it that Borel should choose in his response to discuss, in a different form, the illusion of family fathers, which no one, and Borel no more than Laplace, had really fully dealt with mathematically or by means of common sense, so that his response lost much of its plausibility and its effectiveness. But so it was.

So let us consider, along with Borel, a fair game of heads or tails and a player, Paul, who wins one monetary unit (one euro, let us say) if the coin lands heads and loses the same amount if it lands tails. Let us suppose that Paul begins with zero capital, and let us follow his gains and losses for a large number of tosses. His capital will certainly vary, but we are sure (by our general experience) that after some number of tosses, usually not too many, it will again be zero. After this first return to equilibrium, the player has one chance in two to win a euro. Following Borel, let us call a sequence of tosses in which the player wins right after a return to equilibrium a "good series". Once again we are sure (or nearly sure Borel says, i.e., with an excessively high probability) that such a good series will again come along after a usually small number of bad series. All this is without demonstration, appealing to the common sense of the reader, and has nothing to surprise.

Suppose Paul plays against Pierre and decides to stop when he has obtained a good series.[6] He will thereby win a euro from Pierre for sure (or almost sure), and if he follows the same strategy every day, he will end up bankrupting Pierre completely. But Pierre can use the same martingale equally well and bankrupt Paul in the same way. We have arrived at an "absurd consequence". Borel claims that the error in the argument comes from the supposed certainty of the different phases of the game, whereas a good series may arrive only after a very long time, even if it is certain to finally arrive. The paradox arises only because we "treat a future event as having

[6] Quitting when you are ahead might be the most ancient gambling strategy. Since around 1800, it has been called *faire Charlemagne* in French.

happened, under the pretext that experience has proven it to be extremely probable". Borel announces that he will make his thought more precise later, and he does so in Sect. 18 of the book, where as a mathematical scientist he shows using an absolutely mathematical calculation, that the expected number of returns to equilibrium before the nth toss is approximately equal to

$$\frac{2}{\sqrt{\pi}}\sqrt{n} \approx 1.128\sqrt{n}.$$

In other words, in the course of the first million tosses, the expected number of equilibria is around 1128, and the number of good series half that. This number will grow ever more slowly as the tosses continue, with large deviations becoming more and more likely in the course of time and slowing considerably the returns to zero. Paul, like Pierre, can only hope to win 500 euros in a million tosses, in the course of which they may suffer very significant losses, which will bankrupt them if they do not have a patrimony great enough to absorb these losses and permit them to continue their martingale. Perfectly judicious and original observations that Borel will progressively develop,[7] something Le Dantec's common sense is incapable of doing without calculation. Nevertheless, they have little to do with the radical absurdity of Paul's and Pierre's certain bankruptcy, or of their gains without limit, that Borel had called attention to.

Neither the "absurd consequence" of infinite gains for the two adversaries nor the Borelian response to it goes does to the heart of the problem. Le Dantec did not fail to tell Borel so.[8] Pierre and Paul each follow, for their own account, the same martingale. So each obtains the identical results for themselves. There is no absurdity there. The game is just as favorable to Pierre as to Paul, but at different moments. Moreover, the response to this pseudo absurdity that attributes it all to a mistaken assimilation of very small probability to a zero probability is an argument from authority that makes no sense, because the practical value of the probability calculus is based, very generally, on this assimilation. In his later works, Borel would unreservedly promote the assimilation.[9]

In the end, Borel would admit that the absurdity of the martingale of good series lies elsewhere. He modified his initial formulation slightly but discretely, in accordance with his invariable policy of never conceding anything, finally declaring in *Le Hasard*:

> The absurdity is not that Pierre and Paul both obtain a gain but at different moments. It is that the gain should grow proportionally with time.

Let us agree. But then Pierre's role in the absurdity is not needed. The only absurdity is that the martingale of good series could make Paul gain indefinitely. For Borel,

[7] In [4,5] and successive editions of *Éléments*.
[8] In [20], reprinted in [21]. See also [11].
[9] In [4,6,10], for example. For further discussion, see [22,27].

this comes down to a commonsense absurdity, as everyone knows that martingales do not exist in a fair game. As recalled earlier, this was already Laplace's opinion.

Yet, in the absolutely correct world of mathematics, Paul's capital grows indefinitely with probability one if he persists in his strategy, stopping after each good series and restarting immediately after putting his gain in the bank. We have already seen Borel show, with a more delicate study of the trajectories in the game of heads or tails, that this martingale is suicidal, that the returns to equilibrium and thus Paul's possible gains are too rare for him to be able to get rich during his lifetime, even if he plays every second, night and day. Yet the mathematical result holds for eternity; Paul's gain grows towards infinity. And this is "absurd".

On the other hand, we pointed in the introduction to a much simpler way that this absurdity can be presented. Indeed, as shown by Ampère, and subsequently by Bertrand and Le Dantec as well, the curve of heads or tails returns to zero with probability one, and as Borel pointed out [4, Sect. 21, p. 52, note 1], the same reasoning applies to any gain x, which can also be attained with probability one, no matter how great x is. Which, Borel adds in the same place, "is obviously absurd". The absurdity of the martingale of the player who plays without imagination, playing on until he gains x euros, is much more visible than that of Charlemagne's martingale or its Borelian variants.

There is a contradiction here that Borel did not really know how to resolve, and while waiting for something better, he could only dissimulate. This can at least partly explain how little enthusiasm Borel devoted to developing the theory of denumerable probabilities outside the domain, relatively limited in the end, of pure mathematicians, who are in any case extremely reticent about anything connected with the probability calculus and would not know how to deal with yet another absurdity. But we may well guess that this problem often agitated Borel's rational soul. Let us see.

It was in 1939 that Borel appears to have concerned himself again with the absurdity of martingales. The reason is easy to find. On 9 March 1939, Borel presided over Jean André Ville's defense of his thesis and was so impressed by it that he immediately decided to republish it in the collection of monographs on probability that he directed at Gauthier-Villars [28].

Ville defined a very general notion of a positive martingale that we will not study here, but which comes very close to the notion of a betting system in a game of chance. In order to give a simple example, we restrict ourselves to the game of heads or tails. We toss an abstract well balanced coin, and write 1 if it comes up heads, 0 otherwise. Thus the sequence of possible results is modeled by a sequence of independent random variables equal to 0 or 1 with equal probability. We say that a betting system for the player Paul, with initial capital $S_0 = 1$ euro, consists of a sequence of bets on the next toss, the amount of each bet varying only as a function of the results of the preceding tosses and the resulting capital. Paul can never bet more than he has and must stop playing when he goes bankrupt.

Following Ville, let $\lambda_n(x_1, x_2, \ldots, x_n)$ be the fraction of Paul's current capital that he bets on heads on the $(n+1)$st toss when the first n came out $(x_1, x_2, \ldots, x_n)$, each x being equal to 0 or 1. So if Paul's capital after the nth toss is $s_n(x_1, x_2, \ldots, x_n)$, he bets $\lambda_n s_n$ euros on heads, winning that amount if the coin comes up heads on the following toss and losing it otherwise. His capital after the $(n+1)$st toss is therefore $s_{n+1} = s_n + \lambda_n s_n$ if the coin comes up heads and $s_{n+1} = s_n - \lambda_n s_n$ if it comes up tails.

Let S_n et Λ_n be the random variables equal respectively to Paul's capital after the nth toss and the fraction of that capital that he bets on the $(n+1)$st toss. These are positive functions of $X_1, \ldots, X_n$, and:

$$S_{n+1} = (S_n + \Lambda_n S_n)\mathbf{1}_{X_{n+1}=1} + (S_n - \Lambda_n S_n)\mathbf{1}_{X_{n+1}=0},$$

whence

$$\mathbf{E}(S_{n+1}|X_1, \ldots, X_n) = \frac{1}{2}(S_n + \Lambda_n S_n) + \frac{1}{2}(S_n - \Lambda_n S_n) = S_n, \tag{1}$$

which is Ville's general definition of a martingale for an arbitrary game.

In particular, $\mathbf{E}(S_n) = 1$ for all n. This demonstrates the impossibility on average of a betting system infinitely favorable to the player, but not yet its almost certain impossibility.

Ville shows with complete rigor that for every positive number λ,

$$\mathbf{P}\left(\sup_n S_n \geq \lambda\right) \leq \frac{1}{\lambda}$$

and that this implies

$$\mathbf{P}\left(\sup_n S_n = \infty\right) = 0$$

[28, pp. 100–101]. Even if he follows his martingale for an infinite time, the player cannot enrich himself infinitely. His martingale remains bounded.

Here is a result that could not have failed to delight Borel. Finally a mathematical theory of martingales that is not "absurd". It shows that mathematically absurd martingales are such because they are really absurd, because their conditions are completely unreasonable and impossible to satisfy. They implicitly assume that the player can borrow as much as wants, covering each bet and his losses no matter how enormous they are (i.e., that S_n can be arbitrarily negative, whereas Ville requires that it remain positive). In the framework of Ville's theory, the probability that a betting system enriches infinitely is zero.

So Borel could revisit, now more serenely, the question of probabilistic martingales, all the more so that he had left his main political responsibilities and was about to retire from the Sorbonne. So he took advantage of this free time to edit the final volume of his great *Traité du calcul des probabilités et de ses applications*, entitled *Valeur pratique et philosophie des probabilités* [6].[10]

It would take us too long to study this important work here. All the more so that it seems to have nothing to do with martingales; betting systems are not called

[10] On Borel's treatise, consult [13].

"martingales" and Ville's name is cited only in connection with the theory of collectives, which Borel hardly considered, as he thought the battle of the probability of an isolated event had been won.[11] Nevertheless, as always with Borel, we must pay attention to the slightest allusions. In fact, Chap. IV is devoted to "some errors and paradoxes", and we find two sections devoted to the St. Petersburg paradox.

Pierre plays heads or tails with the understanding that he wins 2^n if heads comes up the first time on the nth toss, which happens with probability $1/2^n$. What price should he pay to participate in this game? Equivalently, what is the expected value of his gain? The answer is easy; Pierre's gain has expected value

$$\sum_{n=1}^{\infty} 2^n \frac{1}{2^n} = \infty.$$

So Pierre should put down an infinite stake. Yet who would risk even a modest amount at this game, which offers astronomic payoffs with only microscopic probabilities? All the scholars discussed this paradox, but Borel had not treated it in his *Éléments*. No doubt he thought everything had been said and the game was, in any case, absurd. It is only in 1939 that he takes up the theme in his and Ville's manner. Borel shows in effect that we can associate with the St. Petersburg game a "very simple martingale", a betting system that is "both theoretically and practically fair", that gives its player, say Paul, the same gains as Pierre in the initial game.

Suppose Paul plays heads or tails and bets on each toss with exponentially growing stakes. He bets $b_n = (n+1)2^{n-1}$ on heads on the nth toss, for $n = 1, 2, 3, \ldots$. So Paul's net gain after the nth toss is

$$M(n) = \sum_{k=1}^{n} b_k Y_k,$$

where (Y_n) is a sequence of independent random variables equal to $+1$ or -1, according to whether the coin comes up heads or comes up tails.

The martingale M is not necessarily positive. It obviously satisfies $\mathbf{E}[M(n)] = 0$, the condition of average fairness, for all n and the martingale condition, Eq. (1):

$$\mathbf{E}(M(n+1)|Y_1, Y_2, \ldots, Y_n) = M(n).$$

In addition, when the first head comes on the nth toss, the sum of Paul's losses and his single gain is, by a simple calculation,

$$(n+1)2^{n-1} = \sum_{k=1}^{n-1}(k+1)2^{k-1} = 2^n,$$

[11] As mentioned earlier in this chapter, Le Dantec had questioned the meaningfulness of the probability of an isolated event. Von Mises had as well. After referring his readers to Ville for critique of von Mises's collectives in Sect. 46 of his 1939 book, Borel went on, in Sect. 48, to argue that an individual can evaluate the probability of an isolated event by contemplating bets.

and this happens with probability $1/2^n$. So Paul obtains the same gain as Pierre does in the St. Petersburg game, with the same probabilities.

Finally, if we set $T = \inf\{n; Y_n = 1\}$, we obtain $\mathbf{P}\{M(T) = 2^n\} = 1/2^n$ and

$$\mathbf{E}[M(T)] = \infty.$$

At time T, which is equal only to 2 on average, the martingale M loses its fairness on average. The absurdity of the infinite St. Petersburg game comes from the absurdity of martingales that are not limited by the finite capital of the players.

Thus we have an explanation of the St. Petersburg paradox based on the more general paradox of mathematical martingales of the type defined here. And it is likely that Borel now thought that a theory of martingales should take this type of behavior into account more precisely, with explicitly stated theorems. For his own part, he was no longer interested in general theories, having long since left the mathematicians to pursue them. For him, it was enough to establish in a practical way that his martingale was illusory, like all the absurd martingales. They all assume, as we have said, that the player can play without any limit on his time and capital.

Borel points out to his reader that this type of absurdity is not particular to the St. Petersburg martingale; it is found equally in the martingale of good series (a.k.a. the Charlemagne martingale) that we saw earlier and in the classical martingale where the player doubles his bet at every loss until he wins. In all case, we are dealing with a martingale of the type $M(n) = \sum_{k=1}^{n} b_k Y_k$, where the sequence of bets (b_n) is a sequence of arbitrary positive numbers.

In the case of the good series, the player never changes his bet; he bets 1 euro on heads for every toss, and $M(n) = \sum_{k=1}^{n} Y_k$ for all n greater than 1. To make this martingale's absurdity obvious, consider the time of the first head after return to equilibrium:

$$T = \inf\{n; M(n) = 1\}.$$

We have seen several times already that T is finite with probability 1 even though its mean is infinite. In this case, $M(T) = 1$ with probability one, violating the martingale's fairness on average, $\mathbf{E}[M(n)] = 0$ for all n. The illusion of the return to equilibrium comes from the same absurdity as that of martingales that are fair on average for fixed times but no longer fair for a random time such as T.

Borel says explicitly that the same holds for the classical doubling martingale. In this case, $b_n = 2^{n-1}$ for all n. The player stops at the time T of the first head. If $T = n$, we have

$$M(T) = M(n) = 2^{n-1} - \sum_{k=1}^{n-1} 2^{k-1} = 1.$$

The martingale M is not fair at time T, where the player is sure to win a euro, and this comes from the same mathematical absurdity.

In general, martingales lose their fairness on average at random times. But as Borel explains at length, those that do so also ruin more rapidly gamblers with limited capital and lifetimes.

Borel could have left it at that, but this would be not to know him. The problem remains: for what "reason" are these martingales with arbitrary signs absurd? Yes, they are certainly absurd for mathematical reasons internal to the theory, but this is not an explanation from the rational point of view. Borel remains attached to the faith of his youth: mathematics are not a pure game of the mind; under certain conditions it tells us something about the outside world. Even if there is nothing to guarantee that mathematics does not deceive us, it has an explanatory power that matters to human society and its science. In the case of absurd martingales, how are these conditions violated? This is as much a metamathematical as a mathematical question, one of those most important to Borel, even as such a mixture of genres is hardly any longer tolerated or understood by his young colleagues.

We could try to follow Borel's reflections on this theme after 1939, but we will not do so, for lack of revealing documents and all the more because Borel's living conditions deteriorated quite seriously during the Occupation [15,23]. We will go directly to the end of the story, in 1949, the year when Borel wrote three notes on martingales for the Academy of Sciences and immediately incorporated two of them in two books that appeared the following year: the final edition of his *Éléments* and a *Que sais-je?* entitled *Probabilité et certitude* [10].[12] What do we find in this ultimate attempt to explain the absurdity of martingales?

We suppose that Pierre plays heads or tails against Paul. Pierre plays the St. Petersburg martingale; on the kth toss, he bets $(k+1)2^{k-1}$ on heads. Pierre leaves the game after the first head, but it is agreed that the game should end after the nth toss at the latest, no matter what happens, even if no head has ever come up. In this case, Paul will win the sum of Pierre's losses, that is

$$\sum_{k=1}^{n}(k+1)2^{k-1} = n2^n.$$

Let us calculate Pierre and Paul's expected gains in this new game. Paul's expected gain is simply

$$\frac{1}{2^n}n2^n = n.$$

Pierre wins 2^k if the first head comes up on the kth toss, which happens with probability $1/2^k$. So his expected gain is

$$\sum_{k=1}^{n}\frac{1}{2^k}2^k = n.$$

The two players have the same expected gain. The game is fair.

[12] The two notes incorporated in his *Éléments* and in *Probabilité et Certitude* were [7,9]. The third, [8], discussed an impractical but mathematically very simple martingale in the game of heads or tails: choose a positive constant a and a constant $\alpha \in [0, 1]$, bet a on the first toss, then multiply your bet by $1+\alpha$ whenever you lose and by $1-\alpha$ whenever you win.

Seen just from Pierre's point of view, Paul's expected gain is his expected loss, so that the sum of Pierre's expected gain and expected loss is zero. Returning to the previous notation for the non-stopped St. Petersburg game, where T is the time of the first head and $M(n)$ is the sum of Pierre's losses and gains at time n, we have the formula

$$\mathbf{E}[M(\inf\{T, n\})] = 0, \tag{2}$$

valid for all n.

Excellent! The St. Petersburg game stopped at n is fair. If Pierre decides to stop for sure at n in order to avoid excessive losses, his martingale gives him no (mathematical) advantage. We easily perceive here a general result: a stopped martingale cannot enrich the player who follows it. Here is a result consistent with both mathematics and gamblers' common sense. It is the natural form of the principle of the impossibility of a betting system, known to all rational gamblers and to some scholars since the eve of time. The illusion of returns to equilibrium and the St. Petersburg paradox and a thousand other things are essentially explained. The paradoxes result from this infinity that we know nothing about and that authorizes everything and its opposite. At a finite distance fixed in advance, they disappear.[13]

But there still remains at least one mystery to consider. Go back to the St. Petersburg martingale M. We saw that both $M(n)$ and $M(\inf\{T, n\})$ have expected value zero, while $\mathbf{E}[M(T)] = \infty$. Here is something that holds no surprise for the modern reader, who knows that passing to the limit under an integral or expectation sign is valid only under sufficient uniformity conditions, but that violates the intuitive principle of continuity, a common sense principle that is usually correct. Here we pass from zero to infinity without warning. How can this phenomenon be rationally explained? Borel simply remarks that Pierre's expected gain is a cumulative sum that grows progressively while Paul's is formed by a single term that will not be present if heads indefinitely fails to appear, which becomes an impossibility quasi-physical for immense values of n, so that it is hardly surprising that the expected infinity of gains has the upper hand indefinitely on the impossible infinity of losses. Here is a Borelian manner of announcing a non-uniformity result, manner that the axiomatic method would reject with contempt, but which Borel considers as the only truly rational one. Reason is a flash of light in the night that no bushel can completely hide [12, Vol. 1, p. 136].

We will say nothing more, save to conclude with Borel that the absurdity of martingales comes from the introduction of denumerable infinity into equations, an infinity merely virtual, not actual as in the Cantorian theory, but which retains enough mathematical tares to corrupt the explanatory value of results deduced from it. Not only does denumerable infinity have no practical value in the probability calculus, unrealized virtual infinity has none either—a Pascalian conclusion (Pensée 233):

[13] Equation (2) is a form of the optional stopping theorem, which Joseph Doob formulated in a very general way in his 1953 book [14]. Doob has told the authors that he had not been aware of the theorem previously. This is an indication of how far ahead Borel was in developing martingale theory 70 years ago.

> We know …the existence and nature of the finite, because we are finite and have extension. We know the existence of the infinite and are ignorant of its nature, because it has an extension like us, but not limits like us. …

Pascal died in Paris 360 years ago, on 19 August 1662 at his sister Gilberte's, rue des Fossés-Saint-Victor, paroisse Saint-Étienne-du-Mont [24,25]. In 1654, Pascal had given the first principles of the mathematics of chance, and in 1658, those of the mathematics of the infinite. Chance, infinity, two Pascalian concepts that mathematical rationality has since tried to deconstruct, with varied results. For lack of an opportunity to celebrate him here, we leave to Pascal the last word, a word about knowledge that closes his Pensée 308:

> From all bodies together, we cannot make one little thought happen; this is impossible, and of another order. From all bodies and minds, we cannot extract one impulse of true charity; this is impossible, and of another and supernatural order.

References

1. Ampère, A.M.: Considérations sur la théorie mathématique du jeu. Perisse, Lyon (1802)
2. Borel, É.: Éléments de la théorie des probabilités. Hermann, Paris (1909). English translation of the 1950 edition, *Elements of the Theory of Probability*, Prentice-Hall, 1965
3. Borel, É.: Les probabilités dénombrables et leurs applications arithmétiques. Rendiconti del Circolo Matematico di Palermo **27**, 247–270 (1909)
4. Borel, É.: Le Hasard. Alcan, Paris (1914). The first and second editions both appeared in 1914, with later editions in 1920, 1928, 1932, 1938, and 1948
5. Borel, É.: Principes et formules classiques du Calcul des Probabilités. Gauthier-Villars, Paris (1925). This is the first fascicle of the first volume of Borel's multi-authored Traité du Calcul des Probabilités et ses Applications.
6. Borel, É.: Valeur pratique et philosophie des probabilités. Gauthier-Villars, Paris (1939)
7. Borel, É.: Le paradoxe de Saint-Pétersbourg. C. R. Acad. Sci. Paris **228**, 404–405 (1949)
8. Borel, É.: Sur une martingale mineure. C. R. Acad. Sci. Paris **229**, 1181–1183 (1949)
9. Borel, É.: Sur une propriété singulière de la limite d'une espérance mathématique. C. R. Acad. Sci. Paris **228**, 429–431 (1949)
10. Borel, É.: Probabilité et certitude. Presses Universitaires de France, Paris (1950). An English translation appeared as *Probability and Certainty*, Walker, New York, 1963
11. Bru, B., Bru, M.F., Chung, K.L.: Borel and the St. Petersburg martingale. Electronic Journal for History of Probability and Statistics www.jehps.net **5**(1) (2009)
12. Bru, M.F., Bru, B.: Les jeux de l'infini et du hasard. Presses universitaires de Franche-Comté, Besançon, France (2018). 2 volumes
13. Bustamante, M.-C., M. Cléry, and L. Mazliak. Le *Traité du calcul des probabilités et de ses applications*. Étendue et limites d'un projet borélien de grande envergure (1921–1939). North-Western European Journal of Mathematics, **1**, 111–167 (2015)
14. Doob, J.L.: Stochastic Processes. Wiley, New York (1953)
15. Guiraldenq, P.: Émile Borel 1871–1956. Imprimerie du Progrès, Saint-Affrique (1999)
16. Hahn, R. (ed.): Correspondence of Pierre Simon Laplace (1749–1827). Brepols, Turnhout (2013). 2 volumes
17. Laplace, P.S.: Essai philosophique sur les probabilités, first edn. Courcier, Paris (1814). A second edition also appeared in 1814. The fifth and definitive edition appeared in 1825. A modern edition with commentary was published by Christian Bourgois, Paris, in 1986

18. Le Dantec, F.: La biologie de M. Bergson. Revue du mois **4**, 230–241 (1907)
19. Le Dantec, F.: Le hasard et la question d'échelle. Revue du mois **4**, 257–288 (1907)
20. Le Dantec, F.: Les mathématiciens et la probabilité. Revue philosophique de la France et de l'étranger **70**, 329–360 (1910)
21. Le Dantec, F.: Le Chaos et l'Harmonie universelle. Alcan, Paris (1911)
22. Mazliak, L., Sage, M.: Au delà des réels. Émile Borel et l'approche probabiliste de la réalité. Revue d'histoire des sciences **67**(2), 331–357 (2014)
23. Mazliak, L., Shafer, G.: What does the arrest and release of Émile Borel and his colleagues in 1941 tell us about the German occupation of France? Science in Context **24**(4), 587–623, 625–636 (2011)
24. Mesnard, J.: Les demeures de Pascal à Paris, étude topographique. Paris et Ile de France. Mémoires **4** (1955)
25. Pascal, B.: Blaise Pascal: Œuvres complètes. Desclée de Brouwer (1964–1992). 4 volumes, edited by Jean Mesnard
26. Poincaré, H.: Calcul des probabilités. Leçons professées pendant le deuxième semestre 1893–1894. Gauthier-Villars, Paris (1896)
27. Shafer, G.: From Cournot's principle to market efficiency. In: J.P. Touffut (ed.) Augustin Cournot: Modelling Economics, pp. 55–95. Edward Elgar (2007)
28. Ville, J.: Étude critique de la notion de collectif. Gauthier-Villars, Paris (1939)

The Dawn of Martingale Convergence: Jessen's Theorem and Lévy's Lemma

Salah Eid

Abstract

Jessen's theorem and Lévy's lemma, which both date from 1934, are the earliest known general formulations of the martingale convergence theorem. Børge Jessen worked within Lebesgue's theory of integration; he saw his theorem as an extension of the Fubini-Lebesgue theorem of 1907–1920. Paul Lévy's vision was probabilistic; he saw his lemma as an extension of Borel's strong law of large numbers of 1909. This chapter reviews how the two arrived at their results. Jessen published his theorem in 1934, and it helped inspired Lévy's formulation of his lemma. In letters between the two authors, each wanted to see the other's result as a trivial consequence of their own. Jessen sought a level of abstraction that proved unattainable, but his interaction with Lévy can be seen as the origin of a now standard version of the martingale convergence theorem. This standard version was first stated by Erik Sparre Andersen and Jessen in 1946, and its standard proof relies on ideas that Jessen and Lévy developed in their correspondence. The correspondence is reproduced in the present volume in the chapter entitled "Analysis or Probability? Eight letters between Børge Jessen and Paul Lévy".

Keywords

History of probability · Jessen's theorem · Lévy's lemma · Martingale convergence theorem · Transfer principle

---

This chapter was translated from the French by John Aldrich, University of Southampton (john.aldrich@soton.ac.uk) and abridged and edited by Laurent Mazliak and Glenn Shafer.

S. Eid (✉)
Université Paris Diderot, Paris, France
e-mail: salaheid.h@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*,
Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_4

1 Introduction

Jessen's theorem and Lévy's lemma, which both date from 1934, are the earliest known general versions of the martingale convergence theorem, which is now habitually stated as follows:

> If $(\mathscr{F}_n)$ is a sequence of sub σ-algebras in a probability space $(\Omega, \mathscr{A}, P)$, increasing or decreasing towards a sub σ-algebra $\mathscr{F}$, and X is an integrable random variable, then $E(X/\mathscr{F}_n) \to E(X/\mathscr{F})$ almost surely and in L^1.

It is well known that this theorem had been anticipated much earlier, in two different frameworks, by Lebesgue and Borel. By 1903, Lebesgue had already proven the theorem of almost everywhere differentiation in his new theory of integration. In 1909, Borel had used probabilistic arguments to state and prove the almost sure convergence of frequencies in the game of heads and tails, the first version of the strong law of large numbers in his new theory of denumerable probabilities. Here were two complementary visions of the world, sometimes intimately united, sometimes resolutely antagonistic, in the image of their authors, who, as we will see, inspired Jessen and Lévy in their work in the 1930s.[1]

Lévy, who read little and badly, happened to read (part of) Jessen's article [81] and presented it to Hadamard's seminar in the spring of 1935. Lévy realized that his own results, obtained by completely different methods, rather resembled Jessen's, and a singular correspondence ensued, a kind of dialogue of the deaf between two mathematicians who conceived of mathematics in entirely different ways, who wrote it in languages without visible connection and yet sometimes understood each other better than they admitted. Most of this correspondence has survived in the Archives of the Institute of Mathematical Sciences at the University of Copenhagen, and we

---

[1] In English-language sources, the name *martingale convergence theorem* is also used for statements that refer explicitly to martingales or supermartingales, as in [48, p. 456].

Jessen's theorem appeared in December 1934 in *Acta Mathematica* [81, §§13 and 14]. Lévy's lemma appeared in 1935 in *Bulletin des Sciences Mathématiques* [128, pp. 88–89] and later in [129, pp. 6–7], in [130, 133], and in Lévy's 1937 book, *Théorie de l'addition des variables aléatoires* [134, §41].

Lebesgue's result appeared in [109], and later in [110, pp. 124–125] and [113, p. 13]. It was developed in [114] and incorporated beginning in 1914 in all the major European treatises on analysis. The theorems on differentiation almost everywhere of Lebesgue, La Vallée Poussin, Denjoy, etc. are the first known statements of the theorem on increasing martingales. The theorems predate the term "almost everywhere" which was introduced by in 1904 [110] and then adopted generally (e.g. [110], second edition, 1928, page 179, note 1). See also [88].

See [22] for Borel's 1909 theorem and [97] for Kolmogorov's definitive version. Borel's theorem was officially brought into the framework of the theory of decreasing martingales by Doob in 1948 [45]. The term "almost sure" was not yet standard in the 1930s, when the usual expression was "convergence with probability one." Lévy seems to have been one of the first to have adopted "presque sûrement" after 1930, though Fréchet in his courses at the Institute Henri Poincaré favoured "presque certainement", without quite imposing it [63, p. 225]. The terminology of probability, like the concepts themselves, remained somewhat fluid until the 1950s.

reproduce it with commentary in a chapter of the present volume, under the title "Analysis or Probability? Eight letters between Børge Jessen and Paul Lévy". In the present chapter, we introduce the two protagonists and their relevant work. Børge Jessen, born in 1907 and a student of Harald Bohr, was professor of geometry at the Polytechnic School of Copenhagen in 1934 and was already making a name for himself among analysts. Paul Lévy, born in 1886, a student of Hadamard and Borel and professor of analysis at the Paris École Polytechnique, had been developing the modern theory of probability after his own fashion for the previous fifteen years.

The exchange between Jessen and Lévy may be one of the possible origins of the formulation and and of the proof of the contemporary martingale convergence theorem, as we have stated it above and as it is found in all probability treatises since the beginning of the 1960s. This modern statement, written in probability theory's language of sequences of sub-algebras, appears for the first time in a famous 1946 article by Erik Sparre Andersen and Jessen [6], which Doob included and developed in his great treatise of 1953 [46],[2] and which, one may say without too much exaggeration, closed the quiet and surely forgotten conversation between Jessen and Lévy. As for the proof of this theorem given today, it is not really different from the extremely simple one that Lévy proposed to Jessen in the case of increasing filtrations and set indicators. It consists of supposing that the variable X is $\mathscr{F}$-measurable and can thus be approximated in L^1 by a sequence of (simple) variables X_n, respectively $\mathscr{F}_n$-measurable. The convergence in L^1 then follows easily from the following inequalities, where we follow Lévy by writing E_n for the conditional expectation operator knowing $\mathscr{F}_n$:

$$|E_n(X) - X| = |E_n(X) - X_n + X_n - X| \leq |E_n(X - X_n)| + |X_n - X|,$$

and so

$$||E_n(X) - X||_1 \leq 2\,||X - X_n||_1 \to 0.$$

For almost sure convergence, it is enough to consider any bounded stopping time σ and to write the same inequalities to show convergence in L^1 of $E_\sigma(X)$ towards X along the filter of bounded stopping times and thus convergence a.s.

---

[2] See also [8]. The Andersen-Jessen formulation and proofs are reproduced in [71, Chap. V, SS20–22]. Doob became aware of the Andersen-Jessen formulation as a result of a 1948 letter from Jessen concerning counterexamples to theorems Doob had published in 1938, and we have conjectured that this may have been decisive in bringing Doob's attention back to his own theory of martingales. The correspondence that began with Jessen's 1946 letter to Doob is reproduced in the chapter in the present volume entitled "Counterexamples to Abstract Probability: Ten Letters by Jessen, Doob and Dieudonné".

We sometimes follow Jessen and other contemporaries in referring to "Sparre Andersen" as if this were his last name. In other sources, e.g., the yearbooks of the Danish Academy of Sciences, his name is given as "Andersen, E. S.". We follow this practice in our list of references and use "Andersen" instead of "Sparre Andersen" when this is more convenient.

For the Moscow school, see [98] and [15, Vol. II, p. 469].

This demonstration, thus reduced to its simplest form, brought into the timeless and "pasteurized" ranks of university courses a theorem that took more than half a century to find its rightful place and then contained nearly all the almost sure or almost everywhere results of the time, Birkhoff's ergodic theorem being the notable exception. So it may not be without interest to recall the confused debates that the theorem generated in 1935, when it had hardly emerged from the ocean of unknown, misunderstood, or mislaid theorems.[3]

2 Jessen's Theorem

In the course of his university studies of mathematics in Copenhagen, Jessen embarked on research under the direction of Harald Bohr. The younger brother of Niels Bohr, Harald was born in 1887 and was already the author of important mathematical works, some with Edmund Landau, on classical analysis and analytic number theory, particularly the Riemann zeta function, Dirichlet series, and the theory of almost periodic functions of real or complex variables. The theory of almost periodic functions originated with Dirichlet series, and Bohr was its real creator [16–18]. The theory's fundamental theorem states that an almost periodic function f has a countable number of proper frequencies. If M denotes the mean taken on increasingly large intervals, $a(\lambda) = M(f(x)\mathrm{e}^{-i\lambda x}) = 0$ except for a countable number of

[3] There are simplified proofs of the martingale theorem in [34,47,143],..., and of course [128]. The demonstration given here is found in [53] and is close in spirit to Lévy's proof, which did not isolate the concept of stopping time but nevertheless used it implicitly to great effect.

Halmos [66, p. 213, theorem B] gives Lévy's statement and original proof, expressed in terms slightly pasteurized and less esoteric than in the inimitable original, to which one will nevertheless want to turn to taste the salt and the bitterness of Lévy composition. Lévy's proof is the first "direct" demonstration of the almost sure theorem. Jessen acknowledged this in the end, at least tacitly.

We have borrowed the adjective "pasteurized" from G. Choquet's very beautiful foreword to the abridged edition of the Lebesgue-Borel correspondence [115, p. 5]:

> Mathematical activity cannot be reduced to the pasteurized theorems that sleep in the journals of libraries; their genesis, which would reveal the operation of creative thought, seldom appears in printed statements. The mass of those millions of theorems resembles those coral reefs that grow every day but deteriorate as soon as the living corals that secrete them die.

Pasteurization is necessary to ensure safe distribution on a large scale, but it tends to sterilize the life the historian should tell about. Regulations from Brussels or Geneva can do no good here; a history of pasteurized mathematics misses the point and can be at best only a pasteurized history of mathematics. To recover the life and the creative thought, we must look elsewhere, and letters are an invaluable resource. This is why Gustave Choquet, in particular, fought against winds and tides to have the Lebesgue correspondence published, for without it one can hardly grasp the genesis of one of the most fertile theories of the 20th century, a raw-milk theory that could survive pasteurization only in a discolored form, diminished and tasteless. We should add that Choquet was a rarity among Paris mathematicians in welcoming the new martingale theory into his seminar on potential theory at the end of the 1950s and in encouraging the work of Meyer, Courrège, Dellacherie, and so many others.

values of λ . The function f has a generalized Fourier series, $\sum a(\lambda)e^{i\lambda x}$, which satisfies the Parseval equation of classical Fourier analysis: $M(|f(x)|^2) = \sum |a(\lambda)|^2$. This remarkable result, which was immediately re-derived and extended in various directions by the greatest analysts of the day, was enough to make Bohr's name. It would be the subject of Jessen's "magister" thesis, which marked the end of his schooling in the Danish system.[4]

2.1 Magister Thesis 1929

Bohr's almost periodic functions arise as natural extensions of ordinary periodic functions, which have only one frequency, and the quasi periodic functions of Bohl and Esclangon, which have a finite number. By extending the results of these last authors to the case of a countable number of frequencies [17,18], Bohr showed that if f is an almost periodic function in his sense, it can be written

$$f(x) = F(x, x, \dots, x, \dots),$$

where $F(x_1, x_2, \dots, x_n, \dots)$ is a function in an infinite number of variables, periodic in each variable, or can be uniformly approximated by such functions. It follows that Bohr's almost periodic functions are the only functions that can be approximated uniformly by generalized trigonometric sums. Thus almost periodic functions may be seen as the restriction to the diagonal of the periodic functions on a torus in infinitely many dimensions, an infinite annulus, to which an appropriate Fourier theory might apply. But the generalized Fourier series of almost periodic functions converge in general no more than the Fourier series of ordinary periodic functions do [90,151,152]. To obtain a satisfactory theory, it would be necessary to do for Bohr's theory what Lebesgue had done for Fourier's [111,113].

[4] On Harald Bohr, see the *Dictionary of Scientific Biography* and [83,156]. Bohr was the first director of the University of Copenhagen's Institute of Mathematics, founded in 1934. He was very important not only in the development of Danish mathematics, but also in the beginning of the "internationalization of mathematics". In the early 1930s Harald and his brother Niels were influential advisers to the Rockefeller Foundation. Jessen's theorem, in its way, was a concrete expression of this new way of doing of mathematics, based as it was on contacts among the principal schools of the old and the new worlds [166].

In 1949 [82], Jessen cited proofs of Bohr's theorem by Bochner, Riesz, La Vallée Poussin, Weyl, and Wiener and proposed another. The literature on this subject during the ten years beginning in 1925 is very important. Our bibliography contains only a small sample of titles, but among them is a paper by Ellen Pedersen, Jessen's future wife [150]. Notable are Stepanoff's almost periodic functions [176], Paley's and Wiener's pseudo periodic functions of [147], …. In the 1930s, von Neumann and Weyl showed the links between Bohr's functions and the theory of group representations [179, Chap. VII]. After the war, Bohr's functions were extended to distributions [164, Chap. VIII S9].

Biographers of Harald Bohr never fail to recall that he was a member of the Danish soccer team at the London Olympic Games of 1908, which beat the French team 17-1 in the semi-finals, after beating it 9-0 in the first round. There were five teams in the tournament, and Denmark lost to Great Britain 2-0 in the final.

It is easy to imagine experienced analysts having this idea and dismissing it. Lebesgue measure in itself did not extend to infinite dimensions and Daniell's integral and Gateaux's means could not take its place. So perhaps Bohr did not particularly push Jessen towards this apparent dead end. Jessen's magister thesis begins with an exposition of recent developments in the theory: six chapters that could be considered the thesis proper. But he added (in extremis?) a seventh chapter, independent of the others, "On functions of infinitely many variables," which contained in particular the first version of "Jessen's theorem."

Jessen was 21 years old, and his education was over. He had already published some elegant short articles, and Bohr had thought enough of him to have involved him in his own work on the zeta function, which had introduced him to functions in infinitely many variables.[5] He was brilliant and naïve and able to pursue a new idea, unencumbered by too much knowledge and prejudice.

To produce a Fourier theory for functions $f(x)$ defined on the torus Q_ω in infinitely many dimensions, where the variable $x = (x_1, x_2, \ldots, x_n, \ldots)$ is an infinite sequence of real numbers modulo 1, we must first define the integral of such functions. In the spring of 1929 Jessen knew the Lebesgue integral for one or a finite number of variables[6] but he was completely unaware of the Daniell integral, as he later told Lévy—see below. So he goes off in the first direction that offers itself, as though he had said to himself right off that the simplest procedure is to integrate f successively, starting with the first coordinate and then the next, all the way to infinity. Thus he considers the sequence of "Lebesgue integrals"

$$\int \mathrm{d}x_n \cdots \int \mathrm{d}x_2 \int f(x_1, x_2, \ldots)\mathrm{d}x_1,$$

in which f is a function defined on Q_ω, integrable in the sense of a theory yet to be born. If the sequence converges, in a sense to be made precise, it can only be to the desired integral. This is a first informal formulation of the martingale convergence theorem for downward martingales, but we must not get ahead of ourselves.

---

[5] So Jessen explains in [81, p. 252]: "The present author was led to the theory in connection with some investigations by Bohr concerning the distribution of the values of the Riemann zeta-function, which were carried out in collaboration with the author."

Functions in an infinite number of variables have a long prehistory, which includes Poincaré's theory of infinite determinants [153,160], but their history really begins in 1906 with Hilbert's theory of integral equations [160], in which Hilbert introduced and used the Hilbert space ℓ^2 [42,60,73]. But the idea of using functions in an infinite number of variables to study the Riemann zeta function and almost periodic functions appears to be Bohr's. He was followed by Jessen, who must have exceeded all his mentor's hopes.

[6] Jessen knew the Lebesgue theory from Carathéodory's treatise [32], of which he had made a thorough study, as Christian Berg tells us [12], and also, it seems, from Julius Pal, a Hungarian mathematician established in Copenhagen, who had taught Jessen. Pal worked with the Hungarian School and in particular with Riesz and was very familiar with Riesz's new functional analysis, where the Lebesgue integral played a central role [59].

It was now a matter of formalizing the idea. Jessen had read and studied Riesz's fundamental 1910 article [159], with his teacher Julius Pal or on his own. Riesz presented the theory of L^p spaces and of L^2 in particular, which, by the Riesz-Fischer theorem of 1907, is isomorphic to Hilbert's space ℓ^2. Riesz first treats functions of a real variable, and in a final paragraph on pp. 496–497 he extends the whole theory to functions of n variables, using what he calls a transfer principle ("Übertragungsprinzip" which Jessen translates as "Overförelsesprincip" in his magister thesis [76, p. 44] and as "Transferring principle" in his [81, §7]). The principle establishes (without proof, considered unnecessary) an (almost) bijective and measure-preserving correspondence between any bounded interval on the real line and an n-dimensional cube of the same measure, thus permitting the automatic transfer of all real theorems to the vector case.[7]

[7] Riesz's article [159] is the only reference on the transfer principle that Jessen gives in his magister thesis. But by the end of 1929, and undoubtedly after his visit to Riesz, he knew that Lebesgue and La Vallée Poussin had used such a principle around the same time. In his 1930 doctoral thesis [77, p. 19], he cites Lebesgue [114, p. 402ff], who used the Hilbert curve to transfer his theorems on differentiation, and La Vallée Poussin [102], who did substantially the same. When he corrected the proofs of his 1929 Oslo talk in 1930, Jessen added the same references in a note [78, p. 134, note 1].

But questions of attribution and dating are never simple, especially when they concern a principle like this, which imposes itself naturally. Already in a 1907 note [157], Riesz had used his transfer principle to go from functions of a single variable to functions of two variables. In an 1899 note [106], Lebesgue had used a "principe de transfert" to extend Baire's theorem for functions of one real variable to the case of two variables. For this purpose he used the Peano space-filling curve [149], still without seeing—he had no need to—that this curve preserves the measure that Borel had just defined in his course [21]. Lebesgue returned to the method, extending and improving it, in his famous 1905 article [112, pp. 193–201], where again measure does not intrude. But beginning with the first edition of his *Intégration* in 1904 [110, pp. 116–117], Lebesgue used the Peano curve to construct the plane measure he needed for the "geometric definition" of his integral; see also [108]. This method is used again in the second edition of the same book in 1928, on pp. 137ff. In this same edition of 1928, on p. 44, Lebesgue proposes a construction of a curve filling the square "different from that of MM. Peano and Hilbert, [which] can be used for spaces with an unspecified number of dimensions and even for spaces with a countable infinity of dimensions". This note, written in 1926, could be a version of Jessen's generalized principle of transfer, and as such would be neither the first nor the last, but it could also be a tired old cat's swipe of the paw, to push away that odd bird, Lévy, who had presented at Hadamard's seminar in 1924 related ideas that enter very much into the correspondence we are publishing. This second hypothesis would lend some credence to Lévy's claim of priority that we will be discussing. Perhaps Lebesgue recognized in Lévy's talk precisely the idea that Lévy would claim credit for ten years later, but that neither of them had taken the trouble to write down. Borel had become a mandarin of the radical Republic, and Baire had isolated himself to die on the shore of Lake Léman. They were no longer there to play mathematics with Lebesgue; Lévy was a bad player, like Lebesgue.

On the history of the transfer principle there is the very nice article by Riesz [162]; see particularly pp. 37–38. The principle was put in abstract form as the "isomorphism theorem" by Halmos and von Neumann at the beginning of the 1940s [67] and independently by Rokhlin around 1940 [163]. Halmos gives a formulation in [66, §41], and there are more references in [15, Vol. II, p. 549]. The latter is remarkably erudite, with very interesting historical comments and a bibliography of over 2000 titles.

Jessen's idea was to extend this principle to the infinite-dimensional case. His plan was to define a set function on Q_ω that coincides with Lebesgue measure on cylinder sets whose base depends only on finitely many coordinates, and to construct a correspondence, bijective up to a set of measure zero, between the unit circle, the torus in one dimension, and the torus of infinite dimension, which preserves the (pseudo) measure thus constructed and transfers to it all the properties of Lebesgue measure, in particular countable additivity. The result is a theory of integration with all the properties of the Lebesgue integral. In his magister thesis as in [77,78,81], Jessen sets off from the natural measure of the generalized intervals of $Q_\omega : a_i \leq x_i \leq b_i$, for a finite number of indices i, and by using the compactness of Q_ω and the Borel (-Lebesgue) covering lemma, he constructs a set function on Q_ω by the process of outer and inner measures, closely following Lebesgue and Carathéodory [32,110]. (See also [66, Chap. 2].) This set function is not yet a true measure, but it becomes so (and can consequently be extended in the manner of Carathéodory) by transfer, after Jessen had established a continuous correspondence between the intervals of the circle and the generalized intervals of Q_ω by extending Hilbert's curve [72] to infinite dimensions. Finally and still following Carathéodory, Jessen constructs the integral on Q_ω for "summable" functions, which he writes as $\int_{Q_\omega} f(x)\mathrm{d}w_\omega$.

It remains to establish a link between this integral and the process of successive integrations defined above and deduce from it, if possible, the Fourier-Lebesgue theory for functions on the infinite torus and its applications. This Jessen outlined in his magister thesis and developed brilliantly in successive articles leading up to his [81] in 1934. But before taking a closer look it may be helpful to put Jessen's theory and his principle of transfer into a somewhat broader context.

Retrospectively at least, there is nothing astonishing in Jessen's discovery in the spring of 1929 of the direct transfer of the infinite dimension to dimension one. This had already been Borel's starting point in 1909 when he constructed the infinite game of heads or tails [22]: we can associate with any infinite sequence of heads and tails an expansion in base 2 of a number in the interval [0, 1], and the dyadic subintervals of this interval correspond to sequences for which the first outcomes are fixed. In 1923 Steinhaus had made this rather informal correspondence precise as rigorously as desirable, extended it, and used it to study series of terms whose signs are drawn from an urn "that always contains as many plus signs as minus signs".[8]

It was again Steinhaus (a little after Jessen or even a little before) who extended Borel's transfer principle to the space Q_ω put in quasi-bijective measure preserving mapping with [0, 1], in order to study series with terms drawn at random from the unit circle:[9]

[8] Steinhaus's article [170], which used Sierpinski's axiomatic set-up of [167], played a very important role in the development of mathematical probability, especially in Moscow, where it was followed by Khinchin's [93] and then Kolmogorov's [94,95], etc.

[9] See Steinhaus's [172,173], included and extended in [85, Chap. 4, §7, pp. 134–139], which gives the construction of the correspondence, and [174,175]. Jessen cites [173] in his 1930 thesis [77, p. 29 note] and in [81]. In his letters to Lévy we find Jessen recognizing the independent priority

$$\sum_{k=1}^{\infty} a_k \mathrm{e}^{2\pi i x_k}.$$

It suffices to expand each x_k in $[0, 1[$ in base 2 (for instance), then to reconstruct, starting from these expansions written one below another, a unique x in $[0, 1[$, moving zigzag in the array thus formed. This is one of the traditional ways, but the choice has no importance, provided that it rewrites the initial two-dimensional array on a line. One thus obtains a generalized "Peano curve", an almost bijective correspondence, which evidently preserves measure, or rather transforms Lebesgue measure into the product probability that governs the drawing of the terms from the urn. Indeed, in both cases, linear order or planar array, one is dealing with the same sequence of independent Bernoulli random variables with the same law, except for a rearrangement which leaves the law invariant. The measure on Q_ω is no other than the one that governs a countable sequence of plays of heads or tails, i.e. (if one wants to avoid the notion of chance), again appealing to Borel's principle of transfer, Lebesgue measure on the unit interval (which is above suspicion). So the measurable sets of the interval $[0, 1]$ and those of the space Q_ω that correspond by transfer have the same measure.

Thus the principle of transfer that Jessen developed in an elegant and rigorous way in his own framework should not surprise us. There are earlier, later and contemporaneous versions, as is common in such cases. Yet we will see that Lévy asserted his own paternity of the principle, which he called the "principe de correspondance", having used it since 1924 and perhaps earlier, in an unpublished course given at the Collège de France in 1919 (and at the time of his childhood walks in the Luxembourg garden?). Priority in the matter of this principle is one of the themes in the correspondence published here. Lévy seemed to consider it a secondary point, even as he constantly returned to it. This was evidently not so for the young Jessen, who would surely have wanted his priority recognized for a principle that he had discovered on his own. In any case, more than any else and without the least visible trace of probabilistic intuition or reasoning, Jessen was guided by the theory of the Lebesgue integral and the principle of transfer, which form the basis for his theory.

Back to Jessen's magister thesis. In §5 of the last chapter, Jessen sketches an outline of a theory of the differentiation of set functions associated with the integral

---

of Steinhaus, then that of Lévy (more arguable) and that of others, known or unknown, (Denjoy, Wiener, Cantelli, Mazurkiewicz, etc.).

We may note that in 1936 [174], Steinhaus proposed another correspondence based on the generalised Peano curve and suggested that Jessen's earlier construction in [78], based on the generalized Hilbert curve, seemed less well adapted.

From his construction in [172] Steinhaus deduced that *in general* entire series whose coefficients have arguments chosen at random have a singularity at all points on their circle of convergence. This gave a precise sense to Fabry's prophetic statements [55, pp. 398–399] and Borel's enigmatic ones [19,20]. Steinhaus's results were developed soon after by Paley and Zygmund [148], followed by many others, including Jessen [81]. For more recent references and developments, see [86,87,91,141].

on Q_ω, following an increasingly fine net of generalized intervals. Jessen transfers La Vallée Poussin's method directly.[10] More precisely, if f is integrable on Q_ω, we write following Jessen $F(E) = \int_E f(x)\mathrm{d}w_\omega$, where F is an additive set function on Q_ω established by transfer on the unit circle, with obvious notation: $\int_E f(x)\mathrm{d}w_\omega = \int_e \varphi(t)\mathrm{d}t$.

We define on Q_ω a net of generalized intervals $\overline{I_n}$, corresponding to a net of intervals on the unit circle, and form the associated step functions:

$$\Delta_n(x) = \frac{F(i_n)}{w_\omega(i_n)}$$

if x is in i_n, an element of the net $\overline{I_n}$. By transfer of the Lebesgue-La Vallée Poussin differentiation theorem, it follows that $\Delta_n(x)$ converges almost everywhere to $f(x)$.

In the following section, Jessen transfers Fubini's theorem to infinite dimension, in the obvious way that one expects, by collecting the coordinates of Q_ω in a finite number of packets $(Q_\omega^{(1)}, Q_\omega^{(2)}, \ldots, Q_\omega^{(n)})$ and writing the corresponding Fubini theorem.[11]

Having defined convergence in measure[12] and convergence in L^2 for Q_ω endowed with the Jessen measure, thus complete by transfer, Jessen can undertake the statement and proof of Jessen's theorem.

We have reached §8. Jessen defines an integrable function f on Q_ω. By the Fubini-Jessen theorem applied to the decomposition of Q_ω into two blocks, the block

[10] Jessen quotes [104,105] and uses La Vallée Poussin's terminology and method of "net derivatives" [105, Chap. IV]. La Vallée Poussin's theorem, like Lebesgue's, is a theorem of increasing martingales, La Vallée Poussin's nets of intervals being filtrations of finite type and the $\Delta_n(x)$ being conditional expectations given these filtrations. Jessen finally acknowledged this in 1945.

[11] The first "Fubini's theorem" for the Lebesgue integral in the plane is in Lebesgue's thesis [107, §37–40], and it treats the case of bounded measurable functions. In 1910 Lebesgue returned to the subject, but meanwhile several authors had stated and proved Fubini's theorem in a more or less complete way, in particular Beppo Levi in 1906, Hobson and Fubini in 1907, Tonelli in 1909. Hawkins studys this complex development in detail and finally attributes to Tonelli the first complete demonstration of the theorem stated by Fubini, for integrable functions of two variables [69]. The modern formulation of Fubini's theorem is due to La Vallée Poussin [103,105]. Fubini is credited only for his extension of the theorem to nonmeasurable functions, which is also partly in Lebesgue's thesis [107, §40], as La Vallée Poussin points out in [105, p. 53, note 1].

[12] In a note Jessen indicates that he is following Riesz [158], who was actually the first to name "convergence in measure", but the concept had already been used, without being named, by Borel and Lebesgue in 1903. See [15, Vol. 1, p. 426] for references. The relationships between convergence in measure and the other modes of convergence of the theory of functions—convergence almost everywhere, convergence in mean,...—were published many times by various authors. They are laid out in Fréchet [61], who was one of the first to show that convergence in measure "corresponds" to (without being identified with) convergence in the sense of Bernoulli's, Moivre's and Laplace's probability calculus, the Laplacian double approximation of "très probablement très proche." See Slutzky [168], Fréchet [62], and Cantelli [31]. The two concepts fused in Kolmogorov's axiomatic framework [97] and also in that of Jessen-Steinhaus-Lévy (1930/1937) but they were understood differently for a long time. See Doob [49]. For Jessen, as for Doob, the problem does not arise: it has no sense.

Q_n of the first n coordinates and the block $Q_{n,\omega}$ of the infinitely many following coordinates, if we write $w_{n,\omega}$ for the Jessen measure on $Q_{n,\omega}$, then the integral

$$\int_{Q_{n,\omega}} f(x_1, x_2, \ldots, x_n, x_{n+1}, \ldots) \mathrm{d}w_{n,\omega},$$

is an integrable function on Q_ω equipped with Lebesgue measure.

Jessen states that this sequence of integrals converges in measure towards f, (p. 50). This is the first known version of the theorem of (increasing) martingales in a general framework. It can, at any rate, be made as general as one wants.

Jessen's proof is interesting though still a little awkward. It anticipates the first part of the modern proof, as we recalled it in the introduction (Jessen did not yet see convergence in mean). The idea is to approximate f by a sequence of functions depending only on the n first coordinates so as to obtain a bound in measure. Jessen obviously hesitates over the nature and generality of such an approximation. For this he uses the step functions $\Delta_n(x)$ defined in §5. These functions, by construction, depend only on a finite number of coordinates and converge almost everywhere to f. For want of anything better and pressed by time, Jessen contented himself with convergence in measure of the sequence $\Delta_n(x)$, treating first the case of a bounded function f and then removing the truncation. The proof is correct but unnecessarily complicated and restrictive.

The next section proves the proposition from which Jessen no doubt began in building his theory of integration, the first known form of the theorem of downward martingales. This is still about convergence in measure:

> The sequence of integrals $\int \mathrm{d}x_n \ldots \int \mathrm{d}x_2 \int f(x_1, x_2, \ldots) \mathrm{d}x_1$ converges in measure towards the integral $\int_{Q_\omega} f(x) \mathrm{d}w_\omega$.

Jessen again truncates and uses the inequalities established in the previous section.

The last section treats the Fourier theory of functions on the infinite torus. Jessen establishes a Parseval equation (p. 54) and a Riesz-Fischer theorem for his framework.

In broad outline, the entire theory of [81] is already present in chapter 7 of Jessen's magister thesis, which can truly be called masterly. Bohr was certainly very impressed.

2.2 Doctoral Thesis 1930

After such an achievement and with the support of Bohr with his well-known academic clout, one can imagine that Jessen was propelled at once to the firmament of new mathematical stars, at least in Copenhagen. He was invited to make a presentation to the seventh Congress of Scandinavian Mathematicians, held in Oslo from 19 to 22 August 1929.[13] Jessen presented his theory of integration in German and in a particularly clear way, a very nice exposition with a reproduction of Hilbert's space-filling curve as it appeared in the original article of 1891, showing at first glance that measure is preserved, the curve preserving throughout construction a perfect symmetry between the two axes. Jessen did not state his two theorems, which he undoubtedly considered marginal, but announced a Fourier theory for functions with a countable infinity of periods. The transactions of the congress were published in 1930, so that at the proof stage Jessen could add a reference to Daniell, which he had learned about in the interim. Lévy read this paper in 1934, and we will see that this reading played a very important part in the emergence of his theory of measure as expounded in his great treatise of 1937 [134].

As we learn from [12], Jessen obtained a grant from the Carlsberg Foundation in 1929 to travel in Europe, first to Szeged where he met F. Riesz, one of the principal inspirers of his thesis, then to Göttingen, from where Hilbert reigned over universal mathematics, and to Paris to "see" Lebesgue, who hardly saw him. But Bohr soon recalled him to Copenhagen. A position of docent (university lecturer) in mathematics was about to become vacant at the Royal Veterinary and Agricultural School in Copenhagen. Considering the scarcity of positions and the good health of their occupants, this was an exceptional opportunity that was not likely to be repeated soon and could not be allowed to escape.[14]

Jessen's habilitation thus became an urgent matter. For Bohr there was no doubt that the final chapter of the magister thesis was already a doctoral thesis. It was

[13] In the list of Oslo lecturers is another young Danish mathematician, Georg Rasch, (1901–1980), slightly older than Jessen, who had submitted his magister thesis in 1925 and was due to submit his doctoral thesis in the coming months. He was appreciated and supported by Nørlund, a professor at the University of Copenhagen from 1922 in a chair specially created for him. If one of the very few mathematical positions in Danish universities became vacant, Rasch had a reasonable chance of obtaining it. But Rasch's mathematical career was destroyed at a stroke in the spring 1930 when Jessen, with Bohr's backing, submitted his doctoral thesis—and what a thesis! Undiscouraged, Rasch turned to the new Anglo-Saxon statistics of Fisher, and also of Neyman, Pearson and others. This was an unknown discipline in Denmark, which remained attached to the continental school of statistics and actuarial science. Like Jessen and at the same time, Rasch obtained a Rockefeller scholarship but in his case it was to study with Fisher in England. When he returned he trained in the new methods the leading Danish statisticians of the next generation, including Anders Hald, an important statistician and a remarkable historian of statistics, who became a friend. Rasch was eventually appointed university professor of statistics in Copenhagen, but only in 1962 and only in the Faculty of Social Sciences. It seems a just return that Rasch's posthumous fame stands a good deal higher than Jessen's. Only scholars know Jessen's theorem while Rasch's models are cited, applied, and extended every day. On the life and work of Georg Rasch there is an interesting thesis by Olsen [146].

[14] Bohr's appointment to the University of Copenhagen in 1930 had set off a chain. His old position at the Polytechnic School went to A. F. Andersen, who was docent at the Royal Veterinary and Agricultural School, and Jessen was at appointed to replace him there. Aksel Frederik Andersen (1891–1972), an analyst of Bohr's school, was very interested in mathematics teaching in both school and university. In particular he took part in revising Bohr's and Mollerup's great treatise on analysis. He retired in 1960.

enough to improve the presentation and to print the whole in Copenhagen. The university's approval was needed and this was obtained on 25 March 1930, signed by J. F. Steffensen, professor of actuarial sciences. The thesis was submitted before the deadline and Jessen was appointed docent. He was 22 years old.

Jessen's doctoral thesis is entitled "Contribution to the theory of the integration of functions of an infinity of variables" [77]. It is in three parts. The first treats integration of functions of n variables, by transfer of the Lebesgue integral in one dimension. This part has no new results but serves as canvas for the next part, which treats functions in an infinite number of variables, taking up again all the results from chapter 7 of the magister thesis that we just discussed. The last part develops applications, in particular to the theory of almost periodic functions. Let us briefly examine the second part, which has a new version of Jessen's theorem.

The beginning of the second part repeats the construction of the integral on the infinite torus following the magister thesis. Things only start to change in §13. There, on pp. 37–38, Jessen proves a theorem that he says he owes to F. Riesz, which affirms that the quotients $\Delta_n(x)$ defined above converge "strongly" towards f, in the sense of convergence in L^1. The proof is very simple. If f is bounded by a constant, the sequence $\Delta_n(x)$ is bounded by the same constant and the result follows from the dominated convergence theorem. If not, it is sufficient to truncate f and to take the limit of the truncated versions.[15]

In this way, Jessen can prove in his §§15 and 16 his theorems for convergence in L^1 for the increasing case just as we did in our Section 1, the approximating step functions depending on only a finite number of coordinates being precisely the quotients $\Delta_n(x)$. Thus in the spring of 1930 Jessen has the statement and proof of the martingale theorem in a form that will hardly ever be further improved, except for the framework.[16] Riesz's intervention was no doubt crucial, but the idea was Jessen's.

[15] The theorem on differentiation in L^1 was not given by Lebesgue, who did not use convergence in mean, but it was known to Riesz from 1910 and undoubtedly earlier, as it was obvious once you had convergence almost everywhere (Lebesgue) and convergence in L^1 had been defined (Fischer, Fréchet, Riesz, Schmidt). Just the same, this result does not appear explicitly in the few works that Jessen had read, and he only learned of the role of "strong" convergence during his stay in Szeged; see footnote 20 below.

[16] As we saw in §1, the form of the approximating step functions matters little, but Jessen does not know yet, it seems, that the functions depending only on a finite number of coordinates are dense in L^1, by construction, an argument from measure theory that Lévy grasps and uses at once for his "lemma", but which is still appearing nowhere else in the current mathematical literature. So Jessen, without realizing it, proves the martingale theorem in L^1 twice, first in the form of the theorem on differentiation, following Riesz, and then in the form of Jessen's theorem, which he deduces from it. From the time of Cournot at least, it has been known that this complication is characteristic of mathematics in the wild, which discovers theorems by chance at a bend of a path in the woods and then hastily assembles whatever arguments are available to prove them. It is work of a different kind to produce simple, clear and well-organized proofs; these seldom come first. In this art too Jessen became a master, to the point perhaps that he forgot the thick jungle where hide the new theorems, true or false.

The third part of the 1930 thesis is extremely interesting, but to do it justice would take us too far from our subject. Jessen presents in detail his Fourier theory for functions defined on the infinite torus and proposes various applications which he would develop more fully in 1934 [81]. Among the applications were Weyl's equi-distribution theorem [180], which he might have learned in Göttingen, almost periodic (random) functions of a complex variable, which are represented in the form of a series $f(s,x) = \sum_{k=1}^{\infty} a_k e^{2\pi i x_k} e^{\lambda_k s}$, where s is the complex variable, and $x = (x_1, x_2, \dots, x_k, \dots)$ is a "parameter" formed of an infinite sequence of real numbers modulo one, and the Riemann zeta function, his work with Bohr that first motivated his theory.[17]

2.3 The Acta Article 1934

Jessen married in 1934, and he needed to prepare his courses for the veterinary school. This may have slowed down his work but did not stop it. Just the same, Jessen does not seem to have returned right away to his theory of integration. Was he discouraged by discovering Daniell's work, which had preceded him, and Steinhaus's work, which had done about the same things (aside from Jessen's theorem) at the same time? In any case, Jessen presented the results from the third part of his thesis in his beautiful presentation to the International Congress of Mathematics of Zurich in 1932, completing these results by taking account of the papers by Paley and Zygmund [148]. He referred, naturally, to his work on integration on the infinite torus, (with a geometrical representation) but without emphasizing it or even mentioning Jessen's theorem. He must have thought, not without reason, that the time was not yet ripe—or worse still, that it was of no interest.

For the academic year 1933–1934 Jessen was a Rockefeller fellow with G. H. Hardy in Cambridge, England, and at the newly established Institute for Advanced Study in Princeton. Under the leadership of John von Neumann and Hermann Weyl, the Institute was wresting mathematical supremacy from Göttingen, which was sinking under Hitlerism. Thus Jessen had the opportunity to mix with some of the leading analysts of the day—von Neumann, Hardy, Weyl, Wiener, Daniell, Bochner, Besicovitch, etc.—and brilliant young people from all over the world. Here Jessen found himself in an atmosphere of great mathematical euphoria and intense activity. He no doubt realized that his direct way of tackling the problems and the principle of transfer, in which he had hardly any longer believed, had led him to results that neither

[17] Jessen developed these themes in 1932 and 1934: [79] and the last part of [81]. They were taken up again by Hunt [74] and are still the object of research; footnote 9 above gives references.

Jessen's 1932 article, [79], was inspired in particular by a very fine article by Jensen [75]. It was rumored that Jensen had proved Riemann's conjecture and on Jensen's death, Rasch was charged with seeing what could be found in the many papers he left. Interesting details are related by [146]. It may be recalled that J. L. W. V. Jensen (1859–1925) was chief engineer at the Copenhagen Telephone Company, where A. K. Erlang also worked. For the latter see [27].

analysts nor probabilists,[18] who saw things differently, had yet caught sight of, so that it would be interesting to showcase these results in a work of synthesis written in English and presenting all the theory and applications as a self-contained coherent whole. All the more so because he had improved his theorem and found unexpected applications. This was the rationale for his article, "The theory of integration in a space of an infinite number of dimensions" [81], printed on July 6, 1934, in volume 63 of *Acta Mathematica*, which contains the definitive version of "Jessen's theorem" on Q_ω. We need to examine it.

Jessen began by remarking that in the previous fifteen years the theory of integration in infinite dimensions had been considered by several authors who exploited in various ways a principle of direct extension in the space under consideration.[19] Jessen, for his part, intended to remain faithful to the theory developed in his thesis, based on a "transferring principle" which allowed him to go from the interval $[0, 1[$, where the Lebesgue integral is available, to the infinite-dimensional torus Q_ω, the correspondence automatically transferring any one-dimensional theorem to an infinite-dimensional theorem, and inversely. Jessen uses, as in his thesis, a procedure of finer and finer successive partitions of the space, in *nets* of generalized intervals which are put in correspondence with nets of intervals in $[0, 1[$, so that the construction of the Lebesgue integral passes to infinite dimension by simple translation. Jessen remarks (§9) that this integral enjoys the Lebesgue property of differentiation: in the limit following a sequence of increasingly fine dissections of the space, the derivative of a primitive recovers the function almost everywhere. As we saw, this is an immediate application of the principle of transfer, but it is also the basis for almost all the others (and it is a martingale theorem, as we have already remarked).

All the preceding results had appeared in Danish in Jessen's two theses, but §11, with the title "An Important Lemma", contains a new result with no natural analog

---

[18] In Princeton, Jessen discovered the new theory of probability, especially Kolmogorov's work on series of independent variables, which encompassed his own generalized Fourier series with variable coefficients on the circle (which varied freely even though they were not drawn randomly). He also had occasion to meet Wiener. According to the *Bulletin of the American Mathematical Society* for March 1934 (pp. 177–178), one of the sessions (on December 27) of the Society's meeting at MIT (December 26–29 1933), had been a "Symposium of invited papers on the topic of probability". The invited contributors were: E. Hopf, MIT, "Remarks on causality and probability"; F. Bernstein, Columbia University, "Foundations of probability in the natural sciences"; G. E. Uhlenbeck, University of Michigan, "The probability of position in a canonical ensemble"; and N. Wiener, MIT, "The Brownian motion". A fifth lecture had been planned, "Some analytical problems relating to probability", to be given by Dr. B. Jessen, Institute for Advanced Study, but it did not take place "on account of illness". These lectures (including Jessen's) were published in the *Journal of Math. and Phys.*, 24(1):1–35 (1935). Jessen's lecture (pp. 24–27) is a short summary of his 1934 article. Thus Jessen was now fully aware that the theory of measure in infinite dimensional spaces was related to the new theory of denumerable probabilities. He cites Kolmogorov's *Grundbegriffe* in his 1934 article, but as we will see, he clearly had not had time to study it in detail.

[19] See Daniell [38], who uses the extension of the integral in the manner of W. H. Young and F. Riesz, and Carathéodory [32], Feller and Tornier [58], and Kolmogorov [97], who use the extension of the measure.

in finite dimensions. It will be debated at length in the correspondence presented below. The lemma states that a measurable function defined on Q_ω that takes the same value on any two points differing on only a finite number of coordinates is constant almost everywhere. We see immediately that if this function is the indicator of a set, the lemma is only a special case of Kolmogorov's 0–1 law [95,97], but Jessen does not mention this in 1934. He refers only to a related result by Steinhaus [171]. His proof uses §9's differentiation theorem with suitable dissections. It seems that it was this lemma that first attracted the attention of Lévy, who at once re-obtained it by a "direct" probabilistic method. Lévy did not refer to Kolmogorov's work either. It would have been possible for him to know it, as he had cited Kolmogorov's 1928 [95] in 1931 [123], but he had not really read it, as he acknowledged in his intellectual autobiography [135, p. 87]. We will return to this point.

Then §§12, 13 and 14 treat variations on "Fubini's theorem" in infinite dimensions. We pass over the statement in §12 of Fubini's theorem itself, which had already appeared in the magister thesis, to the following sections, which contain at last the almost everywhere version of "Jessen's theorem."[20] In §13 Jessen states that

$$\int_{Q_\omega} f(x)\mathrm{d}w_\omega = \lim_{n\to\infty} \int \mathrm{d}x_n \dots \int \mathrm{d}x_2 \int f(x_1, x_2, \dots)\mathrm{d}x_1 \text{ for almost all } x \text{ in } Q_\omega.$$

Jessen's proof is inspired by a proof of Kolmogorov's [95,96] that uses measure theory, a subject whose richness our author seems to have understood during his stay in Princeton.

The next section, §14, "Representation of a Function as the Limit of an Integral," is the main topic in Jessen's correspondence with Lévy. It is the counterpart to the preceding result: when one integrates less and less beginning from infinity, one recovers the function almost everywhere, or

$$f(x) = \lim_{n\to\infty} \int_{Q_{n,\omega}} f(x)\mathrm{d}w_{n,\omega} \text{ for almost all } x \text{ in } Q_\omega.$$

Jessen's proof is rather complicated. It follows Riesz's proof of a result in Fourier analysis in Q_ω that can be seen as a special case of Jessen's statement and that

---

[20] In a note to §15 (page 278), Jessen indicates that he had originally proven his two "Fubini theorems" for convergence in measure and adds:

> It was pointed out to me by Prof. F. Riesz that the (well-known) argument used above would give the same theorems for the more convenient concept of strong convergence. Finally it was Prof. Daniell who suggested to me that the theorems should be true for convergence almost everywhere.

This suggests that Jessen developed his results during his stays in England and the United States in 1933–1934, with the uncompensated and benevolent assistance of Daniell, who would be thus the first to have had the idea of the almost sure theorem for Jessen's martingales. On Daniell's very remarkable personality, see John Aldrich's beautiful and very complete article [1].

Jessen proves in §18.[21] It relies essentially on §9's differentiation theorem. As their correspondence shows, Lévy tried to persuade Jessen that there is a direct Borelian proof, and we may conjecture that this partly inspired the new proof that Jessen then put to Lévy.

Then, in §15, Jessen shows that the two preceding theorems hold for "strong convergence"—convergence in L^p if f is in L^p, $p \geq 1$. In §16, Jessen establishes the maximal inequalities for martingales in L^p, now known in their definitive form as Doob's inequalities but which Jessen describes as analogues of Hardy and Littlewood's "well known" maximal theorem [68].

The remainder of the article is given over to applications. In spite of their very great richness, we cannot consider them here.[22]

In 1934 then, Jessen had a general theory of martingales in a particular framework, the space Q_ω and its partition into nets, with no mention of probability, random variables or conditional expectations.[23]

---

[21] Concerning Riesz's result, Jessen writes in a note on p. 285:

> A proof of the theorem by means of the differentiation theorem of §9 was given by Prof. F. Riesz and communicated to me by Dr. Kalmár. It was this proof that suggested to me the proof of the theorem in §14. I note from a letter from Prof. Zygmund that a proof on similar lines was given by Paley.

No doubt Jessen had met Laszlo Kalmár at the time of his stay in Göttingen, where Kalmár, a student of Fejér and Riesz, was discovering his vocation in mathematical logic. We do not know where Paley proved his "martingale theorem", but it is a theorem he must have known at least implicitly, given his admiration for Borel's "denumerable probabilities". Jessen probably never met Paley, who died before Jessen arrived in Cambridge.

Symmetrically, F. Riesz recommended Jessen's article [81] "for a detailed exposition [of the principle of transfer using the method of nets] (written for the case of an infinity of variables) and for bibliographical indications" [161, p. 193, note 5]. A faint recommendation that could have only half pleased Jessen.

[22] These applications principally concern Fourier theory in Q_ω, random Fourier series, developments based on articles by Paley, Wiener and Zygmund, but also orthogonal systems (e.g. [85,155]), random almost periodic analytic functions, Dirichlet series [33], etc. For these subjects, see the references in footnotes 9 and 17. On the other hand, the article contains no overtly "probabilistic" applications, unlike the article Jessen published a little later with Wintner [84], described below.

[23] It is not only a matter of difference in language; Jessen's martingale theory knows nothing of stopping times or the stopping theorem, the keystone of the probabilistic theory of Ville and Doob, but also of Lévy. On this subject see Bretagnolle [26, p. 241], who gives a very intelligent modern reading of Lévy's [128].

We know from [8] that Jessen noticed Doob's founding 1940 article [44], which itself does not yet contain the stopping theorem, only after the publication of Sparre Andersen's and Jessen's first joint article in 1946 [6]. Doob, for his part, does not seem to have had access to Jessen's work even in 1948. As for Ville's book [178], Jessen wrote a very short review for the *Mat. Tidsskrift* without, apparently, noticing any connection with his own work. Doob, in contrast, immediately understood the interest and originality of Ville's thesis.

Here is a difference of a philosophical or poetic nature. Jessen did not adhere to (or only very slightly and with evident discomfort) the philosophy of chance that Lévy inherited from his masters

2.4 A Probabilistic Interlude 1934–1935

In Princeton Jessen met Aurel Wintner.[24] Two more different mathematicians can hardly be imagined. Jessen was elegant, reserved, rigorous, scrupulous; Wintner was impassioned, a compulsive eater, overflowing with projects and works in progress, all with great fanfare. Wintner had always been interested in celestial mechanics and of course in almost periodic functions. For some time already, he had been studying the limiting laws of series of independent random variables of the type considered by Steinhaus, Jessen and others, what was called at the time the problem of infinite convolutions, on the line or in a finite-dimensional space. Wintner had obtained interesting results on the subject,[25] which Jessen had also addressed, both by himself and with Bohr. Sometime in 1934, probably in the spring, Jessen and Wintner decided to pool their experience on the topic, and they wrote an article, "Distribution functions and the Riemann zeta function", which was published in 1935 in the *Transactions of the American Mathematical Society*, having been received by the journal on July 9, 1934 and presented to the Society on April 20, 1935. The authors proposed to treat the problem of infinite convolutions by the method of Fourier transforms, which they tell us was first applied by Lévy in his 1925 book. This is inaccurate[26] but at least it indicates that the theory of the Fourier transform (we might call it the Laplace-Fourier-Poisson-Cauchy…transform), long considered suspect by mathematicians, had been revamped by the new analysis and was now well established. We will not examine the main part of the article but only the final two very short sections, where the authors undertake to show that the theory of infinite convolutions can be treated from the viewpoint of Khinchin's, Kolmogorov's, and Lévy's (probabilistic) theory

Bertrand, Poincaré and Borel [29]. Jessen does not draw the arguments in his series at random; they are parameters that he measures. This is more proper, but it deprives him of probabilistic intuition and the mathematical concepts linked to it. One can do the calculus of probability without chance, just as one can do mechanics without force, but, Cournot would add, what is gained in logical clarity is lost in richness of reasoning.

[24] Concerning Aurel Wintner (1903–1958), see the *Dictionary of Scientific Biography*. He is credited with 437 articles and 9 books. No doubt some are of only passing interest (Doeblin is more acid in his notebooks), but they testify to his astonishing publishing activity, and some of them are first class, those written with P. Hartmann for example, and undoubtedly those we are considering here. We do not know when Jessen and Wintner met. Wintner spent part of the academic year 1929–1930 on a postdoctoral scholarship at the Copenhagen Observatory, then directed by E. Strömberg. He certainly met Harald Bohr and may have been present at Jessen's doctoral defense in the spring of 1930. To learn more, it would be necessary to analyze the voluminous Jessen-Wintner correspondence, which we were not able to consult, in the Jessen papers at the Institute of Mathematics of Copenhagen and in the Wintner papers at the Milton S. Eisenhower Library of Johns Hopkins University, which also has correspondence with Lévy and Doob.

[25] The literature on this subject was very important in the 1930s and subsequently: see [89] and other work by Kahane in the bibliography, although these give only a sample of the work on infinite convolutions around 1935. There is a large bibliography in [84], of which we have transcribed only a part.

[26] The Fourier transform was introduced by Laplace in 1810 precisely for the purpose of evaluating the asymptotic laws of sums of independent random variables; see Hald's great work [64].

of sums of independent random variables.[27] There is evidently nothing surprising here, but matters were totally different in 1934, and our authors wanted to show with complete clarity the relationship between the two theories (infinite convolutions and sums of independent variables) in a new way, while placing themselves within the framework (more analytical) of the theory of integration in infinite dimensions. Let us follow them for a moment.

We begin with §15, which gives a brief account of the theory of measure and integration in general product spaces. The authors tell us that Jessen presented the theory in great detail in the particular case of the infinite torus [81] and will publish the general case "in a forthcoming paper". The framework is the theory of abstract measure, a set Q, a "Borel field" of subsets of Q, and a positive and countably additive set function m with total mass 1. Given a countable family of such spaces, $q_1, q_2, \dots q_n, \dots$, each supplied with a probability measure, μ_n , there exists on the product space $Q = (q_1, q_2, \dots q_n, \dots)$ a product measure with natural properties,[28] and one has Jessen's theorems, stated without proof, in particular the "important lemma" of §11, stated this time for set indicators:[29]

> If a measurable set of the product space contains all points differing from one another in only a finite number of coordinates, its measure is 0 or 1.

Naturally there are theorems from §13 and §14 of [81]; one of them (in self-explanatory notation) states that:

> If f is an integrable function defined on Q, and if one puts $t = (t_1, \dots, t_n, t_{n+1}, \dots) = (t_n, t_{n,\omega})$ for an arbitrary point of Q, then,
>
> $$f_n(t) = \int_{Q_{n,\omega}} f(t_n, t_{n,\omega}) m_{n,\omega}(\mathrm{d}t_{n,\omega}) \to f(t)$$
>
> for almost all t in Q.

---

[27] Jessen and Wintner cite [94–96, 123], works with which they were only superficially acquainted.

[28] The existence theorem for an infinite product measure in an abstract framework was stated for the first time, with an incomplete proof, by Łomnicki and Ulam in 1934 [139]. Von Neumann gave a complete proof in 1934 in his course of Princeton, although this was only published in 1950. It is likely that Jessen, who was in Princeton from September 1933 to July 1934, attended von Neumann's course. In any case, he published his own proof in 1939, [4, 80]. Other authors gave proofs around the same time, in particular Doob [43] (the validity of which Jessen challenged) and Kakutani [92]. However, Jessen's proof, translated into English in [6, §23–24], is very simple and is valid for a general set of indices. It served as the model for the authors of the treatises of the 1950s, particularly [66, §58, pp.157–158] and [138, Chap. I, §4.2]. For a history and references, see [6, p. 22, note 1] and [7, §3].

[29] Jessen and Wintner do not refer to Kolmogorov at this point. Thus when they were correcting the proofs in February 1935, they knew neither Kolmogorov's 0-1 law, nor for that matter the *Grundbegriffe*, which they do not quote. We will see Jessen belatedly quoting this law in his correspondence with Lévy.

See [51] for other abstract versions of the 0-1 law and of Jessen's Fubini theorems.

In this general version of Jessen's martingale theorem, the concept of conditional expectation does not appear. We remain with product spaces, where Fubini's theorem allows us to do without it.

Jessen and Wintner did not actually need such a general formulation since in the following section, §16, they consider a sequence of independent random variables $x_1(\tau_1), \ldots, x_n(\tau_n), \ldots$ with values in $\mathbb{R}^k$, which they define as measurable functions on the abstract spaces $q_1, \ldots, q_n, \ldots$ each equipped with a probability measure. Kolmogorov, Lévy and no doubt others (including Jessen and Wintner) knew that such variables can be defined on Q_ω (or the unit interval provided with Lebesgue measure), so that Jessen's 1934 theory [81] is amply sufficient, but our authors are seeking generality, especially since they are unaware of the axiomatics of Kolmogorov, and it is to best to be careful.

A fundamental theorem of the theory of sums of independent variables asserts an equivalence for such sums between convergence in law (the convergence of infinite convolutions), convergence in probability ("in measure") and almost sure (almost everywhere) convergence. Jessen and Wintner prove it as follows.

Equivalence between convergence in law and convergence in probability is easy; it is enough to work in the sense of Cauchy. The only real difficulty is to show that convergence of probability entails almost sure convergence. Following our authors, write $t = (\tau_1, \tau_2, \ldots)$ and $s(t) = x_1(\tau_1) + \cdots + x_n(\tau_n) + \cdots$ (in probability) and $f(t) = \mathrm{e}^{is(t)y}$, where y is fixed. The function f is bounded in absolute value and so integrable in Q, to which one can apply the theorem of Jessen stated above. The authors find that the integral taken starting from n can be written

$$f_n(t) = \mathrm{e}^{ix_1(\tau_1)y} \cdots \mathrm{e}^{ix_n(\tau_n)y} a_n(y),$$

where, for all y, the sequence of constants $a_n \to 1$, when $n \to \infty$. Whence it follows that, for all y, $\mathrm{e}^{is(t)y} = \mathrm{e}^{ix_1(\tau_1)y}\mathrm{e}^{ix_2(\tau_2)y} \cdots$, for almost all t and finally that the series $x_1(\tau_1) + \cdots + x_n(\tau_n) + \cdots$ converges to $s(t)$ for almost all t.

This is, to our knowledge, the first probabilistic application of the martingale theorem.[30] It dates from 1934 and was published in 1935 in the same journal where

[30] On the other hand, as is well known, the concept of martingale is as old as the theory of probability, under the generic and polysemous name of "fair game". The method of martingales, which associates a fair game with any unfair game, is in Pascal and especially in Moivre. Starting with an unfair game of heads or tails, the latter constructed a martingale (exponential) very similar to Jessen and Wintner's martingale f_n. Of course, Moivre did not use Jessen's theorem, of which he could have had no conception. He used the Borel-Doob stopping theorem, in the inverse direction, but without the justification that would truly appear only in the 1950s, around three centuries after the probability calculus began. For these matters, see the three editions of Moivre's *Doctrine of Chances* [39], Bertrand's treatise [14], and also Salah Eid's thesis [54], which gives all the references.

Jessen and Wintner [84, §16] add two applications of Jessen's (abstract) 0-1 law. The first shows that the probability of convergence of a series of independent variables is 0 or 1, a result well known to "probabilists". The second, on the other hand, is original: it is the Jessen-Wintner law of pure types, which states that infinite convolutions of probability laws are pure—i.e., they are discrete, singular, or absolutely continuous and not mixtures of the three. See [25, Chap. 3, §5].

Doob published his first article on the theory of martingales in 1940, with neither Doob nor Jessen noticing the connection.

2.5 After 1934

What happened to the theorems of 1934, at least in Jessen's own later work? To treat the question adequately would take us too far from our subject, but here is a summary.

As we have just seen that, Jessen already envisioned in the spring 1934 a comprehensive article that would present his theorem in the greatest possible generality. He told Lévy so in one of his letters. But the project seems to have been repeatedly put back. Jessen was appointed professor at the Polytechnic School of Copenhagen, directed by his father-in-law P. O. Pedersen.[31] He had to prepare his courses, which he did with meticulous care. Less was at stake for his academic career, and he was no longer in Princeton's atmosphere of high-speed mathematics. Jessen could pause and hope that by investing the necessary time he would obtain still more powerful theorems in a yet more general framework, with arbitrary index sets, for example, or in spaces that are no longer products of measure spaces. The matter was not simple, as he must have realized rather quickly, and difficulties appeared at every turn and accumulated. The new theory of abstract measure, which was being born here and there, conceals under the simplicity and generality of its concepts and statements awful traps, and the majority of papers go wrong in one way or another. So it was necessary to begin by imposing some order on this proliferation before trying to place his results in their natural abstract framework. So between 1934 and 1947 Jessen published a series of articles in Danish in the journal he edited, *Mat. Tidsskrift*. These were chapters for a treatise on abstract measure theory, and Jessen assembled them in a volume he published in 1947. One suspects that an English translation was planned but that the project was abandoned because of other pressures. The exposition was remarkably clear and certainly as good as the works that appeared in the 1950s and that it inspired on a number of points.

Of this collection of articles, we will attend only to the fourth, published in 1939, which touched directly on Jessen's theorem, and which was undoubtedly a provisional version of the forthcoming article promised in his 1935 article with Wintner [84, §15]. Here Jessen showed the existence of a product measure for an arbitrary family of probability spaces (§4.4). The demonstration extends the natural set function defined on the cylinder sets. So the principle of transfer is not used and Jessen must modify the proofs in his [81]. Nevertheless, he establishes Jessen's theorem for the decreasing case for set functions in §4.7 and succeeds in establishing it for the increasing case, as stated above following the 1935 article with Wintner [84, §16], in §4.8. We will see that Jessen had a first version of this proof when he wrote to

[31] Peder Oluf Pedersen (1874–1941), Danish engineer and physicist, specialized in electrical engineering. He was director of the Polytechnic School from 1922 to his death in 1941.

Lévy. It was reproduced in English by Sparre Andersen and Jessen in [6, §26], and we will not go into details.[32]

The tragic situation of the world between 1939 and 1945 explains the long silence that follow but perhaps Jessen was still trying unsuccessfully to improve his theorems. In the absence of convincing documents, we can at least imagine the issues involved. It would be necessary, in one way or another, to leave countable product spaces and work within a framework adapted to families of dependent random variables, taking abstract values and indexed by directed sets (increasing or decreasing) as general as possible. Lévy, as we will see, had already relaxed the assumption of independence, while staying with a countable family of real variables, in his 1937 treatise. This potential development raises two delicate questions. The first, posed by many around 1935, is whether the Daniell-Kolmogorov theorem can be placed within an abstract framework, or further, whether with every compatible family of abstract measures defined on the system of finite cylinders of an arbitrary product of measurable spaces, one can associate a measure of which they are the marginals. The second question is posed particularly for Jessen: can his theorem be extended to this new framework, assuming that the set of indices filters towards infinity without being completely ordered. It turns out that the answers to both questions is no,[33] and therefore that one cannot obtain abstract generalizations of Jessen's "Fubini's theorems" in complete generality. Things had to be viewed differently.

---

[32] The articles by Sparre Andersen and Jessen are exposited clearly by Hewitt and Stromberg [71, Chap. VI, §22]. Sparre Andersen and Jessen's 1946 article [6] reproduces Jessen's fourth article of 1939. Their first 1948 article [8] re-expresses everything in the framework of set functions, which is natural and simplifies things considerably. It also settles a minor dispute that seems to have arisen between Doob and Jessen, who had just realized that they had treated the same theorem without knowing it (Lévy being in a different category and being mostly used, when quoted, to push back against the other's exaggerated claims). See the account by Doob [46, 630–632] and also that by Moy [145]. Shu-Teh Chen Moy (1920–1969) is an interesting mathematician in more than one way, and her article on Jessen-Doob is quite clear. The theorems of Doob, Jessen and Lévy are essentially equivalent.

[33] With regard to the first question, all the attempts to extend the Daniell-Kolmogorov theorem [97, pp. 24–30] to an abstract framework proved to contain errors, including those of Doob [43, pp. 90–93, 96–97], Halmos [65, p. 390], and Sparre Andersen [2]. The first counter-examples to such an extension are due to Sparre Andersen and Jessen [7] and Dieudonné [40]. The case of dependent variables with values in an abstract set cannot be treated in general like that of real variables, and this impossibility is related to the non-existence of conditional probabilities, or of "disintegrations" in an abstract framework. For a current view of these questions, with interesting historical notes, see [50] and [15, vol II, Chap. 10].

The correspondence between Doob and Jessen and between Jessen and Dieudonné bearing on the simultaneous publication of their counter-examples to the abstract Daniell-Kolmogorov theorem, preserved in the Archives of the Institute of Mathematics in Copenhagen, is reproduced with commentary in the chapter in the present volume edited by Salah Eid and entitled, "Counterexamples to Abstract Probability: Ten Letters by Jessen, Doob, and Dieudonné".

In the very nice historical note in the last of their *Integration* volumes, published in 1969 [24, p. 121, note 12], Bourbaki remarks:

The situation appears to have really evolved only after the war. In a famous article written in collaboration with Erik Sparre Andersen in 1946 [6] and clarified and extended in 1948 [8], Jessen could state his two theorems in a satisfactorily general framework, the modern framework we mentioned in Section 1. One obtains an absolutely general result by forgetting the product structure and Fubini's theorem and using Kolmogorov's general concept of conditional expectation. One suspects that this very simple idea came from the young Sparre Andersen, although the proofs are almost identical to those of 1939 and these, we will see, Jessen partly explained to Lévy in 1935.[34]

At all events, Jessen could at last write, in 1946 [6, §1]:

> The present paper deals with two limit theorems on integrals in an abstract set. The first limit theorem [that of [81, §14]] is a generalization of the well-known theorem on differentiation on a net, the net being replaced by an increasing sequence of σ-fields. The second limit theorem [that of §13] is a sort of counterpart of the first, the sequence of σ-fields being now decreasing. The proofs follow the lines of the proof of the theorem on differentiation on a net.

> It seems that it is the absence of a satisfactory theory of disintegrations that marks the limit of the theory of "abstract" measure. This difficulty reappears in an insistent way in probability theory in connection with conditional probabilities.

This judgment is not false in the abstract, but without doubt it is erroneous in the life and in the history of the probability calculus, which seems not to have particularly suffered from this insistent difficulty in second half of the 20th century, and no more at all in 1969.

The second question is more subtle. Jessen's theorem passes without difficulty to the case of decreasing filters; see for example [71, note 38]. The increasing filter case was put in jeopardy by a serious difficulty found by Dieudonné in 1950 [41] and persisting in the theory of the filter martingales and the differentiation theory of the 1950s and 1960s [99,100]. One can however obtain relatively general Jessen theorems in the increasing case; for this see the beautiful article by Dorothy Maharam [140].

[34] Erik Sparre Andersen (1919–2003) studied mathematics at the University of Copenhagen, where he worked chiefly with B. Jessen. In 1945 he became an actuary, a career in which he continued for a long time; his [5] is a classic of the actuarial literature. In parallel he published mathematical work that was always very original. Beginning in 1948, after a stay at Cornell University with Feller, he worked on fluctuations of sums of independent random variables (an actuarial topic). His results were truly astonishing for the time, and his combinatorial methods revived the theory in the United States; see e.g. [3,4,56,57,169]. He was appointed professor of mathematics at the University of Aarhus in 1958 and then at Copenhagen in 1966.

One can assume that, during his stay at Cornell in 1948-1949, Sparre Andersen had occasion to present Jessen's martingale theory and conversely to take note of Doob's work. For a moment a collaboration between Doob and Jessen was contemplated [7, p. 5], but it did not happen. How could it? The theory of stochastic processes was the great affair of Doob's life, as it was for Lévy's, but it did not interest Jessen.

> In case of integrals in an infinite product set the theorems lead to known results, when for the nth σ-field of the sequence we take either the system of measurable sets depending on the n first coordinates only, or the system of measurable sets depending on all except the n first coordinates.
>
> If the abstract theory of integration is interpreted as probability theory, our theorems lead to two theorems concerning conditional mean values.

3 Lévy's Lemma

Paul Lévy's life and work are relatively familiar, and we do not need to rehearse them.[35] His interest in probability theory went back to 1919, when he was teaching at the École Polytechnique [11,136]. His first works were mainly concerned with the theory of stable laws, where only convergence in law is involved. He worked essentially within the framework of a finite-dimensional space, endowed with a positive additive set function of unit mass obeying Lebesgue's theory, which he transferred if necessary to the unit cube endowed with Lebesgue measure. Infinite dimensions as such were not involved, nor what Borel called "denumerable probabilities", i.e. probabilities for events depending on a countable infinity of random trials. Thus for ten years, until around 1929, Lévy seemed to limit himself to "Bernoulli's viewpoint", as one said in the 1930s. How can we explain this limitation and then the thunderous turn towards denumerable probabilities, which would continue without pause to produce one of the greatest bodies of work of 20th century probability, including in particular the lemma we are about to consider?

3.1 Before 1930

First a clarification: between 1925 and 1929 Lévy produced no important publications on probability theory, a subject he seems to have abandoned completely after producing his first probability book in 1925 [117], itself based on earlier articles [116]. No denumerable probabilities, no finite probabilities, nothing! Lévy published a few papers in analysis, on the Riesz-Fischer theorem, divergent series, entire series, the Riemann zeta function, doubtless all based on presentations to Hadamard's seminar, in which he took an active part. His contribution to the Congress of mathematicians at Bologna in September 1928 was on "functions of regular growth and iteration of fractional order" [118], an esoteric theme in his work that he described as idealist.[36] And then suddenly in 1929, as noted above, began an almost uninterrupted flow of contributions of the highest rank, all related in one way or another to denu-

[35] See, in particular, his intellectual autobiography [135], the recollections of his son-in-law Laurent Schwartz [165], Bernard Locker's thesis [137], and the chapter by Laurent Mazliak in the present volume. The annotated correspondence between Lévy and Fréchet [10] is invaluable, and having it at hand would help one follow the story.

[36] See [127, p. 61], where Lévy wrote:

merable probabilities, including—besides much else—the 1935 article containing Lévy's lemma.

One can suggest at least two hypotheses to explain Lévy's return to probability.

1. Fréchet's arrival in Paris at the end of 1928, when he was appointed professor at the Institut Henri Poincaré on the recommendation of Borel, who planned to create a major probability center in Paris. From the beginning of 1929, Borel organized, with Fréchet's help, general lectures on probability theory and theoretical physics to be given by the principal proponents of these theories in Europe. Lévy, whose only mathematical contacts had been with speakers at Hadamard's seminar and the Société mathématique de France, who were not much concerned with probability theory, now had in Fréchet an interlocutor who knew all the probability literature and maintained relations with most of the analysts of the time.[37]
2. The Congress of Bologna in September 1928, where the principal "probabilists" of the day met for the first time and where Lévy, who was registered in the analysis section, discovered that probability theory was not only a subject taught at the École Polytechnique but a flourishing field [28]. No one could be unaware of denumerable probabilities after Bologna, for it was the occasion of the famous dispute between Cantelli and Slutsky concerning the paternity of the crown jewel of the theory of denumerable probabilities, the strong law of large numbers for heads or tails. Cantelli argued, with some energy and not without reason, that Borel's proof was fundamentally incomplete and that he himself was the author of the first truly probabilistic proof of this new law of large numbers [30], while Slutsky believed, not without reason, that it was all in Borel's original article of 1909 and moreover had since been repeated by a number of authors.[38] Did

> This contribution plays a particular role in my mathematical work, for not only did I deliberately use idealist reasoning there, I also admitted without demonstration the compatibility of a certain number of axioms, which appeared to me to be essential for intuitive reasons, and consequently essentially subjective.

Some apply these adjectives to all of Lévy's work, especially in probability. This is clearly an exaggeration, as one can see from the example of Lévy's lemma. See [26] and especially Locker, in [137] and in his chapter in the present volume. Lévy returns to this article in his late correspondence with Fréchet [10, letter 104, pp. 199–201].

[37] See [10]. One of the first probability lecturers in the Institut Henri Poincaré's great seminar, in March 1929, was Lévy himself, who excused himself for having so little to say, coming after "Mr. Pólya's remarkable lectures" [120, 154].

[38] Borel gave two versions of his theorem in the 1909 article [22], which he and others of his time considered distinct yet related. According to the analytic version, normal numbers (and even absolutely normal numbers) are of measure one on the unit interval. According to the probabilistic version, the frequency of heads in the game of heads and tails converges to one-half with probability one. His proofs, were cavalier, but he never condescended to change them. The analytic version was at once proved convincingly by a great number of mathematicians, including Lebesgue in 1909, Faber in 1910, Hausdorff, Hardy-Littlewood, Rademacher, Sierpinski etc. The probabilistic version, which was harder to put into a recognised mathematical framework, was stated and proved independently by Cantelli (for the first time in 1917 in his own framework), and in the framework of

> Cantelli's claims clash with Lévy's patriotic feelings? Did they make him read Borel's article more closely? Or did he suddenly realise that there was an immense field of which he had been unaware and where he could revive his own faltering mathematical work?

We may count these complementary facts sufficient to explain Lévy's return to probability around 1929.

Lévy's withdrawal from probability theory in 1925, and especially his failure to notice denumerable probabilities for the ten years before 1929, is more difficult to explain, even hypothetically. This failure may seem especially inexplicable when we recognize how Lévy's work on functional analysis in the 1920s influenced probability. Concerning this work and its influence we refer the reader to the detailed account by Laurent Mazliak [142], who explains how Lévy took charge of editing the manuscripts of René Gateaux, who was killed at the start of the war of 1914–1918, and how Lévy's work was at the origin of Wiener's development of the mathematical model of Brownian movement in 1923. As Mazliak explains, Wiener's construction of his differential space systematically exploited Daniell's integral and the approach to integration in infinite dimension proposed by Gateaux for functionals defined on the space of continuous functions. Lévy thus knew, before 1925, both Daniell's integral and Wiener's measure (at least the name).[39] This fact may help us understand a little better that rather long part of his correspondence with Jessen where he tries to persuade the latter that he had known for a long time about integration in infinite dimensions, the principle of transfer, and a thousand other things as well, even if he did nothing but see them in the distance (or in the fog according to what one wishes to grant him), too far from the forefront of research in the area, and thus missing opportunities seized by others, more perspicacious.

Borel or their own, by Pólya, Steinhaus, Khinchin, Mazurkiewicz, Slutsky, etc. So by 1928, Borel's theorem was a classic known to all (except Lévy, it seems).

[39] In 1934, in his astonishing article on processes with general independent increments, Lévy constructs his processes (Lévy processes) by interpolation and states that, at the end of this construction, "the probability appears as a Daniell integral." The method of interpolation is preferable, according to him (and one cannot disagree), to Wiener's "differential" method, which consists in dividing the time interval into n small intervals, considering the law of the differences of the values of the process on these intervals and then letting n tend to infinity. This last method (of Gateaux-Wiener) lends itself well, by a passage to a suitable limit, to the calculation of probabilities and of "probable values" of functionals given by a simple analytical expression, [125, p. 344]. Lévy adds in a note:

> In his paper on differential space, without introducing this concept from the beginning as we do here, Wiener clearly showed that the mean of a uniformly bounded continuous functional is a Daniell integral.

This revenge on Wiener is also a revenge on Lévy, since the article of 1934 makes it possible to show how the stable laws of Lévy-1920 are interpreted naturally within the framework of the denumerable probabilities of Lévy-1934.

So it was that from 1919 to 1929 Lévy worked mainly on analysis in infinite dimension, on the functional calculus of Hadamard's school, and on probability theory in finite dimensions, an elaboration of the theory of the errors in the École Polytechnique's curriculum. Here Lévy was the first in France, after Laplace, Fourier, Poisson, Cauchy and Poincaré, to develop the method of characteristic functions and its applications to convergence in law, far from the work of Borel who would hasten to let him know that the probability calculus did not require such analytical sophistication, and all that was good for nothing.[40] The probability calculus must have a "practical value", either in application to the physical sciences or else in application to real mathematics-analysis or the theory of numbers.

Was Lévy acting out of pique, out of a desire for revenge, or out of loyalty to Hadamard? In any case, in the 20s he did not seem to be interested in practical probabilities, nor in Borel's denumerable probabilities. Borel, for his part, was certainly never interested in Lévy's stable laws, however useful they might be in times of crisis.[41]

Things changed, as we have noted, after 1928. Lévy then seems to have noticed that there is a difference between the weak and the strong laws of large numbers. He admitted this to Fréchet [10, pp. 104–105, letter 34, 8 January 1937]:

> It may be that before 1928 I confused Bernoulli's viewpoint with that of the strong law of large numbers. But since 1929 and in any case since 1930 I can tell you that, except for an always possible lapse, I did not ...

Lévy is undoubtedly referring his first partly denumerable articles in 1929 and 1930 [119,121,122]. The first is on the metric theory of continued fractions: a number between 0 and 1 is chosen at random, what are the limiting laws of its quotients complete and incomplete when it is developed as a continued fraction?[42] Lévy returned to this topic on several occasions, using it to test his increasingly elaborate methods (e.g. [132]). Borel had already been tackled this topic from the viewpoint of denumerable probabilities in his famous 1909 article [22], which Lévy quotes. Just the same, Lévy is still mainly working from Bernoulli's viewpoint; he treats convergence in law, and denumerable probabilities are not involved. But the end of

[40] Lévy answers Borel in the introduction of his 1925 book. One could undoubtedly add some subjective reasons to the objective reasons for Lévy's neglect of denumerable probabilities, but this would not be very interesting. Lévy's brain refused to move to denumerable probabilities and hardly saw any noticeable difference between convergence in law and almost sure convergence. In this way one can partly explain his particularly injudicious rejection of Bachelier's work in the 1920s. See on this point [37].

[41] See especially Barbut's beautiful book [9].

[42] Lévy announced [119] in a note presented in the Académie des Sciences on March 10, 1930. It was thus written in 1930. It establishes a celebrated formula given by Gauss to Laplace without proof. Kuzmin had proved it at the Bologna conference of 1928 [101], without Lévy realizing or at least recalling. It is easy to speculate that it was from a conversation in Bologna that Lévy became aware of Gauss's problem, although he says in his intellectual autobiography that the idea came to him "one day" without warning [135, pp. 88–89].

the article is concerned with the frequencies of different incomplete quotients of a number taken at random and tries to establish a strong law of large numbers for variables that are not independent. To this end, Lévy uses (for the first time it seems) one of the basic principles of his theory of dependent variables: one goes from the case of independent variables to that of dependent variables by replacing prior with posterior probability.[43] The interesting point is elsewhere, and it will help us to bring to a close this introduction to Lévy's denumerable silence. Lévy states, pp. 190–191:

> Finally there is an essential point whose interest was long unperceived and which has been highlighted by Mr. Cantelli and Mlle Mezzanotte: it is not merely a matter of considering each value of n independently of the others and showing that the difference between the frequency and the average probability is almost surely less than a function of n tending towards zero, but also and more especially of considering all the trials and showing that this difference almost surely tends towards zero, i.e. becomes and *remains* almost surely below any given positive number.

One should not be content to show convergence in probability (or convergence in mean square); one should show almost sure convergence; this was the "essential point" that was "long unperceived" (especially by Lévy) and that Mlle Mezzanotte

[43] This principle can be found already in Borel [23] and Bernstein [13], but strong laws of large numbers for dependent variables were not yet known in 1930. Lévy is clearly improvising (or rather he had read or understood poorly the article by Anna Mezzanotte discussed in the next footnote). He acknowledged this in 1937 [134, §69, p. 252], where he writes in a note, in connection with the article of 1929:

> With an application in mind, I only stated this theorem, which I believed could be regarded as known.

It was not known, and Lévy was ahead of the theory of denumerable probabilities without realizing it. In 1935 [126, 130], Lévy proved the strong law of large numbers that he had "applied" to continued fractions in 1929, and he included it in his 1937 book [134] in the section cited.

On the other hand, several weak laws of large numbers for dependent variables were known by 1930. The earliest date from the work of Markov beginning in 1907 and developed after the 1914–1918 war by Sergei Bernstein and then absorbed in the new theory of "Markov chains" (also consecrated at Bologna), which would lead to the Markovian strong laws of large numbers beginning in 1936 with celebrated contributions by Kolmogorov and Doeblin.

and Mr. Cantelli highlighted.[44] Lévy came to it late, but for him it finally opened the door to denumerable probabilities—around 1930 and not earlier.

3.2 Lévy's Denumerable Probabilities

So in 1930, Lévy was launched on his study of denumerable probabilities for dependent events. He would stick to this path until 1935. We will take notice of a couple steps, first his initial work in 1930 and then his 0-1 law in 1934.

One of Lévy's first "denumerable" notes in 1930, [122], takes up Borel's disputed proof of the law of large numbers and tries to make it rigorous. Recall that Borel's lemma, in its 1909 version, is stated as follows: we carry out a countable infinity of successive trials, assumed independent, and the probability of a favorable outcome for the n-th trial is p_n. The probability that the favourable outcome is produced infinitely often is zero or one, depending on whether the series of the p_n converges or diverges. Here we have a 0-1 law, and it is in this form that Lévy understood

[44] On Francesco Paolo Cantelli, (1875–1966), there are important articles by E. Regazzini and by M. Benzi. Cantelli is truly the first modern "probabilist", and he proclaimed himself as such. There is hardly any doubt that it was he who converted Lévy to denumerable probabilities.

Anna Mezzanotte published, between 1928 and 1938, some interesting actuarial and probabilistic work, in particular a 1928 paper [144] that seems to be the main source for Lévy. There exists, to our knowledge, no biography of Mezzanotte. We know nothing about her. Lévy adored Italy and often spent his holidays there, until 1938 when the fascist law on the Italian Jews was promulgated. Did he meet Mezzanotte in Bologna or elsewhere, or did Cantelli send him her 1928 paper? We do not know, but this article, of which Lévy read at least the first two or three pages, indicates very clearly, on p. 333, that convergence in probability ("nel senso del calcolo delle probabilità") does not imply convergence with probability one. This made a strong impression on Lévy, who had obviously believed the opposite. We are very grateful to E. Regazzini for obtaining a copy for us of Mezzanotte's remarkable piece. She was thus one of Lévy's inspirers, and not one of the least.

In his intellectual autobiography, Lévy tells a different story, without much conviction. He recognized that he had not done much before 1929, adding [135, p. 85]:

> I believe, without being able to affirm it, that it is thanks to Noaillon that I started to think of the various modes of convergence of probability theory. Up to then, I used the name "convergence in the sense of the probability calculus" for the convergence that holds for a sequence X_n of independent random variables with zero expected values under the condition $\sum E(X_n^2) < \infty$, and I considered it obvious that this involves almost sure convergence.

The second part of this quotation is certainly correct, but the first is hardly so, like everything that relates to that silent period that Lévy no longer comprehends and in any case regrets and tries to mask with a touching innocence, as when for example he claims that he became interested in the phenomena of contagion after hearing Pólya's lectures in 1927 or 1928, at the Institute Henri Poincaré [135, p. 88], which was then under construction. Pólya gave these lectures in March 1929, i.e. probably after Lévy's return to probability [154]. Noaillon is thus a decoy to hide something else. We will stick with the 1930 version, which does justice to the incomparable Mezzanotte.

Mezzanotte's intervention thus seems to have been decisive, and we will say no more.

it, just as he would understand Lévy's lemma four or five years later. This type of law had no surprise for Lévy; he had known since 1918 that in infinite dimension, non-trivial volumes have a natural tendency to be null or infinite. Borel's completely innovative and brilliant idea, which remained unnoticed by Lévy for so long, was, we repeat, to regard the event "the favorable outcome occurs an infinite number of times" as worthy of interest. The entire theory of denumerable probability is there, and, around 1930, Lévy finally understood it. So much so that from then on he would regard Borel as his master, officially on a par with Hadamard, but with a more and more pronounced secret preference for Borel.

As we have seen, Lévy had come into the story only after Bologna. In his intellectual autobiography[45] he tells us that this is when he read Borel's 1909 article [22]. There, to prove his law of large numbers, Borel applied his lemma, valid for independent events, to events that are not independent, events involving the numbers X_n of heads from n throws. Lévy reports that he was stopped for "several months" by this difficulty and that he finally overcame it by considering, in the sequence of all trials, subsequences sufficiently distant from one another that the cumulative results are independent or almost so, and that is enough to justify Borel's proof.[46] Lévy was thus led naturally to the law of the iterated logarithm, which he published in 1930 [122], not knowing that Khinchin had published it six years earlier in the same journal [93].

[45] [135, p. 90-92], where he dates his return to the infinite game of heads or tails to 1929. On page 22 of the same book Lévy indicates that he might have read Borel's 1909 article [22] "around 1922" and he realized at the time that he had known it all for a long time (since 1902). This seems doubtful. The reasoning he gives on p. 23 to show the recurrence of the play of heads or tails, which he tells us is equivalent to his work of 1902 (he was then about fifteen years old), does not depend on Borel's lemmas nor on his law of large numbers but on Bertrand's well-known method of successive doubling [14]. If we followed Lévy on this point, we would have to make Bertrand rather than Lévy the true father of denumerable probabilities, a case strengthened because Bertrand was a proven source for Borel too [29].

[46] Lévy makes this reasoning precise in [124, theorem II], which develops his note [122]. He show that if a sequence of constants c_n is given, and if one writes X_n for the number of heads in n tosses, then the probability P that the inequality $X_n > c_n$ holds infinitely often can be only 0 or 1. So that Borel's lemma applies to Borel's theorem, in spite of the non-independence of the events $X_n > c_n$. This type of 0-1 law was extended after 1935 and especially after the Second World War in several directions, in particular to sequences of exchangeable variables by Hewitt and Savage [70] (see [25, Chap. 3, §7]), also to the study of recurrence in random walks, by Chung for example [35,36], and to Markov chains by Kolmogorov, Doeblin, etc.

In [128, p. 89, note 1], Lévy observed that theorem II of [124] is a simple consequence of Lévy's lemma, which covers all possible 0-1 laws. See [10, p. 96, note 108].

We may recall incidentally that there are complete demonstrations of Borel's theorem that follow Borel's "indications", for example [63,177],…,[35]. But that hardly matters, in view of the not easily contested fact that it was Borel who posed the right question in the right way at the right time; according to Cantor's thesis III, that is all that really counts. Cantor was Borel's only master (before he detached himself from Cantor, as he detached himself from free mathematics) [52, p. 120, note 143].

Laurent Mazliak's chapter in the present volume on the relationship between Lévy and Jean Ville discusses Lévy's investigation of the convergence of the sums of independent or dependent random variables beginning in 1934. We refer the reader to this chapter for details. Let us recall only that these fundamental works led Lévy to state results directly related to the theory of martingales, in particular his theorem on convergence generalizing Borel's and Kolmogorov's 0-1 laws.

This result is announced in the first section of an article Lévy wrote in 1934 and published in 1935 [128]. This section, not visibly related to the rest of the article, carries the title "§1. Denumerable probabilities and the theory of measure." It derives directly, as the correspondence between Lévy and Jessen makes clear, from Lévy's reading Jessen's early work [78]. Here Lévy specifies in as detailed a way as possible and for the first time, the theoretical framework in which he works when he considers denumerable probabilities for series of independent variables. (This section is taken up and extended in §39 of his 1937 treatise.) And it is here (pp. 86–88), in this explicit mathematical framework, that Lévy's lemma is first stated, along with a very simple and perfectly acceptable proof.

The lemma concerns a sequence $x_1, x_2, \ldots, x_n, \ldots$ of independent random vari-i-ables, each with the same uniform distribution on $[0, 1]$ (defined in the sense of Jessen-Lévy), and an event E which "depends" on this sequence. Lévy writes $P(E)$ for E's probability at the outset and $P_n(E)$ for E's after the determination of $x_1, x_2, \ldots, x_n$, as a function of the values of these variables, assumed known. One has then:

Lemma I: If an event E has probability α, the sequences realizing this event, except in cases of zero probability, satisfy also the condition $\lim_{n\to\infty} P_n(E) = 1$.

In §41 of his 1937 treatise, Lévy places Lemma I in the most general framework of his theory of denumerable probabilities. Consider a "sequence X_n of variables independent or not", defined on the interval $[0, 1]$ equipped with the uniform measure, so that they appear truly as measurable functions defined on the unit interval. Lévy had shown in §39 of the treatise that this is always possible, in the independent case of his 1935 article [128] as in the dependent case.

Consider a property E of the sequence X_n—that is to say, Lévy explains, the set E of real numbers in $[0, 1]$ for which this property holds. Write $\Pr.(E)$ for its probability (that is to say, Lévy further explains, for the measure of the set E) and $\Pr_n(E)$ for its conditional probability evaluated as a function of $X_1, X_2, \ldots, X_n$ assumed known (a concept that Lévy defined in §23 following his earlier article [131]). One has then:

Theorem 41. — Except in cases of which the probability is null, if $\Pr.(E)$ is determined, $\Pr_n(E)$ tends, for n infinite, towards one if the sequence $X_1, X_2, \ldots, X_n, \ldots$ satisfies the property E, and towards zero in the contrary case.

In today's language, one writes, as we recalled in Sec. 1, in an obvious notation:

$$E_n(\mathbf{1}_E) \to \mathbf{1}_E \text{ a.s.}$$

And this 0-1 law contains all known 0-1 laws, in particular the one Lévy attributed to Jessen in 1935 [128, p. 89, note 1] and that he now [134, p. 130] gives back to Kolmogorov: If $\mathrm{Pr}_n(E) = \mathrm{Pr}.(E)$ for infinitely many n, then $\mathrm{Pr}.(E)$ can only be 0 or 1. This applies when E is an asymptotic event relative to a sequence $X_1, X_2, \ldots, X_n, \ldots$ of independent variables—for example when E is Borel's event "the outcome is favorable infinitely many times". Borel's lemma thus finally unites with Lévy's lemma.

Let us summarize. In the spring 1934, before Jessen and Lévy began their epistolary relationship, the first saw his theorem as an extension of the Fubini-Lebesgue theorem of 1907–1910, and the second saw his lemma as an extension of Borel's lemma of 1909.

Acknowledgements The Jessen archives are described on the website of the Jessen Archives at the Institute for Mathematical Sciences at the University of Copenhagen. It was from this well constructed site that we learned of the existence of the Jessen-Lévy correspondence. We are very grateful to the authors of the site, particularly K. Ramskov and S. Elkjær, and also to Jesper Lützen and the Committee of the Archives of the Institute of Mathematics of the University of Copenhagen. Christian Berg, professor at the University of Copenhagen and former student of Jessen, has very kindly provided us with very readable copies of all of Lévy's letters and the drafts of Jessen's replies. We are infinitely grateful to him. We also thank Glenn Shafer and Niels Keiding who put us in touch with Christian Berg. Christian Berg and Glenn Shafer have also very kindly provided us with copies of Jessen's Danish articles from 1929 to 1947 and these have helped us to reconstruct the genesis of the article of 1934. They have also read our manuscript and made very interesting suggestions which we have incorporated in this final version.

References

1. Aldrich, J.: "But you have to remember P. J. Daniell of Sheffield". Electronic Journal for History of Probability and Statistics **3**(2) (2007)
2. Andersen, E.S.: Indhold og Maal i Produktmaengder. Mat. Tidsskrift, B pp. 19–23 (1944)
3. Andersen, E.S.: On the number of positive sums of random variables. Skand. Aktuarietidskrift **32**, 27–36 (1949)
4. Andersen, E.S.: On the fluctuations of sums of random variables. Math. Scand. **1**, 263–285 (1953). 2:195–223 1954
5. Andersen, E.S.: On the collective theory of risk in case of contagion between the claims. In: Transactions XVth International Congress of Actuaries, New York, vol. II, pp. 219–229 (1957)
6. Andersen, E.S., Jessen, B.: Some limits theorems on integrals in an abstract set. D. Kgl. Danske Vidensk Selskab Mat. -fys. Medd. **22**(14) (1946)
7. Andersen, E.S., Jessen, B.: On the introduction of measures in infinite product sets. D. Kgl. Danske Vidensk Selskab Mat. -fys. Medd. **25**(4) (1948)
8. Andersen, E.S., Jessen, B.: Some limit theorems of set-functions. D. Kgl. Danske Vidensk Selskab Mat. -fys. Medd. **25**(5) (1948)
9. Barbut, M.: La mesure des inégalités. Ambiguïtés et paradoxes. Droz, Genève (2007)
10. Barbut, M., Locker, B., Mazliak, L.: Paul Lévy and Maurice Fréchet. 50 years of correspondence in 107 letters. Springer (2014)
11. Barbut, M., Mazliak, L.: Commentary on Lévy's lecture notes to the École Polytechnique. Electronic Journal for History of Probability and Statistics **4**(1) (2008)
12. Berg, C.: Børge Jessen, 19.6.1907–20.3.1993. Electronic Journal for History of Probability and Statistics **5**(1) (2009)

13. Bernstein, S.N.: Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. Math. Ann. **97**, 1–59 (1926)
14. Bertrand, J.: Calcul des probabilités. Gauthier-Villars, Paris (1888). 2nd edition 1907
15. Bogachev, V.I.: Measure Theory. Springer, New York (2006). 2 vol.
16. Bohr, H.: Sur les fonctions presque périodiques. Comptes rendus Acad. Sci. Paris **177**, 737–739 (1923)
17. Bohr, H.: Sur l'approximation des fonctions presque périodiques par des sommes trigonométriques. Comptes rendus Acad. Sci. Paris **177**, 1090–1092 (1923)
18. Bohr, H.: Zur Theorie der fastperiodischen Funktionen, I. Acta Math. **45**, 29–127 (1924). II, 46:101–214 1925; III, 47:237–281 1926
19. Borel, É.: Sur les séries de Taylor. Comptes rendus Acad. Sci. Paris **123**, 1051–1052 (1896)
20. Borel, É.: Sur les séries de Taylor. Acta Math. **21**, 243–247 (1897)
21. Borel, É.: Leçons sur la Théorie des fonctions. Gauthier-Villars, Paris (1898). 2nd edition 1914, 3rd edition 1928
22. Borel, É.: Les probabilités dénombrables et leurs applications arithmétiques. Rend. Circ. Mat. Palermo **27**, 247–271 (1909)
23. Borel, É.: Sur un problème de probabilités relatives aux fractions continues. Math. Ann. **72**(4), 578–584 (1912)
24. Bourbaki, N.: Intégration chapitre IX. Hermann, Paris (1969)
25. Breiman, L.: Probability. Addison-Wesley, Reading (1968)
26. Bretagnolle, J.: Notes de lecture sur l'æuvre de Paul Lévy. Ann. Inst. H. Poincaré, B **23**(S2), 239–243 (1987)
27. Brockmeyer, E., Halström, H.L., Jensen, B.: The Life and Works of A. K. Erlang. The Copenhagen Telephone Company, Copenhagen (1948)
28. Bru, B.: Souvenirs de Bologne. Journal de la Société Française de Statistique **144**, 135–226 (2003)
29. Bru, B.: Les leçons de calcul des probabilités de Joseph Bertrand. Electronic Journal for History of Probability and Statistics **2**(2) (2006)
30. Cantelli, P.: Sulle probabilità comme limite della frequenza. Rend. Acc. Lincei **26**, 39–45 (1917)
31. Cantelli, P.: Considérations sur la convergence dans le calcul des probabilités. Ann. Inst. H. Poincaré **5**, 3–50 (1935)
32. Carathéodory, C.: Vorlesungen über reelle Funktionen. Teubner, Leipzig (1918). 2d edition 1927, 3rd edition 1968
33. Carlson, F.: Contributions à la théorie des séries de Dirichlet. Arkiv för matematik, astronomi och fysik **23A**, 1–8 (1933)
34. Chow, Y.S.: Convergence theorems of martingales. Z. Warscheinlichkeitstheorie **1**(4), 340–346 (1963)
35. Chung, K.L.: A Course in Probability Theory, revised edn. Academic Press, New York (2000)
36. Chung, K.L., Hsu, P.: Sur un théorème de probabilités dénombrables. Comptes rendus Acad. Sci. Paris **233**, 467–469 (1946)
37. Courtault, J., Kabanov, Y., Bachelier, L.: Aux origines de la finance mathématique. Presses Universitaires Franc-Comtoises, Besançon (2002)
38. Daniell, P.J.: Integrals in an infinite number of dimensions. Ann. of Math. **20**, 281–288 (1919)
39. De Moivre, A.: The Doctrine of Chances or A Method of Calculating the Probability of Events in Play. W. Pearson, London (1718). 3rd ed. London, Millar, 1756
40. Dieudonné, J.: Sur le théorème de Lebesgue-Nikodym (III). Ann. Univ. Grenoble **23**, 25–53 (1948)
41. Dieudonné, J.: Sur un théorème de Jessen. Fund. Math. **37**, 242–248 (1950)
42. Dieudonné, J.: History of Functional Analysis. North Holland, Amsterdam (1981)
43. Doob, J.L.: Stochastic processes with an integral-valued parameter. Trans. Amer. Math. Soc. **44**, 87–150 (1938)
44. Doob, J.L.: Regularity properties of certain families of chance variables. Trans. Amer. Math. Soc. **47**, 455–486 (1940)

45. Doob, J.L.: Application of the theory of martingales. In: Actes du Colloque International Le Calcul des Probabilités et ses applications (Lyon, 28 juin au 3 juillet 1948), pp. 23–27. CNRS, Paris (1949)
46. Doob, J.L.: Stochastic Processes. Wiley, New York (1953)
47. Doob, J.L.: Notes on martingale theory. Proceedings of the Fourth Berkeley Symposium on Math. Stat. and Prob. **2**, 95–102 (1961)
48. Doob, J.L.: What is a martingale? The American Mathematical Monthly **78**(5), 451–463 (1971)
49. Doob, J.L.: The development of rigor in mathematical probability, (1900–1950). In: J.P. Pier (ed.) Development of Mathematics 1900–1950, pp. 157–169. Birkhäuser, Basel (1994)
50. Dudley, R.M.: Real Analysis and Probability. Cambridge (2002)
51. Dunford, N., Tamarkin, J.D.: A principle of Jessen and general Fubini theorems. Duke Math. J. **8**(4), 743–749 (1941)
52. Décaillot, A.M.: Cantor et la France. Kimé, Paris (2008)
53. Edgar, G.A., Sucheston, L.: Amarts : A class of asymptotic martingales, A, B. J. Multivariate Anal. **6**, 193–221, 572–591 (1976)
54. Eid, S.: Martingales et géométrie borélienne des probabilités. Ph.D. thesis, University of Paris Diderot (2008)
55. Fabry, E.: Sur les points singuliers d'une fonction donnée par son développement en série et l'impossibilité du prolongement analytique dans des cas très généraux. Ann. Sci. ENS **(3) 13**, 367–399 (1896)
56. Feller, W.: An Introduction to Probability Theory and its Applications, Vol. 1. Wiley, New York (1950). 3rd ed. 1968
57. Feller, W.: An Introduction to Probability Theory and its Applications, Vol. 2. Wiley, New York (1966). 2nd ed. 1971
58. Feller, W., Tornier, E.: Mass- und Inhaltstheorie des Baireschen Nullraumes. Math. Ann. **107**, 165–187 (1932)
59. Filep, L., Elkjaer, S.: Pál Gyula – Julius Pal (1881–1946), the Hungarian-Danish mathematician. Acta Math. Acad. Paedagogicae Nyíregyháziensis **16**, 89–94 (2000)
60. Fréchet, M.: Les fonctions d'une infinité de variables. In: Comptes rendus du Congrès des Sociétés savantes en 1909, pp. 44–47. Imprimerie nationale, Paris (1910)
61. Fréchet, M.: Sur divers modes de convergence d'une suite de fonctions d'une variable. Bull. Calcutta Math. Soc. **11**, 187–206 (1921)
62. Fréchet, M.: Sur la convergence en probabilité. Metron **8**, 1–50 (1930)
63. Fréchet, M.: Recherches théoriques modernes sur le calcul des probabilités. Premier livre. Généralités sur les probabilités, variables aléatoires, avec une note de M. Paul Lévy. Gauthier-Villars, Paris (1936). 2nd ed. 1950
64. Hald, A.A.: History of Mathematical Statistics from 1750 to 1930. Wiley, New York (1998)
65. Halmos, P.R.: The decomposition of measures. Duke Math. J. **8**(2), 386–392 (1941)
66. Halmos, P.R.: Measure Theory. Van Nostrand, New York (1950)
67. Halmos, P.R., von Neumann, J.: Operator methods in classical mechanics, Ii. Ann. Math. **43**(2), 332–350 (1942)
68. Hardy, G.H., Littlewood, J.E.: A maximal theorem with function-theoretic applications. Acta Math. **54**, 81–116 (1930)
69. Hawkins, T.: Lebesgue's Theory of Integration: Its Origins and Development. University of Wisconsin Press, Madison (1970). 2nd ed. 1975
70. Hewitt, E., Savage, L.J.: Symmetric measures on Cartesian products. Trans. Amer. Math. Soc. **80**, 470–501 (1955)
71. Hewitt, E., Stromberg, K.: Real and Abstract Analysis. Springer, New York (1965)
72. Hilbert, D.: Ueber die stetige Abbildung einer Linie auf ein Flächenstück. Math. Annalen **38**, 459–460 (1891)
73. Hilbert, D.: Wesen und Ziele einer Analysis der unendlichen unabhängigen Variablen. Rend. Circ. Mat. Palermo **27**, 59–74 (1909)
74. Hunt, G.A.: Random Fourier transforms. Trans. Amer. Math. Soc. **71**, 38–69 (1951)

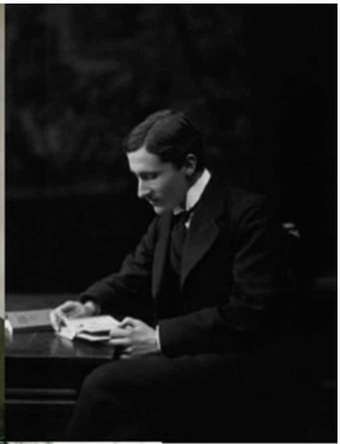
75. Jensen, J.L.W.V.: Sur un nouvel et important théorème de la théorie des fonctions. Acta Math. **22**, 359–364 (1899)
76. Jessen, B.: Hovedsætningerne indenfor de næstenperiodiske Funktioners Teori og deres indbyrdes Sammenhæng. In: Magisterkonferens i Matematik, København, 23/4–21/5 1929 (1929)
77. Jessen, B.: Bidrag til Integralteorien for Funktioner af uendelig mange Variable, (doctoral thesis). G. E. C. Gads Forlag, København (1930)
78. Jessen, B.: über eine Lebesguesche Integrationstheorie für Funktionen unendlich vieler Veränderlichen. In: Comptes rendus du septième Congrès des Mathématiciens scandinaves tenu à Oslo, 19-22 août 1929, pp. 127–138. A. W. Broggers Boktrykkeri, Oslo (1930)
79. Jessen, B.: über die Nullstellen einer analytischen fastperiodischen Funktion. Eine Verallgemeinerung der Jensenschen formel. Math. Ann. **108**, 485–516 (1932)
80. Jessen, B.: Abstrakt maal- og integralteori, 1–10. Mat. Tidsskr. B pp. 73–84 (1934). Also 1935 pp. 60–74, 1938 pp. 13–26, 1939 pp. 7–21, 1942 p.p 43–53, 1944 pp. 28–34, 35–37, 1947 pp. 1–20, 21–26, 27–36; collected in one volume with the same title, København, Matematisk Forening, 1947
81. Jessen, B.: The theory of integration in a space of an infinite number of dimensions. Acta Math. **63**, 249–323 (1934)
82. Jessen, B.: On the proofs of the fundamental theorem on almost periodic functions. D. Kgl. Danske Vidensk Selskab Mat. -fys. Medd. **25**(8) (1949)
83. Jessen, B.: Harald Bohr, 22 April 1887–22 January 1951. Acta Math. **86**, I–XXIII (1951)
84. Jessen, B., Wintner, A.: Distribution functions and the Riemann zeta function. Trans. Amer. Math. Soc. **38**(1), 48–88 (1935)
85. Kaczmarz, S., Steinhaus, H.: Theorie der Orthogonalreihen. Narodowy Fundusz Kultury, Warszawa-Lwów (1935). Monografje Matematyczne VI
86. Kahane, J.P.: Séries de Fourier aléatoires. Presses de l'Université de Montréal (1963). 2nd ed. 1966
87. Kahane, J.P.: Some random series of functions, 2nd edn. Cambridge (1985)
88. Kahane, J.P.: Une théorie de Denjoy des martingales dyadiques. Enseign. Math. **(2) 34**, 255–268 (1988)
89. Kahane, J.P.: Naissance et postérité de l'intégrale de lebesgue. Gazette des mathématiciens **89**, 5–20 (2001)
90. Kahane, J.P.: The youth of Andrei Nikolaevich and Fourier series. In: É. Charpentier, A. Lesne, N.K. Nikolski (eds.) Kolmogorov's Heritage in Mathematics, pp. 7–18. Springer (2007)
91. Kahane, J.P., Lemarié-Rieusset, G.: Séries de Fourier et ondelettes. Cassini, Paris (1998)
92. Kakutani, S.: Notes on infinite product measure spaces, i, ii. Proceedings of the Imperial Academy, Japan **19**(3), 148–151, 184–188 (1943)
93. Khinchin, A.Y.: Sur un théorème général relatif aux probabilités dénombrables. Comptes rendus Acad. Sci. Paris **178**, 617–619 (1924)
94. Khinchin, A.Y., Kolmogorov, A.N.: über Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden. Matematicheskii Sbornik **32**, 668–677 (1925)
95. Kolmogorov, A.N.: über die Summen durch den Zufall bestimmter unabhängiger Grössen. Math. Ann. **99**, 309–319 (1928). 102:484–488 1929
96. Kolmogorov, A.N.: Sur la loi forte des grands nombres. Comptes rendus Acad. Sci. Paris **191**, 910–911 (1930)
97. Kolmogorov, A.N.: Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer, Berlin (1933). Ergebnisse der Mathematik und ihrer Grenzgebiete 3; English translation: Foundations of the Theory of Probability, New York, Chelsea, 1950
98. Kolmogorov, A.N.: A theorem on convergence of conditional expectations (in Russian). In: First Congress of the Hungarian Mathematicians, pp. 367–376. Akadémiai Kiado, Budapest (1950)
99. Krickeberg, K.: Convergence of martingales with a directed index set. Trans. Amer. Math. Soc. **83**, 313–337 (1956)
100. Krickeberg, K., Pauc, C.: Martingales et dérivations. Bull. Soc. Math. Fr. **91**, 455–543 (1963)

101. Kuzmin, R.: Sur un problème de Gauss. In: Comptes rendus du Congrès international des mathématiciens de Bologne, septembre 1928, vol. 6, pp. 83–89 (1932)
102. de La Vallée Poussin, C.: Sur la transformation d'une intégrale multiple en une intégrale simple. Ann. Soc. Sci. Bruxelles **35**, 189–190 (1911)
103. de La Vallée Poussin, C.: Cours d'analyse infinitésimale, vol. I, 3rd edn. Gauthier-Villars, Paris (1914)
104. de La Vallée Poussin, C.: Sur l'intégrale de Lebesgue. Trans. Amer. Math. Soc. **16**, 435–501 (1915)
105. de La Vallée Poussin, C.: Intégrales de Lebesgue, fonctions d'ensemble, classes de Baire, leçons professées au Collège de France. Gauthier-Villars, Paris (1916)
106. Lebesgue, H.: Sur les fonctions de plusieurs variables. Comptes rendus Acad. Sci. Paris **128**, 811–813 (1899)
107. Lebesgue, H.: Intégrale, longueur, aire. Annali di Matematica Pura ed Applicata **7**(1), 231–359 (1902). Thèse sci. math., Paris
108. Lebesgue, H.: Sur le problème des aires. Bulletin de la S. M. F. **31**, 197–203 (1903). Correction, ibid., 33 (1905), pp. 273–274
109. Lebesgue, H.: Sur l'existence des dérivées. Comptes rendus Acad. Sci. Paris **136**, 659–661 (1903)
110. Lebesgue, H.: Leçons sur l'intégration et la recherche des fonctions primitives. Gauthier-Villars, Paris (1904). 2nd ed. 1928
111. Lebesgue, H.: Recherches sur la convergence des séries de Fourier. Math. Ann. **61**, 251–260 (1905). æuvres III, pp. 181–210
112. Lebesgue, H.: Sur les fonctions représentables analytiquement. J. Math. Pures Appl. **(6), 1**, 139–216 (1905)
113. Lebesgue, H.: Leçons sur les séries trigonométriques. Gauthier-Villars, Paris (1906)
114. Lebesgue, H.: Sur l'intégration des fonctions discontinues. Ann. Sci. ENS **(3) 27**, 361–450 (1910). æuvres II, pp. 185–274
115. Lebesgue, H.: Les lendemains de l'intégrale : lettres à Émile Borel. Vuibert, Paris (2004). Edited by Bernard Bru and Dugac, with preface by Gustave Choquet
116. Lévy, P.: Théorie des erreurs. La loi de Gauss et les lois exceptionnelles. Bulletin de la S. M. F. **52**, 49–85 (1924)
117. Lévy, P.: Calcul des probabilités. Gauthier-Villars, Paris (1925)
118. Lévy, P.: Fonctions à croissance régulière et itération d'ordre fractionnaire. Annali di Matematica **2**, 277–282 (1928). Also Atti del Congresso Internatiozionale dei Matematici, Bologna, 1928, Bologne, Zanichelli, 1932
119. Lévy, P.: Sur les lois de probabilité dont dépendent les quotients complets et incomplets d'une fraction continue. Bulletin de la S. M. F. **57**, 178–194 (1929)
120. Lévy, P.: Le théorème fondamental de la théorie des erreurs. Ann. Inst. H. Poincaré **1**, 163–175 (1930)
121. Lévy, P.: Sur la croissance des fonctions entières. Bulletin de la S. M. F. **58**, 29–59, 127–149 (1930)
122. Lévy, P.: Sur la loi forte des grands nombres. Comptes rendus Acad. Sci. Paris **191**, 983–984 (1930)
123. Lévy, P.: Sur les séries dont les termes sont des variables éventuelles indépendantes. Studia Math. **3**, 119–155 (1931)
124. Lévy, P.: Sur un théorème de M. Khintchine. Bull. Sci. Math. **55**, 145–160 (1931)
125. Lévy, P.: Sur les intégrales dont les éléments sont des variables aléatoires indépendantes. Ann. R. Sc. Norm. Sup. Pisa **(2) 3**, 337–366 (1934). Observations sur le mémoire précédent, ibid., 4 (1935), pp. 217–218
126. Lévy, P.: La loi forte des grands nombres pour les variables enchaînées. Comptes rendus Acad. Sci. Paris **201**, 493–495, 800 (1935)
127. Lévy, P.: Notice sur les travaux scientifiques de M. Paul Lévy. Hermann, Paris (1935)
128. Lévy, P.: Propriétés asymptotiques des sommes de variables aléatoires enchaînées. Bull. Sci. Math. **(2) 59**, 84–96, 109–128 (1935)

129. Lévy, P.: Sur la sommabilité des séries aléatoires divergentes. Bulletin de la S. M. F. **63**, 1–35 (1935)
130. Lévy, P.: La loi forte des grands nombres pour les variables enchaînées. J. Math. Pures Appl. **15**, 11–24 (1936)
131. Lévy, P.: Sur la notion de probabilité conditionnelle. Bull. Sci. Math. **(2) 60**, 66–71 (1936)
132. Lévy, P.: Sur le développement en fraction continue d'un nombre choisi au hasard. Compositio Mathematica **3**, 286–303 (1936)
133. Lévy, P.: Sur quelques points de la théorie des probabilités dénombrables. Ann. Inst. H. Poincaré **6**, 153–184 (1936)
134. Lévy, P.: Théorie de l'addition des variables aléatoires. Gauthier-Villars, Paris (1937). Page numbers refer to the 2nd edition, 1954
135. Lévy, P.: Quelques aspects de la pensée d'un mathématicien. Blanchard, Paris (1970)
136. Lévy, P.: Cours de calcul des probabilités de Paul Lévy à l'École Polytechnique (1919). Electronic Journal for History of Probability and Statistics **4**(1) (2008)
137. Locker, B.: Paul Lévy : "La période de guerre". Intégrales stochastiques et mouvement brownien. Ph.D. thesis, Univ. Paris 5 (2001)
138. Loève, M.: Theory of Probability. Wiley, New York (1953)
139. Łomnicki, A., Ulam, S.: Sur la théorie de la mesure dans les espaces combinatoires et son application au calcul des probabilités I. Variables indépendantes. Fund. Math. **23**, 237–278 (1934)
140. Maraham, D.: On two theorems of Jessen. Proc. Amer. Math. Soc. **9**, 995–999 (1958)
141. Marcus, M.B., Pisier, G.: Random Fourier series with applications to harmonic analysis. Princeton University Press (1981). Ann. of Math. Studies
142. Mazliak, L.: The ghosts of the École Normale. Statistical Science **30**(3), 391–412 (2015)
143. Meyer, P.A.: Probabilités et potentiel. Hermann, Paris (1966). Revised with C. Dellacherie, 4 vol., Paris, Hermann, 1975–1992
144. Mezzanotte, A.: Estensione di un teorema sulla oscillazione delle frequenze alla probabilità, alle medie di variabili casuali più generali. Rend. Circ. Mat. Palermo **(1) 52**, 331–344 (1928)
145. Moy, S.T.C.: Measure extensions and the martingale convergence theorem. Proc. Amer. Math. Soc. **4**, 902–907 (1953)
146. Olsen, L.W.: Essays on Georg Rasch and his contributions to statistics. Ph.D. thesis, Institute of Economics, University of Copenhagen (2003). Available at rasch.org
147. Paley, R., Wiener, N.: Fourier Transforms in the Complex Domain. Amer. Math. Soc. Colloquium Pub., 19 (1934)
148. Paley, R., Zygmund, A.: On some series of functions (1), (2), (3). Proc. Camb. Phil. Soc. **26**, 337–357, 458–474 (1930). 28 (1932), pp. 190–205
149. Peano, G.: Sur une courbe, qui remplit toute une aire plane. Math. Ann. **36**, 157–160 (1890)
150. Pedersen, E.: über einige besondere Klassen von fastperiodischen Funktionen. D. Kgl. Danske Vidensk Selskab Mat. -fys. Medd. **8**(6) (1928)
151. Pier, J.P.: L'analyse harmonique : son développement historique. Masson, Paris (1990)
152. Pier, J.P. (ed.): Development of Mathematics 1900–1950. Birkhäuser, Basel (1994)
153. Poincaré, H.: Sur les déterminants d'ordre infini. Bulletin de la S. M. F. **14**, 77–90 (1886)
154. Polya, G.: Sur quelques points de la théorie des probabilités. Ann. I.H.P. **1**, 117–161 (1930)
155. Rademacher, H.: Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen. Math. Ann. **87**, 112–138 (1922)
156. Ramskov, K.: Matematikeren Harald Bohr (199). Licentiatafhandling, Institut for de Eksakte Videnskabers Historie, Aarhus Universitet
157. Riesz, F.: Sur les systèmes orthogonaux de fonctions et l'équation de Fredholm. Comptes rendus Acad. Sci. Paris **144**, 734–736 (1907)
158. Riesz, F.: Sur les suites de fonctions mesurables. Comptes rendus Acad. Sci. Paris **148**, 1303–1305 (1909)
159. Riesz, F.: Untersuchungen über Systeme integrierbarer Funktionen. Math. Ann. **69**, 449–497 (1910)

160. Riesz, F.: Les systèmes d'équations linéaires à une infinité d'inconnues. Gauthier-Villars, Paris (1913)
161. Riesz, F.: Sur l'intégrale de Lebesgue comme opération inverse de la dérivation. Ann. R. Sc. Norm. Sup. Pisa **(2) 5**, 191–212 (1936)
162. Riesz, F.: L'évolution de la notion d'intégrale depuis Lebesgue. Ann. Inst. Fourier **1**, 29–42 (1949)
163. Rokhlin, V.A.: On the fundamental ideas of measure theory. Mat. Sbornik **25**, 107–150 (1949). English transl. Amer. Math. Soc. Transl. 1952, n° 71, reissued in AMS Translations (1) 10 (1962), pp. 1–54
164. Schwartz, L.: Théorie des distributions. Hermann, Paris (1950). 2 volumes, 2nd ed. 1966
165. Schwartz, L.: Un mathématicien aux prises avec le siècle. Odile Jacob, Paris (1997)
166. Siegmund-Schultze, R.: Rockefeller and the Internationalization of Mathematics Between the Two World Wars. Birkhäuser, Basel (2001)
167. Sierpinski, W.: Sur une définition axiomatique des ensembles mesurables (L). Bull. Acad. Sci. Cracovie pp. 173–178 (1919)
168. Slutzky, E.E.: Sur un critérium de la convergence stochastique des ensembles de variables éventuelles. Comptes rendus Acad. Sci. Paris **187**, 370–373 (1928)
169. Spitzer, F.: Principles of Random Walks. Van Nostrand, Princeton (1964)
170. Steinhaus, H.: Les probabilités dénombrables et leurs rapports avec la théorie de la mesure. Fund. Math. **4**, 286–310 (1923)
171. Steinhaus, H.: Sur la portée pratique et théorique de quelques théorèmes sur la mesure des ensembles de droites. In: Comptes-rendus du premier Congrès des mathématiciens des pays slaves, Warszawa 1929, pp. 348–354. Warszawa (1930)
172. Steinhaus, H.: Sur la probabilité de la convergence des séries. Studia Math. **2**, 21–39 (1930)
173. Steinhaus, H.: über die Wahrscheinlichkeit dafür, dass des Konvergenzkreiss einer Potenzreihe ihre natürlich Grenze ist. Math. Zeitschrift **31**, 408–416 (1930)
174. Steinhaus, H.: La courbe de Peano et les fonctions indépendantes. Comptes rendus Acad. Sci. Paris **202**, 1961–1963 (1936)
175. Steinhaus, H.: La théorie et les applications des fonctions indépendantes au sens stochastique. In: Colloque consacré à la théorie des probabilités et présidé par M. Maurice Fréchet, Genève 11 au 16 octobre 1937, cinquième partie, (Actualités Sci. Ind. 738), pp. 58–73. Hermann, Paris (1938)
176. Stepanoff, W.: über einige Verallgemeinerungen der fast periodischen Funktionen. Math. Ann. **95**, 473–498 (1926)
177. Uspensky, J.V.: Introduction to Mathematical Probability. McGraw-Hill, New York (1937)
178. Ville, J.: Étude critique de la notion de collectif. Gauthier-Villars, Paris (1939)
179. Weil, A.: L'intégration dans les groupes topologiques. Hermann (Actualités Sci. Ind. 869), Paris (1940). 2nd ed., Actualités Sci. Ind. 1145, 1965
180. Weyl, H.: über die Gleichverteilung von Zahlen mod. Eins. Math. Ann. **77**(3), 313–352 (1916)

Part II Ville, Lévy and Doob



Did Jean Ville Invent Martingales?

Glenn Shafer

Abstract

Jean Ville, we are sometimes told, brought martingales into mathematical probability. In what way is this true? To respond thoughtfully, we must try to see probability theory as Ville and his contemporaries saw it in the 1930s and excavate the martingales already hidden in that theory. Then we may see Ville's contribution as he himself saw it, as the use of a game to cast light on the denumerable probabilities that Émile Borel had introduced in 1909 and that evolved into the measure-theoretic framework for probability used by mathematicians today. We may then also better understand how Ville's martingales contributed to two other complementary perspectives on mathematical probability: the understanding of randomness in terms of complexity, and the game-theoretic foundation for probability.

Keywords

Algorithmic randomness · Jean Ville · Martingale · Game-theoretic probability · Kolmogorov complexity · Measure-theoretic probability · Game theory · Ville's inequality · Ville's theorem

1 Introduction

Jean Ville's martingales were an irruption of game theory into probability theory. To evaluate his contribution, we need to understand the state of all these ideas in the 1930s, when Ville brought them together. What did mathematical probability mean in France at that time? What was the state of game theory? What role did martingales play in either theory?

G. Shafer (✉)
Rutgers Business School, 1 Washington Place, Newark, NJ 07102, USA
e-mail: gshafer@business.rutgers.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*,
Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_5

Ville was born in 1910 and died in 1989. In Sect. 2, I recount how he discovered martingales. Then I consider mathematical probability as Ville encountered it in Paris in the early 1930s (Sect. 3), martingales' hidden role in probability at the time (Sect. 4), and how Ville connected game theory with denumerable probability (Sect. 5). To conclude, I sketch the legacy of Ville's encounter with martingales (Sect. 6) and ask why he himself abandoned the topic (Sect. 7).

2 A Glimpse of Jean Ville

In 1934, Ville left Paris for Vienna in search of a topic for a thesis that would qualify him to be a university professor in France. He wanted to contribute to mathematical probability. For his mentors in Paris, his immediate adviser Maurice Fréchet and the grand master Émile Borel, this meant contributing to mathematical analysis. Ville was more interested in foundations. In 1934, Richard von Mises's collectives were at the center of this topic. For von Mises, mathematical probability should begin with the idea of an infinite random sequence, a *Kollektiv* as he called it. To speak of probability one-half, for example, one should begin with a sequence of 0s and 1s, say, such that the cumulative fraction of ones tends to $1/2$ in the whole sequence and in any infinite subsequence selected without foreknowledge—i.e., under the condition that the inclusion of an element depends only on its predecessors. This invariance under selection without foreknowledge was von Mises's *axiom of irregularity*. Ville had spent 1933–1934 in Berlin, where von Mises had been a professor, and he had hoped to study both collectives and analysis there. But von Mises had fled from the Nazis to Istanbul, and the great analyst Erhard Schmidt no longer dared take a French student. So Ville had found no mentor in Berlin.

As Ville later explained to Pierre Crépel, he found a different world in Karl Menger's seminar in Vienna.[1] Here the emphasis was not on analysis but on newer ideas, especially from logic and economics. Participants in the seminar during Ville's year there included Kurt Gödel, Albert Tarski, and Karl Popper. The star of the seminar was Abraham Wald, who turned 32 in 1934. Unable to aspire to the university position in Austria he merited because he was Jewish, Wald made a meager living at Oskar Morgenstern's econometric institute and had made Menger's seminar his intellectual home.

When Popper talked about his version of collectives in Menger's seminar, Ville set about to find counterexamples—sequences that satisfied Popper's conditions but were clearly very regular rather than random.[2] But only two weeks after Popper's talk, Wald showed how to overcome the trivial objection to von Mises's formulation that had stymied its acceptance. A brilliant mathematician, but applied rather than

[1] Chapter 17 of the present volume provides a reconstruction in English of Crépel's 1984 interview of Ville and an English translation of their correspondence.

[2] Conditions equivalent to Popper's were independently proposed by Hans Reichenbach and Aaron Copeland. For details, see the chapter in the present volume by Laurent Bienvenu, Glenn Shafer, and Alexander Shen.

pure, von Mises was comfortable with the notion of selection without foreknowledge. Pure mathematicians found it insufficiently formal. For any given sequence of 0s and 1s, they were quick to point out, there are rules that select just the 1s and others that select just the 0s. Once such a rule itself is selected, you can say that it does its work without foreknowledge. Wald countered by pointing out that a mathematical language has only a finite number of symbols. Hence it can express only a countable number of rules for selecting subsequences without foreknowledge. Wald showed that for any such countable number of rules, you can construct a sequence that resists them all.

Here Ville found his own opening. Von Mises, in order to justify his axiom of irregularity, had appealed to a simple version of the cliché that betting systems cannot succeed: you cannot win by deciding as you go along which tosses of a coin or spins of a roulette wheel to bet on. Ville had a slightly more sophisticated knowledge of betting systems; he knew that these systems, often called martingales, also varied the amount of the bet. He soon showed that some of Wald's collectives could be defeated by such systems, but that Wald's idea could be generalized to establish the existence of collectives they could not defeat.

Back in Paris in 1935–1936, Ville sought to make his ideas into a doctoral thesis. In January of 1936, Wald sent a note on his results on collectives to Borel, who promptly inserted it in the Academy of Sciences' *Comptes rendus*, where it was customary for members to announce without proof new mathematical results by themselves or others. This opened the way for Ville to prepare similar notes: his note on Popper, Reichenbach, and Copeland's sequences appeared in April, his announcement of his results using martingales in July.[3] By the end of 1936, he later told Crépel, he had a thesis ready to defend. Or so he thought. Fréchet found the work insufficient for a mathematical doctorate. Where was the analysis?

Ville had enough time to satisfy Fréchet's expectations. For three academic years, from the fall of 1935 through the spring of 1938, he held fellowships from France's newly formed CNRS (Centre National de Recherches Scientifiques). But he had plenty of other things to do in Paris. He helped Borel teach game theory, edited Borel's course for publication, and inserted in it a mathematical result that later proved more useful to his career than his thesis: a new, simpler, and more insightful proof of John von Neumann's minimax theorem. He wrote up an ill-fated note on the convergence of medians.[4] He and his brilliant fellow student Wolfgang Doeblin began a seminar on probability. He talked about his thesis in the seminar. Another fellow student, Robert Fortet, later recalled his also talking about second-order probabilities.[5] This was also a time for Ville to enjoy himself in Paris; he had many academic friends, and he and his wife Lucie frequented artistic and Bohemian circles.

---

[3] See [40,41,46].

[4] Inserted in the *Comptes rendus* by Borel in December 1936 [39]; Laurent Mazliak discusses the irritation it caused Paul Lévy in the next chapter of the present volume.

[5] In [16]. Fortet suggested that we instead take uncertainty about probabilities into account by assessing upper and lower probabilities. Fortet and Ville remained lifelong friends. Doeblin perished in the war [3,10].

The July 1937 issue of the *Transactions of the American Mathematical Society* included a paper by the American Joseph Doob devoted to making Andrei Kolmogorov's measure-theoretic probability usable for stochastic processes that vary continuously in time. By this time, Fréchet favored Kolmogorov's approach over von Mises's. So Ville undertook to study continuous-time martingales within Kolmogorov's and Doob's framework.[6] In March 1938, Borel inserted Ville's announcement of continuous-time results in the *Comptes rendus* [42]. But Fréchet remained unsatisfied. Despairing of a university appointment, Ville took a position at the Lycée Clemenceau in Nantes. By the end of 1938, with the conflict with Germany looming, military mobilization was added to his teaching duties. He was a second lieutenant, having already served his year of active duty in 1932–1933. Fréchet finally relented, perhaps because Ville's situation made it unlikely that he would do much more. Ville recalled, talking with Crépel, that a nudge from Borel had been required. The university formally approved the printing of the thesis on 22 October 1938. Ville defended it on 9 March 1939.

In the end, both Borel and Fréchet accepted the value of Ville's thesis. In the proceedings of the celebrated 1937 Geneva conference on probability, Fréchet praised it as a nail in the coffin of collectives: it was now clear that mathematicians should use Kolmogorov's measure-theoretic axioms instead.[7] Borel quickly arranged for Gauthier-Villars to publish an expanded version of the thesis. This book [43] had the same title as the thesis [44], but it replaced the thesis's brief introduction with two rather philosophical chapters, a 17-page introduction and an 11-page conclusion. These new chapters pointed away from the thesis's accomplishment to the conclusion that probability should be founded on axioms rather than on any notion of random sequences—Fréchet's view precisely. The additional chapters were written before the thesis was defended; Fréchet devoted a sentence to them in his official report. Their non-mathematical character accounts, no doubt, for their being omitted from the printed thesis.

Ville never returned to martingales. He spent 1939–1940 at the front, 1940–1941 as a prisoner of war. When he returned to France in 1941, he launched both a university career and a consulting career. His work ranged across probability, mathematical statistics, signal theory, information theory, mathematical economics, operations research, and computing. In the late 1950s he was appointed to a chair in econometrics at the University of Paris, largely on the strength of von Neumann's acknowledgement of the importance of his proof of the minimax theorem. He retired from the university in 1978.

In 1997, Ville's fellow normalian André Blanc-Lapierre, president of the Academy of Sciences in 1985, remembered Ville for his unusual intelligence. Ville, he said,

[6] Apparently he had already completed his work on discrete-time martingales, where he used Lévy's notion of conditional probability [43, pp. 94–111].

[7] The piece in which Fréchet praised Ville's accomplishment [17], was mostly written after the conference, probably just before the proceedings' publication in 1938. See also [18,20].

> was a very subtle man who produced many ideas. I was always impressed by his notable analytic intelligence. He always gave the appearance of dilettantism, and he came out with things that made people say, "It's amazing that he produced that." A sort of cleverness.[8]

Yet of all his many ideas, Ville knew that his contribution to martingales held a special place. In 2009, when I visited the house in Loir-et-Cher to which he and his wife had retired, the only trace of his academic career that I found was a stack of cheaply reproduced copies of his 1939 book.

3 Probability as Ville Encountered It in the Early 1930s

When Ville had begun his undergraduate study at the University of Paris and the École Normale Superièure in 1929, he had encountered two giants of mathematical probability: Borel, who had entangled probability with both measure theory and number theory in his 1909 paper on denumerable probabilities,[9] and Fréchet, who had just arrived back in Paris after a decade in Strasbourg and was discovering how probability's limit theorems were connected with his own pioneering work in analysis decades earlier.[10]

In his 1909 paper, Borel had worked out consequences of embedding the game of heads and tails in the interval of real numbers [0, 1], representing each infinite sequence of heads and tails as the binary expansion of a number in the interval. Among these consequences was a new kind of limit theorem: the fraction of heads tended to its probability $1/2$ with probability one—i.e., except for the sequences in a subset of [0, 1] of measure zero. As Marie-France Bru and Bernard Bru explain in their chapter in the present volume, this new theory of "denumerable" probability helped Borel convince himself of the importance of probability theory to both mathematics and science. As they also explain, Borel soon realized that his infinities gave no help to scientific and other practical applications of probability, but it remained his mission to promote both the mathematics of probabilities and their applications.

Today, Kolmogorov's 1933 *Grundbegriffe der Wahrscheinlichkeitsrechnung* [22] is seen as measure-theoretic probability's founding document. As Ville told Crépel, the *Grundbegriffe* had not yet attracted much attention in Paris or Vienna when he began his work on martingales.[11] But probability was already measure-theoretic in Paris. Borel's 1909 paper was the oldest paper in the *Grundbegriffe*'s bibliography, and as Kolmogorov noted in its preface, Fréchet had already done the work to make the theories of measure and integration independent of geometric elements so that they could serve as an abstract framework for probability.

---

[8] Condensed and translated by the author from [26].

[9] In both French and English, denumerable is a synonym for countably infinite. Here it refers to a countably infinite sequence of trials.

[10] Ville's youth is described in detail in [31]. Borel's denumerable probability and its roots are studied thoroughly in the first volume of [5]. Fréchet's contributions are discussed in [1,35].

[11] Fréchet and Lévy may have overlooked it until towards the end of 1935 [1, p. 96].

4 Martingales in Probability Before Ville

In the last analysis, we cannot credit any 20th century mathematician with the introduction of martingales into mathematical probability—not Jean Ville, not Paul Lévy, not Joseph Doob. Martingales were there from the beginning—from before the beginning.

We can discern two threads in the early history of mathematical expectation, one associated with Pierre Fermat in his 1654 correspondence with Blaise Pascal, the other with Pascal and Christiaan Huygens. In betting games with multiple rounds, Fermat priced a player's payoff by counting cases. Pascal and Huygens after him priced it by considering the investment required to replicate it by successive bets. This required investment is the value of the payoff, or as Huygens put it in Latin, the value of the *expectatio*.[12]

When we talk about the investment required to replicate a payoff with successive bets, we are talking about a betting strategy. The payoff is the final value of this strategy's capital process—i.e., the final result of a martingale. This picture was quickly obscured, first by the "classical" picture that emphasized counting cases à la Fermat, and then by the 20th-century picture in which a count of cases is merely an example of measure.

After martingales had been naturalized in measure theory à la Doob, mathematicians sometimes noticed their presence in the work of the early authors. In 1983, Eugene Seneta called attention to the martingale in the solution of the gambler's ruin problem that De Moivre published in 1711 [8,30], and we now call it *De Moivre's martingale* [14, pp. 114, 271–272]. In 2003, Yves Derriennic introduced the name *martingale de Pascal* (Pascal's martingale) for the martingale defined by the backward recursion that Pascal explained to Fermat in his celebrated letter of 29 July 1654 [9]. But for the early authors, betting strategies were more than tools for calculating expectations. For Pascal, Huygens, and perhaps even De Moivre, they were embedded in the very meaning of mathematical expectation.

In 1985, Ville reminded Crépel of the classical link between probabilities and games of chance, and cited Bruno de Finetti, "who *defined* probability as the *inverse* of the payment, if the event happens, for staking 1 on the event now." Ville did not claim to have been the first to see the connection between martingales and probability. But he was proud of what he did with the connection.

[12] For accounts of Pascal's and Huygens's arguments, see [19,27,32]. Similar arguments have been found in manuscripts written by Italian teachers of commercial arithmetic in the early 15th century [25]. Whereas Huygens's *expectatio* designated the payoff itself, *expectation* is now used to mean the value of the payoff. Abraham De Moivre's *Doctrine of Chances*, first published in 1718 [7], can be credited with establishing this usage in English. The terms that became established in the 18th and 19th centuries in French (*espérance*) and German (*Hoffnung*) would usually be translated into English as "hope".

5 Combining Game Theory with Denumerable Probability

What Ville was so proud of? He told Crépel at the end of his 1985 letter: given an infinite sequence of outcomes and an event depending on them that has probability zero, he had shown how to find a betting system (a.k.a. martingale) whose capital process is nonnegative and tends to infinity when the event happens. "I insist on the point," he wrote in French, "because it took me so long to make this way of proceeding understood." He could not manage to make people understand what he wanted to say, he writes earlier. In retrospect he found this unsurprising. These people had been nourished on analysis, not on algorithms. They could have followed his mathematical reasoning if they had tried. They just did not see the point.

The point was as philosophical as mathematical. Ville is telling us why an event E of probability zero will not happen. It will not happen because we cannot beat the odds. We can beat the odds if E happens. We start with a finite amount of money (a tiny amount, if you please), risk no more than this (the capital process is nonnegative), and get infinitely rich.

Who did get the point? Borel was always sympathetic, from his Olympian distance. Fréchet had finally recognized value in Ville's argument. Wald understood the argument but found it beside the point so far as collectives were concerned; the purpose of the theory of collectives was to understand probability as frequency, not to conform with the concepts of denumerable probabilities. Von Mises agreed with Wald. Only de Finetti readily understood the argument and gave it its due; in a *Zentralblatt* review of Ville's thesis, he wrote that it gave, in a certain sense, the most general solution of the problem of collectives.[13]

William Feller doubtlessly understood the mathematical claims in Ville's 1936 announcement [40], but he did not see why they were being made. He wrote in his *Zentralblatt* review that Ville inexplicably (*Aus unerfindlichen Gründen*) wanted to alter Wald's idea of a countable set S of selection rules so that every set of measure zero can be ruled out by an acceptable S.

Did Joseph Doob understand what Ville was doing? In his review of Ville's 1939 book [11], Doob called the book "an interesting and valuable discussion of the concept of a collective" and noted that Ville discussed "systems of play in detail" and generalized the notion of a system to that of a "martingale". But after complaining, justly, about the carelessness of Ville's writing, he added:

> The author's main theorem on systems is not as strong as earlier results with which he is apparently unfamiliar. (Cf. Z. W. Birnbaum, J. Schreier, Studia Mathematica, vol. 4 (1933), pp. 85–89; J. L. Doob, Annals of Mathematics, (2), vol. 37 (1936), pp. 363–367.)

[13] For more on Wald's and von Mises's reactions, see de Finetti's summary of the Geneva conference, which he wrote after consulting some of the speakers for their additional thoughts [15], and the chapter by Laurent Bienvenu, Glenn Shafer, and Alexander Shen in the present volume.

This makes no sense. In this chapter of his book, Ville proved two theorems on systems. Theorem 1 (p. 90) tells us that for every set of sequences that has measure zero, there is a nonnegative martingale that is unbounded on all the sequences in the set. Theorem 2 (p. 92) tells us that for every nonnegative martingale, the set of sequences on which it is unbounded has measure zero. The two papers Doob cites do not prove stronger results; they are concerned with other questions. Birnbaum and Schreier's 1933 paper proves a version of Borel's strong law of large numbers. Doob's 1936 paper was his own way of making sense of von Mises; it showed that when you select without foreknowledge an infinite subsequence of a sequence of independent and identically distributed random variables, you do not change the joint distribution. If Doob did bother to understand this aspect of Ville's work, he did not manage to hold it in mind long enough finish his review. He was not interested.

As Laurent Mazliak recounts in his chapter in the present volume, Lévy also did not understand what Ville was doing. Lévy explained to Fréchet in 1964 that he never understood Ville's first definition—and Michel Loève and Aleksandr Khinchin had told him that they did not understand it either.

Did Ville give more than one definition of martingales in his book? We need to be careful with the word *define*. Ville had only one notion of a martingale. For him, a martingale was a betting system. More precisely, it was a strategy for betting in a casino game, where making a bet requires putting all the money the bet can lose on the table. A martingale starts with a finite amount of money (for convenience, Ville took this to be one monetary unit), and it specifies successive bets based on previous outcomes, with the constraint that no bet can risk more than the initial unit capital plus the net gain so far. What Lévy perceived as two definitions, Ville perceived as two ways of specifying a strategy, the second more general than the first. Ville formulated the first only for the case where one bets repeatedly on an event with constant probability not equal to zero or one; you can do this in a casino, and it provides the framework for perfecting von Mises's concept of a collective. The second applies whenever one bets successively on random variables for which successive conditional probabilities are defined.

1. **Betting on successive trials of an event with constant probability p not equal to zero or one.** On each trial, the bettor can bet both for and against the event. Ville wrote $s_0, s_1, s_2, \dots$ for the martingale's *capital process*, s_n being the amount of money the bettor has after the nth trial. Under Ville's assumptions, $s_0 = 1$, and each of the other s_n is a nonnegative function of the outcomes of the first n trials. If the bettor's capital $s_n(x_1, \dots, x_n)$ for a particular string $x_1, \dots, x_n$ of outcomes is zero, then his capital remains zero thereafter, because he can no longer bet. Here, Ville noted, the martingale is "sufficiently defined" by the sequence of functions (s_n).[14] We can recover the betting strategy from this sequence.

[14] Reference [43, p. 88]: "Le système de jeu de A est donc suffisament défini par la donnée de la suite de fonctions $\{s_n\}$."

2. **Starting with conditional probabilities.** Suppose that $X_1, X_2, \ldots$ is a sequence of random variables, that $s_0 = 1$, that $s_1(x_1), s_2(x_1, x_2), \ldots$ are nonnegative functions, and that

$$\mathbf{E}\left(s_n(x_1, \ldots, x_{n-1}, X_n) \mid x_1, \ldots, x_{n-1}\right) = s_{n-1}(x_1, \ldots, x_{n-1}) \qquad (1)$$

for $n = 1, 2 \ldots$. Now we have left the casino for an idealized world of betting, where X_n's outcome x_n might be any real number. But here too, according to Ville, the (s_n) can be taken to describe a martingale in a betting game: on the nth round, the bettor puts $s_{n-1}(x_1, \ldots, x_{n-1})$ on the table, on the understanding that he will get back $s_n(x_1, \ldots, x_{n-1}, x_n)$ after $X_n = x_n$ is observed.[15] This formalism allows the game to continue even when the bettor's capital is zero; we can construct examples where (1) holds and yet there are outcomes $x_1, \ldots, x_n$ such that $s_{n-1}(x_1, \ldots, x_{n-1}) = 0$ and $s_n(x_1, \ldots, x_{n-1}, x_n) > 0$.

The first definition was incomprehensible for Lévy, because he did not formalize betting. The second definition, or at least Eq. (1), was quite comprehensible for Lévy. Ville had said explicitly that the equation used Lévy's concept of conditional probability [43, p. 96].

Ville told Crépel that he had argued with Lévy about how probabilities for a random sequence should be defined. Ville thought the probabilities should be given in sequence: the probability for the first outcome, then the probability for the second given the first, etc. Certainly, he told Crépel, this is appropriate for the study of martingales. As Bernard Locker recounts the next chapter of the present volume, Lévy understood the picture in which probabilities are given successively; it was central to his intuition. But like Kolmogorov and Doob, Lévy insisted that formal theory give joint probabilities, from which conditional probabilities are derived, albeit sometimes incompletely. As a doctoral candidate, Ville needed Lévy's support, and so he was obliged to use Lévy's measure-theoretic concept of conditional probability for Eq. (1).

When Ville moved on from a sequence of random variables to a continuous stochastic process, he used the obvious generalization of (1) to continuous time [43, pp. 111–129]. This severed the direct connection with betting, because a bettor cannot bet continuously. So he simply explored the consequences of the equation, without using the word *martingale*. He used the Kolmogorov/Doob concept of conditional probability, because Lévy had not extended his concept to continuous time.

[15] Reference [43, p. 99]: "Dans ces conditions, nous dirons que la suite $\{s_n\}$ définit une martingale ou un jeu équitable." By equating the martingale with a "jeu équitable" or fair game, Ville is presuming that the conditional probabilities are somehow correct or true. This is a subtle shift from his first definition, where the martingale serves to test the validity of a collective. It is contrasts more sharply with the meaning of *martingale* in the casino, where we assume that the house has an advantage; see §5 of my chapter "Martingales at the casino" in the present volume.

Once we have decided, with Ville, to use capital processes as our exclusive way of describing martingales, it is natural to simplify the mathematical language by identifying the martingale with the capital process. But this step was taken only a decade later by Doob. Perhaps Ville did it when talking about his work, but we do not see it in his writing. This writing was all student writing, of course, and it might have been too assertive to change the meaning of a word his teachers used. Borel was still using *martingale* to mean betting strategy in 1949 [2]. Nevertheless, it was Ville's second "definition" that moved martingales from the casino into the world of conditional probability. Perhaps against his will, he even made this world of conditional probability measure-theoretic, à la Lévy, Kolmogorov, and Doob.[16]

Around the time Ville consulted Lévy about his thesis work, Lévy was working intensely on sums of the form $S_n = Y_1 + Y_2 + \cdots$, where the Y_n are random variables satisfying

$$\mathbf{E}\,(Y_n \mid Y_1, Y_2, \ldots, Y_{n-1}) = 0. \tag{2}$$

He called Eq. (2) *condition* $\mathscr{C}$ [23, p. 238]. It appears that he saw no connection between this and what Ville was doing when Ville consulted him,[17] but much later, when he heard talk about Ville's martingales, he learned that his sequences (S_n) were martingales. So for Lévy, Ville was just naming something Lévy had already been doing. Ville never thought about the matter this way. Lévy had studied such sequences, but so had Pascal, Huygens, De Moivre, and everyone else. What Ville had done was show how such sequences could help us understand the new concept of probability zero that Borel had introduced with his theory of denumerable probabilities. Ville told Crépel that he had been surprised when—in the late 1940s at Toulon following a lecture he gave on signal theory—someone told him that Lévy had invented martingales.

The mathematicians could not digest Ville's first definition because it was explicitly about strategies in a game, or about algorithms if you prefer. In the 1930s, games and algorithms were not mathematics. Borel and John von Neumann were interested in games; Borel had studied mixed strategies in games in the early 1920s, and von Neumann had proven the minimax theorem in 1928. But Borel and von Neumann were striking exceptions. For the functional analysts of the 1930s, computing might exist in the real world, but not in the platonic world of mathematics. Betting was an important source of intuition for mathematical probability, but not something fit for mathematization.

[16] As already noted, Ville had argued against the measure-theoretic treatment of conditional probability, preferring instead to take a system of conditional probabilities as the starting point. His 1938 note on continuous martingales in the *Comptes rendus* [42] suggests that he could have used his preferred approach even in continuous time.

[17] Laurent Mazliak, in the next chapter of the present volume, reports that Lévy arrived at his condition $\mathscr{C}$ only late in 1935. It seems conceivable, if unlikely, that the conversation with Ville helped lead him to it.

6 Legacy

As out-of-time as it may have been in 1939, Ville's perspective on martingales has proven remarkably fruitful in the subsequent 80 years. It has left a legacy in three distinct fields: measure-theoretic probability, algorithmic randomness, and game-theoretic probability.

Measure-Theoretic Probability

Ville's influence on measure-theoretic probability was mediated almost exclusively through Doob. This mediation is described in detail by Bernard Locker in his chapter in the present volume. As Locker recounts, Doob used and extended Ville's ideas almost immediately, in a paper that appeared in 1940 [12].

Doob did not use the word *martingale* in his early papers on the topic. The 1940 paper began with a condition on a sequence $S_1, S_2, \ldots$ of random variables that he called "property $\mathscr{E}$":

$$\mathbf{E}(S_{n+1} \mid S_1, \ldots, S_n) = S_n. \tag{3}$$

The name deliberately directed attention to Lévy and his "condition $\mathscr{C}$". But beginning in 1948 in Lyon, Doob borrowed the name *martingale* from Ville, with a twist. For Doob, a martingale was not a strategy that produced a nonnegative capital process. It was a sequence satisfying (3), nonnegative or not. Doob often credited the colorful name *martingale* for the success of martingales in probability theory [38, p. 253].

Doob was particularly interested in an inequality on p. 100 of Ville's book: if (S_n) is a nonnegative martingale, $S_0 = 1$, and $\lambda > 0$, then

$$\Pr\left(\sup_n S_n \geq \lambda\right) \leq \frac{1}{\lambda}. \tag{4}$$

In his review of Ville's book in 1939, Doob had called this Ville's "main theorem on martingales". It implies that a nonnegative martingale is bounded with probability one. When it stays bounded, how does it behave? Does it wander up and down by some amount forever, or does it converge? As Doob showed in the 1940 paper, it converges with probability one. This is Doob's celebrated martingale convergence theorem. Among many related results in the 1940 paper was a theory for the case where the index n takes negative values; in this case there is convergence when n tends to $-\infty$.

Doob appreciation of the word *martingale* did not extend to an appreciation of Ville's use of it as the name for betting strategies. As Doob put it, in a formulation often repeated by his students, Ville "did not formally define a martingale" [37, p. 307]. Gambling could only be an application, after you had defined the notion of a martingale in measure-theoretic terms [13].

Algorithmic Randomness

The history of the study of randomness is discussed in detail by Bienvenu, Shafer, and Shen in their chapter in the present volume. Here I summarize how Ville's martingales entered into the story.

After World War II, with measure-theoretic probability ascendant, mathematicians momentarily lost interest in von Mises's project of characterizing random sequences. Attention returned to the topic in the 1960s, however, as a result of the discovery of algorithmic complexity, now called *Kolmogorov complexity*, by Kolmogorov and others.

Despite his role in establishing the measure-theoretic framework for the study of probability as pure mathematics, Kolmogorov had not seen this framework as a foundation for applications. He thought applications should instead rely on a finite version of von Mises's picture, even if such a finite version could not be made into rigorous mathematics. In the early 1960s, he realized that a rigorous finite version of von Mises might be given in terms of algorithmic complexity: the random finite sequences are those whose frequencies are not changed very much by relatively simple selection rules. Soon afterwards, he realized that the notion of algorithmic complexity could be used more directly to define randomness: a finite sequence is random if it is sufficiently complex.

In 1966, the Swedish mathematician Per Martin-Löf combined Kolmogorov's notion of algorithmic complexity with measure theory to produce a definition of randomness for an infinite sequence of 0s and 1s: roughly speaking, the sequence is random if it does not belong to any set for which there is an algorithm that proves the set has measure zero. In April 1966, as he was publishing this idea, he also discovered Ville's work and discussed it in lectures at the University of Erlangen-Nürnberg. The following year, the young German mathematician Claus-Peter Schnorr studied notes of these lectures.

It was Schnorr who brought Ville's notion of a martingale back into the picture, using it to reformulate and refine Martin-Löf's ideas. As Schnorr showed, an infinite binary sequence is random in Martin-Löf's sense if and only if no lower semicomputable nonnegative supermartingale becomes unboundedly rich when played against it. Schnorr used this formulation to make nonrandomness quantitative: a sequence is more nonrandom when the supermartingale becomes rich faster.

Game-Theoretic Probability

In the 1970s and 1980s, the successful analyses of randomness in terms of complexity and in terms of betting led many researchers to ask whether these analyses could cast light on applications of probability.

One example of this thinking appeared in A. Philip Dawid's work on probability forecasting in the 1980s. If we give probabilities for a sequence of outcomes not as a single joint distribution but rather, as Ville advocated in his conversation with Lévy, as a sequence consisting of probabilities for the first outcome and conditional probabilities for each successive outcome, then we can think of this information as a strategy for a probability forecaster. Dawid called it a *probability forecasting system*. This led Dawid to martingales and to Ville. In a 1985 article [6, p. 1270], Dawid noted that the capital process for a strategy for betting against a forecasting system is a martingale. Writing c_n for the strategy's bet on the nth round and f_n for its capital at the end of the nth round, he asserted that

> …we can test the validity of this forecasting system by requiring that (f_n) should "look like" a martingale realisation. One way of formalising this, generalizing ideas of Ville (1939) and Schnorr (1971) for the Bernoulli case, is as follows. Consider an opponent who starts off with unit capital $f_0 = 1$. At any time he may choose c_n, as a function of past data, subject to the restriction that he must always have enough capital to meet his debt if he loses. …We can therefore impose, as a new validity criterion, the requirement that ... [the] fortune sequence …is bounded above. …This may be shown to be essentially the same as the requirements of Howard (1975) and Martin-Löf (1966). This martingale criterion says that, when betting at "correct" odds, it is impossible to make an unlimited fortune out of a finite initial capital. As a basis for a theory of probability, it has much in common with (and is as soundly established in practice as) the principle of the impossibility of a perpetual motion machine as a basis for physics.[18]

Françoise Seillier-Moiseiwitsch and Dawid subsequently used this framework to derive a central limit theorem that permits testing the forecaster using only the realized sequence of forecasts and outcomes [29].

Another researcher who became interested in relating Ville's game-theoretic ideas to applications was Vladimir Vovk, who began his study of probability with Andrei Kolmogorov in the 1980s. In the early 1990s, Vovk realized that the techniques he had learned in complexity theory and used, for example, to establish the law of the iterated logarithm for random Kolmogorov sequences [45] could be adapted to Ville's picture.

Subsequent research on these ideas has led to what might appear, in retrospect, as an obvious way of developing Ville's picture. Ville's main result was that an event determined by a sequence of trials has probability zero if and only if there is a nonnegative martingale that tends to infinity when that event happens. This is easily generalized, in the framework of Ville's betting game, to the conclusion that the expected value of a bounded random variable determined by a sequence of trials is the least amount we need to stake in order to get the random variable as a payoff. More or less equivalently, it is the initial value of a bounded martingale that produces the random variable in the limit. In *Game-Theoretic Foundations for Probability and Finance*, Vovk and I call a very general version of this last statement *Ville's theorem* [36, p. 178]. It can be thought of as a generalization of Huygens's definition of expected value.

The game-theoretic picture also generalizes to the case, presaged by Dawid's formulation, where the forecaster and his opponent may decide how to play as they go along rather than following strategies fixed in advance, and further to the case where the individual forecasts may price only some of the payoffs on the individual trials.

[18] [43], [28], [21], [24]. In fact, Ville himself had already explicitly generalized beyond the Bernoulli case in this chapter.

Whereas Schnorr's martingale-based theory of algorithmic randomness lives in the idealized world of infinite sequences, game-theoretic probability lends itself to applications in the finite world. In this finite world, we can test statistical hypotheses by betting against them, rejecting them if we multiply the money we risk by a large amount—i.e., if a nonnegative martingale starting at 1 becomes very large [33]. According to Ville's inequality, (4), this is consistent with the classical notion of testing by singling out a rejection region of small probability. But it is a stronger and more flexible concept, because it allows us to bet opportunistically and to claim the full weight of the final value of our martingale rather than a level λ fixed in advance.

The very name *Ville's inequality* is evidence of the resurgence of Ville's game-theoretic picture. Ville himself had called it the "theorem of gamblers' ruin". If you go into a casino with 1 unit of capital, and the casino's capital is some large amount $\lambda - 1$, then your chance of breaking the bank before losing your own 1 unit is no more than $1/\lambda$, no matter how cleverly and how long you bet [43, p. 100]. In his 1947 book on sequential analysis [47, p. 146], Wald stated and proved the inequality using the vocabulary of likelihood ratios, without mentioning Ville or the word *martingale*. Doob, even though he had recognized the inequality as Ville's main result on martingales in his review of Ville's book, buried it in his magisterial 1953 book on stochastic processes; there it was merely a consequence of a special case of one of many inequalities concerning martingales and supermartingales. If it was mentioned at all in the following half century, it was apt to be attributed to Doob himself. But in the past decade it has become *Ville's inequality*.[19]

7 A Final Question

When we realize how powerful Ville's understanding of martingales has turned out to be, we feel compelled to ask why he abandoned it. Why did he never take up martingales again after returning to France? Marie-France Bru, Bernard Bru, and Kai Lai Chung have suggested that Ville gave up martingales because Doob and Lévy were too far ahead by the time he returned from German captivity in 1941 [4, p. 39]. The diversity of Ville's later publications and activities, together with Blanc-Lapierre's description of his personality, suggests a complementary hypothesis. Perhaps he was bored with martingales. Perhaps Ville was bored by measure theory and pure mathematics in general. Surely Doob's punctilious development of conditional expectation must have bored him. Surely replicating the results of the axiomatic theory in a different way would have bored him. He had given the world one flash of insight. It was time to move on.

Acknowledgements This chapter has benefited from advice and critical comments by Laurent Mazliak, Nell Painter, Aaditya Ramdas, and Vladimir Vovk.

[19] Ville's inequality follows from inequality (3.4′) on p. 314 of Doob's 1953 book [13]. Shafer and Vovk call it "Doob's inequality" on p. 56 and p. 196 of [34].

References

1. Barbut, M., Locker, B., Mazliak, L.: Paul Lévy and Maurice Fréchet: 50 years of correspondence in 107 letters. Springer (2014)
2. Borel, É.: Sur une martingale mineure. Comptes rendus **229**(23), 1181–1183 (1949)
3. Bru, B.: Doeblin's life and work from his correspondence. In: Doeblin and Modern Probability, pp. 1–64. American Mathematical Society (1993)
4. Bru, B., Bru, M.F., Chung, K.L.: Borel and the St. Petersburg martingale. Electronic Journal for History of Probability and Statistics **5**(1) (2009)
5. Bru, M.F., Bru, B.: Les jeux de l'infini et du hasard. Presses universitaires de Franche-Comté, Besançon, France (2018). Two volumes
6. Dawid, A.P.: Calibration-based empirical probability. Annals of Statistics **13**(4), 1251–1273 (1985)
7. De Moivre, A.: The Doctrine of Chances: or, A Method of Calculating the Probabilities of Events in Play. Pearson, London (1718)
8. De Moivre, A.: A. de Moivre: 'De Mensura Sortis' or 'On the Measurement of Chance. International Statistical Review **52**(3), 229–262 (1984)
9. Derriennic, Y.: Pascal et les problèmes du chevalier de Méré. De l'origine du calcul des probabilités aux mathématiques financiéres d'aujourd'hui. Gazette des mathématiciens **97**, 45–71 (2003)
10. Doeblin, W.: Œvres Complètes; Collected Works. Springer (2020). Edited by Marc Yor and Bernard Bru
11. Doob, J.L.: Review of [43]. Bulletin of the American Mathematical Society **45**(11), 824 (1939)
12. Doob, J.L.: Regularity properties of certain families of chance variables. Transactions of the American Mathematical Society **47**(3), 455–486 (1940)
13. Doob, J.L.: Stochastic Processes. Wiley, New York (1953)
14. Ethier, S.N.: The Doctrine of Chances: Probabilistic Aspects of Gambling. Springer (2010)
15. de Finetti, B.: Compte rendu critique du colloque de Genève sur la théorie des probabilités. No. 766 in Actualités Scientifiques et Industrielles. Hermann, Paris (1939)
16. Fortet, R.M.: Faut-il élargir les axiomes du calcul des probabilités. Actualités Scientifiques et Industrielles **1146**, 35–47 (1951)
17. Fréchet, M.: Exposé et discussion de quelques recherches récentes sur les fondements du calcul des probabilités. In: R. Wavre (ed.) Les fondements du calcul des probabilités, No. 735 in Actualités Scientifiques et Industrielles, pp. 7–23. Hermann, Paris (1938)
18. Fréchet, M.: The diverse definitions of probability. Journal of Unified Science (Erkenntis) **8**, 7–23 (1939)
19. Freudenthal, H.: Huygens' foundations of probability. Historia Mathematica **7**, 113–117 (1980)
20. Hempel, C.G.: Review of [18]. Journal of Symbolic Logic **5**(3), 122–123 (1940)
21. Howard, J.V.: Computable explanations. Mathematical Logic Quarterly **21**, 215–224 (1975)
22. Kolmogorov, A.N.: Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer, Berlin (1933). An English translation by Nathan Morrison appeared under the title *Foundations of the Theory of Probability* (Chelsea, New York) in 1950, with a second edition in 1956.
23. Lévy, P.: Théorie de l'addition des variables aléatoires. Gauthier-Villars, Paris (1937). Second edition: 1954.
24. Martin-Löf, P.: The definition of random sequences. Information and Control **9**, 602–619 (1966)
25. Meusnier, N.: Le problème des partis bouge... de plus en plus. Electronic Journal for History of Probability and Statistics **3**(1) (2007)
26. Mounier-Kuhn, P.E.: L'informatique en France, de la Seconde Guerre mondiale au Plan Calcul: Science, Industrie, Politiques publiques. Conservatoire National des Arts et Métiers - Centre Science, Technologie et Société, Paris (1999)
27. Schneider, I.: Christiaan Huygens' non-probabilistic approach to a calculus of games of chance. De zeventiende eeuw **12**(1), 171–185 (1996)
28. Schnorr, C.P.: A unified approach to the definition of random sequences. Math. Systems. Theory **5**, 246–258 (1971)

29. Seillier-Moiseiwitsch, F., Dawid, A.: On testing the validity of sequential probability forecasts. Journal of the American Statistical Association **88**(421), 355–359 (1993)
30. Seneta, E.: Modern probabilistic concepts in the work of E. Abbe and A. De Moivre. Mathematical Scientist **8**(2), 75–80 (1983)
31. Shafer, G.: The education of Jean André Ville. Electronic Journal for History of Probability and Statistics **5**(1) (2009)
32. Shafer, G.: Pascal's and Huygens's game-theoretic foundations for probability. Sartoniana **32**, 117–145 (2019)
33. Shafer, G.: Testing by betting: A strategy for statistical and scientific communication, with discussion. Journal of the Royal Statistical Society, Series A **184**(2), 407–478 (2021)
34. Shafer, G., Vovk, V.: Probability and Finance: It's Only a Game! Wiley, New York (2001)
35. Shafer, G., Vovk, V.: The sources of Kolmogorov's Grundbegriffe. Statistical Science **21**(1), 70–98 (2006)
36. Shafer, G., Vovk, V.: Game-Theoretic Foundations for Probability and Finance. Wiley, Hoboken, New Jersey (2019)
37. Snell, J.L.: A conversation with Joe Doob. Statistical Science **12**(4), 301–311 (1997)
38. Snell, J.L.: Obituary: Joseph Leonard Doob. Journal of Applied Probability **42**, 247–256 (2006)
39. Ville, J.: Sur la convergence de la médiane des n premiers résultats d'une suite infinie d'épreuves indépendantes. Comptes rendus **203**, 1309–1310 (1936)
40. Ville, J.: Sur la notion de collectif. Comptes rendus **203**, 26–27 (1936)
41. Ville, J.: Sur les suites indifférentes. Comptes rendus **202**, 1393–1394 (1936)
42. Ville, J.: Sur un jeu continu. Comptes rendus **206**, 968–969 (1938)
43. Ville, J.: Étude critique de la notion de collectif. Gauthier-Villars, Paris (1939)
44. Ville, J.: Thèses présentées à la Faculté des Sciences de Paris pour obtenir le grade de Docteur ès Sciences mathématique. Première thèse. — Étude critique de la notion de collectif. Deuxième thèse. — La transformation de Laplace. Gauthier-Villars, Paris (1939)
45. Vovk, V.: The law of the iterated logarithm for random Kolmogorov, or chaotic, sequences. Theory of probability and its applications **32**, 413–425 (1987)
46. Wald, A.: Sur la notion de collectif dans le calcul des probabilités. Comptes rendus **202**, 180–183 (1936)
47. Wald, A.: Sequential Analysis. Wiley (1947)

Paul Lévy's Perspective on Jean Ville and Martingales

Laurent Mazliak

Abstract

As the first part of this chapter explains, Paul Lévy's theory of martingales was about extending the law of large numbers and other theorems about sequences of independent random variables to dependent random variables, Lévy showed that this extension is possible when each random variable has mean zero given the preceding ones. Under this condition, the sequence of cumulative sums is a *martingale* as Jean Ville and later Joseph L. Doob used the word, but Lévy never focused on this sequence of cumulative sums as a mathematical object. In this respect, his was not a theory of martingales. Moreover, he never showed much interest in the properties of martingales studied by Ville and Doob. The second part of the chapter describes Lévy's troubled relationship with Ville and his disdain for Ville's mathematical work. We find insights into Lévy's attitude towards Ville in the decades-long correspondence Lévy sustained with Maurice Fréchet.

Keywords

History of probability theory · Martingales · Dependent random variables · Paul Lévy · Jean Ville

AMS Classification: Primary: 01A60, 60-03 · Secondary: 60G42 · 60G44

1 Introduction

Paul Lévy (1886–1971) was one of the major figures on the probabilistic scene of the 20th century, and his research on limit theorems for sums of dependent random variables in the mid 1930s had considerable influence on martingale theory. However,

L. Mazliak (✉)
Statistiques et Modélisation, Sorbonne Université. Laboratoire de Probabilités, Paris, France
e-mail: laurent.mazliak@upmc.fr

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*,
Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_6

Lévy was never interested in the viewpoint of Jean Ville (1910–1989), who thought of martingales as capital processes. For Lévy, the condition on individual random variables that makes the sequence of their cumulative sums a martingale always remained a technical condition. Added to Lévy's personal mathematical disdain for Ville, for which we will provide some explanation, this disinterest explains why Lévy remained away from the flowering of martingale theory after World War II.

The first part of this chapter is about Lévy's important research on topics connected with martingales: how he became interested in his martingale condition and how he used it. After first recalling the singular path followed by Lévy towards probability after the Great War, we discuss the kind of problems he considered and their origin. In particular, we emphasize the important question of the probabilistic study of continuous fractions which, from the very beginning of 20th century (especially in Borel's studies) had been a source of inspiration for major developments in probability. We then describe several works by Lévy in which he introduced conditions related to martingales. We discuss in detail Chap. VIII of his seminal 1937 book [32], where he collected the results he obtained in the 1930s about the extension of limit theorems to dependent random variables. The picture he set out in Chap. VIII remained Lévy's vision of martingales for the rest of his life.

The second part focuses on Lévy's troubled relationship with Ville and tries to explain his constant misunderstanding of the significance of Ville's work. Here we draw on the letters from Lévy to Maurice Fréchet (1878–1973) that are reproduced in [2] and in the Chap. 18 of the present volume. These letters show that the gap between the two mathematicians was widened by an unfortunate combination of circumstances, including a clumsy publication by Ville in 1936, Lévy's taste for quick and final judgments on people, and later the troubled times of the war and the Occupation. Lévy never thought highly of Ville, and this is repeatedly demonstrated by scornful comments in his correspondence with Fréchet. We do not know exactly to what extent this disdain had an effect on Ville—but it probably had some. In any case, this complicated situation casts some light on the creation of fundamental tools of modern probability theory.

2 Lévy and His Martingale Condition

This section begins by briefly recalling why and how Lévy, who had never been interested in probability theory before the Great War, was suddenly captivated by the subject to the point of becoming the unchallenged leading French probabilist of the interwar period.[1] Then, following the account given by Pierre Crépel in a seminar in Rennes in 1984 [18], we discuss the genesis of Lévy's martingale condition. We conclude with our discussion of Chap. VIII of Lévy's 1937 book.

[1] For more detailed treatments of this topic, see Lévy's comments in his autobiography [2,3,34,35,39].

2.1 Lévy's Growing Interest in Probability

Lévy's first encounter with probability as a professional mathematician happened merely by chance. In 1919, illness prevented Georges Humbert from giving part of his course at the Ecole Polytechnique where he was professor of mathematical analysis. Lévy, who had been a *répétiteur* (lecturer) since 1913 at the Polytechnique, a school where he himself had been an outstanding student 12 years earlier, was asked to replace Humbert on the spot for some lectures. Among them were three lectures on probability theory. We luckily have the lecture notes; they were published along with commentary in 2008 in the Electronic Journal for History of Probability and Statistics [3]. A renewed interest in probability in the Polytechnique's curriculum had resulted from the experience of the war, where some basic probabilistic techniques had been used on a very large scale. In particular, least squares had been used to improve the precision of gunfire.

Lévy's engagement with probability might have stopped with elementary teaching, but at the same moment he was resuming his research into functional analysis having finally been freed from his military obligations. (He later wrote that he had mainly worked on anti-aircraft defense during the war [34, pp. 54–55].) His prominence as a probabilist now obscures the fact that before becoming a specialist in probability, he had been a follower of Volterra and Hadamard's techniques of functions of lines for the functional analysis of general electric charge distributions. In 1911, he had defended a brilliant thesis in which he studied Green functions as functions of lines that are solutions of integro-differential equations. After the war, as explained in [39], Hadamard asked Lévy to prepare a posthumous edition of Gateaux's papers. The young French mathematician René Gateaux (1889–1914) had been killed on the front in October 1914. In the previous months, he had collected material for a thesis on functional analysis in which he began to construct an original theory of infinite dimensional integration. Hadamard's request played a major role in Lévy's evolution, for Lévy realized that a probabilistic framework was well adapted to Gateaux's problems. A letter written to Fréchet much later, in April 1945 [2, Letter 57], testifies to the *mathematical technology transfer* operated by Lévy during those years between probability and functional analysis.

> As for myself, I learned the first elements of probability during the spring of 1919 thanks to Carvallo (the director of studies at the Ecole Polytechnique) who asked me to give three lectures on that topic to the students there. In addition, in three weeks, I succeeded in proving new results. And never will I claim for my work in probability a date before 1919. I can even add, and I told Mr. Borel so, that I had not really seen before 1929 the importance of the new problems implied by the theory of denumerable probabilities. But I was prepared by functional calculus for the study of functions with an infinite number of variables, and many of my ideas in functional analysis became without effort ideas which could be applied in probability.

In fact, a first trace of the probabilistic vision can be found in the Lévy-Fréchet correspondence as early as January 1919 (so even before Lévy really became involved in probability…) when Lévy wrote to Fréchet [2, p. 59],

> For example, I think to limit the oscillations and irregularities of the functions by bounding an integral I such as $\int u'^2(t)dt$, or at least by considering as "less probable" the functions for which I would be too large.

The new probabilistic oriented mind proved especially spectacular in Lévy's 1922 book on functional analysis [27], in particular in the third part of the book. In [39], I have studied in detail how Norbert Wiener, an enthusiastic reader of [27], found there the right tools for the infinite-dimensional integration he needed to complete his construction of the Brownian motion measure.

2.2 Genesis of Lévy's Martingale Condition

As Pierre Crépel mentions in [18], the Soviet mathematician Sergei N. Bernstein (1880–1968) had studied, during the 1920s and the beginning of the 1930s, conditions related to martingales or at least approximate martingales but had not singled out the notion of a martingale and given it an autonomous mathematical definition.

In 1927, in a special issue of Matematische Annalen in honor of Riemann's centennial, Bernstein published a long paper, 59 dense pages, devoted to the extension of the central limit theorem to sums of dependent variables [5]. Bernstein had become interested in probability theory some years after beginning a brilliant career in analysis. The change of orientation was mostly due to the opportunity to work as a statistician during the hard years following the revolution.[2] Being interested in applications of mathematics, Berstein was an eager reader of the works of the so-called Petersburg school of probability (Chebyshev, Markov, Liapunov), and he became at that moment the best expert in this group's results on limit theorems, approximation and stability theory, and also events in chain (called by Bernstein the 'chains of A. Markoff'). At that point, he was one of the few to remember Markov's contributions to this topic. In the Annalen paper, Bernstein set out conditions under which random variables $S_n/\sqrt{B_n}$ converge towards the standard normal distribution, where S_n is the sum of dependent variables x_i and $B_n = E(S_n^2)$. He used such a condition, for example, in his theorem B on p. 24 in order to conveniently control the growth of the quadratic expectation of $S_{i+k} - S_i$ conditional on $x_1, \ldots, x_i$.[3]

We may ask what exactly Lévy knew about Bernstein's work before he himself introduced a martingale condition. It is hard to answer this question definitively, but the evidence suggests that Bernstein's influence on Lévy at that moment was quite limited. First because, as Lévy himself often said, he was not very fond of reading the works of others. Such an assertion is not necessarily to be taken at face value, but in Lévy's case it is corroborated by converging information. It is also striking that Bernstein's name appears only very late in Lévy's correspondence with Fréchet, at least in the letters found at the Paris Academy of Science and published in [2],

[2] See for instance [43].

[3] See [20, pp. 253ff] for details about Bernstein's extensions of the central limit theorem to sums of dependent variables.

in contrast with the names of other Soviet scientists such as Andrei N. Kolmogorov (1903–1987) and Aleksandr Y. Khinchin (1894–1959). The first mention of Bernstein occurred in 1942. The correspondence is not complete and Bernstein may certainly have been quoted before. But in his letter dated 4 November 1942, Lévy explained that he asked Loève to give him a description of Bernstein's 1932 talk at the international congress of mathematicians in Zürich, which seems to reveal that he had at most a superficial knowledge of it. It is true that Lévy wrote at the very beginning of his 1935 paper [30] that Bernstein's 1927 paper [5] was an *important step* in the study of sums of dependent variables. But Lévy does not refer to this work of Bernstein before 1935, and it is possible that he was not acquainted with it at all before someone told him that Bernstein had dealt with questions similar to the ones he was considering. Fréchet, who read everything published, often played this role of bibliographical source for Lévy. Our hypothesis is therefore that Lévy's work on martingales was not inspired or substantially influenced by Bernstein's work.

A first trace of Lévy's use of a martingale condition in a primitive setting can be found in a paper he wrote in 1929 [29] about the decomposition of a real number in continued fractions. The role of continued fractions in the development of probabilistic ideas at the beginning of the 20th century has been investigated in several studies. In particular [14,40] study their influence on Borel's evolution towards probability through his acquaintance with the paper of the Lund number theorist Anders Wiman (1865–1959) [49] about Gylden's statistical treatment of the mean motion of planets around the sun. Borel's amazement about Wiman's use of σ-additivity to calculate the distribution of the coefficients of the continued fraction expansion of a random element of [0, 1] led to his own first probability paper [6], in which he used the new Lebesgue measure and integral to formulate probabilistic situations.

Borel always saw the example of continued fractions as a fundamental source of randomness. This example was particularly important in Borel's seminal 1909 article [7] where he presented the application of denumerable probabilities to the decomposition of real numbers, both in decimal and in continued fraction developments. Here Borel introduced the notion of almost sure convergence and a first version of the strong law of large numbers, thus inaugurating the technique of proving existence by a probability computation, which became a hallmark of Borelian reasoning. This reasoning came directly from how he had introduced the measure of sets in his thesis 15 years earlier. To prove the existence of an arc of a circle on which a certain series was uniformly convergent, Borel proved that he could choose the center of such an arc in the complement of a set which he had proved to be of measure zero [24]. So from the very beginning of his probabilistic life, Borel used the proof that an event has probability 1 as a proof of existence. A good example of the technique is given in [7, Sect. II.13], where Borel commented on the proof that almost every real number is *absolutely normal*. Recall that a number is said to be *normal* if each figure between 0 and 9 appears with a frequency 1/10 in its decimal decomposition; it is absolutely normal if the same property is true with the d-basis decomposition (with a frequence $1/d$) for each integer d. Borel wrote,

> In the present state of science, the effective determination of an absolutely normal number seems to be the most difficult problem; it would be interesting to solve it, either by constructing an absolutely normal number or by proving that, among numbers that can be effectively defined, none is absolutely normal. However paradoxical may this proposition seem, it is not the least incompatible with the fact that the probability for a number to be absolutely normal is equal to one.

The strangeness of this kind of existence proof probably explains why, as Jan von Plato observes, the strong law of large numbers and denumerable probabilities seem to have caught mathematicians by surprise and attracted a number of uncomprehending reactions [41, p. 57]. A vigorous reaction came in 1912 from Felix Bernstein (1878–1956) when he revisited Gylden's approach to the problem of secular perturbations by a systematic use of the 'measure of sets of E. Borel and H. Lebesgue' [4, p. 421].[4] Bernstein contested the result obtained by Borel in [7] concerning the asymptotic order of the quotients in a continued fraction and thought he had found a contradiction with his own results. As Bernstein wrote,

> For the continued fractions, [Borel] established the following result : if one considers only quotients which have an influence on $\overline{\lim}\ a_n$, then their growth order is smaller than $\varphi(n)$ with denumerable probability 1 if $\sum \frac{1}{\varphi(n)}$ converges, and larger than $\varphi(n)$ if $\sum \frac{1}{\varphi(n)}$ diverges.
>
> The last part of the theorem is contained in the second part of theorem 4.[5] On the contrary, the first part is in contradiction with the result obtained in theorem 4. The reason for this contradiction is of crucial importance and we shall explain it precisely. The following fact is true : *for geometrical probabilities under consideration, the independence of the elementary cases is not realized.*

So for Bernstein, the source of the contradiction was Borel's application of his (Borel-Cantelli) lemma to a non-independent case. Several weeks later, Borel replied in a note published in the same journal [9]. He emphasized that Bernstein's result is in no way contradictory with his own, but admitted that in [7] he had not taken care to formulate his lemma for dependent variables such as the quotients a_n. On p. 579 of his note he gave a new proof, in which he assumed that the conditional probability p_n of the nth event given the preceding ones satisfies $p'_n \leq p_n \leq p''_n$ where the series p'_n and p''_n have the same behavior (convergence or divergence). Borel does not give any hint of how one may obtain the two terms p'_n and p''_n. Moreover he limits the proof (of this conditional Borel-Cantelli lemma) to the case when p'_n and p''_n are convergent series, asserting without elaboration that the proof would be the same in the divergent case (an unfortunate claim as it is false in the non-independent divergent case!). Nevertheless, one may detect in this proof (where Borel considers the evolution of the conditional means) a first use of a martingale convergence theorem. This is today used as a common tool for obtaining the conditional version of the Borel-Cantelli lemma (see for instance [1, p. 35]).

[4] Bernstein's interest in secular perturbations was inspired by a paper published by Bohl in 1909.

[5] That is to say, Bernstein's own Theorem 4 in [4].

At the same moment, Borel was revisiting the card shuffling problem Poincaré had discussed in the second edition of his textbook on probability [42], proposing in the note [8] a probabilistic proof of convergence to the uniform distribution (ergodic theorem) by consideration of the evolution of the means. This was the first appearance of a probabilistic proof of convergence of a Markov chain, apart from Markov's original proof, which remained completely unnoticed by the French until much later. Borel's note also remained unnoticed, and his proof was rediscovered and extended by Lévy, Hadamard, Hostinský and others at the end of the 1920s [13,37,38].

In [9], Borel underlines Felix Bernstein's confusion; for him, Bernstein did not understand that in the convergent case it may fail to be true that there is a number N such that with probability one the inequality $a_n \geq \varphi(n)$ is never true for n larger than N. The inequality will stop being true with probability one, but the point at which this happens may be random.

Still more interesting is Borel's comment on this axiom that Felix Berstein had stated on p. 419 of [4]:

> When one relates the values of an experimentally measured quantity to the scale of all the reals, one can exclude in advance from the latter any set of measure zero, and one should expect only such consequences of the observed events that are maintained when the observed value is changed to one of the remaining ones that lies within the uncertainty of the observation.

Borel wrote [9, pp. 583–584]:

> I have often thought about the same kind of considerations and, like Mr. Bernstein, I am convinced that the theory of measure, and especially of measure zero, is fated to play a major role in the questions of statistical mechanics.

Perhaps in Bernstein's text Borel found a first formulation of what he called much later (in [11]) the *unique law of chance*; for Borel, the significance of probability is related to the events with small probability, which are the only ones for which probability has a practical and objective meaning: these events have to be considered as impossible.

As said above, Lévy considered continuous fractions in his 1929 paper [29]. His general problem was to look for properties that the sequence of coefficients had in common with a sequence of independent random variables. On p. 190, he wrote,

> In an unlimited series of experiments giving probabilities $\alpha_1, \alpha_2, \dots, \alpha_n, \dots$ to an event A, its frequency during the first n experiments differs from the mean probability
>
> $$\alpha'_n = \frac{\alpha_1 + \cdots + \alpha_n}{n}$$
>
> by a quantity almost surely small for n infinite, that is to say that it converges to zero, except in cases of total probability inferior to any given positive quantity.
>
> It must be observed that this property does not suppose the existence of a limit for α_n: it is besides of little importance whether or not the considered probabilities are independent; if they form a succession, every probability α_n being estimated at the moment of the experiment on the basis of the previous experiments, the theorem clearly remains true.

As seen, Lévy expressed himself in a rather loose way, making an assertion rather than providing any proof. Only several years later did he feel the need to provide a complete proof, when writing, from 1934 to 1936, a series of papers devoted to the study of limit theorems for sequences (and series) of dependent variables. In the introduction of his 1936 paper [31, pp. 11–12], he explains how he considered his new work on the strong law of large numbers an extension of the intuition he had had in 1929. The fundamental idea is that for a sequence of random variables $u_1, u_2, \ldots, u_n, \ldots$, many limit theorems obtained when the sequence is composed of independent variables can be adapted to the situation when for each of these variables u_n, the conditional distribution given $u_1, u_2, \ldots, u_{n-1}$ is provided.

> The simplest application of this observation leads one to think that, under slightly restrictive conditions, one obtains a good evaluation of the sum
>
> $$S_n = u_1 + u_2 + \cdots + u_n$$
>
> when each term u_ν is replaced, not by $\mathscr{E}\{u_\nu\}$, but by $\mathscr{E}_{\nu-1}\{u_\nu\}$. One probably will object that the so-obtained approximated value is a random variable, and does not have the practical value of an *a priori* evaluation. But in the calculus of probability, at least in a general theory, one can only specify the probable relation between the probability distribution and the result of the experiment, between the cause and the effect; the obtained assertions lead to more precise conclusions only in particular cases where one specifies how the conditions of each experiment depend on the results of the previous ones. The already mentioned application to the study of continued fractions is sufficient to justify interest in the method.

In the same paper, in a footnote on p. 13, Lévy commented on the loose presentation he provided in 1929.

> If I limited myself to a statement without proof, it was partly not to interrupt a paper devoted to continued fractions by too long a digression, and partly because, being unsure of having read all the published works on the strong law of large numbers, I thought that so simple a result may have been already known; since then I came to the conclusion that it was a new result, and I do not think that its proof had been published before.

As Crépel pointed out, Lévy's explanation is convincing but suggests that Lévy's lack of precision also reveals that in 1929 he had not yet understood that he could formulate a precise condition that would guarantee the validity of the theorem.

Lévy formulated his martingale condition in a paper published in 1935 [30], though not at its beginning. This paper is devoted to the extension of the strong law to the case of dependent variables. In Lévy's mind, such an extension was a continuation of the theory of Markov chains. His main tool for working with general sequences of random variables was to see them as points in the infinite-dimensional cube $[0, 1]^{\mathbb{N}}$ equipped with the "Lebesgue" measure; here we see a direct inheritance of his first probabilistic work on infinite dimensional spaces. He proves a version of a 0-1 law that he states in the following way [30, p. 88]:

> …$P(E)$ and $P_n(E)$ represent respectively the probability of an event E before the determination of the x_ν, and after the determination of $x_1, x_2, \ldots, x_n$ and as a function of these known variables. This event E depends on the indefinite sequence of the x_ν. …

Lemma 1 *If an event E has a probability α, the sequences satisfying this event, except in cases of probability zero, also satisfy the condition* $\lim_{n\to+\infty} P_n(E) = 1$.

In modern terms, one recognizes a particular case of a martingale convergence theorem asserting that if $(\mathscr{F}_n)$ is a filtration such that $\mathscr{F}_n \uparrow \mathscr{F}_\infty$ and z is a random variable, then $E(z/\mathscr{F}_n) \to E(z/\mathscr{F}_\infty)$ a.s. (the theorem is considered here with $z = \mathbb{I}_E$).

Crépel quotes Loève's enthusiastic comment in [36]. For Loève, the previous lemma is the first convergence theorem of martingales and *perhaps one of the most beautiful results of probability theory*. Later on [34, p. 93], Lévy made this comment, writing α_n for $P_n(E)$:

> This theorem has an important particular case. If α_n is independent of n, and so equal to the *a priori* probability $\alpha = \alpha_0$ of the event E, α is equal to zero or one (otherwise $\alpha_n = \alpha$ could not tend towards one of these possible limits). This is Kolmogorov's theorem of zero-one alternative. It is anterior to my 1934 work, but I did not know it when I wrote this paper, which appeared in 1935.

Lévy's comment is confirmed by what he wrote to Fréchet about the same result in January, 1936, when they discussed Kolmogorov's measure-theoretic proof of the 0-1 law in [25]

> [Kolmogorov's] proof is very simple and correct. One must get rid of the impression that it is a conjuring trick. It uses the following essential notion : the probability of the unlimited sequence of the x_ν cannot be considered well defined unless it appears as the limit (in the sense of convergence in probability) of the probability of a property of the set of the first n variables - which implies the studied property with a probability close to one, if it is realized for very large n. The desired consequence is immediate. My own proof, I think, better highlights these ideas. But one can feel them implicitly in Kolomogorov's.

On Kolmogorov's axiomatic version of probabilities, and in particular his proof of the 0-1 law, and the connection with Lévy's vision, see [17,44].

The first appearance of an explicit martingale condition is placed later in the paper under the name *condition* $\mathscr{C}$. It is stated on p. 93 as

$$E_{n-1}(u_n) = 0. \qquad (\mathscr{C})$$

It is unclear what Lévy had in mind with this letter 'C'. Maybe 'centered', maybe 'convergence', maybe simply 'condition'.

As a main use of condition $\mathscr{C}$, Lévy proposes the following theorem which can be seen as an extension of Kolmogorov's theorem for the independent case.

Theorem 1 *If the sequence (u_n) satisfies condition $\mathscr{C}$ and is uniformly bounded by a number U, then $\sum u_n$ and $\sum E_{n-1}(u_n)^2$ have the same nature (convergent or divergent) with probability 1.*

In his review of the paper for the Zentrablatt, the Czech mathematician Bohuslav Hostinský (1884–1951) paraphrased condition $\mathscr{C}$ as *the probable value of* u_n, *evaluated when one knows* $u_1, u_2, \ldots, u_{n-1}$, *is equal to zero*.

What was the genesis of this condition? Unfortunately, the years when Lévy formulated it are precisely those when the major gap in the Lévy-Fréchet correspondence is found, between 1931 and 1936! However, it is possible to formulate some hypotheses connecting Lévy's thinking at the beginning of the 1930s and the aforementioned works after 1934. Two letters from 1931 presented in Chap. 18 of the present volume seem to give us some insight into how Lévy became aware of some fundamental problems related to σ-additivity. Recall his statement to Fréchet that he "had not really seen before 1929 the importance of the new problems implied by the theory of denumerable probabilities." The letters from 1931 show that he discovered Khinchin's and Kolmogorov's results about series of independent random variables with some delay through Fréchet, at a time when the latter (maybe because he was preparing a new course for the Institut Henri Poincaré) was reading Lévy's 1925 book [28] carefully and had difficulties with several of Lévy's assertions about the distribution of the sums of independent random variables. In Chap. III of the book, Lévy systematically used the distribution function F for representing the distribution of a random variable. This led Fréchet to some doubts about the conditions of validity of expressions like $\int_{-\infty}^{+\infty} \lambda(x) dF(x)$ where $\lambda(x)$ is the conditional distribution of some event given the random variable X with distribution function F (see Letters 27 and 27b in Chap. 18 of the present volume). The existence of a regular version of the conditional distribution (in modern terms) was therefore at stake. Lévy tried to justify the possibility of extending the probability to non-measurable sets by an erratic argument immediately rejected by Fréchet. This argument was also contradicted by several works of the Polish school in Lwow in the 1920s (Steinhaus, Banach, Kuratowski, Ulam…), as Lévy himself mentioned in the next Letter 27c. By reading these works, he realized that a step-by-step extension of a finitely additive probability to a σ-additive probability was not possible as contradictions may occur at the limit: Lévy gives the simple example of non-contradictory choices of attributing a probability p_n to each set A_n of a non-increasing sequence of sets with empty intersection with the condition that the non-increasing sequence of reals $(p_n)_{n\geq 1}$ tends to $p > 0$. As for the case of the 0-1 law we commented on above, it is seen here how the systematic measure-theoretic approach promoted by Kolmogorov for probability radically simplified presentation of the "new problems implied by the theory of denumerable probabilities".

It was at this time, moreover, that Lévy was looking for extensions of properties of the sum of random variables to cases more general than independent variables. There is a hint of this at the end of Letter 27b where he mentioned to Fréchet the possibility of proving an upper limit for the point masses of the distribution of the sum of two random variables $u + v$ if there is such a control for the conditional distribution of v when u is given. Lévy probably tried to generalize his study of the random walk of the payoffs in a head-and-tails game (see Letters 26 and 27) to sums of n random non-independent variables $S_n = u_1 + \cdots + u_n$. In particular, he probably looked for a simple condition on the general term u_n, expressed through the conditional distribution of u_n given $u_1, \dots, u_{n-1}$, to guarantee a convergence. Observe that the condition ($\mathscr{C}$) is indeed stated on the general term and was never expressed by Lévy as an intrinsic property of the sequence $(S_n)_{n\geq 1}$. Lévy always took this point of view and never considered a martingale condition formulated as a property of a sequence of random variables.

2.3 Chapter VIII of the Book *Théorie de l'addition des variables aléatoires*

Lévy's most famous book [32] was published in 1937 and was mostly completed during Summer 1936. It played an important role in making known several fundamental tools of modern probability theory, including the Lévy-Khinchin decomposition formula, and is now considered a classic. Lévy himself was probably convinced of the special importance of the results about sums of random variables he had obtained between 1934 and 1936. This could explain why he decided so quickly to collect them in a book. It is not impossible that his meeting with Doeblin (Lévy first met him during Spring 1936) influenced him. It is known that Doeblin made a great impression on the rather scarcely accessible Lévy (on Doeblin's beginnings in probability see [12,37]). And in a letter to Fréchet on 21 December 1936 [2, Letter 30], Lévy mentioned that he prepared a copy of the manuscript for the 21-year-old Doeblin.

Chapter VIII of [32] is entitled *Various questions related to sums of variables in chain*. In a footnote, Lévy presents the chapter as an investigation for 'chained' (dependent) variables of the questions he had studied in previous chapters for independent variables. It collects results obtained by Lévy in previous years about the extension of limit theorems to dependent variables. It remained his vision of what martingales were about, and so a more detailed description will advance our understanding of this vision. This survey will emphasize two main ideas, already mentioned above. First, for Lévy condition $\mathscr{C}$ was merely a technical condition on the general term of a series that allows the extension of the classical limit theorems. He never considered it as a property of the sequence itself. Second, he saw this chapter as a kind of conclusion to his research on series of random variables. This also may explain why he was not really interested when Ville and Doob later launched a full theory of martingales.

2.3.1 Representation of a Sequence of Dependent Variables

Lévy begins Chap. VIII by explaining what is for him the *General problem of chained probability* (Sect. 64, p. 225). In general, 'chained probability' is a term covering any sequence of (dependent) random variables $X_1, X_2, \dots, X_n, \dots$ and Lévy wants to explain how the distribution of the sequence may be constructed. The main tool, he explains, is to obtain a representation of the following kind : $X_n = G_n(Y_1, Y_2, \dots, Y_n)$ where (Y_n) is a sequence of independent random variables with uniform distribution on $[0, 1]$. The Y_n may be defined as $Y_n = F_n(X_1, X_2, \dots, X_n)$

where $F_n(X_1, X_2, \ldots, X_{n-1}, z)$ is the distribution function of the conditional distribution of X_n when $X_1, X_2, \ldots, X_{n-1}$ are given.

2.3.2 Markov Chains

In Sect. 65 (p. 227), Lévy concentrates on the most important case, Markov chains. After having presented the Chapman-Smoluchowski equations describing the evolution of the transition probabilities, Lévy provides interesting considerations justifying the importance of the Markov case. There are, Lévy writes, situations in physics where one cannot know all the parameters defining the state of a system. One must deal with the 'apparent' parameters and neglect the 'hidden' parameters. Two of these situations are particularly important.

In the first particularly important situation, knowledge of the past compensates for the ignorance of the present values of the hidden parameters, and hence allows prediction. This is the theory of *hereditary phenomena* developed by Volterra, for whom the analytical tool is given by integro-differential equations. In the second particularly important situation, only the present value of the (apparent) parameters is known. One then cannot do better than describe the probabilities of the future states (as a simple example, Lévy cites gambling systems). For this situation, the natural analytical tool is Markov chains, for which the Huygens principle is expressed by the Chapman-Smoluchowski equations: for given times $t_0 < t_1 < t_2$, one can equivalently determine the situation at time t_2 by looking at the direct evolution from t_0 to t_2 or by looking first at the evolution from t_0 to t_1 and then from t_1 to t_2. Lévy's connection between Volterra's theory and Markov chains is a direct interpretation of the early story of Markov chains at the end of the 1920s, and in particular of Hostinský's considerations. It is indeed probably from his studies on Volterra's integro-differential equations that Hostinský was led to propose a first continuous-state Markov-chain model in 1928.[6]

Lévy then develops the classical model of card shuffling proposed by Hadamard to describe the mixing of two liquids, subsequently studied by Poincaré, Borel and Hostinský. It has already been mentioned that Lévy had considered this model in his 1925 book [28], but without connecting it to a general situation (see [13] and the letters from November 1928 in [2]). Lévy takes advantage of his new book to develop the proof of convergence towards the uniform distribution of the cards (ergodic principle) which was only sketched in the earlier book. He had already written down the proof earlier on Fréchet's request—see [2, Letters 18 and 19]).

2.3.3 The 'Martingale' Condition: Condition $\mathscr{C}$

After this long introduction about Markov chains, Lévy presents Sect. 66, entitled *extension of Bernoulli theorem and of Chebyshev's method to sums of chained variables*. He begins by looking for conditions under which the variance of the sum S_n of centered random variables is equal to the sum of their variances. It suffices, Lévy

[6] On Hostinský's beginnings in probability, see in particular [23].

writes, that $\mathscr{M}'(X_j)$ equals 0 for each $i < j$ where $\mathscr{M}'(X_j)$ is the probable value of X_j when X_i is known (conditional expectation). This is obviously implied by the more restrictive hypothesis

$$\mathscr{M}_{\nu-1}(X_\nu) = 0, \nu = 1, 2, 3, \dots \qquad (\mathscr{C})$$

where $\mathscr{M}_i$ is *the probable value calculated as a function of* $X_1, X_2, \dots, X_i$ *supposed given*. And Lévy adds : *This hypothesis will play a major role in the sequel*. If X_n does not satisfy $\mathscr{C}$, one can consider the new sequence $Y_n = X_n - \mathscr{M}_{n-1}(X_n)$. In the same way, writing

$$S_n - \mathscr{M}(S_n) = \sum_1^n (\mathscr{M}_\nu(S_n) - \mathscr{M}_{\nu-1}(S_n)),$$

allows to control the approximation of S_n by $\mathscr{M}(S_n)$ with an error of order $\sqrt{n}$ if the *influence of the* ν*th experiment is small on the nth experiment when* $n - \nu$ *is large (for instance when* $\sum_{h=0}^{p} \mathscr{M}_\nu(X_{\nu+h}) - \mathscr{M}_{\nu-1}(X_{\nu+h})$ *is bounded independently of* ν *and* p*)*.

2.3.4 Consequences of Condition $\mathscr{C}$: Central Limit Theorem

Section 67 is devoted to the central limit theorem for sums of dependent variables. The proof is presented as an extension of Lindeberg's method for random variables that are *small with respect to the dispersion of their sum*. In addition to $\mathscr{C}$, Lévy introduces two more hypotheses

$$\mathscr{M}_{\nu-1}(X_\nu^2) = \sigma_\nu^2 = \mathscr{M}(X_\nu^2) \qquad (\mathscr{C}_1) \tag{1}$$

$$| X_\nu | < \varepsilon b_n, \text{ where } b_n^2 = \sum_{i=1}^n \sigma_\nu^2. \qquad (\mathscr{C}') \tag{2}$$

He observes that $\mathscr{C}_1$ implies that the conditional expectation of X_ν^2 is not dependent on $X_1, X_2, \dots, X_{\nu-1}$. Under these hypotheses, Lévy proves that

$$P\left(\frac{S_n}{b_n} < x\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du,$$

along the lines of Lindeberg's proof. In a second part of the section (p. 242), he proposes to weaken condition $\mathscr{C}_1$, replacing it by the requirement that the probability of divergence of $\sum \sigma_\nu^2$ be positive.

Section 68 is devoted to the general problem of convergence of series with non-independent terms. As Lévy stipulates, the *essential hypothesis is that condition* $\mathscr{C}$ *is satisfied* and the second moments of X_ν are finite. Lévy begins by showing that

Kolmogorov's inequality can be extended to that case, which allows him to prove that the series $\sum X_\nu$ and $\mathscr{M}_{\nu-1}(X_\nu^2)$ have the same behaviour. This in particular proves the conditional generalization of the Borel-Cantelli lemma (called by Lévy *the lemma of Mr. Borel*). Sections 69–72 are devoted to the extension of the strong law of large numbers and of the law of the iterated logarithm. These parts are quite technical and we shall not enter into details. Let us only note that Lévy's approach is always the same: extending previous results (generally Khinchin's and Kolmogorov's) under condition $\mathscr{C}$.

3 Lévy Versus Ville

We now consider the complicated relationship between Lévy and Ville.

Ville's name appears surprisingly often in the Lévy–Fréchet correspondence in [2] and in Chap. 18 of the present volume. He is mentioned 17 times, first in 1936 (in a letter following the aforementioned letter of December 1936 where Doeblin is mentioned for the first time) and finally in 1964. These mentions of Ville are almost always associated with criticisms, sometimes even rather derogatory remarks. It is well known that Lévy could be scathing; he never hesitated to show disdain for works he considered uninteresting or without originality. But in his letters to Fréchet he repeatedly expressed particular negativity towards Ville.

It is interesting to have first a closer look at the last letter in which Ville is mentioned [2, Letter 101]. Lévy wrote it on 28 April 1964, at the age of 78, when he had just conquered a long desired seat at the Paris Academy of Science, succeeding the almost centenarian Hadamard.[7] As may be imagined, one of the most urgent tasks of a new Academician is to think about future candidates to replace the next dead Immortal, and Lévy's letter can be explained by the hypothesis that Fréchet had suggested that they consider a possible application from Ville. Lévy wrote,

> I have never quite understood Ville's first definition of collectives; Loève and Khinchin[8] had told me and written to me that they had not understood. It was in 1950, in Berkeley, that I learned from Loève that the processes called martingales are those I had considered as early as 1935; according to your letter, his second definition, p. 99, coincides perfectly with mine.

[7] The tortuous story of Fréchet and Lévy's elections to the Academy can be followed in detail in [2].

[8] We do not know when Khinchin had an occasion to discuss the matter with Lévy. The appearance of Khinchin's name is interesting because beginning in the 1920s he had been one of the first readers and critics of von Mises' collectives, which, despite some regrettable idealistic tendencies, were considered the approach to probability most compatible with the young USSR's dialectical materialism. See [45] for a translation with commentary of the 1929 text by Khinchin on the subject. As late as 1952, in the icy final period of the Stalinist era, Khinchin again came back to this question in the rather controversial and ideological book *Philosophical questions of contemporary physics* with a chapter entitled *The method of arbitrary functions and the battle against idealism in probability theory* [19].

> Naturally, I did not use this word, which I did not know in 1937, in the 1954 re-edition of my 1937 book; in order to permit a photographic reproduction, I had only corrected a few mistakes and added two notes.
>
> But condition $\mathscr{C}$, introduced p. 238, comes down to saying that the sequence of X_ν is a martingale. This condition appears subsequently: theorems 67,1; 67,2; 67,3; 68; n° 69 1° and 2°. In this way I sketched a theory, developed afterwards by Doob, and which generalizes the sequences of independent random variables with expected values equal to zero.
>
> As for the theory of collectives, despite all the merits I attribute to von Mises, I have always found it absurd, and I did not hide this from Wald when he presented it in Geneva. I am grateful to Ville for having helped me fight this theory. But this is not enough to place him at the same level as…say Fortet and Dugué, to speak only of probabilists from the Sorbonne.

From the last sentence, it seems that for Lévy anyone could have been preferable to Ville for election at the Academy. And the way he insists on quoting all the theorems from Chap. VIII of [32] where the condition $\mathscr{C}$ was used is probably a sign of irritation against what may have seem to him Ville's undue claim of having constructed a new mathematical concept. Lévy's assertion that it was only in 1950 that he learned about the theory of martingales is probably true (though he was present in Lyon in 1948 and listened to Doob's lecture—perhaps the language made it difficult for him to understand it. Lévy had never been a great reader and often selected only papers that were connected with his current research. However, as the word had been introduced by Ville in the 1930s, his observation also provides renewed evidence of his disinterest in Ville's contribution. To this we may add the irony in Lévy's going astray with the definition of martingale, calling the sequence X_ν a martingale, not the sequence of the partial sums. We have already observed that Lévy never considered his condition $\mathscr{C}$ more than a technical condition on the general term of a series that allows the extension of limit theorems. The small confusion here is probably related to this fact.

Lévy's first comments on Ville in his correspondence with Fréchet came in 1936. The name was quoted for the first time on 23 December, but most of the previous letter on December 21 is devoted to demeaning comments on a note by Ville presented to the Academy of Science by Borel on 14 December 1936 [46]. The title of the note is *On the convergence of the median of the first n outcomes of an infinite sequence of independent trials*. It was Ville's third note that year (all presented by Borel) but the two others concerned Ville's study of collectives. It is not clear why Ville decided to publish this relatively elementary result. That Borel presented it is not so surprising as Borel's opinion on Ville was very positive; Ville had been a brilliant student at the École Normale Supérieure, and Borel seems never to have been very particular about the notes he transmitted to the Academy. When Ville became closely associated with Borel is also an interesting question. Ville claimed later he had been writing up Borel's lectures on games in October 1937 when Fréchet wanted him to go to Geneva; Doeblin went instead. Perhaps it was during the winter term of 1936–1937 that Borel gave the course and Ville was taking notes. The lectures were published as [10].[9] But one may ask whether Ville asked Fréchet's opinion about his project on note. The results Ville obtained could be seen as a consequence of Glivenko-Cantelli's theorem on the uniform convergence of the empirical distribution functions. This theorem had been stated and published in 1933 in an issue of the Italian journal of actuaries (*Giornale italiano degli attuari*, whose director was Cantelli). This issue of the journal contained three independent papers with the result, by Cantelli, Glivenko and Kolmogorov (who was surprisingly forgotten when naming the theorem). A striking fact is that the three papers [16,22,26] had the same title 'On the empirical determination of a probability distribution'. In 1936, the result was well known among probabilists and statisticians. Fréchet devoted to the theorem two pages of his volume [21] published the same year. Ville knew Fréchet's book: he mentioned it as a reference for (Kolmogorov's) strong law of large numbers at the beginning of [46]. It is very likely that he did not make the connection between his result and Glivenko-Cantelli theorem. Ville had learned probability with Fréchet at the beginning of the 1930s. It is also possible that he failed to realize that there were new topics in [21]. Besides, after two years abroad in Berlin (1933–34) and then in Vienna (1934–35), and, back in Paris, his interests for collectives and game theory in the years 1935–37 had marginalized Ville in the small group dealing with Markov chains around Fréchet at the Institut Henri Poincaré, where Doeblin became the leader. So, it is not obvious that Fréchet paid much attention to what Ville was doing, and his attempt to support Ville possibly resulted from his conscientiousness about doctorate students, and from a kind of tradition of inter-generational solidarity at the École Normale.

From Lévy's letter of 21 December 1936 [2, Letter 30], it appears that Fréchet had tried to justify Ville's submission to the Academy. Lévy's reaction was rather contemptuous.

> Let me come back to yesterday's conversation. Certainly one can sometimes find important and easy theorems that escaped earlier researchers, and to say that a theorem is easy is not to condemn it. But when we are talking about a particular case of a general problem solved for a long time, and not about a difficult particular case that has been studied recently, I frankly think it would be quite ridiculous to look for a particular case of the classical theorem to make a big deal of. (…) Such is the case with the median. (…) The role of the median has been elucidated for a long time; it is an obvious consequence of Borel's and Cantelli's results.

Fréchet immediately answered on December 22, probably trying once again to soften Lévy's opinion. But in a further letter on December 23, extended by a kind of postscript on December 24, Lévy drove his point home. First, he wrote a complete elementary proof of Ville's result (based on the Glivenko-Cantelli theorem, about which he referred Fréchet to his own book [21]). Second, he took the opportunity to

[9] This book is part of the great Borelian project of the interwar period, the *Treaty of probability and its application*, which Borel launched at the beginning of the 1920s and published in successive volumes until 1939. In [15], the authors study the origins and the development of the Borelian project, and how Borel convened much of his network of past students of the École Normale to publish his lecture notes. About Ville, see in particular Sects. 2.2.5 and 3.1.5 in [15].

expound on his vision of mathematics and explain how different it was from Fréchet's vision, in not so agreeable a tone. In the postscript [2, Letter 32], he wrote

> In the case under consideration, I see only two fundamental ideas: uniform convergence, which is well known; and the strong law of large numbers of Borel-Cantelli. Once these two are known, all the theorems of Polya Glivenko Cantelli and Ville do not seem to me to surpass what Darmois proposes to his students as an examination test for the *licence*.[10]

The subject was closed with this letter, but it certainly convinced Ville not to go forward in that direction, and persuaded Lévy, who liked to have a definite opinion on people (think about the difficult relationship he entertained with Bachelier), that Ville was a dull mathematician. Let us observe moreover that Ville was particularly unlucky with the (unexpected) confrontation with Lévy about the median at the precise moment when the latter was brilliantly making use of it to prove convergence results for sums of random variables.

As we have seen, Lévy wrote that he was grateful to Ville for having fought von Mises' collectives. But it is evident that Lévy never absorbed the content of Ville's thesis, even though in the document itself Ville thanks him for having read part of it and given advice [47, p. 2]. When Ville was interviewed by Crépel in 1984 (see Crépel's chapter in the present volume), he said that

> Paul Lévy had not read his thesis. 'I don't read' he told Ville. Aside from his aversion to reading other mathematicians, Lévy was displeased that Ville's thesis had been printed by the *Rendiconti del Circolo Matematico di Palermo*. 'You had your thesis printed by the fascists' he objected. 'I didn't have any money', Ville responded.

Lévy's supposed comment about Italian fascists must not be overinterpreted and, if it is true, it is probably related to the particular situation in 1939 with the outbreak of WWII. It does not seem that in the 1920s Lévy had harbored particularly hostile feelings against Mussolini's regime (and ironically, when he was in semi-clandestinity during the war, he found, with other Jews, a relative security in the Italian occupation zone in France). In any case, Lévy never explicitly mentioned Ville's thesis in his letters to Fréchet.

The question of Ville's contribution to martingales also came up in January 1964, when Lévy was preparing his candidacy to Paris Academy of Sciences, after Hadamard's death in 1963. At that time, martingales had played a central role in recent developments in probability theory, and so Lévy probably wanted to emphasize his priority on the concept, Once again it was an occasion for him to prove his disdain for Ville's work. On 21 January 1964 (Letter 99b in Chap. 18 in the present volume), Lévy wrote to Fréchet :

> I add another word on martingales, which I introduced in a memoir of 1935, I believe, then in my book of 1937, and which Ville baptized. Doob attached enough importance to the concept

[10] This means for their graduation.

> to devote a whole chapter to them in his Stochastic Processes, between that on Markov Processes and that on Additive Processes. He then created the theory of submartingales. There are now also super martingales; I'e seen it in the Sorbonne's course programs, and I don't even know what it is; but I think I can guess.

He came back to the subject on 2 April (Letter 99f in Chap. 18 of the present volume), mentioning that in Doob's treatise on stochastic processes, there is

> a historical appendix, in which the chapter on martingales begins with the following sentence:
>
> "Martingales have been studied by many authors, referred to below. See particularly Lévy (Théorie de l'addition des variables aléatoires, 1937), Ville (Étude critique de la notion de collectif), J. L. Doob (Regularity properties of certain families of chance variables, 1940)

And he added, with obvious pleasure,

> In the 6 pages that follow and give more details, I find the name of Ville only once; I am cited 6 times, and Doob himself 11 times. The other authors cited are Andersen, Jessen, Zygmund, Marcinkiewicz, de Possel.
>
> Even granting that Doob has indulged himself by writing a chapter that owes so much to his personal work (a remark that can also be made for Markov processes), this clearly shows the importance of martingales; and I don't think one can question my priority. My 1937 book was preceded by 2 memoirs from 1935 that contain the ideas taken up again in the book. Ville's thesis (1939) must have been preceded by 1 or 2 notes, probably in 1938, at least after my book. …
>
> I add that Doob's bibliography includes 8 of my books or articles; this number is only exceeded by Doob himself (13) and Kolmogrorov (12).

As mentioned at the beginning of this section, the last of the letters to Fréchet in which Lévy commented on Ville, dated 28 April 1964 [2, Letter 101], included the admission that he had never understood Ville's first definition of collectives. When he wrote that Doob extended *his* theory of martingales, Lévy probably honestly thought that Ville had not substantially strayed from his own picture. However, as we have noticed before, Lévy never considered the martingale property as an intrinsic property of a sequence of random variables. And it is in Ville that Doob found the germ of his future ideas on martingales.

What was Ville's "first definition of collectives"? As Glenn Shafer explains in this volume, Ville began with the centuries-old notion that a martingale is a gambling strategy. In Chap. IV of his thesis, Ville adopted this definition in the context von Mises had considered: successive trials of an event with probability p. In this context, Ville took *martingale* (or *système de jeu*) to mean any strategy for betting for or against the event on the successive trials, always at odds $p : (1 - p)$, that begins with unit capital and risks no more than this. In the spirit of von Mises's and Wald's earlier definitions, Ville's proposed that a sequence of outcomes $X_1, X_2, \ldots$ be called a collective if it does not allow any of a given countable set of martingales to become infinitely rich.

Ville further noted that the martingale is uniquely determined once you know, for each round n and each sequence of possible outcomes $X_1, \ldots, X_n$, the capital $s_n(X_1, \ldots, X_n)$ that the gambler would have after $X_1, \ldots, X_n$ happens. The sequence of such functions (s_n) is thus another way of specifying the martingale. In Chap. V of his thesis, Ville generalized his picture from the case of a repeated event with constant probability to an arbitrary sequence of random variables. In order to do this, he noted that (s_n) qualifies as a martingale if and only if at the beginning of each round n, the expected value of s_n is equal to the current capital s_{n-1}, and he undertook to make this precise using Lévy's definition of conditional probability. Of course, saying that $s_n(X_1, X_2, \ldots, X_n)$ has expected value equal to the current capital $s_{n-1}(X_1, X_2, \ldots, X_{n-1})$ is the same as saying that $s_n(X_1, X_2, \ldots, X_n) - s_{n-1}(X_1, X_2, \ldots, X_{n-1})$ has expected value zero—i.e., that it satisfies Lévy's condition $\mathscr{C}$. So this "second definition" of Ville's, as Lévy called it, does agree with Lévy's own definition.

When Ville generalized his picture to continuous time, in the second part of Chap. V of his thesis, he generalized the condition for being a martingale in the obvious way: functions (s_t) of the underlying continuous process (X_t) form a martingale if whenever $t_0 < t_1$, the expected value of s_{t_1} conditional on the values of s_t for $t \leq t_0$ is s_{t_0}. It was this continuous-time picture that engaged Doob's imagination.

Ville tried to prove the gambler's ruin inequality in the framework of Doob's 1937 paper on stochastic processes with a continuous parameter. He failed, because he tried to use as probability space the outsize set of all functions of time instead of the topologically suitable set of continuous functions, but he gave Doob a fundamental new tool and a new project.

It is remarkable that Lévy kept in touch with Ville during the Occupation, when he lived near Grenoble. Probably, if Lévy had a bad opinion about Ville, the latter had on the contrary a great admiration for Lévy and wished to stay in contact with him. However, he had a second scientific misfortune with Lévy, this time about the recurrence property of Brownian motion. In 1942, Ville published a note in the Comptes-Rendus on the subject [48] and was preparing a related paper when he was informed by Fréchet that Lévy had already published some of his results in his great 1940 memoir to the American Mathematical Society about Brownian motion [33]. Ville decided in 1943 to withdraw his own paper (maybe also because he knew that Lévy could not submit any paper at the time because of Vichy racial laws).

All this did not help Lévy change his opinion on Ville as a poor mathematician, but maybe made him feel some sympathy for the young man. He certainly considered him a serious and capable reader. In the long letter Lévy wrote to Fréchet on 27 September 1943 [2, Letter 50], Lévy mentioned that he would be happy if Fréchet chose Ville to examine his new manuscript about random derivatives. Lévy wrote

> If I gave you the impression that I have little admiration for his works (and actually they never seem very original to me, he is above all a good pupil) I realize that he is very serious, has a great sense of rigor and thoroughly understands the questions he deals with. I will fully trust him to review my manuscript.

In fact, Fréchet chose Loève for the work, maybe to be on the safe side because he was concerned about Lévy's difficult character. And, after the Liberation, Lévy returned to his former disdain. On 12 March 1945 [2, Letter 56], Lévy again explains to Fréchet that Ville's 1936 note on the medians was not original. However, this time, Lévy had made a mistake, probably because he wrongly remembered Ville's note. A week before, he had copied on a sheet of paper a theorem from [32] (theorem 43.2 which says that if S_n is a sequence of random variables converging in probability to S, then any converging sequence of medians of S_n converges to a median of S); Lévy asserted that Ville's result was a direct consequence of this theorem. However, this consequence was only indirect because Ville considered empirical medians, a fact Lévy had been well aware of in 1936. This was probably what Fréchet had replied to him. At the end of the letter, Fréchet had written with a pencil: *Replied on March 5 that it is a different theorem from Ville's*. Nevertheless, Lévy, made the following not-so-kind comment

> I was amazed when I received your letter. I was always confused about the result of Ville's that you mentioned to me in 1936 when it was published.
>
> I am sorry about that, but it does not change much my opinion about the note's lack of originality. The strong law of large numbers (…) had been known for a long time (1917 or even 1909).
>
> Moreover, in my theorem 43.2, it is of little importance whether the distributions be theoretical probability distributions or empirical ones. (…) Taking into account the strong law of large numbers, Ville's result appears therefore as an application of my theorem 43.2.
>
> Obviously, I cannot blame Ville for not having known my book at the time when I was correcting the proofs. But my theorem 43.2 has always been, in my opinion, an obvious observation that I stated explicitly only because I needed it. In the same way, Ville's theorem is for me only an obvious corollary of the strong law of large numbers.

That was still not enough; two years later, the subject came up again and Lévy expressed his exasperation (20 August 1947 [2, Letter 61]). He wrote to Fréchet: *Let me frankly tell you that there are details to which I cannot give as much importance as you do*. He later added,

> I sometimes make the mistake of not making clear results that seem obvious to me but are not obvious to others. I have also missed several priorities that I am not in a position to claim afterwards. In the case at hand, the only thing I told you is that I had known Ville's result for a long time. But, due to the fact that it is an obvious corollary to Glivenko-Cantelli's result, I did not try to take any credit for it, or to call it 'my theorem'.

This letter seemed to have completed the discussion, and (if we consider the set of Lévy's letters to Fréchet reasonably complete up to 1965), it was the last time Ville was quoted in the correspondence except for the 1964 letters, mentioned above, at the moment of Lévy's election to the Academy.

Clearly Lévy had only a superficial knowledge of Ville's works, including his thesis. He remained convinced that his Chap. VIII of [32] was the last word on 'martingales' before Doob. And he thought that even Doob's work was mostly based on his own.

4 Conclusion

Fréchet's persistence in promoting Ville in the letters with Lévy during and after the war may have been motivated in part by Fréchet's hope that Lévy's could support Ville in a search for an appropriate academic position. That support never came. In the event, Ville was passed over for several positions for which he was probably the most qualified candidate, and for ten years, from 1946 to 1956, he had only secondary academic positions along with his industrial work.

After his bad judgment on Ville's 1936 note on the median, [46] Lévy never changed his opinion of him. In particular, he was not interested in the ideas about martingales in Ville's thesis. Lévy later claimed that it was only in the 1950s, when he went to USA, that he learned by chance from Loève that Doob had devised a theory of martingales. Lévy's disinterest was not only due to his bad opinion on Ville. A deeper reason was that he was convinced of having presented in [32], especially Chap. VIII with its condition $\mathscr{C}$, a rather complete version of how these processes could be defined and studied. Lévy never had the idea of considering 'martingales' that did not begin as successive sums of random variables, because his basic interest was to study extensions of the law of large numbers and central limit theorem. He was not seduced by Ville, and he was not really seduced by Doob either, though he later admitted that Doob's methods had proven more powerful than his own. Had Lévy studied with more care and attention what Ville had proposed, maybe some martingale techniques would have arrived sooner in France after WWII and under a different shape. This may be a good subject for an alternate history study.

References

1. Baldi, P., Mazliak, L., Priouret, P.: Martingales and Markov Chains. Chapman & Hall/CRC (2002)
2. Barbut, M., Locker, B., Mazliak, L.: Paul Lévy - Maurice Fréchet, 50 years of Correspondence in 107 letters. Springer (2014)
3. Barbut, M., Mazliak, L.: Commentary on Lévy's lecture notes to the Ecole Polytechnique (1919). Electronic Journal for History of Probability and Statistics **4**(1) (2008)
4. Bernstein, F.: Über eine Anwendung der Mengenlehre auf ein aus der Theorie des säkularen Störungen herrührendes Problem. Math. Ann. **71**, 417–439 (1912)
5. Bernstein, S.: Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. Math. Ann **97**, 1–59 (1927)
6. Borel, É.: Remarques sur certaines questions de probabilités. Bull. SMF **33**, 123–128 (1905)
7. Borel, É.: Les probabilités dénombrables et leurs applications arithmétiques. Rend. Circ. Palermo **27**, 247–271 (1909)
8. Borel, É.: Sur le battage des cartes. CRAS **154**, 23–25 (1912)
9. Borel, É.: Sur un problème de probabilités relatif aux fractions continues. Math. Annalen **72**, 578–584 (1912)
10. Borel, É.: Applications aux jeux de hasard (J. Ville, rédacteur). Gauthier-Villars (1938). Volume VI, fascicle II of *Traité du calcul des probabilités et de ses applications*, par É. Borel
11. Borel, É.: Les probabilités et la vie. Presses Universitaires de France. (1943)
12. Bru, B.: Doeblin's life and work from his correspondence. In: Doeblin and Modern Probability, pp. 1–64. American Mathematical Society (1993)
13. Bru, B.: Souvenirs de Bologne. Jour. Soc. Fr. Stat **144**, 135–226 (2003)

14. Bru, M.F., Bru, B.: Les jeux de l'infini et du hasard. Presses Universitaires de Franche-Comté (2018). Collection "Sciences: concepts et problèmes"
15. Bustamante, M.C., Clery, M., Mazliak, L.: Le Traité du calcul des probabilités et de ses applications. étendue et limites d'un projet borélien de grande envergure (1921-1939). North-Western European Journal of Mathematics **1**, 85–123 (2015)
16. Cantelli, F.P.: Sulla determinazione empirica delle leggi di probabilità. Giornale Ist. Ital. Attuari **4**, 421–424 (1933)
17. Chaumont, L.i., Mazliak, L., Yor, M.: Some aspects of the probabilistic works. In: E. Charpentier, A. Lesne, N.K. Nikolski (eds.) Kolmogorov's heritage in mathematics. Springer, London (2007)
18. Crépel, P.: Quelques matériaux pour l'histoire de la théorie des martingales (1920-1940). Publications mathématiques et informatique de Rennes (1) (1984). Talk:1; available online at Numdam
19. А. Я. Хинчин: метод произвольных функций и борьба иротив идеализма в теории вероятностей. In: Философские вопросы современной физики. Академия наук, Москва(1952)
20. Fischer, H.: A History of the Central Limit Theorem: From Classical to Modern Probability. Springer (2010). Sources and Studies in the history of mathematics and physical sciences
21. Frechet, M.: Recherches théoriques modernes sur le calcul des probabilités, Livre I. Gauthier-Villars (1937). Volume I, fascicle III of *Traité du calcul des probabilités et de ses applications*, par É. Borel
22. Glivenko, V.I.: Sulla determinazione empirica delle leggi di probabilità. Giornale Ist. Ital. Attuari **4**, 92–99 (1933)
23. Havlova, V., Mazliak, L., Šišma, P.: Le début des relations mathématiques franco-tchécoslovaques vu à travers la correspondance Hostinský-Fréchet. Electronic Journal for History of Probability and Statistics **1**(1) (2005)
24. Hawkins, T.: Lebesgue's Theory of Integration: Its Origins and Development. University of Wisconsin Press (1970)
25. Kolmogorov, A.N.: Grundbegriffe der Warscheinlichkeitsrechnung. Springer (1933)
26. Kolmogorov, A.N.: Sulla determinazione empirica delle leggi di probabilità. Giornale Ist. Ital. Attuari **4**, 83–91 (1933)
27. Lévy, P.: Leçons d'analyse fonctionnelle. Gauthier-Villars (1922)
28. Lévy, P.: Calcul des probabilités. Gauthier-Villars (1925)
29. Lévy, P.: Sur les lois de probabilité dont dépendent les quotients complets et incomplets d'une fraction continue. Bull. SMF **57**, 178–194 (1929)
30. Lévy, P.: Propriétés asymptotiques des sommes de variables aléatoires enchaînées. Bull. Sci. Math **59**, 84–96, 109–128 (1935)
31. Lévy, P.: La loi forte des grands nombres pour les variables aléatoires enchaînées. J. Math. Pures et Appl. **15**, 11–24 (1936)
32. Lévy, P.: Théorie de l'addition des variables aléatoires. Gauthier-Villars (1937)
33. Lévy, P.: Le mouvement brownien plan. AMS Journal **62**, 487–550 (1940)
34. Lévy, P.: Quelques aspects de la pensée d'un mathématicien. Blanchard (1970)
35. Locker, B.: Paul Lévy, la période de guerre (2001). Thesis, Université Paris 5
36. Loève, M.: Paul Lévy, 1886–1971. Annals Proba. **1**(1), 1–18 (1973)
37. Mazliak, L.: On the exchanges between Wolfgang Doeblin and Bohuslav Hostinský. Revue Hist. Math. **13**, 155–180 (2008)
38. Mazliak, L.: Poincaré's odds. In: B. Duplantier, V. Rivasseau (eds.) 'Poincaré, 1912–2012', Poincaré Seminar XVI, 24 November 2012. Birkhäuser-Science (2014). Progress in Mathematical Physics
39. Mazliak, L.: The Ghosts of the Ecole Normale. Statistical Science **30**(3), 391–412 (2015)
40. Mazliak, L., Sage, M.: Altered states. Borel and the probabilistic approach to reality. In: P. Cantù, G. Schiemer (eds.) Logic, Epistemology, and Scientific Theories. From Peano to the Vienna Circle. Springer (2021). Vienna Circle Institute Yearbook Series
41. von Plato, J.: Creating Modern Probability. Cambridge University Press (1994)

42. Poincaré, H.: Calcul des Probabilités, 2 edn. Gauthier-Villars (1912)
43. Seneta, E.: Bernstein, Sergei Natanovich. In: S. Kotz, N.L. Johnson (eds.) Encyclopedia of Statistical Sciences, vol. I, pp. 221–223. Wiley, New York (1982)
44. Shafer, G., Vovk, V.: The sources of Kolmogorov's Grundbegriffe. Statist. Sci. **21**(1), 70–98 (2006)
45. Verburgt, L.M.: Khinchin's 1929 paper on von Mises's frequency theory of probability. Statistical Science (2021)
46. Ville, J.: Sur la convergence des médianes des *n* premiers résultats d'une suite infinie d'épreuves indépendantes. CRAS **203**, 1309–1310 (1936)
47. Ville, J.: Étude critique de la notion de collectif (1939). Thesis, Université Paris 5
48. Ville, J.: Sur un problème de géométrie suggéré par l'étude du mouvement brownien. CRAS **215**, 51–52 (1942)
49. Wiman, A.: Über eine Wahrscheinlichkeitsaufgabe bei Kettenbruchentwickelungen. Stockh. Öfv. **57**, 829–841 (1900)

Doob at Lyon: Bringing Martingales Back to France

Bernard Locker and Laurent Mazliak

Abstract

The evolution of Joseph Leo Doob's work on probability is examined from the vantage point of his lecture on applications of the theory of martingales at the colloquium on probability at Lyon in 1948. During the 1940s, Doob had developed the theory of stochastic processes in Kolmogorov's framework. In particular, he had built on Paul Lévy's and Jean Ville's work to develop a theory of martingales. At Lyon, he used his martingale convergence theorem, which he had already proven in 1940, to derive the strong law of large numbers and the almost sure consistency of Bayesian estimation. This article discusses the inception of the Lyon colloquium, Lévy's and Ville's work on martingales, Doob's work in the 1940s, and finally the interactions at the colloquium and Doob's lecture there. This lecture can be seen as bringing martingales back to France, the country of their origin. But the time was not yet ripe for French mathematicians to rise to the challenge they presented.

Bernard Locker (1946–2018), historian of mathematics, taught at Université Paris Descartes. In 2001, he defended a beautiful thesis devoted to Lévy's work during the German Occupation of France.
Laurent Mazliak, Sorbonne Université, LPSM, Paris, France took care of a new version of Bernard's 2008 text (new references, some corrections...).
This chapter was translated from the French by Ronald Sverdlove, Yeshiva University (e-mail: mathmus@gmail.com).

B. Locker
Université Paris Descartes, Paris, France

L. Mazliak (✉)
Sorbonne Université, LPSM, Paris, France
e-mail: laurent.mazliak@upmc.fr

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*,
Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_7

Keywords

History of probability · Martingales · Paul Lévy · Jean Ville · Joseph Doob · Lyon colloquium · Martingale convergence theorem · Strong law of large numbers · Bayesian estimation

1 The Colloquium

From June 28 to July 3, 1948, an international colloquium on probability theory and its applications was held at Lyon by France's National Center for Scientific Research (CNRS) with financial support from the Rockefeller Foundation. In 1946, in the framework of its postwar program for France, the Foundation had decided to give two grants to the CNRS, the first for $250,000 to be used for materials, the second for $200,000 to be used for organizing special conferences, about ten a year for three years, that would include foreign scholars as participants. Warren Weaver, representing the Foundation, expected the conferences to be small and informal, "the attendance of mature contributors restricted to say 15" [78, p. 9]. They were to cover all fields of scientific research, and one third of them were to take place in the provinces. In the end, thirty-six colloquia were held between 1947 and 1952. For mathematics, the colloquia began in 1947 with harmonic analysis at Nancy (15–22 June) and algebraic topology at Paris (26 June–2 July). Probability theory and its applications was among the subjects chosen for 1948.[1]

Maurice Fréchet (1878–1973), effectively in charge of organizing the colloquium at Lyon, wrote the preface to the proceedings, which appeared in 1949 [32]. There he began by justifying the choice of Lyon:

> Lyon was chosen for the site of the colloquium dedicated to this subject because the University of Lyon has steadfastly supported the initiatives of Professor Eyraud in this area. It liberally welcomes articles on the 'probability calculus' in the mathematical section of the 'Works of the University of Lyon'; it has created and continues to support a very useful Institute of Financial Sciences and Insurance (ISFA); finally it has created a certificate of higher education in economics, thus taking the lead of the movement that, despite resistance, moves political economy toward the use of mathematics.

This was not merely politeness toward Henri Eyraud (1892–1994). According to Michel Armatte [3], ISFA owed its success to Eyraud's course, and to the support of Fréchet and Émile Borel (1871–1956) since its creation in 1930.

There was also there a nod to Georges Darmois (1888–1960), who had been director of studies at the Institute of Statistics of the University of Paris since 1945. Though

[1] The CNRS (Centre national de la recherche scientifique) is a research agency of the French government. Zallen recounts the history of these grants, with interesting details on the roles of Weaver and Louis Rapkine, as well as Frédéric Joliot, Georges Teissier and Pierre Auger [78]. Dosso particularly emphasizes Rapkine's role [28]. See also the archives of the CNRS: International Colloquia Supported by the Rockefeller Foundation or the CNRS 1946–1967 [ART 141–173].

ten years younger than Fréchet, Darmois was already an old hand at statistics and its instruction. His activism alongside Borel and Fréchet's is described by Meusnier [51] and Catellier and Mazliak [10]. We add that Fréchet and Darmois had participated in the meeting of the Commission on Econometrics on 16 December 1946, which created two seminars in econometrics, one being located in the provinces. Lyon was chosen over three other provincial cities, Rennes, Strasbourg, and Lille [8]. The choice of Lyon fit perfectly into Fréchet's persistent activism, sustained since his course at Strasbourg in 1919 by a vision of the interconnections among probability, statistics, social sciences, research, and education [2,4,34,37,48,66].

Counting Fréchet, Darmois, and Eyraud, sixteen participants produced papers for the proceedings. As Fréchet explained,

> In addition to the sixteen scholars slated to present communications…, more than thirty people (from all parts of France and the world, most specialists in the questions treated) asked to listen to the lectures and participated very helpfully in their discussion.

Fréchet thus respected perfectly Weaver's wishes to keep the colloquia small. This was not always the case in the other colloquia, particularly in physics. Not initially on the list of speakers, Fréchet had prepared a lecture [33] in case one of the invited lecturers was absent. Calyampudi Radhakrishna Rao (born in 1920), who did not present a lecture, participated in all the meetings and intervened several times.

The proceedings appeared in 1949 as *Le calcul des probabilités et ses applications, Lyon—28 juin au 3 juillet 1948*, Colloques Internationaux du Centre National de la Recherche Scientifique, XIII, Paris. Here we list the titles, translated into English when necessary, of the sixteen contributions. Five of the contributors, including the American Joseph Leo Doob (1910–2004), were foreigners:

1. G. Ottaviani (Italy): *The uniform law of large numbers in the spirit of the classical theory of probabilities*.
2. J. L. Doob (USA): *Application of the theory of martingales*.
3. D. van Dantzig (Holland): *On the method of generating functions*.
4. H. Wold (Sweden): *On stationary point processes*.
5. J. Wishart (UK): *Test of homogeneity of regression coefficients*.

Ten were French:

1. G. Darmois: *On certain forms of probabilistic dependence*.
2. M. Fréchet: *Typical values of order zero or infinity of a random number and their generalization*.
3. P. Lévy: *Double Markov processes*.
4. A. Blanc Lapierre: *Considerations on the harmonic analysis of random functions*.
5. J. Kampé de Fériet: *Stationary random functions and transformation groups in an abstract space*.

6. E. Halphen: *On the problem of estimation*.
7. P. Delaporte: *On the systematic use of mathematical statistics in factor analysis*.
8. R. Fortet: *Probability of loss of a telephone call*.
9. J. Ville: *Random functions and transmission of information*.
10. G. Malécot: *Stochastic processes and genetics*.
11. H. Eyraud: *Pure economics. Credit and speculation*.

For the history of martingales, we will remember that Paul Lévy (1886–1971) and Jean Ville (1910–1989), along with Doob, were among the contributors. Martingales, of which Lévy and Ville were the pioneers in prewar France, thus returned from America after the war as a theory elaborated by Doob in a framework for probability and stochastic processes inspired by the 1933 *Grundbegriffe* of Andrei Kolmogorov (1903–1987), a framework that Doob had continually developed and applied since 1935. In his contribution to the colloquium, entitled "Application of the theory of martingales", Doob would show how this theory of martingales could be applied to the strong law of large numbers and to statistical estimation.

Yet the history of martingales must record that Lévy did not recognize his own prewar techniques in Doob's martingales at Lyon, and that Ville did not meet Doob there. Although Fréchet's preface leaves the impression that all the contributors to the proceedings had been present at Lyon, Ville was not. He had been unable to make the journey, he told Fréchet a few days later, because of his duties examining candidates to the École Polytechnique. The other contributors sent him a card regretting his absence. It appears that Ville never met Doob. So Lyon symbolizes a missed opportunity. Of the three pioneers of martingales who contributed to the colloquium, two were left at the side of the road, while the third made martingales one of the masterpieces of his work.[2]

[2] We learn about the card and Ville's excuse for his absence in a letter he wrote to Fréchet, dated 9 July 1948 and preserved in the Fonds Fréchet in the archives of the Academy of Sciences. In a letter Ville wrote to Pierre Crépel in 1985, reproduced in translation in Crépel's Chap. 6 in the present volume, Ville erroneously claimed that he had not been invited to the colloquium and that he had not met Doob for this reason.

Colloque International sur le Calcul des Probabilités, Lyon 1948.

First row left: Paul Lévy and Maurice Fréchet. Doob is directly behind Fréchet. On the picture one also can find among others R.Fortet, D. Van Dantzig, E.Mourier, J.Kampé de Fériet, A.Blanc-Lapierre…

(Photo: ©Private collection. Courtesy of F.Lederer)

2 Paul Lévy

From the time of his thesis of 1911 until "turning to probability" in 1919, Lévy was an analyst, working mainly on functional analysis and on the calculus of variations in infinite dimensions [4]. For the period that interests us, which goes until 1937, let us say that Lévy knew Fréchet's integral in "abstract spaces" [31], that he had become familiar with Daniell's integral [13] with some delay (he says so in [43]), and that his means on the L^2 sphere had brought him very close to Wiener's measure. In 1923 Norbert Wiener (1894–1964) cited him along with Gateaux and Daniell in his fundamental article on Brownian motion [76], and in 1924 Lévy helped Wiener edit a French article that translated Wiener's "differential space" into the language of Borel's "denumerable probabilities" [77].[3] In 1924, Lévy created his own approach to measure with his "theory of partitions", which he then extended to "abstract

[3] On this, consult [50]

spaces" and presented in Chap. II of [43] as possibly a foundation for a general procedure for effectively constructing all probability laws on sets having the power of the continuum. Satisfied with this theory, he never used it in practice. He felt the need to protect himself from the

> prejudices of certain analysts, and not the least of them, with respect to the theory of probability, or at least with respect to probabilists, supposed not to have the sense of rigor.

In his preface to the first edition of his book on the addition of random variables [43], he responded in advance to these analysts by recalling his contributions since 1919 to "translating well known theorems of analysis into the language of probability" and marveling

> …that one could think that an argument, to be rigorous, needed to be translated from one language to another… …this sounds to me like saying that my French text had to be translated into German in order for my arguments to appear rigorous.

He added that his translation was "within reach of a beginner".

If Lévy did not isolate the concept of a martingale, it was his work on sums of dependent variables (variables in chains or *variables enchaînées*, as they were called in French at the time) that links him to the history of martingales before Ville and Doob. On the place of Lévy in the history of martingales, we refer to two chapters in the present book: Laurent Mazliak's chapter on Ville and Lévy and Salah Eid's chapter concerning the correspondence between Lévy and Jessen on the rarely discussed relation between Lévy's Lemma and Jessen's Theorem.[4] Here we recall only that from 1935 to 1936, Lévy extended the strong law of large numbers, Kolmogorov's three-series theorem, and certain results on convergence to the normal distribution beyond the case of independent variables to dependent variables satisfying his "condition $\mathscr{C}$" and auxiliary conditions. In the notation Lévy used in [41,43], condition $\mathscr{C}$ says that

$$\mathscr{M}_{\nu-1}\{X_\nu\} = 0, \tag{1}$$

where $\mathscr{M}_{\nu-1}\{X_\nu\}$ is "the probable value of X_ν, given the values of $X_0, \ldots, X_{\nu-1}$", that is, the conditional expectation relative to these variables.

In 1935, 1936, and 1937, Lévy studied cumulative sums

$$S_n = \sum_{\nu=1}^{n} X_\nu, \tag{2}$$

where the dependent variables X_ν were subject to his condition $\mathscr{C}$. These sums clearly form a martingale with respect to the filtration associated with both the X_ν

[4] Given an integrable random variable X and a filtration (F_n), the martingale $E(X|F_n)$ converges to X almost surely in L^1. Lévy proved this for the case where X is the indicator of a set; this is often called "Levy's Lemma" or "Lévy's zero-one law".

and the S_ν. In Lévy's notation:

$$\mathscr{M}_{n-1}\{S_n\} = S_{n-1}. \tag{3}$$

For Lévy in 1937 [43], the conditional probabilities permitting the definition of conditional expectations such as those in (1) and (3) are understood without reference to what is now called the Radon–Nikodym theorem, even though Radon published it in French in 1930 in its most general and classical form, nor even any allusion to Kolmogorov's presentation "à la Radon–Nikodym" in his 1933 *Grundbegriffe* [39]. Lévy was far from such a presentation. We know that he read little and preferred to rediscover for himself results he heard discussed, as the need arose and using his own methods. But it is appropriate to add that with Lévy, when the probabilist took precedence over the analyst, the results of measure theory were arranged in a strange catalog, dedicated only to his needs in probability, translated once and for all into his language of probability.[5]

The French probabilists took little or no notice of Kolmogorov's *Grundbegriffe* when it appeared. This changed after Fréchet's journey to the Soviet Union in late 1935, when he visited Kolmogorov and other mathematicians in Moscow, as well as his daughter and her husband Edgar Lederer in Leningrad where Lederer worked as a chemist. In a letter from Lévy to Fréchet dated 29 January 1936 [4, Letter 29], we see the two discussing Kolmogorov's very abstract zero-one law, with its equally abstract proof based on the monotone class theorem. Lévy contrasts this abstraction with the intuitive statement and geometric proof of his own zero-one law, which says that an event's conditional probability converges almost surely to one or zero depending on whether the event happens or not. Lévy's deeply geometric and visual mind always favored the effective construction of manipulated objects, and this attitude succeeded spectacularly in his later work on Brownian motion and related processes, which he saw as being built at every moment, pathwise, rather than emerging as a whole from an existence theorem.[6]

At the core of the contrast between the Kolmogorov's and Lévy's approaches was the contrast between their definitions of conditional probability, Kolmogorov's ruthlessly abstract definition versus Lévy's geometric definition relying on the representation of a random variable as a transform of the uniform distribution on $[0, 1]$, so that the conditional distribution of Y given X is reduced to drawing a point from the unit square $[0, 1] \times [0, 1]$. Without mentioning Kolmogorov but surely stimulated by his encounter with the *Grundbegriffe*, Lévy spells out his viewpoint on conditional probability in a brief article that appeared in March 1936 [42]. He insists it was

[5] Concerning Lévy's psychology and method of working, see [4] and especially [46]. In the 1930s, Lévy still used the somewhat archaic term *valeur probable* instead of the more popular *espérance* for *expectation*. These two terms appear in proximity on many pages of [43]. See [6,52] for more on Radon [61] and Nikodym [57], which repeats the communication presented by Otton Nikodym in September 1929 to the *First Congress of Mathematicians of the Slavic Countries*.

[6] Concerning Kolmogorov's and Lévy's zero-one laws and their proofs, see [11, pp. 44–46]. See also Chap. 6 in the present volume.

the viewpoint he has always taken in his research on dependent variables and will continue to take. It enables him to show, for an event A of probability α, that if its probability becomes $\lambda(x)$ when one knows $X = x$, then $\alpha = \mathscr{E}(\lambda(x))$ (Here he uses $\mathscr{E}$ instead of $\mathscr{M}$ for expectation). The measurability requirements for conditional probability are specified, and the whole is presented in terms of Lebesgue's integral and measure.

As he often did, he dresses his argument up with an appeal to authority, preferably to one of the great Russian probabilists [48]; in this case Sergei Bernstein is invoked. He enlarges on his viewpoint the following year in his book on the addition of random variables [43, §§22–23]. The co-existence between determination in a single experiment and successive determination was very dear to Lévy and permeated all his considerations having to do with conditional probabilities and/or dependent variables in [43]. It is to be put in parallel with his vision of stochastic processes, in which chance intervenes at each instant t after having constructed the trajectory before t [4, pp. 22–48]. See also [41], strangely misplaced in the last volume (volume VI) of his complete works.

In Lévy's 1937 book [43, §23], it is the Lebesgue–Stieltjes integral that comes to the forefront. Given an event B, perhaps of the form $\{Y < y\}$, and a random variable X with distribution function F, the "conditional probability of B on the hypothesis $X = x$" must then be a function $g(x)$ admitting a Lebesgue–Stieltjes integral with respect to F satisfying

$$\int_{-\infty}^{x-0} g(x)dF(x) = \text{Pr.}\{B \text{ and } X < x\}. \tag{4}$$

The existence and the properties of g are obtained, up to a set of F-probability zero, via a method of differentiating the function obtained from (4) by the change of variables $\xi = F(x)$ [43, §23]. This reflects Lévy's preference for distribution functions and his earlier use of the decomposition of functions of bounded variation, where he adapted Lebesgue's theorem on the almost everywhere differentiability of nondecreasing functions and the Lebesgue–Stieltjes integral to decompose a distribution function F into three parts, the sum F_1 of jumps, the absolutely continuous part F_2, and the difference $(F - F_1 - F_2)$ [43, §12]. After deriving (4) when the joint distribution of Y and X is known, he treats the case of determination by "two successive trials", giving measurability conditions for reconstituting from a function $g(x)$ a "well determined probability" for a pair Y, X of random variables such that (4) is valid and g is thus a conditional probability.

Thus supplied with the conditional probability $\text{Pr.}\{Y < y|X = x\}$, Lévy designates by $\mathscr{M}_X$ "a probable value calculated when X is known" and adds that "it is therefore a function of X". Everywhere in his reasoning the measurability of $\mathscr{M}_X\{Y\}$ is taken for granted. The conditional probabilities $\text{Pr.}\{X_\nu|X_0, \ldots, X_{\nu-1}\}$ and the corresponding conditional expectations $\mathscr{M}_{\nu-1}\{X_\nu|X_0, \ldots, X_{\nu-1}\}$ are obtained by iteration [43, §64] but also by recourse to denumerable probabilities and the representation of all the random variables involved as measurable functions of a single variable. All the properties of conditional probabilities and expectations that Lévy will use in [43] he assumes to flow from his "notion of conditional probability".

He would later remain ignorant of the counterexamples given by Jean Dieudonné in 1948 [15], although Fréchet was made aware of them by a letter of November 1951 from Robert Fortet, recounting that Dieudonné had showed that conditional probabilities do not always exist [48].

Lévy always remained faithful to his notion of conditional probability, never bothering in the rest of his work (even for the difficult questions he attacked, in the theory of Markov processes in continuous time for example) to check his bearings, as if it were obvious that the properties he had derived were universally and eternally valid. He returned to the point a letter to Fréchet on 9 January 1962 [4, Letter 90], insisting on the consistency of his "notion" of 1936–1937 with Dobb's viewpoint and its divergence from de Finetti's viewpoint.

With respect to the foundations of probability, Lévy sat uncomfortably between Richard von Mises's collectives and Bruno de Finetti's subjectivism. Von Mises (1883–1953) had explained his collectives in lectures at the Institut Henri Poincaré in November 1931 [55]. To demonstrate their superiority, he criticized conceptions about random variables that he attributed to "Mr. Fréchet and some others", who "do not give an exact definition of this new notion [and] claim it is known a priori" in order to draw from "ideas established by Mr. Borel …a sort of mathematical probability whose object does not belong to the real world". De Finetti (1906–1985), on the other side of the philosophical and logico-mathematical chessboard of the foundations of probability, took his turn at the Institut Henri Poincaré in five lectures in May 1935 [14], developing a viewpoint that he himself called "the most extreme solution on the side of subjectivism", rejecting as "illusory" the idea that "the impossibility of making the relation between probabilities and frequencies precise is analogous to the practical impossibility encountered in the experimental sciences of connecting a theory's abstract notions exactly with empirical realities". This idea he attributed to "modern" treatises: [9,34,40,53]. Lévy's "moderate subjectivist" position, which accommodated itself to a realistic interpretation of frequencies, was squeezed on both sides, and this may explain his self-contradictions concerning von Mises [54] and Ville [75], perhaps even his negative opinion of Ville. In 1939, he wrote to Fréchet that collectives should not be rejected completely, but in another letter to Fréchet 25 years later he claimed that he had always found them absurd [4, Letters 40,101].

3 Jean Ville

It was in his thesis, *Étude critique de la notion de collectif*, quickly published as a book by Borel [75], that Ville introduced the term *martingale*. In the debates on foundations, Ville's martingales played a significant role in the opposition to von Mises's "frequentist" theory of collectives. We point out the very useful commentaries on Ville's book in [58] and [60, Chap. 6]. An analysis of the book and a note on Ville and the Lyon Colloquium are included in [7]. On all that concerns Ville and his book, the reader should refer first to Glenn Shafer's work [65] and his Chap. 5 in the present volume. Here we recall, with Bernard Bru, Marie-France Bru, and Kai Lai Chung [7, p. 20] that Ville

> begins by defining the general notion of a (positive) martingale adapted to an arbitrary sequence (X_n) of random variables using the now classical martingale property (e.g. [56]), the conditional expectation being "defined" in the sense "indicated by Mr. Paul Lévy".

In their well-argued article, these three authors find in Ville's text a "type of reasoning 'by stopping' …used by all sound authors" when he proves his martingale inequality in discrete time. We will see this inequality again below, because it did not escape Doob, who generalized it in [22].[7]

We add here some details on the framework in which Ville defined his conditional expectations. First, section I of his Chap. V, which concerns martingales in discrete time, relies exclusively on Lévy's 1937 definition [43, pp. 96–99], to the point of reproducing identically Lévy's "proof" of the existence of conditional probability that we just described. Then in Sect. II (pp. 111–129), Ville discusses the generalization of his inequality to (positive) martingales in continuous time, and for this he works in the space, which he calls E_0, of all real functions of a real variable t representing time. Citing Kolmogorov [39], he writes that

> a distribution function will then be a completely additive function ≥ 0 defined on certain subsets of E_0 …. Letting $P(L)$ be this function, we naturally assume that $P(E_0) = 1$.

Among the conditions that he required a positive functional $\{S_\tau\}_{\tau\geq 0}$ to satisfy to qualify as a martingale was his Condition (d): "the mean value of $S_{\tau+\tau'}(X)$ when one knows that $X(t) = X_0(t)$ for $t \leq \tau$ is equal to $S_\tau(X_0)$". On p. 112, he cites Doob [19, p. 123], writing.

> Condition (d) brings in the notion of conditional mean, which we base on the notion of conditional probability as it was defined by Mr. Doob …; this notion is the generalization to a function space of the notion of conditional probability (due to Mr. P. Lévy) that we already used (p. 87).

He adds that the existence of the conditional probability that he defines using Fréchet's integral "results from a proof of Mr. Nikodym (p. 168–179)". So his references are now Fréchet [31], Nikodym [57], Kolmogorov [39], and finally Doob [20]. This presentation of martingales in continuous time did not escape Doob in his review of Ville's book [21].

Ville had been mobilized for the war by the time he defended his thesis in 1939, and he was among the officers captured by the Germans in the spring of 1940. After his return to France from a prisoner-of-war camp in the spring of 1941, he considered the application of martingales to the geometry of vector Brownian motion but soon found that Lévy had left him behind on this topic. In 1946, after the faculty at Lyon chose Gustave Malécot over him for a chair, he left the university. Remaining on very cool terms with the academics at Lyon, he moved on to other topics, including

[7] Ville's inequality, proven on pp. 100–101 of his book, says that if (S_n) is a nonnegative martingale and $\lambda > 0$, then $\Pr(\sup_n s_n \geq \lambda) \leq 1/\lambda$.

the transmission of information [71] and existence conditions for a total utility and a price index [35, 70]. His written contribution to the Lyon colloquium was entitled "Random Functions and the Transmission of Information" [72]. A list of Ville's teaching engagements and some of his consulting work up to 1956 is provided in [73], and a comprehensive list of his publications is provided in [64].

4 Joseph Doob

It is at the top of the first page of [22], published in 1940, that Doob announces that he will study "certain families of chance variables" x_t having the property he denotes $\mathscr{E}$—i.e., verifying for all $t_1 < \cdots < t_{n+1}$ and with probability 1 the relation (written here in his own notation)

$$E[x_{t_1}, \ldots, x_{t_n}; x_{t_{n+1}}] = x_{t_n}. \tag{5}$$

This equation is followed immediately by a footnote:

> We shall use the notation $E[y]$ for the expectation of the chance variable y, and $E[y_1, \ldots, y_n; y]$ for the conditional expectation of y for given $y_1, \ldots, y_n$, a function of $y_1, \ldots, y_n$. If the y_i are not finite in number, the notation will be modified accordingly. We shall assume the definitions of Kolmogoroff …for these conditional expectations.

As in [19, 20], the invocation of Kolmogorov's name refers to the conditional expectations Kolmogorov constructed using Nikodym's theorem in his 1933 *Grundbegriffe* [39]. In 1940, Doob still used the term *chance variable*, which he later replaced with *random variable*. The expectation of X given A is denoted $E[A; X]$ here and $\mathrm{E}\{A\backslash X\}$ in [24]. This may confuse today's readers, because it is now standard to put X first, as in $E[X|A]$.

Still on this first page, Doob makes precise the sense he wants to give to his random variables and to the probability measure P, and that he will use throughout:

> In the following, we shall always suppose that the x_t are measurable functions defined on a space Ω, on certain sets of which a measure function is defined. That this can always be done, and how this is to be done, was shown by Kolmogoroff …. The space Ω, following Kolmogoroff, will be taken to be the space of real-valued functions of t. … The qualification "with probability 1" will be used interchangeably with "almost everywhere on Ω."

The conditioning in (5) and the property $\mathscr{E}$ are immediately generalized on the next page to an infinite set of indices T (which can be $\mathbb{N}$, $-\mathbb{N}$, a section of $\mathbb{Z}$ or of $\mathbb{R}$) by considering the "Borel field" of "x_t-sets" generated by all finite sets of the variables considered. The "chance variables with property $\mathscr{E}$" of this paper will be called "martingales" in Doob's talk at Lyon [24].[8]

[8] "Borel field" translates the French *corps de Borel*, then the standard name for the Borelian ancestor of "σ-algebras". In 1937, in the first edition of [43], Lévy manipulated and constructed *corps de*

In all the preceding cases, Doob's martingale, in modern terms, is adapted to the family of σ-algebras $(F_t, t \in T)$ generated by the random variables with indices smaller than t, and this family is increasing in t. When the set of indices is a subset of $\mathbb{Z}$ bounded above, an obvious translation of the index leads to the set of nonnegative integers, and the results Doob obtains when the index tends to $-\infty$ are concerned rather with what are now widely called "backward martingales".[9]

So in this article, consisting of three sections, Doob is going to rely—clearly, permanently, and explicitly—on measure theory, as he had never stopped doing since his first article on probability in 1934 [16]. This first article had been followed in the same vein by Doob's [17–20], and also by his 1940 article with Ambrose [27], which slightly preceded the article we are discussing [22].[10] All this together formed the framework that would permit him, after having established in the first section the now classical theorems on discrete-time martingales (relative to uniform integrability, convergence and closure), to put in place in the last two sections, the tools and the first results for martingales in continuous time. Given our focus on the state of affairs before the Lyon colloquium, we will remain with the first section, in which Ville and Lévy are cited or commented on four times.

For chance variables $\ldots, x_{-1}, x_0, \ldots, x$, Doob proves the inequalities

$$\int_{\Lambda \cdot N} x dP \geq kP\{\Lambda \cdot N\} \tag{6}$$

and

$$\int_{M \cdot N} x dP \leq kP\{M \cdot N\} \tag{7}$$

where $\Lambda = \left\{\sup_{m \leq j \leq n} x_j \geq k\right\}$, $M = \left\{\inf_{m \leq j \leq n} x_j \leq k\right\}$, N is in the σ-algebra generated by the finite dimensional random vectors $(x_{p_1}, \ldots, x_{p_n})$ where $p_1, \ldots, p_n$ are distinct integers no greater than m, and the dots represent intersection. Lévy and Ville are cited in a footnote [22, p. 458]:

> These inequalities …are implicit in the work of Ville [75, pp. 100–101], who discussed sequences of non-negative chance variables with the property $\mathscr{E}$. The method of proof we use was used by Lévy [43, p. 129], in a related discussion.

The inequalities (6) and (7) are important because Doob can deduce from them, after some gymnastics, a theorem numbered 1.2 (p. 458) for sequences satisfying $\mathscr{E}$ and

---

Borel in "abstract sets" (terminology borrowed from Fréchet). See p. 17 of the 1954 edition of [43] for his later distinction between "Borel fields" and "closed Borel fields".

[9] Recall that a backward martingale indexed by the nonnegative integers is basically a sequence $(Y_n; n \geq 0)$ of random variables adapted to a decreasing sequence $(G_n; n \geq 0)$ of σ-algebras satisfying $E[Y_n|G_{n+1}] = Y_{n+1}$. When we set $X_{-n} = Y_n$ and $F_{-n} = G_n$, this becomes $E[X_{-n}|F_{-n-1}] = X_{-n-1}$ and the sequence of σ-algebras $(F_{-n}; -n \leq 0)$ is increasing $(F_{-n-1} \subseteq F_{-n})$, giving a martingale that is ordinary but indexed by the nonpositive integers.

[10] The article with Ambrose [27] was received by the *Annals of Mathematics* on 25 September 1939; [22] was received by the *Transactions of the American Mathematical Society* on 11 December 1939.

indexed by the nonpositive integers (a "backward martingale" theorem as mentioned earlier):

> Let $\ldots, x_{-1}, x_0$ be a sequence of chance variables with the property $\mathscr{E}$. Then $\lim_{n\to-\infty} x_n = x$ exists with probability 1, and the chance variables $x, \ldots, x_{-1}, x_0$ have the property $\mathscr{E}$. The chance variables $\{x_j\}$ are uniformly integrable, and $\mathrm{E}|x_0| \geq E|x_{-1}| \geq \cdots \geq E|x|$; $E|x_n| \to E|x|$.

Ville had first used a form analogous to (6) in the case of a martingale indexed by the natural numbers and with $k = 1$. He had extended it to the continuous case as early as 1938 in a note in the *Comptes rendus* [7,74]. In 1937 [43], Lévy also uses (6) for his martingale (3), aiming especially to extend Kolmogorov's inequality from the case of independent variables to his martingale differences satisfying his formula (1)'s condition $\mathscr{C}$. In Lévy's notation, under condition $\mathscr{C}$ for the X_ν with $T_n = \max_{\nu\leq n}|S_\nu|$, $c > 0$, and $b_n^2 = \mathscr{M}\left\{S_n^2\right\}$, one has $\text{Prob}\,\{T_n > cb\} < 1/c^2$.

We next note Doob's theorem on almost sure convergence of integrable martingales and closure in the equi-integrable case that today is called "regular" (Theorem 1.3, p. 460):

> Let $x_1, x_2, \ldots$ be a sequence of chance variables with the property $\mathscr{E}$. Then $E|x_1| \leq E|x_2| \leq \cdots$. If $\lim_{n\to\infty} E|x_n| = l < \infty$, then $\lim_{n\to\infty} x_n = x$ exists, with probability 1, and $E|x| \leq l$. If the x_j are uniformly integrable, $\lim_{n\to\infty} x_n = x$ exists, with probability 1, and the chance variables $x_1, x_2, \ldots, x$ have the property $\mathscr{E}$.

Doob makes this comment:

> Ville has studied sequences of non-negative chance variables with the property $\mathscr{E}$. Since, by the corollary to Theorem 0.2, Ville's hypotheses imply that
>
> $$Ex_1 = Ex_2 = \cdots = E|x_1| = E|x_2| = \cdots,$$
>
> the hypotheses of the first part of Theorem 1.3 are satisfied, in Ville's case. Ville proved that in his case $\sup_{j\geq 1}|x_j| < \infty$, with probability 1 (implied by our conclusion that $\lim_{n\to\infty} x_n$ exists with probability 1, and that the limit is integrable) and applied this fact to the study of certain games of chance.

(Doob's corollary to his Theorem 0.2 says that when the x_t have property $\mathscr{E}$, Ex_t is constant and $E|x_t|$ is nondecreasing in t.) As Doob's words suggest, Ville did not establish the almost sure convergence of "Ville's martingales" in [75]. This result is due entirely to Doob. We also note that in this first part (p. 462), Doob cites two more results on almost sure convergence in Lévy's [43]: Lévy's zero-one law and the "martingale version" of Kolmogorov's three-series theorem.

5 At the Colloquium

The hypotheses advanced by Bru, Bru, and Chung [7] to explain Ville's retreat from the martingale scene and his absence from the Lyon colloquium are clear. In particular, he may have lost interest in martingales after reading Doob's 1940 treatment [22], which seemed to close the question; see Laurent Mazliak's Chap. 6 in the present volume and [4, Letter 44]. But we can add another hypothesis. In November 1939 [21], Doob had published a review of Ville's 1939 book, of which the whole second half was very negative:

> It is unfortunate that this book, which contains much material which clarifies the subject, should contain so much careless writing. This ranges from uniformly incorrect page references to mathematical errors. Thus (p. 46) it is claimed (and used in a proof) that every denumerable set is a G_δ. The author's main theorem on systems is not as strong as earlier results with which he is apparently unfamiliar. (Cf. Z. W. Birnbaum, J. Schreier, Studia Mathematica, vol. 4 (1933), pp. 85–89; J. L. Doob, Annals of Mathematics, (2), vol. 37 (1936), pp. 363–367.) His discussion of random functions is inadequate and obscure, for example, his demonstration that his main theorem on martingales does not go over to the continuous process uses as an example a measure on function space not in accordance with the usual definition of probability measures on this space. A specialist who can overlook such slips will find many stimulating ideas in this book. Other readers can profit by the comparative analysis of the different criteria for collectives, and by the discussion of martingales.

Ville, after that, would have no more wanted to meet Doob than the academics of Lyon.

The error concerning G_δs does occur on the page of Ville's book that Doob flags. Concerning the alleged stronger results by Birnbaum and Schreier and by Doob himself, see Glenn Shafer's Chap. 5 in the present volume. In his celebrated conversation with J. Laurie Snell in 1997 [67], Doob speaks of this review and of the direction in which reading Ville pointed him, but he makes no allusion to this negative second part. Henry Thomas Herbert Piaggio (1884–1967, assistant, then professor of mathematics at the University of Nottingham from 1908 to 1950) also reviewed Ville's book [59] but did not mention the word "martingale".

In his treatment of continuous-time martingales in 1940 [22], Doob simply points out in a footnote on (p. 476) that Ville's "discussion of the meaning of a continuous process and the generalized upper bounds is somewhat obscure" before himself extending the inequalities (6) and (8) to continuous time as Ville had done but for nonnegative martingales. The "main theorem on martingales" of Ville's to which Doob alluded in 1939 is no doubt Ville's inequality and the extension of it that he made to continuous time, after having found a counterexample in order to better show hypotheses needed to make this extension valid. In the counterexample, Ville assumed a probability on a set of curves which is no longer the probability P "à la Kolmogorov" that he had defined for the space E_0 of all functions of real variables.

Lévy and Doob certainly did meet at Lyon, but between them it was not at all a question of martingales. The proceedings of the colloquium show that Lévy gave a presentation on 28 June on double Markov processes, where linear time is replaced by curves [45]. Following the presentation, Doob agreed with David van Dantzig that "Mr. Lévy has not made the definition of double Markov processes sufficiently precise". This provoked a response from Lévy on his analytic hypotheses and the conditional independence of two parts of the plane given a curve. Joseph Kampé de Fériet and Fréchet also asked questions, to which Lévy responded in writing on 29 and 30 June. Lévy did not ask Doob any questions about his presentation, in which, we may note, Doob cited only Ville [75], von Mises [54], and himself [22], not Lévy.

Making this all the more extraordinary, the second part of Doob's contribution touched on the application of martingale techniques to the strong law of large numbers, but only in the case of independent variables. Both Lévy and Fréchet knew Lévy's 1937 book [43] very well. Fréchet had the proofs in his hands [4, Letter 30]. But it is clear that neither Lévy nor Fréchet made the connection between this book and Doob's presentation. Lévy clearly had not read Doob's 1940 article [22], where he is cited several times.

It is well known that Lévy read little. This was all the more the case during the war, when he was in hiding from the Nazi hunt for Jews. Freed from hiding, he devoted himself to the publication and extension of the theorems he had proven during this period (stochastic integrals "à la Lévy" and Brownian motion), then to Markov processes (after a return to his "area processes"). He discovered belatedly that he had anticipated Kakutani on "Kakutani's Theorem", and he did not discover Ito's 1944 work [38] until 1954 [4,48].

According to a letter he wrote to Fréchet in 1964 [4, Letter 101], Lévy did not make the connection between his work and martingales until 1950:

> It was in 1950 at Berkeley that I learned from Loève that the processes called martingales were those that I had considered starting in 1935; according to your letter [Ville]'s second definition, p. 99, coincides with mine or at least becomes the same when constants are added.)

In the introduction to the second edition of [44], published that same year, Lévy writes that he has "renounced introducing the important notion of separable processes and speaking of martingales" and refers the reader to Doob [25].

The Lyon colloquium ended on Saturday, 3 July 1948. Several participants were to get together for lunch in Paris. Lévy, who had finished writing *Stochastic Processes and Brownian Motion* [44], which was supposed to appear at the beginning of the new academic year, returned to Paris and left on vacation. Doob quickly left France; we find him again at a Congress in Madison, Wisconsin, on 7 September 1948 [23]. Four more years would pass before the appearance of Doob's *Stochastic Processes* [25], with the 100 pages of its Chap. VII devoted to martingales. Martingales would need more round trips to America before a new generation of French probabilists took them up in their turn.

Michel Loève was not at the Lyon colloquium, much to the regret of Fréchet [32]. Several years later, right after the appearance of Doob's *Stochastic Processes* in 1953 [25], Loève would visit Paris and would take Paul-André Meyer to the USA, where the connection would be made that signaled a new development of martingales (and also probabilistic potential theory) in France.

6 Doob's Lecture

At Lyon, the "families of chance variables with property $\mathscr{E}$" of 1940 [22] became "martingales" relative to their natural filtration.

Doob's paper for the colloquium [24], has been reprinted at the end of [49]. It consisted of three parts, the first a review of Doob's definition of martingales and four results in his 1940 paper [22], including the two we mentioned above. These would be applied in the two following parts to the strong law of large numbers and inverse probability. These two topics are themes that run through the history of probability theory and statistics. The law of large numbers lies at the heart of controversies about foundations of probability, and inverse probability lies at the center of polemics on the use of probability in statistics. The strong law of large numbers and its connections with foundations were also discussed at Lyon by Ottaviani, who advocated "Cantelli's classical theory" against von Mises' collectives.[11]

Doob had already tried, in a single 1934 article that cast probability and statistics together in the mold of analysis [16], to make the law of large numbers a consequence of Birkhoff's ergodic theorem and to propose "for the first time a complete proof of the validity of maximum likelihood of R. A. Fisher." In 1936 he had to return to Fisher, giving conditions for the consistency of the maximum likelihood estimator [18].[12]

6.1 Strong Law of Large Numbers

Doob's proof at Lyon of the strong law of large numbers for identically distributed, independent, and integrable variables is a nice example of the kind of "spectacular application of martingale theory" that Doob often took pleasure in presenting, in particular in 1949–1950 at Feller's seminar at Cornell [12].

Given a sequence (u_n) of independent identically distributed variables, and assuming that their common mean exists, Doob defines another sequence of random variables indexed by the negative integers $-n$:

$$x_{-n} = \mathrm{E}\{\ldots, y_{-n-1}, y_{-n} \backslash y_{-1}\}, \tag{8}$$

[11] It is not possible to give a complete bibliography on these themes, so much have they been studied by historians of probability and statistics. But the author's "heartthrobs" include Stephen Stigler's [68, Chap. 3], the introduction to Christian Robert's [62], and Stephen Fienberg's [30]. For the history of the strong law, which had already been put forward by Borel in 1909, see Eugene Seneta's [63]. Let us add that the first appearance of the term "strong law" was in French ("loi forte"), in a note by Aleksandr Khinchin in the *Comptes rendus* for 30 January 1928; see [48].

[12] On Fisher and maximum likelihood, see especially [1,5,69]. Edwards [29] discusses the sense in which Fisher uses "inverse probability". Though published in 1936, [18] was presented in December 1934 and received by the journal 4 April 1935. In 1971 [26, p. 454], Doob returned to the relations between ergodic theory and martingales, affirming that "in a reasonable sense there are only two qualitative convergence theorems in measure theory, the ergodic theorem and the martingale convergence theorem."

where y_{-n} is the partial sum $u_1 + \cdots + u_n$ (as before, $\mathrm{E}\{A\backslash X\}$ is the conditional expectation of X given A). Because the u_i are independent and identically distributed,

$$x_{-n} = \frac{u_1 + \cdots u_n}{n}. \tag{9}$$

The sequence $(x_{-n})_{n\geq 0}$ is a martingale and it converges almost surely to a random variable with expectation $E(u_1)$ by Theorem 1.2 of [22], recalled by Doob at Lyon as property (ii) in the first section of his paper. Doob adds that the limit is easily identified as $E(u_1)$ (by (9), for example).

6.2 Inverse Probability

As the application of martingales to inverse probability at Lyon, we have a result of (almost sure) consistency of Bayesian parametric estimation. The parameter θ is governed by a priori density $f(\theta)$. For each θ the variable Y admits a law $F(\theta, y)$ with density $f(\theta, y)$ (Hypothesis A in Doob's text), and the mapping that associates each θ with its law is injective (Hypothesis B). The frequency $v_n(y)$ of observations smaller than y in a sample of size n tends almost surely to $F(\theta, y)$ (for each value y) when n tends to infinity (by a direct application of the iid law of large numbers), which allows us to view θ as a function $\hat{\theta}$ defined (up to a negligible set) on the sequences $(y_1, y_2, \ldots, y_j, \ldots)$.[13]

Doob remarks that the measurability assumption made in his Hypothesis A and Hypothesis B implies his Preliminary Hypothesis C, which says that "$\hat{\theta}$ is a measurable function of y_j sequences, that is a random variable on $y_1, y_2, \ldots, y_j, \ldots$ sample space". He finds this "somewhat surprising". (Notice that the "chance variables" of [22] have now become "random variables".) The almost sure convergence of $E[\hat{\theta}|y_1, y_2, \ldots, y_n]$ (here, we use the modern notation for conditional expectation) to $\hat{\theta}$ and the almost sure convergence of the conditional variance to 0 are obtained by application of the martingale results in [22], recalled in the first part of the paper. The remainder is a discussion of the hypotheses and their interpretation.

After Doob's presentation, only Rao rose to ask questions, on the possibility of applying the method without a prior distribution for θ, as in the nonparametric case. One can read Doob's responses in two parts, (i) and (ii), in the discussion that concludes Doob's paper. For a complementary analysis and for the connection with more recent results and the extensions of Doob's consistency theorem, see [36] and, for the nonparametric case [47].

[13] On Bayesian inference, see [30]. In 1936 [18], Doob had defined the consistency of a sample statistic as convergence in probability to the "true value" of the parameter. He had also drawn attention to the interest of almost sure convergence.

References

1. Aldrich, J.: R. A. Fisher and the making of maximum likelihood 1912–1922. Statistical Science **12**(3), 162–176 (1997)
2. Armatte, M.: Maurice Fréchet statisticien, enquêteur et agitateur public. Revue d'histoire des mathématiques **7**(1), 7–65 (2001)
3. Armatte, M.: L'enseignement de la statistique économique (1885-1925); présentation de quelques documents. Electronic Journal for History of Probability and Statistics **2**(2) (2006)
4. Barbut, M., Locker, B., Mazliak, L.: Paul Lévy and Maurice Fréchet: 50 years of correspondence in 107 letters. Springer, London (2014)
5. Bartlett, M.: R. A. Fisher and the last fifty years of statistical methodology. Journal of the American Statistical Association **60**(310), 395–409 (1965)
6. Bourbaki, N.: Éléments d'histoire des mathématiques. Hermann, Paris (1969)
7. Bru, B., Bru, M.F., Chung, K.L.: Borel and the St. Petersburg martingale. Electronic Journal for History of Probability and Statistics **5**(1) (2009)
8. Bungener, M., Joël, M.E.: L'essor de l'économétrie au CNRS. Cahiers pour l'histoire du CNRS **4**, 45–78 (1989)
9. Castelnuovo, G.: Calcolo delle probabilità. Zanichelli (1928)
10. Catellier, R., Mazliak, L.: The emergence of French probabilistic statistics. Borel and the Institut Henri Poincaré in the 1920s. Rev. Hist. Math. **18**(2), 271–335 (2012)
11. Chaumont, L., Mazliak, L., Yor, M.: Some aspects of the probabilistic work. In: A. Charpentier, A. Lesne, N.K. Nikolski (eds.) Kolmogorov's heritage in mathematics, pp. 41–66. Springer (2007)
12. Chung, K.L.: Probability and Doob. The American Mathematical Monthly **105**(1), 28–35 (1998)
13. Daniell, P.J.: A general form of integral. Annals of Mathematics pp. 279–294 (1918)
14. De Finetti, B.: La prévision: ses lois logiques, ses sources subjectives. Annales de l'Institut Henri Poincaré **7**(1), 1–68 (1937)
15. Dieudonné, J.: Sur le théoreme de Radon-Nikodym III. Annales de l'Université de Grenoble **23**, 25–53 (1948)
16. Doob, J.L.: Probability and statistics. Transactions of the American Mathematical Society **36**(4), 759–775 (1934)
17. Doob, J.L.: Note on probability. Annals of Mathematics **36**(2), 363–367 (1936)
18. Doob, J.L.: Statistical estimation. Transactions of the American Mathematical Society **39**(3), 410–421 (1936)
19. Doob, J.L.: Stochastic processes depending on a continuous parameter. Transactions of the American Mathematical Society **42**(1), 107–140 (1937)
20. Doob, J.L.: Stochastic processes with an integral valued parameter. Transactions of the American Mathematical Society **44**(1), 87–150 (1938)
21. Doob, J.L.: Review of [74]. Bulletin of the American Mathematical Society **45**(11), 824 (1939)
22. Doob, J.L.: Regularity properties of certain families of chance variables. Transactions of the American Mathematical Society **47**(3), 455–486 (1940)
23. Doob, J.L.: The transition from stochastic differences to stochastic differential equations. Econometrica **17**(1), 63–78 (1948)
24. Doob, J.L.: Application of the theory of martingales. In: Le Calcul des Probabilités et ses applications, Colloques Internationaux du CNRS XIII, pp. 23–27. CNRS (1949). Reproduced at the end of [49]
25. Doob, J.L.: Stochastic Processes. Wiley, New York (1953)
26. Doob, J.L.: What is a martingale? The American Mathematical Monthly **78**(5), 451–463 (1971)
27. Doob, J.L., Ambrose, W.: On two formulations of the theory of stochastic processes depending upon a continuous parameter. Annals of Mathematics **41**(4), 737–745 (1940)
28. Dosso, D.: Louis Rapkine (1904–1948) et la mobilisation scientifique de la France libre. Ph.D. thesis, Paris Diderot (Paris VII) (1998)
29. Edwards, A.W.F.: What did Fisher mean by "inverse probability" in 1912–1922? Statistical Science **12**(3), 177–184 (1997)

30. Fienberg, S.E.: When did Bayesian inference become Bayesian? Bayesian Analysis **1**(1), 1–40 (2006)
31. Fréchet, M.: Sur l'intégrale d'une fonctionnelle étendue à un ensemble abstrait. Bulletin de la Société Mathématique de France **XLIII**, 248–265 (1915)
32. Fréchet, M.: Avant propos. In: Le Calcul des Probabilités et ses applications, Colloques Internationaux du CNRS XIII, pp. 5–10. CNRS (1949)
33. Fréchet, M.: Les valeurs typiques d'ordre nul ou infini d'un nombre aléatoire et leur généralisation. In: Le Calcul des Probabilités et ses applications, Colloques Internationaux du CNRS XIII, pp. 47–51. CNRS (1949)
34. Fréchet, M., Halbwachs, M.: Le calcul des probabilités à la portée de tous. Presses Universitaires de Strasbourg (2019). 2nd edition, edited by Eric Brian, Hugo Lavenant, and Laurent Mazliak. Original appeared in 1924.
35. Gardes, F., Garrouste, P.: Ville's contribution to the integrability debate: the mystery of a lost theorem. History of Political Economy **38**(5), 87–106 (2004)
36. Ghosal, S.: A review of consistency and convergence rates in posterior distribution in Bayesian inference (1999). Proceedings of Varanasi Symposium, Banaras Hindu University
37. Havlova, V., Mazliak, L., Sisma, P.: Le début des relations franco-tchécoslovaques vu à travers la correspondance Hostinsky-Fréchet. Electronic Journal for History of Probability and Statistics **1**(1) (2005)
38. Ito, K.: Stochastic integrals. Proceedings of Imperial Academy Tokyo **20**, 519–524 (1944)
39. Kolmogorov, A.N.: Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer, Berlin (1933). An English translation by Nathan Morrison appeared under the title *Foundations of the Theory of Probability* (Chelsea, New York) in 1950, with a second edition in 1956.
40. Lévy, P.: Calcul des probabilités. Gauthier-Villars, Paris (1925)
41. Lévy, P.: La loi forte des grands nombres pour les variables aléatoires enchaînées. Journal de mathématiques pures et appliquées **15**, 11–24 (1936)
42. Lévy, P.: Sur la notion de probabilité conditionnelle. Bulletin des sciences mathématiques **60**, 66–71 (1936)
43. Lévy, P.: Théorie de l'addition des variables aléatoires. Gauthier-Villars, Paris (1937)
44. Lévy, P.: Processus stochastiques et mouvement brownien. Gauthier-Villars, Paris (1948)
45. Lévy, P.: Processus doubles de Markoff. In: Le Calcul des Probabilités et ses applications, Colloques Internationaux du CNRS XIII, pp. 53–59. CNRS (1949)
46. Lévy, P.: Quelques aspects de la pensée d'un mathématicien. Blanchard (1970)
47. Lijoi, A., Prünster, I., Walker, S.G.: Extending Doob's consistency theorem to nonparametric densities. Bernoulli **10**(4), 651–663 (2004)
48. Locker, B.: Paul Lévy, la période de guerre. Intégrales stochastiques et mouvement brownien (2001). Thesis, Université Paris 5
49. Locker, B.: Doob at Lyon. Electronic Journal for History of Probability and Statistics **5**(1) (2009)
50. Mazliak, L.: The Ghosts of the Ecole Normale. Statistical Science **30**(3), 391–412 (2015)
51. Meusnier, N.: Sur l'histoire de l'enseignement des probabilités et des statistiques. Electronic Journal for History of Probability and Statistics **2**(2) (2006)
52. Michel, A.: Constitution de la théorie moderne de l'intégration. Vrin (1992)
53. von Mises, R.: Wahrscheinlichkeit, Statistik und Wahrheit. Springer (1928)
54. von Mises, R.: Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik. Deuticke (1931)
55. von Mises, R.: Théorie des probabilités. Fondements et applications. Annales de l'Institut Henri Poincaré **3**(2), 137–190 (1932)
56. Neveu, J.: Martingales à temps discret. Masson, Paris (1972)
57. Nikodym, O.: Sur une généralisation des intégrales de M. J. Radon. Fundamenta Mathematicae **15**(1), 131–179 (1930)
58. Nualart, D.: Les martingales i les seves aplicacions des d'una perspectiva històrica. Butlletí de la Societat Catalana de Matemàtiques **4**, 33–46 (1989)
59. Piaggio, H.T.H.: Review of [74]. The Mathematical Gazette **23**(257), 490–491 (1939)

60. von Plato, J.: Creating Modern Probability. Cambridge University Press (1994)
61. Radon, J.: Theorie und Anwendungen der absolut additiven Mengenfunctionen. Sitz. Akad. Wiss. Wien. **122**, 1295–1438 (1913)
62. Robert, C.P.: Le choix bayésien: Principes et pratique. Springer (2006)
63. Seneta, E.: On the history of the strong law of large numbers and Boole's inequality. Historia Mathematica **19**(1), 24–39 (1992)
64. Shafer, G.: A comprehensive list of the publications of Jean André Ville (1910–1989). Electronic Journal for History of Probability and Statistics **5**(1) (2009)
65. Shafer, G.: The education of Jean André Ville. Electronic Journal for History of Probability and Statistics **5**(1) (2009)
66. Siegmund-Schultze, R.: Maurice Fréchet à Strasbourg: Les mathématiques entre nationalisme et internationalisme. In: E. Crawford, J. Ollf-Nathan (eds.) La science sous influence: L'université de Strasbourg, enjeu des conflits franco-allemands, 1872–1945, pp. 185–196. La Nuée Bleue, Strasbourg (2005)
67. Snell, J.L.: A conversation with Joe Doob. Statistical Science **12**(4), 301–311 (1997)
68. Stigler, S.M.: The History of Statistics, The Measurement of Uncertainty before 1900. Harvard (1986)
69. Stigler, S.M.: The epic story of maximum likelihood. Statistical Science **22**(4), 598–620 (2007)
70. Ville, J.: Sur les conditions d'existence d'une ophélimité totale et d'un indice du niveau des prix. Annales de l'Université de Lyon **9**, 32–39 (1946)
71. Ville, J.: Théorie et application de la notion de signal analytique. Câbles et transmissions **2**(1), 61–74 (1948)
72. Ville, J.: Fonctions aléatoires et transmission de l'information. In: Le Calcul des Probabilités et ses applications, Colloques Internationaux du CNRS XIII, pp. 115–119. CNRS (1949)
73. Ville, J.: Curriculum vitae of Mr. Jean André Ville (1956). Electronic Journal for History of Probability and Statistics **5**(1) (2009)
74. Ville, J.A.: Sur un jeu continu. Comptes rendus **206**, 968–969 (1938)
75. Ville, J.A.: Étude critique de la notion de collectif. Gauthier-Villars, Paris (1939)
76. Wiener, N.: Differential-space. Journal of Mathematics and Physics **2**(1-4), 131–174 (1923)
77. Wiener, N.: Un problème de probabilité dénombrables. Bulletin de la Société mathématique de France **52**, 569–578 (1924)
78. Zallen, D.T.: The Rockefeller Foundation and French research. Cahiers pour l'histoire du CNRS **5**(24), 218 (1989)

Part III Modern Probability

Stochastic Processes in the Decades after 1950

Paul-André Meyer and Glenn Shafer

Abstract

Paul-André Meyer (1934–2003), founder and leader of the Strasbourg school of probability, worked from the 1960s into the 1990s on the theory of stochastic processes, emphasizing processes in continuous time. In this sweeping review, which first appeared in French in 2000, Meyer emphasizes the founding role of Doob's *Stochastic Processes*, published in 1953, which presented tools and topics that fueled probabilistic research for the rest of the century: filtrations, stopping times, martingales, Markov processes, diffusions, Itô's stochastic integral, and stochastic differential equations. The review then concentrates on two periods of research. The period from 1950 to 1965 was dominated by the study of Markov processes and their connections with potential theory and martingales. In the period from 1965 to 1980, martingales became more prominent, along with the stochastic integral, excursions, the general theory of processes, and stochastic mechanics. The review extends into the 1980s, discussing the Malliavin calculus and noncommutative probability theory. The great contributions by Doob and Itô are given their due.

Keywords

General theory of processes · Malliavin calculus · Markov processes · Martingales · Noncommutative probability · Potential theory · Stochastic differential geometry · Stochastic integral · Stochastic processes · Stochastic processes · White noise

---

Paul-André Meyer (1934–2003) worked at the University of Strasbourg. This chapter was first published in French in 2000 [172]. Glenn Shafer, Rutgers University. Translation by Jeanine Sedjro.

P.-A. Meyer · G. Shafer (✉)
Rutgers University, 377 South Harrison Street, Apt. 20B, East Orange, New Jersey 07018, USA
e-mail: gshafer@business.rutgers.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*,
Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_8

1 Introduction

Doing "history of mathematics" about Probability Theory is an undertaking doomed to failure from the outset, hardly less absurd than doing history of physics from a mathematician's viewpoint, neglecting all of experimental physics. We can never say often enough, *Probability Theory is first of all the art of calculating probabilities*, for pleasure and for probabilists to be sure, but also for a large public of users: statisticians, geneticists, epidemiologists, actuaries, economists.... The progress accomplished in fifty years responds to the increasing role of probability in scientific thought in general, and finds its justification in more powerful methods of calculation, which allow us for example to consider the measure associated with a stochastic process as a whole instead of considering only individual distributions of isolated random variables.

It must be acknowledged from the beginning that the "history" below, written by a mathematician, not only ignores the work accomplished by non-mathematicians and published in specialized journals, but also the work accomplished by mathematicians deepening classical problems—sums of independent variables, maxima and minima, fluctuations, the central limit theorem—by classical methods, because daily practice continues to require that these old results be improved, the same way the internal combustion engine continues to be improved to build cars.

Probability has developed many branches since 1950. The schematic description found here concerns only stochastic processes, understood in the fairly narrow sense of random evolutions governed by time, continuous or discrete. Moreover, we must leave aside, for lack of competence, the study of classes of special processes.

I have presented the parts of probability that I myself came in contact with, and their development *as it appeared to me*, trying at most to verify certain points by bibliographical research. In particular, saying that an article or an author is "important" signifies that they have aroused a certain enthusiasm among my colleagues (or in me), that they were the source of some other work, that they enlightened me on this or that subject. I feel especially uncomfortable presenting work that appeared in the East (Japan being part of the West on this occasion). In fact, not only was communication slow between the two political blocs, but probabilists worked in slightly different mindsets, with certain mental as well as linguistic barriers. Even in the West, we can distinguish smaller universes, each with its traditions, tastes and aversions. The balance between pure and applied probability, for example, was very different in the Anglo-Saxon countries, endowed with powerful schools of statisticians, than in France or Japan. The text that follows should therefore be considered as expressing personal opinions, not value judgments.

2 Probability Around 1950

This initial date may be less arbitrary in probability than elsewhere. In fact, it is marked by two works that have reached a broad public, the first one summarizing two centuries of ingenuity, the second one providing tools for future development.

First Feller's book *An Introduction to Probability Theory and Its Applications* [92], without a doubt one of the most beautiful mathematics book ever written, with technical tools barely exceeding the level of secondary school. Next Halmos's *Measure Theory* [110], the first presentation in the West of measure theory free of unnecessary subtleties and well adapted to the teaching of probability according to Kolmogorov's axioms [147]. Then came Loève's 1960 book [158], for many years the prototype for textbooks on the topic. In fact, discussions on the foundations of probability, which had embroiled the previous generation, were over. Mathematicians had made a definitive choice of their axiomatic model, leaving it to philosophers to discuss the relation between it and "reality". This had not happened without resistance. A majority of probabilists, particularly in the United States, had long considered the teaching of the Lebesgue integral not only a waste of time, but also an offense to "probabilistic intuition".

2.1 Early Developments

Let us note a few mathematical events that came just before the period at hand and seeded future developments. The first article published by Itô on the stochastic integral dates from 1944 [123]. Doob worked on the theory of martingales from 1940 to 1950 [61,63], and it was also in a 1945 article by Doob that the strong Markov property was clearly enunciated for the first time and established in a very special case [62]. The theorem giving the structure of strongly continuous semigroups of operators, which greatly influenced Markov process theory, was proven in 1948, independently by Hille [116] and Yosida [232,233]. Great progress in potential theory, which was also destined to influence probability, was achieved by H. Cartan in 1945 and 1946 [40,41], and by Deny in 1950 [55]. In 1944 [140,141], Kakutani published two brief notes on the relations between Brownian motion and harmonic functions, which became the source of Doob's work on this topic and grew into a wide area of research. In 1949 [139], inspired by the Feynman integral, Kac presented the "Feynman-Kac formula", which remained a theme of constant study in various forms. We use this occasion to commemorate this extraordinary lecturer, more fond of flinging ideas all about than of writing polished articles. And in 1948 [156], Paul Lévy published an extremely important book, *Stochastic Processes and Brownian Motion*, which passed in review the entire menagerie of stochastic processes known at the time. Like all of Lévy's work, it was written in the style of explanation rather than proof, and rewriting it in the rigorous language of measure theory was an extremely fruitful exercise for the best probabilists of the time (Itô, Doob). Another example of the depth probabilists reached working with their bare hands was the famous work of Dvoretzky, Erdős and Kakutani on multiple points of Brownian motion in $\mathbb{R}^n$, from 1950 to 1957 [74–76]. It took a long time to notice that although the result was perfectly correct, the proof itself was incomplete!

2.2 "Stochastic Processes"

Doob's book, *Stochastic Processes*, published in 1953 [64], would be the Bible of the new probability, and it merits analysis. Aside from the abundance of his own discoveries, Doob's special stature stems from his familiarity with measure theory, which he adopts as the foundation of probability without any backward glance or mental reservation. But the theory of continuous-time processes poses difficult problems for measure theory. If a particle is subject to random evolution, showing that its trajectory is continuous or bounded requires that all time values be considered, whereas classical measure theory can handle only a *countable* infinity of time values [60]. So probability is not just based on measure theory; *it requires more from measure theory than the rest of analysis does*. Doob's book begins with an abrupt chapter and finishes with a dry supplement—adhering in between to a high-grade austerity, accentuated by a typography that recalls of the great era of *le Monde* but is made pleasant by a style free of pedantry. Starting with Doob, probability, even in the eyes of Bourbaki, will be one of the respectable disciplines.

It is instructive to enumerate the subjects covered in Doob's book. He begins with a discussion of the principles of the theory, particularly the solution to the difficulty mentioned above; here he introduces "separability" for processes. Then a brief exposition on sums of independent variables; the theory of martingales in discrete and continuous time (work by Doob that was still fresh) with many applications; processes with independent increments; Markov processes (Markov chains, resuming Doob's 1945 work [62], diffusions, presenting Itô's stochastic integral with an important addition for further work, and stochastic differential equations). It all appears prophetic now. On the other hand, three subjects are discussed only slightly: Gaussian processes, stationary processes, and prediction theory for second order processes. Each of these branches is being called on to detach itself from the common trunk of process theory and to grow in an autonomous fashion—and we will not talk about them here.

We must comment on one aspect of Doob's book, crucial for the future. Kolmogorov's mathematical model represents the events of the real world by elements of the sigma-algebra $\mathscr{F}$ of a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Intuitively speaking, the set Ω is a giant "urn" from which we draw out a "ball" ω, and the elements of $\mathscr{F}$ describe the various questions that one can ask about ω. Paul Lévy protested against this model, criticizing it for evoking *only one* random draw, whereas chance evidently enters at every moment in a random evolution. Doob resolved this difficulty in the following way: There is a single random draw, but it "reveals" itself progressively. Time t (discrete or continuous) is introduced in the form of an increasing family $(\mathscr{F}_t)$ of sigma-algebras—what is currently called a *filtration*. The sigma-algebra $\mathscr{F}_t$ represents "what is known of ω up to time t". Now let T be the moment where for the *first time* the random evolution is seen to have a certain property—for the insurance company, the first fire of the year 1998, for example. It is a random quantity such that there is no need to look at the evolution beyond t to know if $T \leq t$. In mathematical language, the event $\{T \leq t\}$ belongs to $\mathscr{F}_t$. In fact, to know if there was a fire in January 1998, there is no need to wait until the month of March. Compare this definition to that of the *last* fire of the year 1997: to know if it occurred in November, you need to know that a fire occurred in November *and also that no fire occurred in December*. These "non-anticipatory" random variables are now called *stopping times*. The idea of non-anticipatory knowledge is implicit in French, where (normally) the inflection of a word depends only on words coming before it, but not in German, where the whole meaning of the sentence depends on the final element. The importance of the notion of stopping times comes surely from the work of Doob and of his disciple Snell [208], but it must have a prior history, because it penetrates for example Wald's sequential statistical analysis.

3 The Great Topics of the Years 1950–1965

This period was dominated by the theory of Markov processes and its interaction with potential theory and martingale theory.

3.1 Markov Processes

The efforts of probabilists of the first half of the century had been mostly dedicated (the problem of foundations aside), to the study of independence: sums of independent random variables and corresponding limit distributions. After independence, the simplest type of random evolution is Markovian dependence (named after A. A. Markov, 1906 [162]). An example is given by the successive states of a deck of cards that is being shuffled. For predicting the order of cards after the next shuffle, all useful information is included in the (complete) knowledge of the current state of the deck; if this is known, knowledge of previous states adds nothing to the accuracy of the prediction. Most examples of random evolution given by nature are Markovian, or become Markovian by a suitable interpretation of the words "current state" and "complete knowledge". The theory of Markov processes divides into several subtheories, depending on whether time is discrete or continuous, or on whether the set of possible states is at most countably infinite (the Markov *chain* case) or continuous. (For some authors, a Markov chain is a Markov process in discrete time, regardless of the cardinality of the state space.) On the other hand, the classical theory of sums of independent random variables can be generalized into a branch of Markov process theory where a group structure replaces addition: in discrete time this is called *random walks*, and in continuous time *processes with independent increments*, the most notable of which is *Brownian motion*.

From the probabilistic point of view, a Markov process is determined by its *initial law* and its *transition function* $P_{s,t}(x, A)$, which gives, if we observed the process in state x at time s, the probability that we find it at a later time t in a set A (if we exclude the case of chains, the probability of finding it *exactly* in a given state y is null in general). The transition function is a simple analytical object—and in particular, when it is *stationary*, meaning it only depends on the difference $r = t - s$, we obtain a function $P_r(x, A)$ to which the analytical theory of semigroups, in full flower since

Hille-Yosida's theorem, applies. Hence the interest in Markov processes around the 1950s.

The main question we ask ourselves about these processes is that of their long term evolution. For example, the evolution of animal or human populations can be described by Markovian models that have three types of limiting behavior: extinction, equilibrium, or explosion—the final one, impossible in the real world, nevertheless constitutes a useful mathematical model for a very large population. The study of various states of equilibrium with a stationary regime is related to statistical mechanics.

Continuous-time and finite-state space *Markov chains*, already well known for years, represent perfectly regular random evolution, which stays in a state for a certain period of time (of known law), then jumps into another state drawn at random according to a known law, and so on and so forth indefinitely. But as soon as the number of states becomes infinite, extraordinary phenomena can arise: jumps may accumulate in a finite period of time (and then the process becomes indescribably complicated), or even worse, each state may be occupied from the outset according to a "fractal" set. The problem is of an elementary nature, very easy to raise and not easy at all to resolve. This is why Markov chains have played the role of testing ground for every later development, in the hands of the English school (Kingman, Reuter, Williams…) and of K. L. Chung, whose insistence on a probabilistic rather than analytic attack on the problems has had a considerable influence.

The other area of Markov process theory in full flower was *diffusion theory*. In contrast to Markov chains, which in simple cases progress only by jumps separated by intervals of constant length, diffusions are Markov processes (real, or else with values in $\mathbb{R}^n$ or in a manifold) with *continuous* trajectories. We knew since Kolmogorov [146] that in the most interesting cases the transition function is the solution of a parabolic partial differential equation, the Fokker-Planck equation (in fact of two equations, depending on whether we move forward or backward in time). During the 1950s, we readily tried to construct diffusions with values in manifolds by semigroup methods, but the work that stood out was Feller's analysis of the structure of diffusions in one dimension [93,94]. One of the themes of the following years would be the analogous problem in higher dimensions, where substantial but not definitive results would be obtained.

The ideas introduced by Doob (increasing families of sigma-algebras, stopping times) made it possible to give a precise meaning to what we call the *strong Markov property*: Given a Markov process whose transition function is known (and stationary, for simplicity), the process considered from a random time T is again a Markov process with the same transition function, *provided T is a stopping time*. This had been used long before the notion of stopping time was extracted, in heuristic arguments such as D. André's "reflection principle"[1] —and also in false heuristic arguments, in which T is not really a stopping time. In fact, the first case where a strong Markov property was rigorously asserted and proven is found, it seems, in Doob's 1945 article

[1] Which allows the distribution of the maximum of a Brownian motion to be calculated.

on Markov chains [62], but Doob himself hides the question under a smoke screen in his great article of 1954 [65]. In the case of Brownian motion, the first modern and complete statement in the West is due to Hunt in 1956 [119], while the Moscow school reached in parallel a greater generalization.

3.2 Development of Soviet Probability

While probability was a marginal branch of mathematics in Western countries, it had always been among the strongest points of Russian mathematics, and it had grown with Soviet mathematics. Two generations of extraordinary quality would make of Moscow, then Kiev, Leningrad, Vilnius, probabilistic centers among the most important of the world—before the post-Stalin wave of persecution (mostly antisemitic) brought this boom to a halt, and forced many major figures into internal or external exile. Dynkin himself left in 1977 for the United States. It would take a specialist to tell the whole story. In any case we can identify two dates: 1952, when Dynkin published his first article on Markov processes [77], and 1956, the birth date of the journal *Teoriya Veroyatnosti*, which published in its first issue two still classic articles, by Prokhorov and Skorokhod [189,206], on weak convergence of measures on metric spaces.[2] Skorokhod's classic book on processes, which extended this work, appeared in 1961 [207].

Concerning the theory of Markov processes, which for many years was one of the principal themes (but not the only theme) of Soviet probability, the history of connections between the Russian school (see [78]) and "Western" probability (including the rich Japanese school!) is partly one of misunderstanding. This is probably due to the absence of structured research in the West, and to the systematic character, in contrast, of the publications of Dynkin's seminar, supporting each other, using a rather abstract common language, and giving prominence to Markov processes with nonstationary transition functions. The fact is that the main results on the regularity of trajectories and the strong Markov property have been proven twice: by Dynkin and Yushkevich, and by Hunt and Blumenthal. The situation was repeated much later, when many important Soviet works (on excursions, on "Kuznetsov measures") were understood late in the West, after being partially rediscovered.

After these generalities, we can examine various streams of ideas.

3.3 Classical Potential Theory and Probability

In 1954 [65], developing an idea of Kakutani's, dating from 1944 and taken up again in 1949, Doob published an article on the connection between classical potential theory in $\mathbb{R}^n$ and continuous-time martingale theory. The main idea is the link between the solution of Dirichlet's problem in an open set and the behavior of Brownian

[2] Weak convergence is associated with the integration of bounded continuous functions [14].

motion starting from a point x of this open set: The first moment when a trajectory ω of Brownian motion meets the boundary depends on ω, it is therefore a "random variable". Let us call it $T(\omega)$; let $X(\omega)$ be the position of the trajectory at that moment. It is clear that $X(\omega)$ is a point on the boundary; so if f is a boundary function, $f(X)$ is a random quantity whose expected value (the integral) depends on the initial point x. So let us call this expected value $F(x)$: *this function on the open set solves Dirichlet's problem on the open set with boundary condition f.*

All of this had been known for a long time in the case of simple open sets like balls. But for arbitrary domains Doob had to resolve (relying on potential theory) delicate problems of measurability, and most of all, he established a link between the harmonic and superharmonic functions of potential theory, and martingale theory: *if we compose a harmonic or superharmonic function with Brownian motion, we obtain a martingale or supermartingale with continuous trajectories*. Let us emphasize this continuity: the superharmonic functions are not in general continuous functions, but the Brownian trajectories "do not see" their irregularities. Doob uses this result, along with the theory of martingales, to study the behavior of positive harmonic or superharmonic functions at the boundary of an open set, a subject to which he will devote several articles [67–71].

Perhaps the most striking result of this probabilistic version of potential theory is the intuitive interpretation of the relatively technical notion of the *thinness* of a set, introduced in the study of Dirichlet's problem in an open set. We can always "solve" Dirichlet's problem in a bounded open set with a continuous boundary condition f, but we get a generalized solution that neither necessarily has f as its limiting value *everywhere* nor (where it does have it) necessarily has it in the sense of the *ordinary* topology. There are bad points, and even at the good points one must not approach the boundary too quickly. The notion of thinness makes these two notions precise: "regular" points on the boundary, for example, are those where the complement of the open set is not thin. Now, the probabilistic interpretation of thinness is very intuitive: to say that a set A is thin at the point x means that a Brownian particle placed at x will take (with probability 1) a certain time before returning to A. (We say returning to A rather than finding A, because, if x itself belongs to A, this encounter with A at moment 0 does not count.) A certain number of delicate properties of thinness immediately become evident.

Even though it is not our subject, it is worth pointing out that this immediate post-war period, particularly fruitful in the area of probability, was also fruitful for potential theory. The very abundant and interesting production (never assembled) of mathematicians like M. Brelot and J. Deny bore fruit not merely in potential theory and probability; few people know that distribution theory, for example, was born from a question posed to L. Schwartz on polyharmonic functions.

3.4 Theory of Martingales

We will not give here the definition of martingales, even though it is simple, but only the underlying idea. The archetype of martingales is the capital of a player in the course of a *fair* game. This capital stays constant *on average*, but its actual course can fluctuate considerably; significant but rare gains can compensate for accumulations of small losses (or conversely) [46]. The notion of *super*martingale corresponds similarly to an *un*favorable game (the "super" expressing the point of view of the casino). In continuous time, Brownian motion, meaning the mathematical model describing the motion of a pollen particle in water seen in a microscope, is also a pure fluctuation: *on average*, the particle does not move: the two dimensional Brownian motion is a martingale.[3] If we add a vertical dimension, we lose the martingale property, because the particle will tend to go down if it is denser than water (in this case the vertical component is thus a *super*martingale), and go up otherwise.

After a pre-history where the names of S. Bernstein (1927 [10,11]), P. Lévy (1937 [155]) and J. Ville (1939 [223]) stand out, the biggest name of martingale theory is that of Doob, who proved many fundamental inequalities and the first limit theorems, and linked martingales with the "stopping times" we talked about above, these random variables that represent the "first time" that we observe a phenomenon. Doob gathered in his book so many striking applications of martingale theory that the probabilistic world found itself converted, and the search for "good martingales" became a standard method for approaching numerous probability problems. We have at our disposal a considerable number of results on martingales: conditions under which a martingale diverges to infinity, how to study its limit distributions if it does not diverge, and especially a set of very precise inequalities, allowing us to bound the fluctuation of a martingale based on observable characteristics. We will talk about this more below.

3.5 Markov Processes and Potential

It was clear that the results obtained by Doob for Brownian motion should extend to much more general Markov processes. Doob himself went from classical potential theory to a much less classical theory, that of the potential for heat [66].[4] But the fundamental work in this direction was accomplished by Hunt's very great article, published in three parts in 1957 [120]. This article (preceded by an article by Blumenthal [21] that laid the foundation), contained a wealth of new ideas. The most important for the future, probably, was the direct use in probability (for lack of an already developed potential theory, which Doob already had in his first article) of Choquet's theorems on capacities. But Hunt also established (by a proof that is a real masterpiece) that *any* potential theory satisfying certain axioms stated by Choquet and Deny is susceptible of a probabilistic interpretation. This result unifying analysis and probability contributed to making the latter a respectable field.

[3] Brownian motion happens to be simultaneously a martingale and a Markov process, but these two notions are not related.

[4] Of which the core is the elementary solution of the heat equation, that is, the Brownian transition function itself.

The third part of Hunt's article is also very original, because it provides a substitute for the *symmetry* of Green's function in classical potential theory. The main role is no longer played by a single semigroup, but by a pair of transition semigroups that are "dual" with respect to a measure—in classical potential theory, the Brownian semigroup is its own dual with respect to Lebesgue measure. In this case, we can build a much richer potential theory, but (provisionally) the duality remains devoid of probabilistic interpretation: folklore sees it as a kind of time reversal, but this interpretation is rigorous only in particular cases.

A second aspect of probabilistic potential theory concerns the study of the *Martin boundary*. This is a concept introduced in 1941 in a (magnificent) article by R. S. Martin [163], a mathematician who died shortly afterwards. On one hand, he interpreted the Poisson representation of positive harmonic functions as an integral representation by means of extremal positive harmonic functions; on the other hand, he indicated a method for constructing these functions in any open set: He "normalized" Green's function $G(x, y)$ by dividing it by a fixed function $G(x_0, y)$, then compactified the open set so that all these quotients are extended by continuity; all the extremal harmonic functions are then among these limit functions. This idea was picked up and developed by Brelot in 1948 and 1956 [25,26], and it was partly the origin of Choquet's research on integral representation in convex cones [42]. It was again Doob who, in 1957, discovered the probabilistic meaning of these quotients of harmonic or superharmonic functions. A series of subsequent articles was to extend all this to general Markov processes, by showing that "Martin's boundary" was an advantageous replacement for the "boundaries" introduced earlier to capture the asymptotic behavior of Markov processes. Yet the most decisive step was to be accomplished by Hunt in a brief and schematic article in 1960 [121]—his last publication in this area—that introduced a new way to "reverse time" for Markov processes starting from certain random times, so giving a probabilistic interpretation very useful for Martin's theory. Hunt's article, which concerned only discrete chains, was extended to continuous time by Nagasawa in 1964 [178] and Kunita and T. Watanabe in 1965 [151]. The result of this work is a rigorous probabilistic interpretation of the duality between Markov semigroups.

In two dimensions, Brownian motion is said to be *recurrent*: its trajectories, instead of tending to infinity, come back infinitely often to an arbitrary neighborhood of any point of the plane. It gives rise to the special theory of logarithmic potential. There exists a whole class of Markov processes of the same kind, whose study is related rather to ergodic theory. This is an opportunity to mention Spitzer's 1964 book on recurrent random walks [210], which has had a considerable influence (see also [209]). It opened an important line of research, linking probability, harmonic analysis and group theory (discrete groups and Lie groups). It would merit a special study, which surpasses my own competence.

Work a little remote from this, which deserves to be cited because it concluded years of research on the regularity of trajectories of Markov processes, is an article of D. Ray from 1959 [193]. This article shows (using methods close to those of Hunt) that it is in part an artificial problem. *Any Markov process* can be rendered strongly Markovian and right-continuous by compactifying its state space by adding "fictional states". Ray's article contained an error, corrected by Knight [143], but it involved a very fruitful method, also destined to rejoin Martin's theory of compactification. On this subject there was again a parallel development in the work of the Russian school, but the results are not directly comparable.

The classic book presenting Hunt's theory and its development (with the exception of Martin's boundary) is the 1968 book of Blumenthal and Getoor [22]. Since we will return very infrequently to probabilistic potential theory, let us mention nevertheless that the subject has remained active up to the time of this writing, mainly in the United States (Getoor, Sharpe [104–106]). For modern presentations, see the books by Sharpe in 1988 [200] and Bliedtner and Hansen in 1986 [20]. For interactions between classical potential theory and Brownian motion, the reference is Doob's monumental 1984 treatise [72]. Yet the most active branch currently is that of *Dirichlet spaces*, which we will say a word about later on.

3.6 Special Markov Processes

Hunt's general theory of Markov processes is only one branch of Markov process theory. The 1960s saw extraordinary activity in the study of special processes. First, the very close study of trajectories of classical processes—Hausdorff dimensions, etc., what we would today call their fractal structure. Let us cite for example, other than the works of Dvoretzky, Erdős, and Kakutani [73–76], those of S. J. Taylor [48,220]. Then the study of Markov chains with little regularity, which provides an inexhaustible source of examples and counterexamples (Chung [43]; Neveu [185, 186]). In 1979 [228], Williams called Neveu's 1962 article "the finest paper ever written on chains". Finally a very rich production in the study of *diffusions*, which will find its Bible in the 1965 book (too long awaited) of Itô and McKean [131]. The main problems here concern the structure of diffusions in several dimensions, and in particular the possible behavior at the boundary of an open set of a diffusion whose infinitesimal generator is known in the interior. For example, to take a problem dealt with by Itô and McKean in 1965 [131], find all strongly Markovian processes with continuous trajectories on the positive closed half-line that are Brownian motions in the open half-line. But of course the problem in several dimensions (studied by the Japanese school; we cite for example Motoo in 1964 [176]) is much more difficult. It is a matter of making precise the following idea: the diffusion is formed from an interior process, describing the first trip to the boundary, then the excursions starting and ending on the boundary. It's just that infinitely many small excursions happen in a finite amount of time, and we must manage to describe them and put them back together. It's a difficult and fascinating problem.

3.7 Connections Between Markov Processes and Martingales

It is natural that martingales should be applied to Markov processes. Conversely, methods developed for the study of Markov processes have had an impact on the theory of martingales.

Probabilistic potential theory was developed for a stationary transition function, i.e. for a *semigroup of transition operators* (P_t). This semigroup operates on positive functions. The functions that generalize superharmonic functions, here called *excessive functions*, are measurable positive functions f such that $P_t f \leq f$ for every t (and a minor technical condition). In classical potential theory, it is known how to describe these functions, which decompose into a sum of a positive harmonic function and the Green potential of a positive measure μ. On the other hand, we can associate a Markov process (X_t) with the transition function, and the excessive functions are those for which the process $(f(X_t))$ is a *supermartingale*. In probabilistic theory, there are no measure potentials available, but Dynkin had stated the problem of representing an excessive function f as the potential of an *additive functional* [79]. Without going into technical details, such a functional is given by a family of random variables (A_t) representing the "mass of the measure μ seen by the trajectory of the process between times 0 and t", and the connection with the function f is that for a process starting from x, the expected value of A_∞ is equal to $f(x)$. The Russian school (Volkonskii in 1960 [224], Shur in 1961 [202]) had obtained very interesting partial results. In the West, Meyer, who was working with Doob, was able to improve Shur's result in 1962 [167] by giving a necessary and sufficient condition for an excessive function to be representable in this way (a condition Doob had formulated earlier in potential theory) and to study the uniqueness of the representation.

A little later, in 1962 [167], Meyer noticed that the methods that had just been used in the theory of Markov processes transposed without change to the theory of martingales, to solve an old problem raised by Doob: the decomposition of a *supermartingale* into a difference of a martingale and a process with increasing trajectories—an obvious result in discrete time. We knew that conditions were needed (Ornstein had shown an example where the decomposition did not exist), and the notion of "class (D)" answered the question precisely. From that moment on, methods that had succeeded with Markov processes would be grafted onto the general theory of processes, giving numerous results. In particular, capacitary methods would make their entry into the theory of processes. This was quite hard to accept in an environment that had still been balking at the Lebesgue integral ten years earlier! Whence a certain bad mood, quite noticeable particularly in the United States.

Before resuming the main flow of thought, a few remarks about a very important particular case of the problem of decomposition. One-dimensional Brownian motion admits no positive superharmonic functions but plenty of positive subharmonic functions (the convex positive functions), and the corresponding problem of representation had been solved by hand. One of the marvels of Lévy's work had been the discovery and study of Brownian motion's *local time* at a point, which measures in a certain way the time spent "at that point". In all rigor, this time is zero, but the time spent in a small neighborhood, properly normalized, admits a nontrivial limit. Trotter had made a thorough study of this local time in 1958 [221]. In 1963 [219], Tanaka made the connection between local time and Doob's decomposition of the absolute value of Brownian motion, thus establishing what was henceforth called "Tanaka's formula". The construction of local times for various types of processes (Markovian, Gaussian…) has remained a favorite theme of probabilists. On local times one may consult Azéma and Yor's 1978 collection [4]. See also [194].

The problem of decomposition has had other important extensions. A 1965 article by Itô and Watanabe [132], otherwise devoted to a Markovian problem, introduced the very useful notion of *local martingale*,[5] which allows us to treat the problem of decomposition without any restriction. On the other hand, a 1965 article by Fisk [96], developing work by Orey [187], introduces the notion of *quasi-martingale*, corresponding somewhat to the notion of a function of bounded variation in analysis.

We could choose as the symbolic date to close this period the year 1966, during which the second volume of Feller's treatise appeared [95]. Like the first, it addresses the vast audience of probability users and remains as concrete and elementary as possible. Like the first, it assembles and unifies an enormous mass of practical knowledge, but this time it uses measure theory. Moreover, the period preceding 1966 was a time of synthesis and perfection, Dynkin's second book on Markov processes appeared in 1963 [80], Itô and McKean's book on diffusions in 1965 [131], and Meyer's synthesis of recent works on the general theory of processes in 1966 [169]. See also [84]. Many special cases of stochastic processes are not considered in the present chapter (e.g. branching processes—see [111]).

4 The Period 1965–1980

In this period, Markov processes were less dominant. The stochastic integral and martingales took center stage.

4.1 The Stochastic Integral

Doob's book pointed out already that Itô's stochastic integral theory was not essentially tied to Brownian motion, but could be extended to some square-integrable martingales. As soon as the decomposition of the submartingale square of a martingale was known, this possibility was opened in complete generality (Meyer, 1963 [168]). Thus, two branches of probability were brought back together. We have already talked about martingales; we must back up to talk about the stochastic integral.

A stochastic process X can be considered a function of two variables $X(t, \omega)$ or $X_t(\omega)$, where t is time, and ω is "chance", a parameter drawn randomly from a giant "urn" Ω. The *trajectories* of the process are the functions of time $t \mapsto X_t(\omega)$. In

[5] Technical definition weakening the integrability condition for martingales.

general they are irregular functions, and we cannot define by the methods of analysis an "integral" $\int_0^t f(s)dX_s(\omega)$ for reasonable functions of time, which would be limits of "Riemann sums" on the interval $(0, t)$

$$\sum_i f(s_i)(X_{t_{i+1}} - X_{t_i}),$$

where s_i would be an arbitrary point in the interval (t_i, t_{i+1}). This is all the more impossible if the function $f(s, \omega)$ itself depends on chance. Yet Itô had studied since 1944 the case where X is Brownian motion, and f a process such that *at each instant* t, $f(t, \omega)$ *does not depend on the behavior of the Brownian motion after the instant* t, and where s_i is the *left* endpoint of the interval (t_i, t_{i+1}). In this case, we can show that the Riemann sums converge—not for each ω, but as random variables on Ω—to a quantity that we call the stochastic integral and that has all the desirable properties of an integral.

All this could seem artificial, but the discrete analog shows that it certainly is not. The sums considered in this case are of the form

$$S_n = \sum_{i=1}^{n} f_i(X_{i+1} - X_i).$$

Set $X_{i+1} - X_i = x_i$, and think of S_n as the capital (positive or negative!) of a gambler passing his time in a casino, just after the nth round of play. In this capital, f_i represents the gambler's *stake* on the ith round, whereas x_i is a normalized quantity representing the gain of a gambler who stakes just 1 franc on that round. That f_i only depends on the past then signifies that *the gambler is not a prophet*. Instead of using the language of games of chance, we can use that of financial mathematics, in which the normalized quantities X_t represent *prices*, of stocks for example—and we know this is how Brownian motion made its appearance in mathematics (Bachelier, 1900 [5]).

Another task of great practical importance involving the stochastic integral is modeling the *noise* that disturbs the evolution of a mechanical system. Here we should mention a stream parallel to the purely probabilistic developments: the efforts devoted to this problem by applied mathematicians close to engineers. We should cite the name of McShane, who devoted numerous works to various aspects of the stochastic integral [164]. The only one of these aspects that has a properly mathematical importance is Stratonovich's integral (1966, [213]), which possesses the remarkable property of being the limit of deterministic integrals when we approximate Brownian motion by differentiable curves. Whence in particular a general principle of extension from ordinary differential geometry to stochastic differential geometry.

Itô's most important contribution is not to have defined stochastic integrals—N. Wiener [226] had prepared the way for him—but to have developed their calculus (this is the famous "Itô's formula", which shows how this integral differs from the ordinary integral) and especially to have used them to develop a very complete theory of *stochastic differential equations*—in a style so luminous moreover that these old articles have not aged [122, 124, 125, 127–129].

There is still a lot to say about Itô's differential equations properly speaking, and we will mention them again in connection with stochastic geometry. Here, we will talk about generalizations of this theory.

The theory of the stochastic integral with respect to a square-integrable martingale is the subject of a still famous 1967 article by Kunita and Watanabe [150], otherwise oriented to applications to Markov processes. It is related to a 1964 article by Watanabe [225] that gives a general form to the notion of *Lévy system*, which governs the jumps of a Markov process, and to a 1965 article of Motoo and Watanabe [177]. Kunita and Watanabe's work was taken up in 1967 by Meyer [170], who added complementary ideas, such as the *square bracket* of a martingale (adapted from a notion introduced by Austin in discrete time [3]), the precise form of the dependence only on the past of the integrated process (what are now called *predictable* processes), and finally a still imperfect form of the notion of a *semimartingale* (see below).

This theory would very quickly extend to martingales that are not necessarily square-integrable, on one hand by means of the notion of a local martingale (Itô and Watanabe, 1965 [132]), which leads to the final notion of semimartingale (Doléans-Dade and Meyer, 1970 [57]), and on the other hand by means of new martingale inequalities, which will be discussed later (Millar, 1968 [173]). It would be useless to go into details. Let us consider instead the general ideas.

From a concrete point of view, a semimartingale is a process obtained by superposing a *signal*—that is to say, a process with regular trajectories, say of bounded variation, satisfying the technical condition of being *predictable*—and a *noise*, that is, a meaningless process, a pure fluctuation, modeled by a local martingale. The decomposition theorem, in its final form, says that under minimal integrability conditions (absence of very big jumps), the decomposition of the process into a sum of signal and noise is unique: knowing the law of probability we can filter the noise and recover the signal uniquely. Yet this reading of the signal depends not only on the process, but also on the underlying *filtration*, which represents the knowledge of the observer.

We can extend to all semimartingales the fundamental properties of Itô's stochastic integral, and especially, develop a unified theory of stochastic differential equations with regard to semimartingales. This was accomplished by Doléans-Dade in 1970 [58] for the *exponential equation*, which plays a big role in the statistics of processes, and by Doléans-Dade in 1976 [56] and Protter in 1977 [192] for general equations (Kazamaki had opened the way for the case of continuous trajectories in 1974 [142]). The study of stability (with respect to all parameters at the same time) was carried out in 1978 by Émery [88] and Protter [190, 191]. We can equally extend to these general equations a big part of the theory of *stochastic flows*, which developed after the "Malliavin calculus".

The theory of stochastic differential equations therefore ends up being in complete parallelism with that of ordinary differential equations. Like the latter theory, it can be approached by two types of methods: for the variants of the Lipschitzian case, Picard's method leading to results of existence and uniqueness, and for existence without uniqueness, methods of compactness of Cauchy's type. But there is a distinction specific to the probabilistic case, the distinction between uniqueness of trajectories and *uniqueness in law*. We limit ourselves here to mentioning the 1971 work of Yamada and Watanabe [230].

The possibility of bringing several distinct driving semimartingales, in other words several different "times", into a stochastic differential equation in several dimensions makes them resemble equations with total differentials more than ordinary differential equations, with geometric considerations (properties of Lie algebras) that come into Stroock and Varadhan's 1972 article [217] before reaching their full development in the "Malliavin calculus".

Let us come back for a moment to Itô's integral. We can say that it is not a "true" integral, trajectory by trajectory, but it is one in the sense of *vector measures*. M. Métivier was one of the rare probabilists to know the world of vector measures, and he devoted (with J. Pellaumail) part of his activity to the study of the stochastic integral as a vector measure with values in L^2, then in L^p, then in the non-locally convex vector space L^0 (finite random variables with convergence in measure). Métivier and Pellaumail suspected that semimartingales were *characterized* by the property of admitting a good theory of integration (see their 1977 article [165,166]). This result was established independently in 1979 by Dellacherie and Mokobodzki and by Bichteler [13], who started from the other end, that of vector measures.

It is impossible to take full account here of the abundance and variety of work related to semimartingales. The processes in this class include most of the usual processes, and they have very good stability properties. In particular, if we replace a law on the space Ω by an equivalent law (one with the same null sets) without changing the filtration, the semimartingales for the two laws are the same (whereas the decomposition into "signal plus noise" changes). This remarkable theorem is due, in its final form, to Jacod and Mémin in 1976 [134], but it has a long history, which brings it back, in the particular case of Brownian motion, to *Girsanov's theorem* (1960 [107]; see also [35–38]). It opens the way to a general form of statistics for stochastic processes. Statistics seeks to determine the law of a random phenomenon from observations when we do not know the law a priori. The search for properties of processes that are invariant under changes in the law is therefore very important. See for example [135,157].

The rapid evolution of ideas in probability resulted—this a general phenomenon in mathematics—in the multiplication of informal publications, such as the volumes of the Brelot-Choquet-Deny seminar on potential theory. The advent of Springer's *Lecture Notes* series led to the international distribution of publications of this type, which were at first "in house". In probability, we find the series *Séminaires de Probabilités de Strasbourg* (1967), then the lecture notes of l'École d'Été de St Flour (1970), and finally the *Seminar on Stochastic Processes* in the United States (1981). See also [154] for an example of transfer to Italy.

4.2 Markov Processes

During this whole period, the general theory of Markov processes remained extremely active, but it was no longer the dominant subject in probability as it had been in the preceding period.

We can distinguish a few themes particularly studied.

In the beginning of the theory of Markov processes, various classes of processes were introduced axiomatically, including Dynkin's "standard processes" and "Hunt" processes, which allowed Hunt to develop probabilistic potential theory. A 1970 article by C. T. Shih [201] was at the origin of a movement of ideas that identified a class of Markov processes, the *right processes*, that possess remarkable stability properties. We will limit ourselves here to mentioning the essential role of Ray's compactification in these questions, and to referring to two books of synthesis: Getoor in 1975 [102] and Sharpe in 1988 [200].

A second important theme is the *duality* of Markov processes with respect to a measure. Here we start not from a given pair of Markov semigroups in duality with respect to a measure, but rather from a single semigroup, for which we want to build a dual semigroup. Chung and Walsh, in 1969, contributed the most important article on this question [47].

For understanding duality, Mitro's articles (1970 [175]) had a great impact. They give a construction for adjoining the forward trajectories of one of the Markov processes with the backward trajectories of its dual process, in a way that makes it into a *stationary* process arising at a random moment (possibly $-\infty$) and disappearing at a random moment (possibly $+\infty$). In fact, all of this had already been discovered, in a more general form, in two articles by Dynkin in 1973 [83] and Kuznetsov in 1974 [152], whose discovery (after Dynkin's arrival in the USA!) generated a good number of papers. The importance of these results for potential theory (excessive measures) was progressively understood; see Fitzsimmons and Maisonneuve's 1986 article [98] and Getoor's 1990 book [103]. The connection with the strange processes constructed by Hunt in 1960 [121], completing the understanding of this article, was given by Fitzsimmons in 1988 [97].

It is impossible to do more here than name other important subjects: "Lévy systems" of general Markov processes [9]; local times of Markov processes; various transformations preserving the Markov property (an essential element of Dynkin's program from the beginning). It is better to devote a little time to a particularly fascinating theme, excursion theory.

The original idea of excursion theory is to study the behavior of a Markov process "around" a fixed state a. The simplest example is that of discrete time Markov chains; there the process's successive passage times in a state constitute what is called a *renewal process*, and the structure of these processes (which have countless applications) has long been known. Between successive passages through a, the chain makes "excursions", which (in the most interesting case, where the chain returns to a an infinite number of times) are independent and have the same law. In continuous time, the situation is much more complicated. The model is the in-depth study Lévy had made of the *passages of Brownian motion at* 0. The set of these passages is a perfect set of null measure, riddled with small holes during which Brownian motion makes its excursions. How to enumerate them, how to compare them with one another, in what sense to consider them as independent and equally distributed? The problem arises in fact for all Markov processes (and it is especially interesting for continuous-time Markov chains with little regularity, a case studied by Chung). It is even more difficult to describe when the Markov process is not studied in the neighborhood of a point, but in the neighborhood of a whole "boundary", because then the impact point moves on the boundary, and we must describe how.

Concerning encounters with a single state, the axiomatic characterization of the random sets that can be interpreted as the moments of a Markov process's return to a fixed state was the work (after Kingman's preliminary studies) of Krylov and Yushkevich in 1965 [149], in a difficult article, taken up and greatly simplified by Hoffmann-Jørgensen in 1969 [117]. On excursions themselves, the new idea that clarified the problem came from Itô in 1972 [130], certainly one of the great conceptual achievements of probability, because the excursion, which is a trajectory, is treated as a point, and the succession of excursions is treated as a new random process with a simple description. See also [44]. Finally, on boundary problems, we must limit ourselves to citing a remarkable 1971 article by Dynkin [82], which has been read too little in the West (but see El Karoui and Reinhard 1975 [87]), and Maisonneuve's work 1974 [159].

We must finally mention an important development for the future: the construction of reversible Markov processes (also called symmetric) by the Hilbertian method of *Dirichlet forms*. Introduced in potential theory by Beurling and Deny in 1959 [12], this method came into probability with Silverstein in 1974 [203] and with the 1975 Japanese edition of Fukushima's book [99]. It has become one of the most powerful tools for building Markov processes in infinite dimensions, a subject very much alive because of its possible applications to physics.

4.3 General Theory of Processes

The "general theory of processes" is the development of one of the subjects initiated by Doob's book, that of *increasing families of sigma-algebras* (now called *filtrations*) and of stopping times. This development took place in constant interaction with Markov process theory on one hand and martingale theory on the other, and the division is therefore somewhat artificial.

The beginning of the general theory of processes was dominated by the notion of stopping times, and in particular by numerous results around the strong Markov property, and around the "slightly less strong" variants discovered in Markov chain theory by Chung [43]. Numerous results on filtrations are given as lemmas in articles on Markov processes, Dynkin's in particular. One of the first articles dedicated entirely to the general theory is that of Chung and Doob [45, 1965], and the first—or only—book completely dedicated to it is that of Dellacherie [53, 1972], which introduces the notion of the *dual projection* of an increasing process, particularly useful for applications. Another motivation for the study of the general theory of processes is provided by the study of transformations of Markov processes. See also [54].

A chapter of the general theory of processes that deserves mention, because it is particularly attractive from the viewpoint of the philosophy of probability, is that of *enlargement of filtrations*, whose starting point is a theorem proven independently, in 1978, by Barlow [7] and by Yor and de Sam Lazaro [231], and whose strongest results are due to Jeulin in 1980 [137]. It can be presented as follows. Doob's fundamental idea in martingale theory had been to express mathematically the fact that players are not prophets, using the notion of *filtration*. Can we also describe mathematically what a prophet is? Needless to say, we are concerned with a mathematical abstraction, analogous to conditioning by the value of a random variable, which does not suppose that we really know this value. We describe "the universe plus the prophet" by a second filtration, bigger than the first since we have more knowledge at every moment. The theorems that we establish take therefore the following form: martingales of the small filtration (or only a few of them) become *semi*martingales in the big filtration, which we know how to decompose into "signal plus noise".

The set of all these topics—martingale inequalities, general theory, stochastic integral, enlargement—constitute what we call *Stochastic Calculus*. But the tree carries yet more branches; let us mention some. The use of martingale methods to deal with problems of weak convergence of process laws, a subject illustrated by Aldous' long article in 1978 [2] published only in part and followed by numerous authors; the generalization of martingale convergence theorems in the form of *asymptotic martingales* or *amarts*, in discrete or continuous time (the reader wanting to pursue this question can consult Edgar and Sucheston 1992 [85,86]); the extension of known results on martingales to certain processes in *multidimensional* time (Cairoli 1970 [33], Cairoli and Walsh 1975 [34]). Finally, let us note the development of a "prediction theory" by F. Knight in 1979 and 1992 [144,145]), which shows the tight links uniting the most general possible theory of processes with Markov processes.

Stochastic calculus, despite its relatively abstract character, rapidly found applications. The first ones came from *electrical engineering* laboratories (transmissions of signals in the presence of noise). But the most recent and widest ones concern "financial mathematics", thus going back to the very sources of Brownian motion theory (Bachelier 1900 [5]). These mathematical problems even resuscitated, in the 1990s, branches of stochastic calculus that seemed asleep since 1970.

4.4 Inequalities of Martingales and Analysis

I do not pretend to deal here with all the relations between probability and analysis (harmonic analysis, Banach spaces, fractals…), a subject on which I lack competence.

Relations between martingales and analysis were already present in the work of Doob, who applied martingale convergence theorems to derivation theory on one hand, and to the behavior of harmonic functions at the boundary on the other. This subject remains partly open, by the way, especially with respect to its extensions to multivariable complex functions. But the most fruitful interactions begin with the discovery of *Burkholder's inequalities* (1964, 1966 [27,28]). These inequalities establish an equivalence in norm in L^p, for $p > 1$, between two random variables associated with the martingale: one is the trajectory's absolute upper bound, and the other a Hilbertian-type quantity (the square root of a sum of squares), easy to define in discrete time, more delicate in continuous time. These equivalences in norm were extended by Burkholder and Gundy in 1970 [31] to spaces other than L^p, and by Davis to L^1 norms, thus initiating the theory of the H^1 space of martingales. For a synthesis of various types of inequalities, see Lenglart, Lépingle, and Pratelli in 1980 [153]; see also [29,30,32,100].

On the other hand, the period around 1970 was marked in analysis by a return to direct methods of real variables—rather than abstract methods of functional analysis or classical methods of complex analysis (inherently limited to the plane). The theory of singular integrals and the theory of $H^p(\mathbb{R}^n)$ spaces were developing very fast. Equivalencies between a "maximum" norm and various "quadratic" norms played an important role in these theories. In 1970 Stein's treatise appeared [211], accompanied by a small volume on Littlewood-Paley's theory [212], which appealed directly to martingale methods. The development of H^1 space theory (Fefferman and Stein 1972 [91]), H^p spaces for $p < 1$, the duality between H^1 and BMO (John and Nirenberg 1961 [138], Fefferman 1971 [90]), the "atomic" approach to H^p theories (Coifman 1974 [50]), all this would have consequences and parallels in probability, in the form of H^p spaces of martingales imported by Herz in discrete time, then extended to continuous time [112,113]. The H^1 space in particular took on great importance in stochastic integral theory.

Another aspect, that of Littlewood-Paley theory: in its classical form, it concerns harmonic or holomorphic functions in the unit disc, but it is also used in the half-space $\mathbb{R} \times \mathbb{R}_+$, then extends to $\mathbb{R}^n \times \mathbb{R}_+$, and finally (Stein made the step) to $E \times \mathbb{R}_+$, where E is an abstract space with a Markovian semigroup (P_t). We can therefore introduce martingale methods (Meyer 1976 [171], Varopoulos 1980 [222]). These probabilistic methods have applications in group theory (Varopoulos), and to the analysis of infinite-dimensional semigroups, where they allow us to define notions that correspond to Ricci's curvature in Riemannian geometry, where the semigroup is that of Brownian motion. This is a large subject, which we cannot take up here, but for which we will direct the reader to the bibliography of Bakry's 1994 lectures [6].

This subject is related to *hypercontractivity*, for which on the contrary it is probability that has influenced analysis: the starting point is Nelson's 1973 probabilistic proof (through Wiener chaos), motivated by quantum field theory [181]. Gross' famous 1975 article [109] gave the problem its definitive status, and it is still a subject very much alive.

4.5 Martingale Problems

The universality of martingales during the period we are studying translates into a new notion, that of *martingale problems*, introduced in diffusion theory by Stroock and Varadhan in 1969 [215,216], then used in many other areas, like that of point processes (see Jacod 1979 [133,218]).

Stroock and Varadhan's idea consists in characterizing the law of a stochastic process by a family of processes that we require to be martingales (later local martingales). In the case of diffusions (or more generally Markov processes), these processes are constructed in a simple manner from the infinitesimal generator. The unknown of the martingale problem is thus a probability law, for which we must discuss existence and uniqueness—and for existence, it is quite natural to use a method of *weak convergence*. Stroock and Varadhan use this method to handle the problem of diffusions with coefficients that are assumed only to be continuous, which seems to resist functional analysis methods.

The research undertaken in order to apply the method is even more important than the method itself: criteria of weak compactness criteria using "local characteristics" of semimartingales, problems of constructing *all* the martingales from a given family of martingales by means of stochastic integrals. Here again, space is insufficient to develop these themes, which have a considerable practical importance. Let us only mention articles by Jacod and Yor 1977 [136] and Yor and de Sam Lazaro 1978 [231].

4.6 "Stochastic Mechanics"

This stream of ideas has been relatively narrow but constant in volume of publication. Since the beginning of Quantum Mechanics, the Copenhagen interpretation has encountered opponents, some of whom were determinists, while others sought a classical probabilistic interpretation. We can trace the latter tendency to Schrödinger himself [196]. The majority of physicists have ignored these moods, but in 1967 Nelson [180] made a fascinating presentation of them for probabilists: To any wave function $\psi(t, x)$ of quantum mechanics, solution of a classical Schrödinger equation, Nelson associates a natural diffusion admitting at each moment the probability density $|\psi(t, x)|^2$, predicted by the Copenhagen interpretation. Loosely speaking, the wave function, which is complex and satisfies a linear equation, "codes" two transition functions, one forward, the other backward, which satisfy two coupled nonlinear equations. In 1980 [179], Nagasawa managed to present Nelson's equations, no longer as another interpretation of quantum mechanics, but as general model of equilibrium of "populations" of similar individuals, in which each individual interacts with the population density, that is with its own probability of presence. Thus it became possible to be interested in these equations without taking sides in a theological quarrel. On the other hand, Nelson's book was incomplete from a mathematical point of view: in Nelson's two diffusion generators, the first order term explodes on "the nodal set" where the wave function becomes null, and Nelson could treat rigorously only the case where this set is empty. It is only in the works of Nelson 1985 [184], Zheng 1985 [234] and especially Carlen 1984 [39] that this difficulty was resolved in a satisfying manner. Since then, the publication of books on these questions has never ceased, and we will not carry out an inventory.

4.7 Relations to Physics

Yet Nelson's book was to have a completely different legacy, in the direction of "orthodox" quantum physics. This is a matter of Euclidean field theory, which we can try to describe this way: Despite its extraordinary practical successes, relativistic quantum field theory found itself in a state of intellectual confusion. One of the procedures considered for remedying this consisted of constructing "models" of nontrivial quantum fields satisfying a certain number of axioms that are natural from a physical point of view. Nelson [182, 1973] showed how, by an analytic continuation, these constructions can be reduced to constructions of *probability measures* on a space of distributions $S'(\mathbb{R}^n)$ having a Euclidean invariance (rather than relativistic invariance) property and a form of the Markov property. This is a subject on which we will say very little (by pure ignorance) except:

1. The method succeeds perfectly in dimension 2, where it has stimulated in an extraordinary way the close study of planar Brownian motion, and especially the study of multiple points of the Brownian curve. Let us mention in passing that this is one of the subjects to which Dynkin devoted himself after his departure from Russia in 1977.
2. The problems tied to quantum field theory have also motivated much research on the construction of measures in infinite dimensions.

Let us mention the 1974 and 1979 books by Simon [204,205] and the 1981 book by Glimm and Jaffe [108]. The introduction to Simon's 1974 book contains one of the most beautiful tributes paid to probability by a non-probabilist.

5 After 1980

As with Markov processes, the stochastic calculus also began to fade, following the same pattern: the trunk does not continue to develop, a few branches stay very alive. In the particular case of stochastic calculus, we see it descend from the sky after 1980, reflect itself in remarkable textbooks, become a concrete working tool that allows us to resume Paul Lévy's work and calculate countless laws of processes. The general direction being less clear, I will try to cite a few important directions, in which I myself became more or less seriously interested.

5.1 The "Malliavin Calculus"

The probabilistic theory of diffusions has always appealed to theorems borrowed from the theory of partial differential equations, which permit us to assert that the transition function of a diffusion with a given generator has a sufficiently regular density. This was known in the case of elliptic generators, and also in certain degenerate cases (hypoelliptic), thanks to a famous theorem of Hörmander. This did not lack concrete applications, in the study of stochastic control problems for example.

In 1976, P. Malliavin presented at a conference in Kyoto a *purely probabilistic* method to establish the existence of very regular densities for solutions of certain stochastic differential equations with C^∞ coefficients [160,161]. This was extraordinarily original work, and it took the probabilistic milieu many years to assimilate it. Before talking about its content, we must say a word about the "tradition" in which this article appeared, at the juncture of many streams about which we have said nothing so far.

Malliavin was known foremost as an analyst and a geometer, but in probability he was self-taught, educated by reading Itô's and Doob's books. He became interested in problems of vanishing cohomology—i.e., the nonexistence of nontrivial harmonic forms of certain degrees on certain compact manifolds. Now the nonexistence of nontrivial harmonic *functions* is related to the asymptotic behavior of Brownian motion on the manifold. Can we do the same for forms? This question had been taken up by Bochner [23,24] and Yano. From a probabilistic point of view, the problem is related to two others: the construction of Brownian motion on the manifold (i.e., construction of the diffusion whose generator is the Laplace-Beltrami operator), and how we can "follow" a differential form along Brownian motion trajectories. This takes us into the vast field of *stochastic differential geometry*.

Starting in 1963, Itô had studied the parallel transport of vectors along Brownian motion trajectories, a problem taken up by Dynkin in 1968 [81]. The generator of this operation on forms is nevertheless not the most interesting Laplacian (de Rham), but another Laplacian called *horizontal*, which differs from the former by a first-order term. After Itô and Dynkin, we can mention the works of the English school (Eells, Elworthy), then those of Malliavin himself. One of the results of these efforts was a probabilistic construction of Brownian motion on Riemannian manifolds by stochastic differential equations, without concatenation, by lifting to the frame bundle. Malliavin had all these techniques, unknown to the majority of probabilists.

Second ingredient, the stochastic "calculus of variations", that is the variation of solutions of the equation as a function of the initial conditions. Here again, on this widely studied question, Malliavin brought a new tool (although it is found in part in a little known 1961 article of Blagoveshchenskii and Freidlin [19]): an Itô stochastic differential equation with C^∞ coefficients on $\mathbb{R}^n$ defines a "flow of C^∞ diffeomorphisms" on $\mathbb{R}^n$. This was sure to generate plenty of work on stochastic flows.

Third ingredient, Wiener's 1938 [227], or rather Itô's 1951 [126] use of chaos expansion to introduce a Laplacian in infinite dimensions, the *Ornstein-Uhlenbeck Laplacian*, which is a self-adjoint operator relative to Wiener measure, and in relation to which Malliavin defines *Sobolev spaces in infinite dimensions*, his principal tool being an integration by parts formula for Wiener measure. Here again, he had forerunners: the L^2 space of the Brownian measure is isomorphic, when we use the chaos expansion, to the physicists' *Fock space* (Segal 1956 [199, 1956]), one of the basic objects of quantum field theory, and the idea of defining weakly differentiable functions and Sobolev spaces on Fock space had been widely studied outside of probability (in particular, there were rich works by P. Krée on this subject). But this had all remained rather abstract, whereas Malliavin made a very efficient tool out of it. As for the way Malliavin put these various elements together to establish Hörmander's theorem, it was simultaneously a mathematical tour de force and, for the probabilistic public, a shower of novelties to assimilate.

In these conditions, it is fair to mention the 1981 work of Stroock [214], who (aside from his own original contributions) put all this within reach of probabilists. The 1980 Durham colloquium [229], with its introduction by Williams, also played a big role in the diffusion of the "Malliavin calculus".

Among developments that followed, we will mention only (for lack of space) the work in the 1980s of Bismut, who modified and completed Malliavin's tools, established the complete form of Hörmander's theorem, extended it to diffusions with boundary conditions, carried out a fusion of Malliavin's calculus and large deviations methods—and especially, found a new outlet for them with his probabilistic proof of the Atiyah-Singer *index theorem* [15–18]. But the strongest influence of the "Malliavin calculus" on probability properly speaking comes no doubt from a relatively secondary aspect of his technique: the use of Wiener chaos and the Ornstein-Uhlenbeck process on Wiener space. That has attracted great interest in infinite-dimensional analysis, coming back to certain concerns of theoretical physics.

Perhaps we should mention another entirely different probabilistic approach to the existence of densities: that of Krylov in 1973 [148]. Here there are deep results that remain isolated.

5.2 Stochastic Differential Geometry

This subject is prior to the "Malliavin calculus", but it profited from its diffusion. Here is a sample of problems dealt with during this period: How can we read the local geometry of a Riemannian manifold from the behavior of its Brownian motion over short periods of time? Or conversely, its global behavior from the asymptotic behavior of the Brownian? What is the behavior of trajectories of a process whose generator is the sum of a first-order term and a small second-order term? This last problem is related to the quasi-classical approximation of quantum mechanics when Planck's constant "approaches zero", and to large deviation problems.

Another aspect of stochastic geometry, the study of *semimartingales on manifolds*, inaugurated by L. Schwartz in 1980 [197], and relying on the fact that the class of semimartingales is invariant under class C^2 maps. It is possible in particular to define *continuous martingales* with values in manifolds (and of which the Brownian motion of a Riemannian manifold is an example). Here again, we encounter an extension of the relation between Brownian motion and harmonic functions, in which the Brownian motion takes place in a curved space, and the harmonic functions become harmonic maps between Riemannian manifolds. For lack of space, let us refer to Émery's 1989 book [89].

5.3 Distributions and White Noise

The popularity of the "Malliavin calculus" helped bring back into mainstream probability a subject that had diverged from it quite early. This is the subject of *distributions in infinite dimensions*, whose history will force us to back up.

First, there are the ideas of Gelfand (1955 [101] and Minlos 1959 [174] on random distributions, according to which the most natural way to consider the trajectories of a stochastic process is to regard them as distributions. The law of the stochastic process is then a measure on a space of distributions—and the main spaces of distributions being nuclear, they possess excellent measure-theoretic properties. We go from there to the study of a particularly interesting random distribution, that of *white noise*, which is the derivative of Brownian motion in the sense of distributions, developed starting in 1967 by Hida (see [114] and Hida's 1980 book [115]). Here the essential point is the expansion of functionals of Brownian motion in Wiener chaos, and the definition of classes of *generalized functionals* by non-convergent expansions. Hida's distributions are of interest to physicists, because they provide a rigorous way to understand the analogies between Brownian motion and the Feynman integral, the latter appearing as a distribution on Wiener space. All this has produced a sustained stream of publication, but a little on the fringe of the main streams of probability. See also [49].

The "Malliavin calculus" renewed interest in these problems by introducing a whole family of Sobolev spaces of differentiable functionals, whose duals are quite naturally distribution spaces. This point of view is due primarily to Ikeda and Watanabe. We will not go into details here, but "Wiener analysis" is currently a flourishing branch.

5.4 Large Deviations

I will do nothing but cite Schilder 1966 [195] and the fundamental work of Donsker and Varadhan 1976 [59]. This subject deserves to be treated separately.

5.5 Noncommutative Probability

The axioms of Quantum Mechanics developed by von Neumann in 1932 (before Kolmogorov's axioms!) were in fact—if we exclude the problem of the quantification of classical mechanics—probability axioms, where random variables are called self-adjoint operators, probability laws are called positive self-adjoint operators of unit trace, etc. Later, there appeared the possibility of addressing probability in a more general setting, C^*-algebras for example. Quite naturally, one sought to develop a noncommutative measure theory, at least in the relatively simple case of a tracial law (Segal 1953 [198], Nelson 1974 [183]). Again quite naturally, one posed problems of probabilistic nature, like the validity of the martingale convergence theorem in von Neumann algebras. But the absence of interesting examples, and the impossibility of defining conditioning in sufficient generality, left this stream of research marginal for a long time—among probabilists, because physicists needed models of *quantum noise*. It is just that mathematicians are hardly interested in anything but fundamental physics, whereas here it is rather a matter of applied physics (quantum optics), and so the fields of research remained nearly disjoint. We can point to Cushen and Hudson's definition of a noncommutative Brownian motion in 1971 [51], and to Accardi, Frigerio and Lewis's 1982 article on the general definition of noncommutative stochastic processes [1]. A good reference for this period is Davies' 1976 book [52].

Yet the situation changed completely with the development of a noncommutative form of stochastic calculus, with Streater's and colleague's articles on fermionic Brownian motion and the corresponding theory of the stochastic integral (see Barnett, Streater, and Wilde 1982 [8]), and especially Hudson and Parthasarathy's 1984 article on bosonic Brownian motion [118]. Independently of the value of this article, the reason it had so much impact is that it is accessible: unlike the others, it does not require a heavy background in functional analysis, and it connects much more directly to the classical Itô calculus and to the theory of Wiener chaos.

We will not comment further on this recent trend, except to mention that it had "fallout" in classical probability, by raising beautiful problems about martingales that give rise to a chaotic representation ("Azema's martingales" for example). A good reference is Parthasarathy's 1992 book [188].

5.6 Omissions

The theory of stochastic processes is not all of probability, and I am far from having taken up all aspects of the theory of stochastic processes, or even merely Markov process theory or martingale theory. I had to omit not only works on which I was poorly informed, but also works I know well and I admire. I hope the reader has taken pleasure in the preceding account, and I ask him to be indulgent.

References

1. Accardi, L., Frigerio, A., Lewis, J.T.: Quantum stochastic processes. Publications of the Research Institute for Mathematical Sciences, Kyoto **18**(1), 97–133 (1982)
2. Aldous, D.: Stopping times and tightness. The Annals of Probability **6**(2), 335–340 (1978)
3. Austin, D.G.: A sample function property of martingales. Annals of Mathematical Statistics **37**(5), 1396–1397 (1966)
4. Azéma, J., Yor, M.: Temps locaux. No. 52–53 in Astérisque. Société mathématique de France (1978)
5. Bachelier, L.: Théorie de la spéculation. Annales de l'École Normale Supérieure **17**, 21–86 (1900)
6. Bakry, D.: L'hypercontractivité et son utilisation en théorie des semigroupes. In: Lectures on probability theory, LN 1581, pp. 1–114. Springer (1994)
7. Barlow, M.T.: Study of a filtration expanded to include an honest time. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **44**(4), 307–323 (1978)

8. Barnett, C., Streater, R., Wilde, I.: The Itô-Clifford integral. Journal of Functional Analysis **48**(2), 172–212 (1982)
9. Benveniste, A., Jacod, J.: Systèmes de Lévy des processus de Markov. Inventiones Mathematicae **21**(3), 183–198 (1973)
10. Bernstein, S.: Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. Mathematische Annalen **97**, 1–59 (1927)
11. Bernstein, S.: Equations differentielles stochastiques. Act. Sci. Ind. **738**, 5–31 (1938). Conf. Intern. Sci. Math. Univ. Genève, Hermann
12. Beurling, A., Deny, J.: Dirichlet spaces. Proceedings of the National Academy of Sciences **45**(2), 208–215 (1959)
13. Bichteler, K.: Stochastic integration and L^p-theory of semimartingales. The Annals of Probability **9**(1), 49–89 (1981)
14. Billingsley, P.: Convergence of Probability Measures. Wiley, New York (1968)
15. Bismut, J.M.: Martingales, the Malliavin calculus and hypoellipticity under general Hörmander's conditions. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **56**(4), 469–505 (1981)
16. Bismut, J.M.: The Atiyah-Singer theorems: A probabilistic approach. I. The index theorem. Journal of Functional Analysis **57**(1), 56–99 (1984)
17. Bismut, J.M.: Large Deviations and the Malliavin Calculus. Birkhäuser (1984)
18. Bismut, J.M.: Last exit decompositions and regularity at the boundary of transition probabilities. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **69**(1), 65–98 (1985)
19. Blagoveshchenskii, Y.N., Freidlin, M.I.: Some properties of diffusion processes depending on a parameter. Doklady Akademii Nauk **138**(3), 508–511 (1961)
20. Bliedtner, J., Hansen, W.: Potential Theory: An Analytic and Probabilistic Approach to Balayage. Springer (1986)
21. Blumenthal, R.M.: An extended Markov property. Transactions of the American Mathematical Society **85**(1), 52–72 (1957)
22. Blumenthal, R.M., Getoor, R.K.: Markov processes and potential theory. Academic Press (1968)
23. Bochner, S.: Diffusion equation and stochastic processes. Proceedings of the National Academy of Sciences **35**(7), 368–370 (1949)
24. Bochner, S.: Harmonic analysis and the theory of probability. University of California (1955)
25. Brelot, M.: Sur le principe des singularités positives et la topologie de R. S. Martin. Annales de l'université de Grenoble. Nouvelle série. Section sciences mathématiques et physiques **23**, 113–138 (1948)
26. Brelot, M.: Le problème de Dirichlet, axiomatique et frontière de Martin. Journal de Mathématiques Pures et Appliquées **35**, 297–335 (1956)
27. Burkholder, D.L.: Maximal inequalities as necessary conditions for almost everywhere convergence. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **3**(1), 75–88 (1964)
28. Burkholder, D.L.: Martingale transforms. The Annals of Mathematical Statistics **37**(6), 1494–1504 (1966)
29. Burkholder, D.L.: Distribution function inequalities for martingales. The Annals of Probability **1**(1), 19–42 (1973)
30. Burkholder, D.L., Davis, B.J., Gundy, R.F.: Integral inequalities for convex functions of operators on martingales. In: Sixth Berkeley Symposium on Mathematical Statistics and Probability, vol. 2, pp. 223–240. University of California (1971)
31. Burkholder, D.L., Gundy, R.F.: Extrapolation and interpolation of quasi-linear operators on martingales. Acta Mathematica **124**, 249–304 (1970)
32. Burkholder, D.L., Gundy, R.F.: Distribution function inequalities for the area integral. Studia Mathematica **44**(6), 527–544 (1972)
33. Cairoli, R.: Une inégalité pour martingales à indices multiples et ses applications. Séminaire de probabilités de Strasbourg **4**, 1–27 (1970)
34. Cairoli, R., Walsh, J.B.: Stochastic integrals in the plane. Acta Mathematica **134**, 111–183 (1975)

35. Cameron, R.H., Martin, W.T.: Transformations of Weiner integrals under translations. Annals of Mathematics pp. 386–396 (1944)
36. Cameron, R.H., Martin, W.T.: Fourier-Wiener transforms of analytic functionals. Duke Mathematical Journal **12**(3), 489–507 (1945)
37. Cameron, R.H., Martin, W.T.: Fourier-Wiener transforms of functionals belonging to L_2 over the space C. Duke Mathematical Journal **14**(1), 99–107 (1947)
38. Cameron, R.H., Martin, W.T.: The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals. Annals of Mathematics **48**(2), 385–392 (1947)
39. Carlen, E.A.: Conservative diffusions. Communications in Mathematical Physics **94**(3), 293–315 (1984)
40. Cartan, H.: Théorie du potentiel newtonien: énergie, capacité, suites de potentiels. Bulletin de la Société Mathématique de France **73**, 74–106 (1945)
41. Cartan, H.: Théorie générale du balayage en potentiel newtonien. Annales de l'université de Grenoble. Nouvelle série. Section sciences mathématiques et physiques **22**, 221–280 (1946)
42. Choquet, G.: Theory of capacities. Annales de l'Institut Fourier **5**, 131–295 (1954)
43. Chung, K.L.: Markov Chains with Stationary Transition Probabilities. Springer (1960)
44. Chung, K.L.: Excursions in Brownian motion. Arkiv för Matematik **14**(1), 155–177 (1976)
45. Chung, K.L., Doob, J.L.: Fields, optionality and measurability. American Journal of Mathematics **87**(2), 397–424 (1965)
46. Chung, K.L., Feller, W.: On fluctuations in coin-tossing. Proceedings of the National Academy of Sciences **35**, 605–608 (1949)
47. Chung, K.L., Walsh, J.B.: To reverse a Markov process. Acta Mathematica **123**, 225–251 (1969)
48. Ciesielski, Z., Taylor, S.J.: First passage times and sojourn times for Brownian motion in space and the exact Hausdorff measure of the sample path. Transactions of the American Mathematical Society **103**(3), 434–450 (1962)
49. Clark, J.M.C.: The representation of functionals of Brownian motion by stochastic integrals. The Annals of Mathematical Statistics **41**(4) (1970)
50. Coifman, R.R.: A real variable characterization of H^p. Studia Mathematica **51**(3), 269–274 (1974)
51. Cushen, C.D., Hudson, R.L.: A quantum-mechanical central limit theorem. Journal of Applied Probability **8**(3), 454–469 (1971)
52. Davies, E.B.: Quantum Theory of Open Systems. Academic Press (1976)
53. Dellacherie, C.: Capacités et Processus Stochastiques. Springer (1972)
54. Dellacherie, C., Meyer, P.A.: Probabilités et potentiel, Chapitres I à IV. Hermann (1975)
55. Deny, J.: Les potentiels d'énergie finie. Acta Mathematica **82**, 107–183 (1950)
56. Doléans-Dade, C.: On the existence and unicity of solutions of stochastic integral equations. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **36**(2), 93–101 (1976)
57. Doléans-Dade, C., Meyer, P.A.: Intégrales stochastiques par rapport aux martingales locales. Séminaire de probabilités de Strasbourg **4**, 77–107 (1970)
58. Doléans-Dade, C.: Quelques applications de la formule de changement de variables pour les semimartingales. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **16**(3), 181–194 (1970)
59. Donsker, M.D., Varadhan, S.R.S.: Asymptotic evaluation of certain Markov process expectations for large time. Communications on Pure and Applied Mathematics **28**, 1–47, 279–301, 389–461 (1976)
60. Doob, J.L.: Stochastic processes depending on a continuous parameter. Transactions of the American Mathematical Society **42**(1), 107–140 (1937)
61. Doob, J.L.: Regularity properties of certain families of chance variables. Transactions of the American Mathematical Society **47**(3), 455–486 (1940)
62. Doob, J.L.: Markoff chains—denumerable case. Transactions of the American Mathematical Society **58**(3), 455–473 (1945)
63. Doob, J.L.: Continuous parameter martingales. In: Second Berkeley Symposium on Mathematical Statistics and Probability, pp. 269–177. University of California (1951)

64. Doob, J.L.: Stochastic Processes. Wiley, New York (1953)
65. Doob, J.L.: Semimartingales and subharmonic functions. Transactions of the American Mathematical Society **77**(1), 86–121 (1954)
66. Doob, J.L.: A probability approach to the heat equation. Transactions of the American Mathematical Society **80**(1), 216–280 (1955)
67. Doob, J.L.: Probability methods applied to the first boundary value problem. In: Third Berkeley Symposium on Mathematical Statistics and Probability, vol. 2, pp. 49–80. University of California (1956)
68. Doob, J.L.: Brownian motion on a Green space. Theory of Probability & Its Applications **2**(1), 1–30 (1957)
69. Doob, J.L.: Conditional Brownian motion and the boundary limits of harmonic functions. Bulletin de la Société Mathématique de France **85**, 431–458 (1957)
70. Doob, J.L.: Discrete potential theory and boundaries. Journal of Mathematics and Mechanics **8**(3), 433–458 (1959)
71. Doob, J.L.: Boundary properties of functions with finite Dirichlet integrals. Annales de l'Institut Fourier **12**, 573–621 (1962)
72. Doob, J.L.: Classical Potential Theory and its Probabilistic Counterpart. Springer (1984)
73. Dvoretsky, A., Erdős, P., Kakutani, S.: Nonincrease everywhere of the Brownian motion process. In: Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 4, pp. 103–116. University of California (1961)
74. Dvoretzky, A., Erdős, P., Kakutani, S.: Double points of paths of Brownian motion in n-space. Acta Sci. Math. Szeged **12**, 75–81 (1950)
75. Dvoretzky, A., Erdős, P., Kakutani, S.: Multiple points of paths of Brownian motion in the plane. Bulletin of the Research Council of Israel **3**, 364–371 (1954)
76. Dvoretzky, A., Erdős, P., Kakutani, S., Taylor, S.J.: Triple points of Brownian paths in 3-space. Mathematical Proceedings of the Cambridge Philosophical Society **53**(4), 856–862 (1957)
77. Dynkin, E.B.: Criteria of continuity and of absence of discontinuities of the second kind for trajectories of a Markov random process. Izv. Akad. Nauk Ser. Mat. **16**(6), 563–572 (1952)
78. Dynkin, E.B.: Foundations of the theory of Markov processes. Izd. Fiz.-Mat., Moscow (1959)
79. Dynkin, E.B.: Natural topology and excessive functions connected with a Markov process. Doklady Akademii Nauk **127**(1), 17–19 (1959)
80. Dynkin, E.B.: Markov Processes (in Russian). Izdat. Fiz.-Mat., Moscow (1963)
81. Dynkin, E.B.: Diffusion of tensors. Doklady Akademii Nauk **179**(6), 1264–1267 (1968)
82. Dynkin, E.B.: Wanderings of a Markov process. Theory of Probability & Its Applications **16**(3), 401–428 (1971)
83. Dynkin, E.B.: Regular Markov processes. Russian Mathematical Surveys **28**(2), 33–64 (1973)
84. Dynkin, E.B., Yushkevich, A.A.: Markov Processes: Theorems and Problems. Nauka, Moscow (1967)
85. Edgar, G.A., Sucheston, L.: Amarts: a class of asymptotic martingales. Journal of Multivariate Analysis **6**, 193–221, 572–591 (1976)
86. Edgar, G.A., Sucheston, L.: Stopping times and directed processes. Cambridge (1992)
87. El Karoui, N., Reinhard, H.: Compactification et balayage de processus droits. No. 21 in Astérisque. Société mathématique de France (1975)
88. Émery, M.: Stabilité des solutions des équations différentielles stochastiques application aux intégrales multiplicatives stochastiques. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **41**(3), 241–262 (1978)
89. Émery, M.: Stochastic calculus in manifolds. Springer (1989)
90. Fefferman, C.: Characterizations of bounded mean oscillation. Bulletin of the American Mathematical Society **77**(4), 587–588 (1971)
91. Fefferman, C., Stein, E.M.: H^p spaces of several variables. Acta Mathematica **129**, 137–193 (1972)
92. Feller, W.: An Introduction to Probability Theory and its Applications. Wiley (1950)
93. Feller, W.: Diffusion processes in one dimension. Transactions of the American Mathematical Society **77**(1), 1–31 (1954)

94. Feller, W.: The general diffusion operator and positivity preserving semi-groups in one dimension. Annals of Mathematics **60**(3), 417–436 (1954)
95. Feller, W.: An Introduction to Probability Theory and its Applications II. Wiley (1966)
96. Fisk, D.L.: Quasi-martingales. Transactions of the American Mathematical Society **120**(3), 369–389 (1965)
97. Fitzsimmons, P.J.: On a connection between Kuznetsov processes and quasi-processes. In: Stochastic Processes, 1987 PPS 15, pp. 123–133. Birkhäuser (1988)
98. Fitzsimmons, P.J., Maisonneuve, B.: Excessive measures and Markov processes with random birth and death. Probability Theory and Related Fields **72**(3), 319–336 (1986)
99. Fukushima, M.: Dirichlet Forms and Markov Processes. North Holland (1980)
100. Garsia, A.M.: Martingale Inequalities. Seminar Notes on Recent Progress. Benjamin, Reading, Massachusetts (1973)
101. Gelfand, I.M.: Generalized random processes. Dokl. Akad. Nauk **100**(5), 853–856 (1955). Translated into English on pp. 529-533 of Gel'fand's Collected Papers, Volume III, Springer 1989
102. Getoor, R.K.: Markov Processes, Ray Processes and Right Processes, LN 440. Springer (1975)
103. Getoor, R.K.: Excessive Measures. Birkhäuser (1990)
104. Getoor, R.K., Sharpe, M.J.: Last exit decompositions and distributions. Indiana University Mathematics Journal **23**(5), 377–404 (1973)
105. Getoor, R.K., Sharpe, M.J.: Last exit times and additive functionals. The Annals of Probability **1**(4), 550–569 (1973)
106. Getoor, R.K., Sharpe, M.J.: Balayage and multiplicative functionals. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **28**(2), 139–164 (1974)
107. Girsanov, I.V.: On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. Theory of Probability & Its Applications **5**(3), 285–301 (1960)
108. Glimm, J., Jaffe, A.: Quantum Physics: A Functional Integral Point of View. Springer (1981)
109. Gross, L.: Logarithmic Sobolev inequalities. American Journal of Mathematics **97**(4), 1061–1083 (1975)
110. Halmos, P.R.: Measure theory. Van Nostrand (1950)
111. Harris, T.E.: The theory of branching processes. Springer, Berlin (1963)
112. Herz, C.: Bounded mean oscillation and regulated martingales. Transactions of the American Mathematical Society **193**, 199–215 (1974)
113. Herz, C.: H_p-spaces of martingales, $0<p \leq 1$. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **28**(3), 189–205 (1974)
114. Hida, T.: Finite dimensional approximations to white noise and Brownian motion. Journal of Mathematics and Mechanics **16**(8), 859–866 (1967)
115. Hida, T.: Brownian Motion. Springer (1980)
116. Hille, E.: Functional Analysis and Semi-groups. American Mathematical Society (1948)
117. Hoffmann-Jørgensen, J.: Markov sets. Mathematica Scandinavica **24**(2), 145–166 (1969)
118. Hudson, R.L., Parthasarathy, K.R.: Quantum Itô's formula and stochastic evolutions. Communications in mathematical physics **93**(3), 301–323 (1984)
119. Hunt, G.A.: Some theorems concerning Brownian motion. Transactions of the American Mathematical Society **81**(2), 294–319 (1956)
120. Hunt, G.A.: Markoff processes and potentials I, II, III. Illinois Journal of Mathematics **1**(1), 44–93 (1957). Part II 1(3):316–369; Part III 2(2):151–213
121. Hunt, G.A.: Markoff chains and Martin boundaries. Illinois Journal of Mathematics **4**(3), 313–340 (1960)
122. Ikeda, N., Watanabe, S.: Stochastic Differential Equations and Diffusion Processes. North-Holland (1988)
123. Itô, K.: Stochastic integral. Proceedings of the Imperial Academy **20**(8), 519–524 (1944)
124. Itô, K.: On a stochastic integral equation. Proceedings of the Japan Academy **22**(1-4), 32–35 (1946)
125. Itô, K.: Stochastic differential equations in a differentiable manifold. Nagoya Mathematical Journal **1**, 35–47 (1950)

126. Itô, K.: Multiple Wiener integral. Journal of the Mathematical Society of Japan **3**(1), 157–169 (1951)
127. Itô, K.: On a formula concerning stochastic differentials. Nagoya Mathematical Journal **3**, 55–65 (1951)
128. Itô, K.: On stochastic differential equations. Memoirs of the American Mathematical Society **4**, 51 pp. (1951)
129. Itô, K.: Lectures on Stochastic Processes. Tata Institute, Bombay (1961)
130. Itô, K.: Poisson point processes attached to Markov processes. In: Sixth Berkeley Symposium on Mathematical Statistics and Probability, vol. 3, pp. 225–239 (1972)
131. Itô, K., McKean, H.P.: Diffusion processes and their sample paths. Springer (1965)
132. Itô, K., Watanabe, S.: Transformation of Markov processes by multiplicative functionals. Annales de l'Institut Fourier **15**(1), 13–30 (1965)
133. Jacod, J.: Calcul stochastique et problemes de martingales, LN 714. Springer (1979)
134. Jacod, J., Mémin, J.: Caractéristiques locales et conditions de continuité absolue pour les semi-martingales. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **35**(1), 1–37 (1976)
135. Jacod, J., Shiryaev, A.: Limit theorems for stochastic processes, vol. 288. Springer (1987)
136. Jacod, J., Yor, M.: Étude des solutions extrémales et représentation intégrale des solutions pour certains problèmes de martingales. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **38**(2), 83–125 (1977)
137. Jeulin, T.: Semi-martingales et grossissement d'une filtration, LN 833. Springer (1980)
138. John, F., Nirenberg, L.: On functions of bounded mean oscillation. Communications on Pure and Applied Mathematics **14**(3), 415–426 (1961)
139. Kac, M.: On distributions of certain Wiener functionals. Transactions of the American Mathematical Society **65**(1), 1–13 (1949)
140. Kakutani, S.: On Brownian motions in n-space. Proceedings of the Imperial Academy **20**(9), 648–652 (1944)
141. Kakutani, S.: Two-dimensional Brownian motion and harmonic functions. Proceedings of the Imperial Academy **20**(10), 706–714 (1944)
142. Kazamaki, N.: On a stochastic integral equation with respect to a weak martingale. Tôhoku Mathematical Journal **26**(1), 53–63 (1974)
143. Knight, F.: Note on regularization of Markov processes. Illinois Journal of Mathematics **9**(3), 548–552 (1965)
144. Knight, F.B.: Prediction processes and an autonomous germ-Markov property. The Annals of Probability **7**(3), 385–405 (1979)
145. Knight, F.B.: Foundations of the prediction process, vol. 1. Oxford (1992)
146. Kolmogorov, A.N.: Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. Mathematische Annalen **104**, 415–458 (1931)
147. Kolmogorov, A.N.: Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer, Berlin (1933)
148. Krylov, N.V.: Some estimates in the theory of stochastic integrals. Theory of Probability & Its Applications **18**(1), 54–63 (1973)
149. Krylov, N.V., Yushkevich, A.A.: Markov random sets. Tr. Mosk. Mat. Obs. **13**, 114–135 (1965)
150. Kunita, H., Watanabe, S.: On square integrable martingales. Nagoya Journal of Mathematics **30**, 209–245 (1967)
151. Kunita, H., Watanabe, T.: Markov processes and Martin boundaries Part I. Illinois Journal of Mathematics **9**(3), 485–526 (1965)
152. Kuznetsov, S.E.: Construction of Markov processes with random times of birth and death. Theory of Probability & Its Applications **18**(3), 571–575 (1974)
153. Lenglart, É., Lépingle, D., Pratelli, M.: Présentation unifiée de certaines inégalités de la théorie des martingales. Séminaire de probabilités de Strasbourg **14**, 26–48 (1980)
154. Letta, G.: Martingales et Intégration Stochastique. Edizioni della Normale, Pisa (1984)
155. Lévy, P.: Théorie de l'addition des variables aléatoires. Gauthier-Villars, Paris (1937). Second edition 1954.
156. Lévy, P.: Processus Stochastiques et Mouvement Brownien. Gauthier-Villars, Paris (1948)

157. Liptser, R.S., Shiryaev, A.N.: Statistics of Random Processes. Springer (1977). Two volumes
158. Loève, M.: Probability Theory. Van Nostrand (1960)
159. Maisonneuve, B.: Systèmes régénératifs. No. 15 in Astérisque. Société mathématique de France (1974)
160. Malliavin, P.: Géométrie différentielle stochastique. University of Montréal (1977)
161. Malliavin, P.: Stochastic calculus of variation and hypoelliptic operators. In: K. Itô (ed.) Proc. Internat. Symposium on Stochastic Differential Equations, Kyoto, 1976, pp. 195–263 (1978)
162. Markov, A.A.: Extension of the law of large numbers to dependent quantities (in Russian). Izvestiia Fiz.-Matem. Obsch. Kazan Univ., (2nd Ser.) **15**, 135–156 (1906)
163. Martin, R.S.: Minimal positive harmonic functions. Transactions of the American Mathematical Society **49**(1), 137–172 (1941)
164. McShane, E.J.: Stochastic Calculus and Stochastic Models. Academic Press (1974)
165. Métivier, M., Pellaumail, J.: Mesures stochastiques à valeurs dans des espaces L_0. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **40**(2), 101–114 (1977)
166. Métivier, M., Pellaumail, J.: Stochastic integration. Academic Press (1980)
167. Meyer, P.A.: A decomposition theorem for supermartingales. Illinois Journal of Mathematics **6**(2), 193–205 (1962)
168. Meyer, P.A.: Decomposition of supermartingales: the uniqueness theorem. Illinois Journal of Mathematics **7**(1), 1–17 (1963)
169. Meyer, P.A.: Probability and Potentials. Blaisdell (1966)
170. Meyer, P.A.: Intégrales stochastiques I–IV. Séminaire de probabilités de Strasbourg **1**, 72–162 (1967)
171. Meyer, P. A. Démonstration probabiliste de certaines inégalités de Littlewood-Paley, Séminaire de probabilités de Strasbourg, X, LN511. Springer (1976)
172. Meyer, P.A.: Les processus stochastiques de 1950 à nos jours. In: J.P. Pier (ed.) Development of Mathematics 1950–2000, pp. 813–848. Birkhäuser (2000)
173. Millar, P.W.: Martingale integrals. Transactions of the American Mathematical Society **133**(1), 145–166 (1968)
174. Minlos, R.A.: Generalized random processes and their extension in measure. Trudy Moskovskogo Matematicheskogo Obshchestva **8**, 497–518 (1959)
175. Mitro, J.B.: Dual Markov processes: construction of a useful auxiliary process. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **47**(2), 139–156 (1979)
176. Motoo, M.: The sweeping-out of additive functionals and processes on the boundary. Annals of the Institute of Statistical Mathematics **16**(1), 317–345 (1964)
177. Motoo, M., Watanabe, S.: On a class of additive functionals of Markov processes. Journal of Mathematics of Kyoto University **4**(3), 429–469 (1965)
178. Nagasawa, M.: Time reversions of Markov processes. Nagoya Mathematical Journal **24**, 177–204 (1964)
179. Nagasawa, M.: Segregation of a population in an environment. Journal of Mathematical Biology **9**(3), 213–235 (1980)
180. Nelson, E.: Dynamical Theories of Brownian Motion. Princeton (1967)
181. Nelson, E.: Construction of quantum fields from Markoff fields. Journal of Functional Analysis **12**(1), 97–112 (1973)
182. Nelson, E.: The free Markoff field. Journal of Functional Analysis **12**(2), 211–227 (1973)
183. Nelson, E.: Notes on non-commutative integration. Journal of Functional Analysis **15**(2), 103–116 (1974)
184. Nelson, E.: Quantum Fluctuations. Princeton (1985)
185. Neveu, J.: Entrance, exit and fictitious states for Markov chains. In: Proc. Aarhus Colloq. Combinatorial Probability, pp. 64–68 (1962)
186. Neveu, J.: Sur les états d'entrée et les états fictifs d'un processus de Markov. Annales de l'Institut Henri Poincaré **17**(4), 323–337 (1962)
187. Orey, S.: F-processes. In: Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 2, Pt. 1, pp. 301–313. University of California (1967)
188. Parthasarathy, K.R.: An introduction to Quantum Stochastic Calculus. Birkhäuser (1992)

189. Prokhorov, Y.V.: Convergence of random processes and limit theorems in probability theory. Theory of Probability & Its Applications **1**(2), 157–214 (1956)
190. Protter, P.: $\mathscr{H}^p$ stability of solutions of stochastic differential equations. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **44**(4), 337–352 (1978)
191. Protter, P.: Stochastic Integration and Differential Equations: A New Approach. Springer (1990)
192. Protter, P.E.: On the existence, uniqueness, convergence and explosions of solutions of systems of stochastic integral equations. The Annals of Probability **5**(2), 243–261 (1977)
193. Ray, D.: Resolvents, transition functions, and strongly markovian processes. Annals of Mathematics pp. 43–72 (1959)
194. Revuz, D., Yor, M.: Continuous Martingales and Brownian Motion. Springer, Berlin (1991)
195. Schilder, M.: Some asymptotic formulas for Wiener integrals. Transactions of the American Mathematical Society **125**(1), 63–85 (1966)
196. Schrödinger, E.: Über die Umkehrung der Naturgesetze. de Gruyter, Berlin (1931)
197. Schwartz, L.: Semi-martingales sur des variétés, et martingales conformes sur des variétés analytiques complexes, LN 780. Springer (1980)
198. Segal, I.E.: A non-commutative extension of abstract integration. Annals of Mathematics **57**(3), 401–457 (1953)
199. Segal, I.E.: Tensor algebras over Hilbert spaces. I. Transactions of the American Mathematical Society **81**(1), 106–134 (1956)
200. Sharpe, M.: General Theory of Markov Processes. Academic Press (1988)
201. Shih, C.T.: On extending potential theory to all strong Markov processes. Annales de l'Institut Fourier **20**(1), 303–315 (1970)
202. Shur, M.G.: Continuous additive functionals of Markov processes and excessive functions. Doklady Akademii Nauk **137**(4), 800–803 (1961)
203. Silverstein, M.L.: Symmetric Markov Processes. LN 426. Springer (1974)
204. Simon, B.: The $P(\Phi)_2$ Euclidean (Quantum) Field Theory. Princeton (1974)
205. Simon, B.: Functional Integration and Quantum Physics. Academic Press (1979)
206. Skorokhod, A.V.: Limit theorems for stochastic processes. Theory of Probability & Its Applications **1**(3), 261–290 (1956)
207. Skorokhod, A.V.: Investigations in the theory of random processes. University Press Kiev (1961)
208. Snell, J.L.: Applications of martingale system theorems. Transactions of the American Mathematical Society **73**(2), 293–312 (1952)
209. Spitzer, F.: Recurrent random walk and logarithmic potential. In: Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 2, pp. 515–534. University of California (1961)
210. Spitzer, F.: Principles of random walk. Van Nostrand (1964)
211. Stein, E.M.: Singular Integrals and Differentiability Properties of Functions. Princeton (1970)
212. Stein, E.M.: Topics in Harmonic Analysis Related to the Littlewod-Paley Theory. Princeton (1970)
213. Stratonovich, R.L.: A new representation for stochastic integrals and equations. SIAM Journal on Control **4**(2), 362–371 (1966). Translated from Vestnik Moscow Univ. Ser. I Mat. Mech. **1**,3–12 (1964).
214. Stroock, D.W.: The Malliavin calculus and its application to second order parabolic differential equations: Part I. Mathematical Systems Theory **14**(1), 25–65 (1981)
215. Stroock, D.W., Varadhan, S.R.S.: Diffusion processes with continuous coefficients. Communications on Pure and Applied Mathematics **22**(3,4), 3:345–400, 4:479–530 (1969)
216. Stroock, D.W., Varadhan, S.R.S.: Diffusion processes with continuous coefficients. Communications on Pure and Applied Mathematics **24**(2), 147–225 (1971)
217. Stroock, D.W., Varadhan, S.R.S.: On the support of diffusion processes with applications to the strong maximum principle. In: Sixth Berkeley Symposium on Mathematical Statistics and Probability, vol. 3, pp. 333–359. University of California (1972)
218. Stroock, D.W., Varadhan, S.R.S.: Multidimensional Diffusion Processes. Springer (1979)

219. Tanaka, H.: Note on continuous additive functionals of the 1-dimensional Brownian path. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **1**, 251–257 (1963)
220. Taylor, S.J.: The α-dimensional measure of the graph and set of zeros of a Brownian path. Mathematical Proceedings of the Cambridge Philosophical Society **51**(2), 265–274 (1955)
221. Trotter, H.F.: A property of Brownian motion paths. Illinois Journal of Mathematics **2**(3), 425–433 (1958)
222. Varopoulos, N.T.: Aspects of probabilistic Littlewood-Paley theory. Journal of Functional Analysis **38**(1), 25–60 (1980)
223. Ville, J.: Étude critique de la notion de collectif. Gauthier-Villars, Paris (1939)
224. Volkonskii, V.: Additive functionals of Markov processes. Trudy Moskovskogo Matematicheskogo Obshchestva **9**, 143–189 (1960)
225. Watanabe, S.: On discontinuous additive functionals and Lévy measures of a Markov process. Japanese Journal of Mathematics **34**, 53–70 (1964)
226. Wiener, N.: Differential space. Journal of Mathematics and Physics **2**, 131 (1923)
227. Wiener, N.: The homogeneous chaos. American Journal of Mathematics **60**(4), 897–936 (1938)
228. Williams, D.: Diffusions, Markov Processes, and Martingales. I: Foundations. Wiley, Chichester (1979)
229. Williams, D. (ed.): Stochastic Integrals, Durham, UK, LN 851. Springer (1981)
230. Yamada, T., Watanabe, S.: On the uniqueness of solutions of stochastic differential equations. Journal of Mathematics of Kyoto University **11**(1,3), 1:155–167, 3:553–563 (1971)
231. Yor, M., de Sam Lazaro, J.: Sous-espaces denses dans L^1 ou H^1 et représentation des martingales. Séminaire de probabilités de Strasbourg **12**, 265–309 (1978)
232. Yosida, K.: On the differentiability and the representation of one-parameter semi-group of linear operators. Journal of the Mathematical Society of Japan **1**(1), 15–21 (1948)
233. Yosida, K.: Functional Analysis. Iwanami (1951)
234. Zheng, W.A.: Tightness results for laws of diffusion processes application to stochastic mechanics. Annales de l'Institut Henri Poincaré Probabilités et statistiques **21**(2), 103–124 (1985)

Martingales in Japan

Shinzo Watanabe

Abstract

Japanese mathematicians did not contribute directly to martingale theory before 1960, but many later contributions were based on the stochastic calculus that Kiyosi Itô first introduced in 1942. Itô's collaboration with Henry McKean on the pathwise construction of diffusions attracted wide interest from students in Japan. Subsequent Japanese contributions in the 1960s included adaptations of results on Markov processes to martingales, such as Itô and Watanabe's multiplicative analog of the Doob–Meyer decomposition, which involved the introduction of local martingales, contributions to stochastic integration for square-integrable martingales and semimartingales, and contributions to the representation of martingales. Japanese contributions after 1970 included Itô's reformulation of the stochastic calculus in term of stochastic differentials, Itô's circle operation, the Itô–Tanaka formula, and the Fukushima decomposition.

Keywords

Additive functional · Hunt process · Itô's circle operation · Itô–Tanaka formula · Martingale · Markov process · Multiplicative decomposition · Sample function · Semimartingale · Watanabe decomposition · Fukushima decomposition

1 Before 1960: Itô's Stochastic Analysis

Modern probability theory, as founded and developed by distinguished pioneers such as A. N. Kolmogorov, P. Lévy, N. Wiener, and so on, attracted great interest and attention from Japanese mathematicians, including K. Yosida, S. Kakutani, K. Itô, G. Maruyama, and others. Around 1935, Kiyosi Itô (1915–2008), then a student at the

S. Watanabe (✉)
Kyoto University, Kyoto, Japan
e-mail: watanabe@math.kyoto-u.ac.jp

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*,
Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_9

University of Tokyo, found Kolmogorov's recently published book, *Grundbegriffe der Wahrscheinlichkeitsrechnung* one day in a bookstore. As he often recollected in later years, this fortuitous discovery gave him one of his motivations for devoting his future life to the study of probability theory. See for example the page next to the preface in [18].

Although the study of modern probability theory in Japan certainly started before 1940, the war disrupted communications with other advanced countries. Under these circumstances, Itô completed two important contributions [10, 11] that are now considered the origin of *Itô's stochastic analysis* or *Itô's stochastic calculus*. In the first work, he gave a rigorous proof of what is now called the *Lévy–Itô theorem* for the structure of sample functions of Lévy processes, through which we have a complete understanding of the Lévy–Khinchin formula for canonical forms of infinitely divisible distributions. In the second work, he developed a complete theory of stochastic differential equations determining sample functions of diffusion processes whose laws are described by Kolmogorov's differential equations. In this work, he introduced the important notion of a *stochastic integral* and the basic formula now known as *Itô's formula* or *Itô's lemma* and thus founded a kind of Newton–Leibniz differential and integral calculus for a class of random functions now often called *Itô processes*. As we will see below, this work by Itô was further developed and refined in the martingale framework. Itô himself did not say anything about martingale theory in this work; the theory of martingales began to be noticed in Japan only after the publication of Doob's book [4] in 1953.

Itô published two books in Japanese on modern probability theory, [12] in 1944 and [15] in 1953, in which he introduced the modern theory of probability to the mathematical community in Japan. He thus provided beginning Japanese researchers and students in this field with excellent textbooks. He incorporated his results in [10, 11] into these books, especially into [15], the more advanced of the two. Although the term *martingale* cannot be found in the books, some fundamental ideas and techniques of martingale theory are implicit in them. In particular, von Mises's important ideas of "Stellenauswahl" (selection) from a random sequence and "Regellosigkeit" (irregularity) are explained in [12, Sects. 26–27] and in [15, Sect. 17]. In modern martingale theory, selection is a typical example of a *martingale transform*, a discrete-time version of the stochastic integral (for martingale transforms, cf. e.g., [47, p. 97] and [9, p. 25]). Such ideas and notions are used to obtain an extension of the famous *Kolmogorov maximal inequality* for sums of independent random variables, which plays an important role when we define stochastic integrals. As we know, Doob extended this inequality in his martingale theory, so that it is now well known as the *Kolmogorov–Doob* maximal inequality.

After the war ended and international communications began to recover, Itô sent an expanded and refined English version of [11] to Doob, asking him to help with its publication in United States. Doob was kind enough to arrange its publication in *Memoirs of AMS* in 1951 [14]. When Doob's book [4] appeared in 1953, Itô was impressed by its beautiful theory of martingales and was glad to see its treatment of his stochastic integrals in the martingale framework [43, p. xv]. But Itô was given a chance to visit the Institute for Advanced Study at Princeton University from 1954 to 1956, and during that period and for some time thereafter, his main effort was devoted to the study of one-dimensional diffusion processes jointly with Henry P. McKean [43, p. xv], [19, p. 2].

At Princeton, W. Feller had already almost completed his theory of the analytical description of one-dimensional diffusion processes. At his suggestion, Itô began to work with McKean, who was then a student of Feller's, on a pathwise theoretic construction of one-dimensional diffusion processes.

Apparently this work was done independently of Doob's martingale theory,[1] but in fact it has much to do with martingale theories: the Itô–McKean construction of sample paths of one-dimensional diffusions makes a *random time change* in the paths of a one-dimensional Wiener process (Brownian motion), while, in Doob's theory, a time change is formulated as an *optional sampling*, and, indeed, Doob's optional sampling theorem plays a key role in his theory of martingales. Also, Itô–McKean's time change is based on Brownian local time. Later, a general notion of local time was established by the French school (cf. e.g., [39, p. 206] and [40, p. 96]) in the martingale framework, and it plays an important role in the stochastic calculus of the random functions called semimartingales.

2 Japanese Contributions to Martingales from 1961 to 1970

Itô returned to Kyoto from Princeton in 1956. His and McKean's joint work continued at Kyoto University for several years; there McKean gave a series of lectures that stimulated much younger researchers (T. Hida, N. Ikeda, M. Motoo, M. Nisio, H. Tanaka, T. Ueno, T. Watanabe,...) as well as graduate students (M. Fukushima, H. Kunita, K. Sato, S. Watanabe, T. Yamada,...). Many were interested in the problem of extending the theory of Feller and Itô–McKean from one-dimensional diffusions to multidimensional cases, particularly the problem of diffusion processes with Venttsel's boundary conditions. Worldwide, modern probability theory had been developing from the pioneering works by Kolmogorov, Lévy, Wiener and others. In the theory of Markov processes, the most advanced countries around 1960 were the United States and the Soviet Union. The main themes were Markov processes and related problems in analysis, potential theory in particular, and functionals of sample functions such as additive and multiplicative functionals, as studied by W. Feller, S. Kakutani, J. L. Doob, M. Kac, G. A. Hunt, and many others in United States, E. B. Dynkin and his group in Moscow, A. V. Skorokhod and his group in Kiev, and so on.

The theory of martingales became known very gradually at the time, mostly from the work of Doob and the influence of his book [4]. It was recognized as useful because of the (sub-, super-) martingale convergence theorems and the theorems on the existence of regular modifications of sample functions of (sub-, super-) mar-

[1] The situation at the time was similar in all work on Markov processes, perhaps with the exception of Doob's work. The same objects were given different names in each theory; for example, *stopping times* were called *Markov times* in Markov process theory. Since then, the two theories have gradually mixed together, bringing remarkable progress to each.

tingales in continuous time. I personally came to know of it for the first time in Khinchin's paper [26] treating McMillan's theorem in information theory by Doob's martingale convergence theorem. In Markov process theory, the existence of a nice Markov process (Hunt process) has usually been based on the existence of regular modifications for sample functions of (sub-, super-) martingales in continuous time.

Doob applied his martingale theory to the study of Markov processes and potentials in an essential way. One of his typical ideas was the following: if $u(x)$ is a harmonic or a sub(super)harmonic function in a domain D of $\mathbf{R}^d$, and, if $B(t)$ is a d-dimensional Brownian motion (i.e., Wiener process) starting from a point in D, then the stochastic process $t \in [0, \tau_D) \mapsto u(B(t))$, where τ_D is the first exit time from D of $B(t)$, is a continuous (local) martingale (resp. sub(super)martingale). So, for example, if $u(x)$ is a positive superharmonic function, then the process $t \mapsto u(B(t))$ has bounded and continuous sample functions with probability one, because of a well known result of Doob on sample paths of positive supermartingales. So we can say, for example, that Brownian motion behaves so as to avoid any discontinuity or any point at which a positive superharmonic function assumes an infinite value.

Gradually it began to be understood that there is a deep interplay between Markov process theory and the martingale theory: Many important results in Markov processes, formulated in a more abstract and general framework in the martingale theory, may be regarded as basic and abstract results and principles, so that the original results in Markov processes are just typical applications in a more concrete or special situation. The following subsections are all concerned, more or less, with this kind of progress in martingale theory.

2.1 The Doob–Meyer Decomposition Theorem for Supermartingales

The theory of Markov processes and potentials has advanced a great deal. In particular, G. A. Hunt developed a very general theory of *excessive functions* for a given Markov process in a restricted but reasonably general and convenient class; these Markov processes are now known as *Hunt processes*. In Moscow, Dynkin emphasized the importance of the study of additive and multiplicative functionals of Markov processes in connection with potential theory.[2] (Actually, some of its importance had already been demonstrated in Itô and McKean's work.) P. A. Meyer, who originally studied potential theory in the famous French school guided by Brelot, Choquet and Deny, made a deep study of additive functionals (AFs) of a Hunt process [32] from a potential theoretic point of view. His results, viewed in a martingale framework, could be understood as giving an abstract and general principle in stochastic processes. It is indeed a realization of Doob's idea that a submartingale should be a sum of a martingale and a process with increasing sample paths, so that a supermartingale should have a representation as a martingale minus an increasing process. Meyer

[2] Cf. Introduction of [5], which was originally Dynkin's plenary lecture at ICM 1962, Stockholm.

rewrote his results on AFs of a Hunt process in a framework of martingale theory and thus obtained a general result concerning the representation of a supermartingale as a difference of a martingale and an increasing process [31,33]. This is now called the *Doob–Meyer decomposition of supermartingales*, which is certainly one of the most basic and important results in martingale theory.

K. Itô and S. Watanabe [21] studied, in contrast with the *additive* decomposition of Doob–Meyer, the *multiplicative* decomposition of a positive supermartingale into a product of a positive martingale and a positive decreasing process. This problem originated in the study of multiplicative functionals (MFs) in connection with transformations of Markov processes, a problem much studied at that time. This paper introduced the notion of *local martingales*, which is now a basic tool in localization arguments in martingale theory.

We should mention here some relevant important developments, around that time, in the study of *positive martingales*. In 1970 C. Doléans-Dade [3] obtained a general expression of so-called *exponential martingales*. In 1960 I. V. Girsanov [8] established the *Girsanov theorem*, later generalized and refined by P. A. Meyer and others of the French school, which is concerned with the transformation of the martingale character under an equivalent change of the underlying probability.[3] These cannot be considered Japanese contributions, but G. Maruyama in 1954 [29] and M. Motoo in 1960–61 [36] had already studied, in their works on diffusion processes associated with Kolmogorov differential equations, important examples of exponential martingales and the Girsanov transformations they define, even though they did not state their results in terms of martingale theory.

2.2 Stochastic Integrals for Square-Integrable Martingales and Semimartingales

Stochastic integrals were first introduced by K. Itô in 1942 [11]. J. L. Doob [4] pointed out the martingale character of stochastic integrals and suggested that a unified theory of stochastic integrals should be established in a framework of martingale theory. His program was accomplished by H. Kunita and S. Watanabe [27] and P. A. Meyer [35], among others.

I would like to comment on these works in more detail. Here again, they have their origin in the theory of Markov processes, particularly in the work of M. Motoo and S. Watanbe [37,46] on square-integrable additive functionals (AFs) of a Hunt process having zero expectations.[4] A main aim of the work in [37,46] was to study the structure of the space $\mathbf{M}$ formed by the square-integrable AFs having zero expectations, particularly to understand and generalize a result of A. D. Venttsel' [45] in the case of Brownian motion; if $X(t) = (X_1(t), \dots, X_d(t))$ is a d-

[3] Cf. e.g., [38, p. 109].

[4] For an AF, it is equivalent that it have zero expectation and that it be a martingale with respect to the natural filtration of the process.

dimensional Brownian motion, the space $\mathbf{M}$ consists of AF $A(t)$ represented in the form $A(t) = \sum_{i=1}^{d} \int_0^t f_i(X(s))dX_i(t)$ as a sum of Itô's stochastic integrals.[5] In this study, a fundamental role is played by a *random inner product* $\langle M, N \rangle$, $M, N \in \mathbf{M}$, which is defined to be a continuous AF with almost all sample paths locally of bounded variation. Using this random inner product, important and useful notions such as *stochastic integrals*, *stable subspaces*, *orthogonality* and *projection* of subspaces in $\mathbf{M}$, *basis* of a subspace, and so on, can be introduced and studied. The orthogonal complement $\mathbf{M}_d$ of the subspace $\mathbf{M}_c$ formed of all continuous elements in $\mathbf{M}$ was studied in [46]. There, a random point process was defined by jumps of sample paths of the Hunt process, and its *compensator*, called the *Lévy measure* of the Hunt process, was introduced and studied.

During a period around 1963, H. Kunita and I conceived the idea of extending results in [37,46] to a more general and abstract situation in which the natural filtration associated with the Hunt process is replaced by a general filtration and an AF is replaced by a general càdlàg adapted process. By the Doob–Meyer decomposition theorem, we can still define the random inner product $\langle M, N \rangle$ for square-integrable martingales M and N.[6] Stochastic integrals with respect to a square-integrable martingale can be characterized by this random inner product and can be constructed along the lines of Itô and Doob.

In this period, I was visiting Paris as a scholarship student (*boursier*) of the French Government. Very fortunately, I had an opportunity to attend a lecture on the decomposition of supermartingales by Meyer at the Collège de France, just before he moved from Paris to Strasbourg. After the lecture, he kindly invited me to his home and we exchanged information on our current work.

Thus, the works [27,35], which both finally appeared in 1967, are very much related; indeed, as Meyer kindly stated in [35], his work was motivated by [27]. If we review the work in [27] now from the standpoint of martingale theory, it should be said that, as far as discontinuous stochastic processes are concerned, it is rather restricted and incomplete in many points. As we know, a mathematically complete and satisfactory theory was established by Meyer and his French (Strasbourg) school,[7] and [35] was a starting point for this French contribution.

The class of stochastic processes introduced in Itô's original paper [11] (now often called Itô processes) can be naturally extended to a class of stochastic processes called *semimartingales*.[8] *Itô's formula* or *Itô's lemma* leads to a kind of Newton–Leibniz differential and integral calculus for semimartingales.

---

[5] The f_i are Borel functions on $\mathbf{R}^d$ with certain integrability conditions.

[6] A standard terminology now is *predictable quadratic co-variation* of M and N. Meyer [35] introduced another random inner product $[M, N]$, called the *quadratic co-variation* of M and N, which plays important role in the study of discontinuous semimartingales.

[7] Cf. e.g., [2,22,38] as important texts treating the theory.

[8] The term *semimartingale* and its notion were introduced by Meyer [34,35]. The word was used differently in Doob's book [4]; there *semi-martingale* was used to mean submartingale, and *lower semi-martingale* to mean supermartingale.

For a semimartingale, we have a decomposition of a sample function as the sum of a continuous semimartingale and a discontinuous semimartingale. Roughly speaking, a process is a discontinuous semimartingale if its sample function can be obtained as a *compensated sum* of jumps. The continuous part is a sum of a (locally) square-integrable continuous martingale and a continuous process with sample functions (locally) of bounded variation. So a semimartingale has sample functions similar to those of a Lévy process. We can associate with a semimartingale a system of quantities which correspond, in the case of a Lévy process, to its Lévy–Khinchin characteristic: What corresponds to the covariance of the Gaussian component in the Lévy process is the predictable quadratic co-variation of the continuous martingale part of the semimartingale. What corresponds to the Lévy measure of the Lévy process is the compensator of a point process defined by the size of jumps of sample paths of the semimartingale.[9]

Actually, Lévy processes are a particular case of semimartingales. Indeed, it is the most fundamental case, in which the associated characteristic quantities are deterministic (i.e., non-random). In particular, a d-dimensional Wiener process $X(t)$ is characterized as a d-dimensional continuous martingale $X(t) = (X_1(t), \ldots, X_d(t))$ with the predictable quadratic co-variation satisfying the condition $\langle X_i, X_j \rangle(t) = \delta_{i,j} t$, $i, j = 1, \ldots, d$. In [4], this characterization of the Wiener process in the frame of martingale theory is attributed to P. Lévy. Also, there is a similar martingale characterization for a Poisson process [46] and for Poisson point processes (cf. e.g. [9,22]).

Thus, we can see that Itô's works on Lévy processes in [10] and on stochastic integrals and Itô processes in [11] have grown into a unified general theory of semimartingales. In this framework, many important stochastic models can be defined and constructed by appealing to the theory of stochastic differential equations or the method of martingale problems.

2.3 Martingale Representation Theorems

In the case of a Wiener process, the martingale representation theorem[10] states that *every local martingale with respect to the natural filtration of a Wiener process can be expressed as the sum of a constant and a stochastic integral of a predictable integrand* $f(s)$ *with* $\int_0^t f(s)^2 ds < \infty$ *for every* t*, a.s.* As mentioned above, this kind of representation theorem first appeared in the study of AFs by Venttsel', and its extension to general Hunt processes has been a main motivation of our work in [37]. Its further extension to the case of general square-integrable martingales motivated our work in [27]. In [27], we presented several useful results for the representation

[9] The notion of the compensator for a point process is key in the martingale theoretic approach to point processes. Indeed, it has much to do with semimartingale theory; the discontinuities of a semimartingale define a point process on the real line and, conversely, a point process on a general state space defines a discontinuous semimartingale by a projection of the state space to the real line. Cf. e.g. [9,22,25], for the martingale-theoretic approach to point processes and applications.

[10] We state it in the one-dimensional case; its multi-dimensional extension is straightforward.

of martingales. However, the notion of a *basis* in the sense of [37] could not be stated explicitly. Later, this notion was completely established by M. H. A. Davis and P. Varaiya [1].

The martingale representation theorem for a Wiener process as stated above has played an important role in financial mathematics. In this field, this theorem is very well known as *Itô's representation theorem.*[11] Indeed, this is because an essential part of the proof of this theorem is to prove that every square-integrable functional F of Wiener process paths $\{w(t); 0 \le t \le T\}$ can be represented as $F = E(F) + \int_0^T f(s)dw(s)$ by Itô's stochastic integral. Such a representation can be obtained, as Itô remarked in 1951 [13, Theorem 5.1, p. 168], by expanding F into an orthogonal sum of multiple Wiener integrals and then rewriting the multiple Wiener integrals as iterated Itô stochastic integrals.

3 Japanese Contributions to Martingales After 1971

During this period, stochastic analysis based on semimartingales was developed and used around the world. It became one of the most important and useful methods in probability theory and its applications. Many standard textbooks, including [9, 22, 24, 39, 40], treated stochastic analysis based on semimartingales and martingale methods. Here, I review some work in this period in which we can find some Japanese contributions.

3.1 Fisk–Stratonovich Symmetric Stochastic Integrals. Itô's Circle Operation

In 1975, K. Itô [16], using the general results in [27, 35], reformulated the stochastic calculus in terms of stochastic differentials. This put Itô's formula in a form convenient for applications. The fact that Itô's formula needs extra terms as compared with the standard Newton–Leibniz rule is most interesting and mysterious in the stochastic calculus; it might be a surprise for beginners. This causes a difficulty when we want to apply the stochastic calculus for stochastic processes moving on a differentiable manifold. The process given in each local coordinate is a semimartingale but the rule of the calculus is not a usual one, so that some difficulty always arises when we want to obtain coordinate-free notions and results. For a typical example, see a very troublesome construction of a solution of stochastic differential equations on a manifold in [30].

On the other hand, Stratonovich [41] and Fisk [6] introduced a type of stochastic integral (sometimes called a *symmetric stochastic integral*) different from Itô's. Itô noticed that this kind of stochastic integral can be immediately defined by modifying Itô integrals; for two continuous semimartingales X and Y, the symmetric

[11] Cf. D. W. Stroock's interesting remark [42, p. 180].

stochastic integral of X by Y, denoted as $\int X \circ dY$, is, by definition, a continuous semimartingale Z given by

$$Z(t) = \left(\int_0^t X(s) \circ dY(s) \right) := \int_0^t X(s)dY(s) + \frac{1}{2}\langle M^X, M^Y \rangle(t),$$

where the first term on the right-hand side is Itô's stochastic integral and the second term is the predictable quadratic co-variation of the martingale parts M^X and M^Y of X and Y, respectively. Itô wrote this in stochastic-differential form as $X \circ dY = XdY + \frac{1}{2}dXdY$. This operation on stochastic differentials is often called *Itô's circle operation*. Under this new operation, we have the same rule of transformations as that of Newton–Leibniz in the ordinary differential calculus. In other words, under Itô's circle operation Itô's formula has the same form as in ordinary differential calculus. As it has turned out in many later works, this circle operation is an indispensable tool in the study of random motions on manifolds, producing many fruitful results (cf. e.g., [9,40]).

3.2 Itô–Tanaka's Formula and Local Times

In the Itô–McKean theory, the local time of Brownian motion (i.e., the Brownian local time) plays a fundamental role. The notion of Brownian local time was first introduced by Lévy, and a rigorous and precise result on its existence as a sojourn time density and its continuity on the space variable was obtained by H. F. Trotter [20,44]. However, Trotter's paper was rather hard to follow, at least for beginners.

Around 1962, H. Tanaka was visiting McKean at MIT and he sent a letter to his friends in Japan communicating a nice and much simpler proof of Trotter's theorem. His idea is to use Itô's calculus, particularly Itô's stochastic integral, in an essential way. Tanaka's proof was reproduced in McKean's book [30] and then spread widely.

An essential point in Tanaka's proof was an extension of Itô's formula. Itô's formula is concerned with a transformation of a semimartingale by $\mathscr{C}^2$-functions: If $f(x)$ is a $\mathscr{C}^2$-function and $X(t)$ is a continuous semimartingale, then $f(X(t))$ is also a continuous semimartingale and Itô's formula describes its semimartingale decomposition precisely. If $f(x)$ now is only a convex function, or a difference of two convex functions, it is still true that $f(X(t))$ is a continuous semimartingale. In its semimartingale decomposition, the continuous martingale part has the same form as in the case of $f(x)$ being a $\mathscr{C}^2$-function; it is given by the stochastic integral $\int_0^t f'(X(s))dM^X(s)$ with respect to the continuous martingale part M^X of X. Then all terms in this semimartingale decomposition except the part of the process of bounded variation can be known explicitly, so that this part has a representation as a difference of other terms that are known explicitly. Brownian local time at $a \in \mathbf{R}$ is obtained in this way when $X(t)$ is a one-dimensional Wiener process and $f(x)$ is the convex function given by $f(x) = \max\{x - a, 0\}$. This formula representing Brownian local time is called *Tanaka's formula*.

In a similar way, we can define the local time for every continuous semimartingale. This notion was established around the last half of the seventies by members of the French school, including Meyer, Azema, and Yor (cf. e.g. [39,40]). Le Gall [28] obtained a nice application of this theory to the pathwise uniqueness problem for one-dimensional stochastic differential equations. Itô's formula for continuous semimartingales on $\mathscr{C}^2$-functions can be extended to functions that are differences of two convex functions in which the part of process of bounded variation can be expressed by an integral of local times. Such a formula is often called an *Itô–Tanaka formula*. Thus, we may say that the theory of local times for semimartingales is an important French contribution motivated by a Japanese contribution.

An important idea for extending Itô's formula beyond the Itô–Tanaka formula was given by M. Fukushima in his theory of Dirichlet forms and symmetric Markov processes associated with them [7]. He introduced a class of stochastic processes *with zero energy* and extended Itô's formula using this notion. The notions of semimartigales and semimartingale decomposition, in this case, are thereby extended; the decomposition is now known as the *Fukushima decomposition* and is playing an important role in path-theoretic studies in symmetric Markov processes. In the case of a one-dimensional Wiener process, as such an important process as Brownian local time is defined by the Itô–Tanaka formula, many new important processes can be obtained through the Fukushima decomposition: a typical example is the *Cauchy principal value of Brownian local time*, introduced and studied by T. Yamada [48] and M. Yor [49], among others.

3.3 Problems Concerning Filtrations

The martingale theory is usually developed by fixing a filtration to which the martingale property is referred. So it is very important to see how changing the filtration affects the theory. Among many important problems of this kind, Th. Jeulin and M. Yor, among others, established a theory concerning an enlargement of filtrations [23]. This is certainly a French contribution, but as Yor has often pointed out, his original motivation for this study was work by K. Itô [17]. In this paper, Itô discussed how to give meaning to a class of stochastic integrals by a Wiener process in which the integrands are not adapted to the natural filtration of the Wiener process.

References

1. Davis, M.H.A., Varaiya, P.: The multiplicity of an increasing family of σ-fields. The Annals of Probability **2**(5), 958–963 (1974)
2. Dellacherie, C., Meyer, P.A.: Probabilities and Potential B: Theory of Martingales. North-Holland, Amsterdam (1982). Chapters V–VIII
3. Doléans-Dade, C.: Quelques applications de la formule de changement de variables pour les semimartingales. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **16**(3), 181–194 (1970)
4. Doob, J.L.: Stochastic Processes. Wiley, New York (1953)
5. Dynkin, E.B.: Markov Processes. Springer (1965)
6. Fisk, D.L.: Quasi-martingales. Transactions of the American Mathematical Society **120**(3), 369–389 (1965)

7. Fukushima, M., Oshima, Y., Takeda, M.: Dirichlet forms and symmetric Markov processes. de Gruyter, Berlin (1994)
8. Girsanov, I.V.: On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. Theory of Probability & Its Applications **5**(3), 285–301 (1960)
9. Ikeda, N., Watanabe, S.: Stochastic Differential equations and Diffusion Processes. North-Holland (1988)
10. Itô, K.: On stochastic processes (infinitely divisible laws of probability). Japanese Journal of Mathematics **18**, 261–301 (1942)
11. Itô, K.: Zenkoku sizyo sugaku danwakai-si. Journal Pan-Japan Mathematical Colloquium (1077), 1352–1400 (1942). Translated as "Differential equations determining a Markoff process" on pp. 42–75 of [43]
12. Ito, K.: Foundations of Probability Theory (in Japanese). Iwanami (1944)
13. Itô, K.: Multiple Wiener integral. Journal of the Mathematical Society of Japan **3**(1), 157–169 (1951)
14. Itô, K.: On stochastic differential equations. Memoirs of the American Mathematical Society (4), 51 pp. (1951)
15. Ito, K.: Probability Theory (in Japanese). Iwanami (1953)
16. Ito, K.: Stochastic differentials. Applied Mathematics and Optimization **1**(4), 374–381 (1975)
17. Itô, K.: Extension of stochastic integrals. In: Proceedings of the International Symposium on Stochastic Differential Equations, Kyoto,1976, pp. 95–109. Mathematical Society of Japan (1978)
18. Itô, K.: A note of thanks by K. Itô. In: Stochastic Analysis and Related Topics in Kyoto: In honour of Kiyosi Itô. Mathematical Society of Japan (2004)
19. Itô, K.: Memoirs of my research on stochastic analysis. In: Stochastic Analysis and Applications, pp. 1–5. Springer (2007)
20. Itô, K., McKean Jr., H.P.: Diffusion Processes and their Sample Paths. Springer (1965)
21. Ito, K., Watanabe, S.: Transformation of Markov processes by multiplicative functionals. Annales de l'institut Fourier **15**(1), 13–30 (1965)
22. Jacod, J., Shiryaev, A.N.: Limit Theorems for Stochastic Processes. Springer (1987)
23. Jeulin, T., Yor, M.: Grossissements de filtrations: exemples et applications. Springer (1985)
24. Karatzas, I., Shreve, S.E.: Brownian Motion and Stochastic Calculus. Springer (1988)
25. Kasahara, Y., Watanabe, S.: Limit theorems for point processes and their functionals. Journal of the Mathematical Society of Japan **38**(3), 543–574 (1986)
26. Khinchin, A.Y.: On the basic theorems of information theory. Uspekhi matematicheskikh nauk **11**(1), 17–75 (1956)
27. Kunita, H., Watanabe, S.: On square-integrable martingales. Nagoya Journal of Mathematics **30**, 209–245 (1967)
28. Le Gall, J.F.: Applications du temps local aux équations différentielles stochastiques unidimensionnelles. In: Séminaire de Probabilités XVII 1981/82, pp. 15–31. Springer (1983)
29. Maruyama, G.: On the transition probability functions of the Markov process. Natl. Sci. Rep. Ochanomizu Univ **5**, 10–20 (1954)
30. McKean, H.P.: Stochastic Integrals. Academic Press (1969)
31. Meyer, P.A.: A decomposition theorem for supermartingales. Illinois Journal of Mathematics **6**(2), 193–205 (1962)
32. Meyer, P.A.: Fonctionnelles multiplicatives et additives de Markov. Annales de l'institut Fourier **12**, 125–230 (1962)
33. Meyer, P.A.: Decomposition of supermartingales: the uniqueness theorem. Illinois Journal of Mathematics **7**(1), 1–17 (1963)
34. Meyer, P.A.: Probability and Potentials. Blaisdell (1966)
35. Meyer, P.A.: Intégrales stochastiques I–IV. Séminaire de probabilités de Strasbourg **1**, 72–162 (1967)
36. Motoo, M.: Diffusion process corresponding to $1/2 \sum \partial^2/\partial x^{i2} + \sum b^i(x)\partial/\partial x^i$. Annals of the Institute of Statistical Mathematics **12**(1), 37–61 (1960)

37. Motoo, M., Watanabe, S.: On a class of additive functionals of Markov processes. Journal of Mathematics of Kyoto University **4**(3), 429–469 (1965)
38. Protter, P.: Stochastic Integration and Differential Equations: A New Approach. Springer (1990)
39. Revuz, D., Yor, M.: Continuous Martingales and Brownian Motion, 2nd edn. Springer (1991)
40. Rogers, L.C.G., Williams, D.: Diffusions, Markov Processes, and Martingales, vol. 2: Itô Calculus. Wiley (1987)
41. Stratonovich, R.L.: A new representation for stochastic integrals and equations. SIAM Journal on Control **4**(2), 362–371 (1966). Translated from Russian article that appeared in Vestnik Moscow Univ. Ser. I Mat. Mech. **1**m 3–12 in 1964
42. Stroock, D.W.: Markov Processes from K. Itô's Perspective. Princeton University Press (2003)
43. Stroock, D.W., Varadhan, S.R.S. (eds.): Kiyosi Itô: Selected Papers. Springer, New York (1987)
44. Trotter, H.F.: A property of Brownian motion paths. Illinois Journal of Mathematics **2**(3), 425–433 (1958)
45. Venttsel', A.D.: Additive functionals on a multi-dimensional Wiener process. Doklady Akademii Nauk **139**(1), 13–16 (1961)
46. Watanabe, S.: On discontinuous additive functionals and Lévy measures of a Markov process. Japanese Journal of mathematics **34**, 53–70 (1964)
47. Williams, D.: Probability with Martingales. Cambridge University Press (1991)
48. Yamada, T.: Principal values of Brownian local times and their related topics. In: N. Ikeda, et al. (eds.) Itô's Stochastic Calculus and Probability Theory, pp. 413–422. Springer (1996)
49. Yor, M.: Some Aspects of Brownian Motion: Part II, Some Recent Martingale Problems. Birkhäuser, Basel (1997)

My Encounters with Martingales

Klaus Krickeberg

Abstract

The text below describes the mathematical activities of Klaus Krickeberg during the period when he was working on martingales, along with many other different mathematical subjects. This period extends from 1946 to 1964. His main results about martingales (semi-martingales always admitted) may be summarized as follows: a) processes with a directed index set (Moore-Smith sequences) b) The "Krickeberg-decomposition" of martingales c) The discovery that the covering theorem by the mathematician Giuseppe Vitali in classical analysis implied some form of almost everywhere convergence of martingales d) Based on this he studied systematically sufficient conditions for such convergence which are in some way related to this covering theorem; he called them "Vitali-conditions" e) A convergence theorem including both increasing and decreasing martingales (Børge Jessen problem) f) Stochastic convergence of martingales.

Keywords

Martingales · Vitali conditions

1 Studying at the University of Berlin Right After the War

The University of Berlin had been founded in 1810 during the Napoleonic wars, based on the two ideas of the unity of teaching and research and of classical education for all students. When it reopened in January 1946 after another and by far more devastating

This chapter was previously simultaneously published in the Electronic Journal of History of Probability and Statistics and in Izvestiya NAN Armenii: Matematika, 44 (1), 17–24 (2009).

K. Krickeberg (✉)
Bielefeld, Germany
e-mail: krik@ideenwelt.de

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*,
Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_10

war, most of its buildings lay in ruins but much of the old spirit was alive. In 1949 it was renamed “Humboldt Universität”.

In late summer 1946 I applied for admission as a student of physics. I was only 17 years old and my chances were slim because there were many older applicants who had lost years of their life through the war, and they naturally enjoyed priority. Moreover, since the University was situated in East Berlin and depended on the Soviet and nascent East German administration, there existed already some kind of “affirmative action” in favour of descendants of workers and peasants, whereas my father was a physician. Nevertheless I was admitted, perhaps because I had obtained my “Abitur” (secondary school diploma) in June with the best grade possible.

The material conditions were of course difficult, the worst being hunger. In winter, the temperature inside the large physics auditorium descended sometimes to $-10°$ centigrade. However, there was an enormous enthusiasm for studying as I have seen it later only once more, namely in Hanoi in 1974.

As a student of physics I had to follow the normal basic mathematics curriculum. The analysis (calculus) course was taught by Erhard Schmidt (Gram–Schmidt orthogonalization, Hilbert–Schmidt integral equations). He was already over 70 and retired, but had taken up service again because, as a sequel of the Nazi era and the war, there were not enough mathematics professors in Berlin. He was still the dominating figure among the mathematicians at the University. His lectures were marvelous. He used to prepare them on his way from home to the lecture rooms, never used any notes, and often said in the middle of a proof, “Oh, I think I can prove this in a much better way”, and then he started all over again. He also smuggled in little mistakes to test our attention. After the first semester, in spring 1947, I decided that mathematics was really much more fascinating and switched subjects, physics now becoming my minor.

Schmidt’s five semesters’ (two and a half years’) course covered calculus, functions of a complex variable and elliptic functions. Traditionally, calculus in the second semester in Germany meant differentiation and Riemann integration for functions of two real variables. Instead he started out by presenting set theory, very concretely in the plane but in such a suggestive way that we were prompted to invent by ourselves the axioms of Boolean algebra. Then he did abstract measure theory, following the approach of his longstanding friend Constantin Carathéodory, but again in the concrete setting of the plane, and led us directly to the Lebesgue integral. In the second year he taught, in parallel to functions of a complex variable via the “Riemann” approach, a course on “Complements to calculus” which consisted in fact of functional analysis and some other advanced topics.

Erhard Schmidt was really a geometer. He could “see” what was happening in infinite-dimensional spaces. When he described a projection in a Hilbert space (a *Perpendikel* in his vivid terminology), he showed it to us with his hands. It is this geometric approach that I have taken later when dealing with martingales, after his teaching of measure theory had lead me into probability theory via some detours.

2 Collecting Building Blocks for Martingale Theory

Schmidt retired in 1950, this time for good. Thus he could not advise my doctoral thesis, but it was definitely inspired by him and dealt with geometric measure theory and locally Lipschitzian manifolds. I obtained the German doctoral degree in 1952 and became a lecturer at Humboldt University and a scientific collaborator of the German Academy of Sciences at Berlin. There I came into contact with the group that ran the reviewing journal *Zentralblatt für Mathematik und ihre Grenzgebiete*. During the war and the first years after many mathematical journals had ceased to arrive in Germany. Hence it was decided to publish a *Lückenband* (gap volume) where these missing papers would be reviewed. For this volume I was asked to review among others two papers by Børge Jessen and Erik Sparre Andersen on limit theorems for "integrals" [9] and "set functions" [10], respectively. It was about martingales as we know now, but the word "martingale" did not appear, nor had I ever heard of it. Later on I became aware of a paper of de la Vallée-Poussin [21] from the year 1915 in which he had already proved almost sure convergence of martingales formed by discrete random variables.

In 1950 and 1952 I attended the annual meetings of the German Mathematical Association (DMV). There I met Otto Haupt, a versatile mathematician from Erlangen who had co-authored with Georg Aumann an advanced text on differential- and integral calculus. He was about to prepare a new edition together with a third author, Christian Pauc from France. Pauc had come to Germany in the early forties as a prisoner of war, and Haupt had managed to get him out of the prison camp and bring him to Erlangen to do abstract mathematics with him. After the war, Pauc was promptly accused of "collaboration with the enemy" and could not get a position in France, so he had to exile himself with his family to South Africa. He finally came back some years later to take up a professorship in Nantes (where the street in which the university's School of Technology is situated is now named after him!).

Haupt asked me to read the typewritten manuscript of the new edition of the "Haupt–Aumann–Pauc" [6]. I agreed. It arrived in small installments and I sent back fairly long comments. The authors tried to treat not only integration, i.e. measure theory, but also differentiation in the most abstract setting possible. Motivated by early work of de Possel [3], Pauc had worked on differentiation of generalized interval functions (cell functions) in South Africa with C. A. Hayes [7]. The main issue was a general version of Lebesgue's theorem to the effect that every function of bounded variation defined in an interval is almost everywhere differentiable. These versions in abstract spaces were based on generalizations of Vitali's covering theorem. I then proved [11] that conditions of the Vitali type were also necessary in order that certain statements on upper and lower derivatives be valid.

In May 1953 I left Berlin and moved to Würzburg together with the managing director of the *Zentralblatt*. He had accepted a full professorship there and created a small research team in which I got a temporary position. I read a lot and noticed that probability theory and measure theory were not unrelated. In particular I perused the book by J. L. Doob [5] that had just appeared, where he presented the martingale convergence theorems that he had obtained independently of Jessen and Sparre

Andersen. Both versions are essentially equivalent and he analysed the fine differences. Doob's approach via "upcrossing inequalities" permitted him to treat also semi-martingales.

Next I discovered the counterexample by Dieudonné [4]. It concerned increasing martingales that are countable *Moore-Smith* sequences instead of ordinary sequences, i.e. they are indexed by a general countable *directed* set instead of the positive integers. It showed that even under the usual boundedness conditions, such a martingale need not converge almost surely. On the other hand, I realized that Lebesgue's differentiation theorem for a function f could be formulated as a theorem on the almost sure convergence of the martingale whose (non-denumerable) index set consisted of all decompositions of the interval in which f was defined, into a finite number of subintervals. The semi-order relation in this set which makes it "directed" is "to be a subdivision", and the sigma-algebra whose index is a given decomposition is the one generated by it. I then tried to formulate a generalized Vitali condition concerning *any increasing* martingale with a directed index set, countable or not, that would imply almost sure convergence (leaving aside a technical discussion of"separability"). This was indeed possible. It turned out that in the case of an increasing martingale of bounded variation (i.e. bounded in $\mathbf{L}^1$) with a *totally* (linearly) ordered index set, this condition was trivially satisfied, which yielded another proof of Doob's theorem for discrete or continuous parameters (indices) without using upcrossing inequalities. In the context of classical differentiation theory, the Vitali condition was satisfied by Vitali's covering theorem, which gave Lebesgue's theorem as a particular case.

When I started this work, two technical questions were still open. Firstly, my proof worked only for positive martingales. By generalizing the Jordan-decomposition of functions of bounded variation I showed that every martingale of bounded variation is the difference of two positive ones (Krickeberg decomposition). The second question concerned passing to the limit under an integral sign. By chance, *Zentralblatt* assigned to me for reviewing a book in Italian by Cafiero [1] on set functions, which, although marred by several basic errors, contained a lot on uniform integrability that solved my problem.

At the International Congress of Mathematicians in Amsterdam in 1954 I met Børge Jessen and told him about my preoccupations. He then stated the problem of finding a *single* proof of the almost sure convergence of both increasing and decreasing martingales. Six years later I had the pleasure of giving a lecture in Copenhagen in his presence (in Danish) on the solution of this problem. This lecture took place at the end of my stay at the University of Aarhus (Denmark) as a visiting professor during the academic year 1959/60. Erik Sparre Andersen was then a professor in Aarhus but no longer interested in martingales.

3 A Year in Illinois

In Amsterdam I had also talked to Jerzy Neyman about possibilities of spending some time in the United States. He advised me very kindly regarding studies in Berkeley, but I finally decided to apply for a Research Associateship at the University of Illinois in order to work with Doob. I obtained it and also a Fulbright "travel only"' grant, which paid my voyage in fall 1955 on board of the French steamship "Liberté" (which had been German before the war). I had just got my *Habilitation* at the University of Würzburg, a degree that might vaguely be described as a little bit above the Ph.D., and this entitled me to a First Class ticket for the "Liberté".

I still vividly remember my first meeting with Doob in his office in the University at Urbana, Illinois. When I presented my results to him he said "Oh, that's interesting" and suggested that I publish them in the Transactions of the American Mathematical Society [12]. He also told me about the new work of Paul-André Meyer on stopping times and martingales, which initiated the well-known development that was to exert a dominating influence on the theory of stochastic processes indexed by a time parameter.

I continued being interested in martingales indexed by a directed set and ran a seminar on the topic. A young mathematician from Taiwan, Y. S. Chow, who was at the University of Illinois with a research grant, refined the theory based on Vitali conditions in several ways [2]. The thesis by Helms [8] went into a different direction. He obtained the *mean* convergence in $\mathbf{L}^p$ for any uniformly integrable increasing martingale with a directed index set. I then proved a theorem which, in any systematic presentation, ought to be stated *before* tackling questions of almost sure convergence under such and such condition, namely that *every* $\mathbf{L}^1$-bounded increasing semi-martingale with a directed index set, countable or not, converges *stochastically* [13]. As a curiosity it might be mentioned that this contains the Radon–Nikodym theorem as a particular case, the "Radon–Nikodym integrand" being obtained by "stochastic differentiation".

During this academic year 1955/56 I met L. C. Young from the University of Wisconsin who asked me to spend the following year with him, again as a Research Associate. I liked the idea of abandoning martingales for a while and going back to subjects related to my thesis. It became a very fruitful year, too, with work on geometric measure theory, Laurent Schwartz distributions in $\mathbb{R}^n$ for $n > 1$, and the like.

In the summer of 1957 I sailed back to Le Havre, spent a few days in Paris as arranged by Pauc, and then returned to Würzburg and to martingales.

4 Final Work Till 1964

In the following papers, in order to get rid of the complicated discussions around the concept of a separable stochastic process indexed by a non-denumerable set, I dealt with essential convergence instead of almost sure convergence; in the separable case and in particular for a denumerable index set, the two are trivially equivalent. Proving convergence theorems for decreasing martingales and semi-martingales analogous to those obtained before in the increasing case was not very hard, including a counterexample along the lines of the one that Dieudonné had constructed in the increasing case [14].

The problem of the *necessity* of Vitali conditions for essential convergence was treated in [15]. There, a whole family of Vitali conditions was defined, each of them corresponding to an $\mathbf{L}^p$ space and more generally, following a suggestion by Leopold Schmetterer made in 1958 at a colloquium in Paris, to an Orlicz space. Finally, the paper [20] written jointly with Pauc gave a survey on the whole area of increasing or decreasing martingales and their relations with the finer differentiation theory of functions of several real variables and with differentiation in general spaces.

Schmetterer was, among others, a number theorist and statistician. At his invitation I spent the summer 1958 as a senior assistant (*Oberassistent*) at the University of Hamburg. In the fall of that year, I was offered a full professorship for probability theory and statistics at both the Universities of Cologne and Heidelberg of which I accepted the latter. This meant of course a lot of new work. I had to learn about modern statistics; my only previous experience had been guiding a physician in Würzburg through very simple clinical trials. I also came into contact with many other facets of probability theory and did not want to stay with martingales.

However, there still was Børge Jessen's problem. In order to solve it I looked at any family (Moore-Smith sequence) of sigma-algebras indexed by a directed set but no longer necessarily increasing or decreasing as it is in the case of a family underlying a martingale. Each sigma-algebra defines a corresponding conditional expectation operator in a suitable space $\mathbf{L}$ of random variables, e.g. in $\mathbf{L}^1$. It turned out that one can define a Vitali condition (whose form depends on $\mathbf{L}$) which implies, for every random variable X in $\mathbf{L}$, essential convergence of the corresponding "trajectory", i.e. of the family of the conditional expectations of X with respect to the underlying sigma-algebras. This condition is trivially satisfied if the family of sigma-algebras is increasing or decreasing. Thus, the classical martingale convergence theorems by Jessen, Sparre Andersen and Doob, both in the increasing and decreasing case, are indeed particular cases of this general theorem.

In 1963 I was invited to the All-Union Congress of the Soviet probabilists and mathematical statisticians in Tbilisi. A. N. Kolmogorov chaired the session where I spoke (in Russian) about my general convergence theorem, which I had presented three year before in Copenhagen but not yet published; it then appeared in [16]. The congress was memorable in many respects. Kolmogorov gave a statistical analysis of Pasternak's poetry although the latter had no odour of sanctity with Soviet authorities. I also met H. Cramér for the first time.

The All-Union Congress marked farewell to martingales for me. Much later I started working on point processes. The theory of point processes on the real line makes much use of martingales, but again, I was interested in processes in more general spaces and thus got into stochastic geometry and geometrical statistics [17–19]. I am glad I was never tempted to get involved in the applications of martingales to the theory and, worse, the practice of financial speculations that have contributed in no small measure to the 2008 crisis in the world's money markets and economy.

References

1. Cafiero, F.: Funzioni additive d'insieme ed integrazione negli spazi astratti. Editrice Liguori (1953)
2. Chow, Y.S.: The theory of martingales in an s-finite measure space indexed by directed sets. Ph.D. thesis, University of Illinois (1958)
3. De Possel, R.: Sur la dérivation abstraite des fonctions d'ensemble. Journal de mathématiques pures et appliquées **15**, 391–409 (1936)
4. Dieudonné, J.: Sur un théorème de Jessen. Fundamenta Mathematicae **37**(1), 242–248 (1950)
5. Doob, J.L.: Stochastic Processes. Wiley, New York (1953)
6. Haupt, O., Aumann, G., Pauc, C.Y.: Differential- und Integralrechnung III, 2nd edn. de Gruyter, Berlin (1955)
7. Hayes, C.A., Pauc, C.Y.: Full individual and class differentiation theorems in their relations to halo and Vitali properties. Canadian Journal of Mathematics **7**, 221–274 (1955)
8. Helms, L.L.: Convergence properties of martingales indexed by directed sets. Ph.D. thesis, Purdue University (1956)
9. Jessen, B., Sparre Andersen, E.: Some limit theorems on integrals in an abstract set. Matematisk-fysiske Meddelelser **22**(14), 29 pp. (1946)
10. Jessen, B., Sparre Andersen, E.: Some limit theorems on set-functions. Matematisk-fysiske Meddelelser **25**(5), 8 pp. (1948)
11. Krickeberg, K.: La nécessité de certaines hypothèses de Vitali fortes dans la théorie de la dérivation extrème de fonctions d'intervalle. Comptes rendus **238**, 764–766 (1954)
12. Krickeberg, K.: Convergence of martingales with a directed index set. Transactions of the American Mathematical Society **83**(2), 313–337 (1956)
13. Krickeberg, K.: Stochastische Konvergenz von Semimartingalen. Mathematische Zeitschrift **66**(1), 470–486 (1956)
14. Krickeberg, K.: Absteigende Semimartingale mit filtrierendem Parameterbereich. Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg **24**(1), 109–125 (1960)
15. Krickeberg, K.: Notwendige Konvergenzbedingungen bei Martingalen und verwandten Prozessen. In: Transactions of the Second Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, 1–6 June 1959, pp. 279–305. Czechoslovak Academy of Sciences (1960)
16. Krickeberg, K.: Convergence of conditional expectation operators. Theory of Probability & Its Applications **9**(4), 538–549 (1964)
17. Krickeberg, K.: Invariance properties of the correlation measure of line processes. Izvestija Akad. Nauk. Armjan. SSR, Ser. Fiz.-Mat. Nauk **5**(3), 251–262 (1970). Reprinted in Stochastic Geometry (ed. E. F. Harding and D. G. Kendall), pp. 76–88. Wiley, New York (1974)
18. Krickeberg, K.: The Cox process. In: Symposia Mathematica 9: Calcolo delle probabilità, Istituto nazionale di alta matematica, Roma, 1971, pp. 151–167. Academic Press (1972)
19. Krickeberg, K.: Lectures on point processes (in Vietnamese). Mathematical Institute, Hanoi (1976)
20. Krickeberg, K., Pauc, C.: Martingales et dérivation. Bulletin de la Société Mathématique de France **91**, 455–543 (1963)
21. de la Vallée Poussin, C.: Sur l'intégrale de Lebesgue. Transactions of the American Mathematical Society **16**(4), 435–501 (1915)

Part IV Modern Applications

Martingales in the Study of Randomness

Laurent Bienvenu, Glenn Shafer and Alexander Shen

Abstract

Martingales played an important role in the study of randomness in the twentieth century. In the 1930s, Jean Ville used martingales to improve Richard von Mises's and Abraham Wald's concept of an infinite random sequence, or collective. After the development of algorithmic randomness by Andrei Kolmogorov, Ray Solomonoff, Gregory Chaitin, and Per Martin-Löf in the 1960s, Claus-Peter Schnorr developed Ville's concept in this new context. Along with Schnorr, Leonid Levin was a key figure in the development in the 1970s. While Schnorr worked with algorithmic martingales and supermartingales, Levin worked with the closely related concept of a semimeasure. In order to characterize the randomness of an infinite sequence in terms of the complexity of its prefixes, they introduced new ways of measuring complexity: monotone complexity (Schnorr and Levin) and prefix complexity (Levin and Chaitin).

Keywords

Martingale · Collective · Complexity · Randomness · Semimeasure

L. Bienvenu
LaBRI (Laboratoire Bordelais de Recherche en Informatique), CNRS, Bordeaux, France
e-mail: Laurent.Bienvenu@computability.fr

G. Shafer (✉)
Rutgers Business School, 1 Washington Park, Newark, New Jersey, USA
e-mail: gshafer@business.rutgers.edu

A. Shen
LIRMM, University of Montpellier, CNRS, Montpellier, France
e-mail: alexander.shen@lirmm.fr

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*,
Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_11

1 Introduction

In the 1930s, Jean Ville used martingales to improve Richard von Mises's concept of a random sequence, or collective. When the study of random sequences was revived by Andrei Kolmogorov and others in the 1960s, martingales again found their place.

In its broadest outlines, the story we tell here is about different mathematical formulations of the notion that an individual sequence is random. Richard von Mises proposed to define this notion in terms of limiting frequency and selection rules. Then Ville showed that martingales—i.e., capital processes for gambling strategies—can do the job more completely with respect to the measure-theoretic understanding of probabilities for infinite binary sequences.

The contributions of von Mises, Ville, and Abraham Wald in the 1930s were neglected in the 1940s and 1950s, along with the very notion of an individual random object, as measure theory emerged as an adequate and convenient foundation for probability's theory and applications.

In the 1960s, tools from the theory of computation permitted the revival of the study of randomness. Kolmogorov (and later Gregory Chaitin) proposed to define random objects as objects of maximal complexity. Per Martin-Löf showed that the notion of measure zero can also be made algorithmic. His work on algorithmic measure zero inspired Claus-Peter Schnorr's work on algorithmic martingales. The relations between the definitions of randomness that use complexity, effective measure and martingales were established in the 1970s by Schnorr, Leonid Levin and others. These results now form the basis of algorithmic randomness theory.

We begin the article by reviewing the contributions of von Mises, Wald, and Ville. In Sect. 2, we recall von Mises's concept of a collective, which he first published in 1919. In Sect. 3, we review how Wald, in the 1930s, simplified the concept, clarified it, and demonstrated its consistency. In Sect. 4, we review how Ville, in the thesis and book he published in 1939, defined a stronger concept based on martingales.

After pausing, in Sect. 5, to consider how collectives fell out of fashion in the 1950s, we describe developments in the 1960s and 1970s. In Sect. 6, we review the invention of the concept of algorithmic complexity, describing work by Ray Solomonoff, Kolmogorov, and Chaitin in the 1960s. In Sect. 7, we explain how Martin-Löf came to define randomness for infinite sequences in the mid 1960s. In Sect. 8, we review Schnorr's introduction of algorithmic martingales around 1970. In Sect. 9, we discuss semimeasures, introduced by Levin in a paper with Alexander Zvonkin in 1970, and their relation to martingales. In Sect. 10, we discuss how Schnorr and Levin related randomness to complexity using monotone complexity (discovered by Schnorr and Levin) and prefix complexity (introduced by Levin and rediscovered and made popular by Chaitin). In Sect. 11, we discuss subsequent developments in algorithmic complexity and martingales, particularly those related to von Mises's original project of providing a foundation for probability and its applications.

2 Richard von Mises's Collectives

In two articles published in the *Mathematische Zeitschrift* in 1919 [54,55], the German-Austrian mathematician and philosopher Richard von Mises (1883–1953) proposed basing mathematical probability on infinite imagined sequences, which he called *Kollektiv*, or collectives.

An accomplished and wide-ranging applied mathematician now turning his attention to probability theory, von Mises borrowed the word *Kollektiv* from German scientists and statisticians who had used it more or less as English statisticians used *population*. The astronomer Heinrich Bruns, for example, used the term *Kollektivgegenstand*, or collective object, which he defined as a collection of similar things that can be statistically classified according to a variable label.[1]

As his first example, von Mises mentioned draws from an urn:

> Examples of such infinite imagined sequences are draws from an urn, leading in the case of the usual lottery to a 5-dimensional label space (the coordinates being the five numbers on the ticket drawn), or the molecules of a gas with the 3-dimensional label space of their velocities…[2]

Imagining draws from an urn is one of the oldest ways of thinking about probability. Laplace talked about drawing balls from an infinite urn.[3] Von Mises wanted to turn this talk into something still imaginary but mathematically defined. The idea of a urn with infinitely many balls half black and half white does not make mathematical sense. But the idea of an infinite sequence of 1s and 0s having a limiting relative frequency of 1s equal to $1/2$ does make mathematical sense, even though the infinitude remains imagined. So instead of an urn with infinitely many objects, let's have an infinite sequence of objects—a collective. Instead of a system of urns, let's have a system of collectives. The usual rules for combining probabilities—adding them, multiplying them, Bayes' rule, etc.—will be derived from corresponding operations on collectives. This was the starting point of von Mises's theory of probability and statistics.

Each object a_i in a collective $a_1, a_2, \ldots$ has a label (*Merkmal*). In von Mises's gas example the label consists of 3 real numbers, and so the label space is $\mathbb{R}^3$. Given a subset of E of $\mathbb{R}^3$, von Mises assumed that the proportion of $a_1, \ldots, a_N$'s labels that fall in E converges as N goes to infinity. The number to which it converges is the probability the collective assigns to E. Thus the collective determines a probability distribution on the label space.

---

[1] *Ein Kollektivgegenstand ist eine Vielheit von gleichartigen Dingen, die nach einem veränderlichen Merkmal statistisch geordnet werden kann.* [8, p. 96], in italics in the original.

[2] In the original: Beispiele solcher unendlich gedachter Folgen sind die Ziehungen aus einer Urne, die etwa beim gewöhnlichen Lotto zu einem 5-dimensionalen Merkmalraum führen (Koordinaten sind die 5 Nummern einer Ziehung), oder die Moleküle eines Gases mit dem 3-dimensionalen Merkmalraum ihrer Geschwindigkeiten…[54, p. 70]

[3] See for example §2 of Marie-France Bru and Bernard Bru's chapter, "Borel's denumerable martingales, 1909–1949", in the present volume.

Collectives are imaginary. So are the lines and planes of geometry and the idealized objects of theoretical physics. In von Mises's view, the applied mathematician uses such idealizations to make predictions in the empirical world, retaining only those idealizations that predict well enough to be useful.

In the second 1919 paper in *Mathematische Zeitschrift* [55, p. 57], von Mises formulated two axioms for collectives. For simplicity, we state them for a collective in which the label space has two elements, say 1 and 0 or heads and tails:

I There exists a limiting frequency: if s_N is the number of heads among the first N elements, the ratio s_N/N converges to some real number p as N tends to infinity.

II This limiting frequency is stable: if we select an infinite subsequence using the indices and not the labels of the selected elements, then this subsequence has the same limiting frequency.

In later formulations (in [56], for example), von Mises expanded the second axiom to allow the decision whether to include an element in the subsequence to use knowledge of the labels of preceding elements.

Axiom I reminded many mathematicians of Émile Borel's work on denumerable probabilities, published in 1909 [6]. In this classical article, which introduced a measure-theoretic understanding of infinite sequences of random outcomes, Borel had shown that s_N/N converges to p with probability 1 in the case of independent trials. Axiom II is also appealing. Suppose someone tells you that flipping a coin produced the sequence

$$10101010101010101010101010101\ldots$$

where 1 (heads) and 0 (tails) alternate. Would you believe this? Probably not. The limiting frequency of 1s exists and is equal to 1/2, but the sequence is too regular. This is where axiom II comes in: if one selects from this sequence the bits in even positions, one gets the sequence

$$11111111111111111111111111111\ldots$$

in which the frequency of 1s is different (1 instead of 1/2). We can win for sure by betting only on the trials in this subsequence. By ruling this out, von Mises explained, the second axiom expresses a classical principle, the principle that excludes systems for beating the odds.

Von Mises was influential and persuasive, and within a few years there were textbooks that used his ideas to explain the rules of the probability calculus—e.g., [17, 23]. But many mathematicians realized that the set-theoretic concept of mathematical existence did not allow his second axiom to be satisfied for a value of p strictly between 0 and 1. No matter where the 1s are in an infinite sequence, there exists a mathematical rule that picks out their indices and thus identifies a subsequence in which the limiting frequency of 1s is 1 instead. Georg Pólya and Felix Hausdorff immediately raised this issue, and many others piled on. He met vigorous opposition whenever he presented his ideas to pure mathematicians.[4]

Von Mises never abandoned collectives, confident that there was a core of truth in the idea. His book on the topic, *Wahrscheinlichkeit, Statistik und Wahrheit* [56], based on lectures he had first given at Staßburg in 1914, was published by Springer in Vienna in 1928, with subsequent editions in 1936 and 1951. In 1931 he published a treatise on probability and statistics based on collectives [57]. After fleeing from Berlin to Turkey in 1933, he emigrated to the United States in 1939, where he became a professor at Harvard. The first English edition of *Wahrscheinlichkeit, Statistik und Wahrheit*, entitled *Probability, Statistics, and Truth*, appeared in 1939. After his death in 1953, his widow Hilda Geiringer completed editing the translation into English of the 1951 edition, which appeared as the second English edition in 1957. She also completed a treatise on probability and statistics in English, which appeared in 1964 [60].

Although von Mises found little sympathy with most pure mathematicians, there were mathematicians and philosophers who undertook to reformulate his second axiom in a way that the mathematical existence of collectives could be demonstrated. Among them were the United States mathematician Arthur Copeland (1898–1970), the German philosophers Hans Reichenbach (1891–1953), and the Viennese philosopher Karl Popper (1902–1994). Reichenbach was a colleague of von Mises in Berlin and also emigrated to Turkey and then to the United States. Popper emigrated to New Zealand in 1937 and then to England in 1949. The three authors, Copeland in 1928 [18], Reichenbach in 1932 [67], and Popper in 1935 [66], made suggestions that turned to out to be equivalent to each other and closely related to the concept of a normal number, already developed in Borel's 1909 article. Their suggestions boiled down to requiring stability of the limiting frequency for subsequence selection rules that select just the trials for which the r preceding trials, for a specified r, match a specified string of 1s and 0s of length r. They showed how to construct sequences whose limiting frequencies are not affected by such selections.[5]

Von Mises found this suggestion interesting and instructive, but he did not agree that it captured fully the notion of a collective. In the second edition of *Wahrscheinlichkeit, Statistik und Wahrheit*, published in 1936 [56, 2nd ed., p. 119], he acknowledged that the notion of a subsequence selection rule should be restricted in some way. But he did not think that being able to construct the collective is an advan-

[4] The letters between von Mises and Hausdorff on the topic are reproduced in [31, pp. 825–829]. Hausdorff also had difficulties with von Mises first axiom, but he thought the second axiom raised the greatest logical difficulties (p. 826). Jean Ville recalled that Maurice Fréchet also had the most difficulty with the second axiom; see the chapter of the present volume devoted to Pierre Crépel's interview and correspondence with Ville. Reinhard Siegmund-Schultze has reviewed Pólya and Hausdorff's objections and von Mises's extensive interaction with Pólya [81]. Von Mises received a very negative reception at Göttingen in 1931 [81, p. 493]. We shortly discuss his reception at Geneva in 1937. His 1940 debate with the United States mathematician Joseph Doob, at a meeting of the Institute of Mathematical Statistics at Dartmouth, New Hampshire, is also notable [59,61].

[5] Popper's exposition was not as mathematically precise as that of the other two authors, and the assertion that his proposal was exactly identical with theirs may be a simplification.

tage, because the construction itself would define a selection rule that changes the frequencies. He took what he considered a practical attitude:

> The stipulation that every subsequence selection rule in a collective should leave the limiting frequency unchanged says nothing other than this: We agree that when a collective in a concrete problem is to be subjected to a particular subsequence selection rule, we intend to assume that this subsequence selection would not change the limiting value of relative frequency. My irregularity axiom contains nothing more than this.[6]

This passage was later quoted by Wald [89, p. 82].

Of course, one can question whether any story about an infinite sequence is really practical. This was Kolmogorov's reservation about von Mises's viewpoint; see Kolmogorov's letter to Maurice Fréchet in the chapter "Andrei Kolmogorov and Leonid Levin on Randomness" in the present volume.

For further discussion of the work stimulated by von Mises during the 1920s and 1930s, see Martin-Löf [49], van Lambalgen [36], Chapter 6 of von Plato [65], and S. D. Chatterji's commentary in Hausdorff's collected works [31, pp. 829–833].

3 Abraham Wald's Clarification

Abraham Wald (1902–1950), the most prolific contributor to Karl Menger's mathematical seminar in Vienna, reformulated von Mises's second axiom in a way that was widely accepted as definitive.

Karl Menger (1902–1985), son of the Austrian economist Carl Menger, was associated with Moritz Schlick's seminar on the philosophy of science, which became known in retrospect as the Vienna circle [4,53]. After earning his doctorate in mathematics in 1924, Menger worked for two years on topology with L. E. J. Brouwer in Amsterdam before returning to Vienna, where he eventually obtained a post in the university and organized his own seminar on mathematics. The seminar's proceedings for eight academic years, from 1928/1929 through 1935/1936, were published as a journal, the *Ergebnisse eines Mathematischen Kolloquiums*. Prominent contributors included Nachman Aronszajn, Kurt Gödel, Marston Morse, John von Neumann, Albert Tarski, and Norbert Wiener. In 1998, Springer reprinted the proceedings, along with several introductory articles in English, in a single volume [20].

Wald was the same age as Menger but was a latecomer to the university. He had been born in Transylvania, where his father was an orthodox Jewish baker, and the family had come to Vienna after the Romanian annexation of the region during World War I. Unable to afford study at a Viennese *Gymnasium*, he passed the examination

[6] In the original: Die Festsetzung daß in einem Kollektiv jede Stellenauswahl die Grenzhäufigkeit unverändert läßt, besagt nichts anders als dies: Wir verabreden, daß, wenn in einer konkreten Aufgabe ein Kollektiv einer bestimmten Stellenauswahl unterworfen wird, wir annehmen wollen, diese Stellenauswahl ändere nichts an den Grenzwerten der relativen Häufigkeiten. Nichts darüber hinaus enthält mein Regellosigkeitsaxiom.

for entrance to the university after attending an engineering school and being tutored by his brother in mathematics [62,93]. By the time he completed his doctorate with Menger in 1931, he was contributing to almost every topic in the seminar. Unable to obtain a university post of his own, he continued to participate in the seminar while earning a living in Oskar Morgenstern's economics institute, until the worsening political situation forced Menger to end the seminar in 1936.

Collectives came into Menger's seminar by way of Schlick's, where Menger had been intrigued by a semi-technical exposition of Karl Popper's ideas on the topic. Schlick was averse to inviting Popper into his seminar [4, p. 118], and Menger had not read Popper's *Logik der Forschung* [66], which treated collectives but was just appearing at the time. So Menger asked Popper to come to his own seminar to give a more precise mathematical explanation [52]. According to the seminar's proceedings for 1934/1935 (vol. 7, p. 12; [20, p. 330]), Popper and Wald both spoke in the session of 6 February 1935. Popper's title was "Über nachwirkungsfreie Folgen" (On sequences without aftereffects); Wald's was "Über den Kollektivbegriff" (On the notion of a collective). Wald continued the following week, with the title "Allgemeines Existenztheorem für Kollektiva bezüglich beliebiger Auswahlvorschriften und Merkmalmengen" (General existence theorem for collectives with respect to arbitrary selection procedures and label sets). Wald's work appeared in 1937, in Volume 8 of the seminar's proceedings [88]. In the meantime, he stated the main results without proof in a note that Borel inserted in the Paris *Comptes rendus* for the session of 20 January 1936 [87].[7]

For the binary case, where the collective is a sequence of 1s and 0s, Wald's conclusion was very simple: for any *countable* family of selection rules and for any $p \in (0, 1)$ there exists a continuum of sequences that satisfy axioms I and II with limiting frequency p. This strengthened the result obtained by Copeland, Reichenbach, and Popper, for the class of selection rules they considered was countable. As von Mises wrote in 1936, Wald's results were in the same direction as those of Copeland, Reichenbach, and Popper, but were much more far-reaching.[8]

Wald's results on collectives brought together two mathematical ideas that were in flux in 1935—probability measure and computability. It would be at least another decade before there was a wide consensus among mathematicians in favor of countable additivity for probabilities and in favor of Church's thesis for defining effective calculability. But Wald knew the state of mathematical play when we was writing, and his language and reasoning were precise and appropriate relative to that state of play.

---

[7] In a footnote to the *Comptes rendus* note, Wald says that he will give proofs in the seventh volume of Menger's *Ergebnisse*, the volume for 1934–35. In a letter to von Mises dated 27 April 1936 (Papers of Richard von Mises, HUG 4574.5 Box 3, Folder 1936, Harvard University Archives), Wald states that the article with the proofs will appear in July 1936. But the notice of Wald's talk on 13 February 1935, in the seventh volume of the *Ergebnisse*, states correctly that it is in the eighth volume, which did not appear in print until 1937 [88].

[8] This comment appears in an endnote on p. 274 of the second German edition of *Wahrscheinlichkeit, Statistik und Wahrheit* [56]: "In der gleichen Richtung noch wesentlich weiter gehende Resultate bei A. Wald,…". Wald is not otherwise mentioned in the book.

Von Mises, taking the viewpoint of a statistician rather than that of a pure mathematician, had thought of a collective as a sequence of objects, which are labelled with their values for certain variables. The axioms concerned relative frequencies of these labels. On the advice of Menger and Tarski, Wald simplified this setup, taking the sequence of labels itself to be the collective. Given a label space M, Wald made the notion of a collective relative to

- a set $\mathfrak{M}$ of subsets of M, and
- a set $\mathfrak{S}$ of selection procedures,[9] each selection procedure consisting of a sequence $f = f_1, f_2, \ldots$, where f_i is a mapping from M^{i-1} to $\{0, 1\}$.

A selection procedure f, applied to a sequence $m_1, m_2, \ldots$ of elements of M, selects the subsequence consisting of the m_i for which $f_i(m_1, \ldots, m_{i-1}) = 1$. A sequence $m_1, m_2, \ldots$ of elements of M is a collective with respect to $\mathfrak{M}$ and $\mathfrak{S}$ if for every element L of $\mathfrak{M}$, there is a number p_L that is the limit of the relative frequency of L in the sequence itself and in every infinite subsequence that is produced by one of the selection procedures in $\mathfrak{S}$.

With these definitions, Wald showed the existence of collectives for the probability models used in applications, assuming only that the set $\mathfrak{S}$ of selection procedures is countable. More precisely, he demonstrated this existence when $\mathfrak{M}$ is a set of subsets of M containing M, μ is a nonnegative additive set function on $\mathfrak{M}$ satisfying $\mu(M) = 1$, and either μ is concentrated on countably many points or, more generally, there is a countable algebra $\mathfrak{R}$ of elements of $\mathfrak{M}$ such that

$$\mu(A) = \sup_{R \in \mathfrak{R}\ \&\ R \subseteq A} \mu(R) = \inf_{R \in \mathfrak{R}\ \&\ R \supseteq A} \mu(R) \tag{1}$$

for all $A \in \mathfrak{M}$. This condition is satisfied by the usual probability distributions on Euclidean space.

The class of probability distributions that satisfy Wald's assumptions neither contains nor is contained in the class of Kolmogorov's countably additive probability measures. On the one hand, Wald's second assumption is satisfied by some probability measures that are finitely but not countably additive. On the other hand, it is not satisfied when $\mathfrak{M}$ includes all Lebesgue-measurable sets.

In the decade after Wald's work, a consensus formed in favor of Kolmogorov's assumption of countable additivity. Under this assumption, as William Feller noted in 1939 [24], it is much easier to prove Wald's existence theorems. It is obvious (and was proven by Doob in 1936 [21]) that if you apply a selection procedure s of Wald's type to a sequence of independent random variables $m_1, m_2 \ldots$ on M, each

[9] There is a subtle shift in terminology here. Von Mises had called a subsequence selection rule a *Stellenauswahl*, or "place selection". Wald calls it instead an *Auswahlvorschrift*, or "selection procedure" (in the note in French he used *procédé de choix*). The word *Vorschrift*, which can also be translated as "instruction", puts an emphasis on the computational aspect of the selection that was missing from von Mises's terminology.

governed by the countably additive measure μ, you obtain a subsequence with the same distribution. So for any particular such s and any particular measurable subset A of M, the strong law of large numbers says that the frequency with which A happens in the subsequence selected by s fails to converge to $\mu(A)$ with probability zero. It follows by countable additivity that for any given countable set $\mathfrak{A}$ of measurable subsets and any countable set $\mathfrak{S}$ of selection procedures, the total probability that the convergence fails for some pair (s, A), $s \in \mathfrak{S}$ and $A \in \mathfrak{A}$, is zero. So the complement of this event has probability one and therefore has the cardinality of the continuum.

When can a collective be constructed? To answer this question, Wald introduced the notion of a mathematical object being *constructively defined* (konstruktiv definiert). According to Wald's definition, a sequence of labels $m_1, m_2, \ldots$ is constructively defined if each m_i can be calculated in finitely many steps. Similarly, a selection procedure $f_1, f_2, \ldots$ is constructively defined if each $f_i(m_1, \ldots, m_{i-1})$ can be calculated in finitely many steps, and a set function μ on a sequence $A_1, A_2, \ldots$ of sets is constructively defined if each $\mu(A_i)$ can be calculated in finitely many steps. Wald showed that if $\mathfrak{S}$ is a countable set of constructively defined selection procedures, and $\mathfrak{M}$ is a nonnegative additive set function on M with $\mu(M) = 1$, then there is a constructively defined collective with respect to $\mathfrak{M}$ whose probabilities are identical with μ provided that either

- M is finite and $\mathfrak{M}$ includes all subsets of M, or
- μ is constructively defined on a countable algebra $\mathfrak{R}$ of elements of $\mathfrak{M}$ such that (1) holds for all $A \in \mathfrak{M}$.

Let us explain Wald's proof in the simple case where we consider only a finite system of selection procedures, say a set S consisting of n selection procedures, and we want to construct a collective ω consisting of 1s and 0s with limiting frequency 1/2. Suppose we have constructed $\omega_1 \ldots \omega_{i-1}$ and now want to specify ω_i. Let S_i be the subset of S consisting of the procedures in S that will include the ith entry of ω in the subsequence they select when applied to ω:

$$S_i = \{s \in S \mid s(\omega_1 \ldots \omega_{i-1}) = 1\}.$$

Because we have already constructed $\omega_1 \ldots \omega_{i-1}$, we have determined S_i and also the preceding S_j (those for $1 \leq j < i$). Let k_i be the number of the preceding S_j that are equal to S_i, and set

$$\omega_i = \begin{cases} 1 & \text{if } k_i \text{ is even,} \\ 0 & \text{if } k_i \text{ is odd.} \end{cases}$$

(In particular, $\omega_1 = 1$, because $k_1 = 0$; there are no j satisfying $1 \leq j < 1$.) If we fix a subset A of S and select from ω the subsequence consisting of the ω_i for which $S_i = A$, we get the alternating sequence 101010 . . . . By considering the 2^n different subsets A of S, we partition ω into 2^n subsequences, all equal to 101010 . . . . Each of these has limiting frequency 1/2, and so ω does as well. If we apply a selection procedure $s \in S$ to ω, we pick up the entries in half these 2^n subsequences, those corresponding to the subsets of S that contain s, and the limiting frequency will still be $1/2$.

The proof for countably many selection procedures builds on this simple picture: we add new procedures one by one at intervals so great that the boundary effects cannot affect the limiting frequency.

Now that we have modern notions of computability (partial recursivity, Turing machines, etc.), we can call Wald's concept of constructivity informal. But his argument is quite rigorous from the modern point of view when we replace "constructively defined" by "computable".

Wald notes that the requirement of having only countably many selection procedures is rather weak—so weak that it is not interesting to try weakening it further [88, p. 47]. For example, it is satisfied if we consider selection rules that are definable in some formal theory like *Principia Mathematica*. He notes that if in some application we need only selection procedures from some specific class (as von Mises suggested [56, 2nd ed., p. 117]), and this class is constructively defined, then the existence of a constructively defined collective (proved by Wald) may be useful, notwithstanding the fact that for this collective (as for every constructively defined collective) there is a constructively defined selection procedure that changes the frequencies. He further agrees with von Mises that this picture is not appropriate for games of chance and similar phenomena, where no gambling system should change the frequencies.

In October 1937, Wald presented his results in a celebrated colloquium on probability at Geneva. This colloquium, chaired by Maurice Fréchet, brought together most of the world's leading probabilists for the last time before the war. In addition to Wald and Fréchet, attendees included Harald Cramér, Wolfgang Doeblin, William Feller, Bruno de Finetti, Werner Heisenberg, Eberhard Hopf, Bohuslav Hostinsky, Paul Lévy, Jerzy Neyman, George Polya, and Hugo Steinhaus. The session on foundations was remembered for its lively discussion of collectives, which were criticized by Feller, Fréchet, Lévy, and others. The second installment of the proceedings, published in 1938, included articles on collectives by Wald [89] and von Mises [58]. Wald, still writing in German, stated the theorems he had proven in his *Ergebnisse* article and refuted some of the criticisms. Von Mises, who had not been at the colloquium but wrote in French, embraced Wald's analysis, seeing it as a vindication of his confidence that his axioms were logically consistent.

Wald's introduction of constructivity into the discussion of collectives coincided with a debate among logicians concerning how this notion should be made precise. The debate was motivated by David Hilbert's question of whether there exists a procedure for separating mathematical truths from falsehoods, and it was eventually settled by a consensus around *Church's thesis*, the thesis that effective calculability should be identified with a precise concept that had been given different but equivalent definitions by Alonzo Church and his students, by Gödel, and by Alan Turing (see, e.g., [19]). In 1940 [16], Church suggested using this new precise concept of effective calculability, now usually called simply *computability*, to define collectives. Under Church's definition, a sequence of 1s and 0s is a collective with probability p if the limiting frequency of 1s is p in the sequence and in any subsequence selected by a computable selection rule. With this definition, as Church explained, the existence of collectives can be proven using Wald's construction or Doob's measure-theoretic argument. Wald's construction does not provide a constructive (computable) collective in Church's sense—computable collectives do not exist for obvious reasons. The explanation of this paradox: Wald's construction starts with the enumeration of all selection rules from a given class $\mathfrak{S}$. But the set of all computable selection rules, while being a part of a effectively enumerable set of all partial computable functions (and therefore countable), is not effectively enumerable.

4 Jean Ville's Martingales

Jean André Ville (1910–1989) was participating in Menger's seminar when Karl Popper and Abraham Wald gave their talks on collectives in February 1935. The most brilliant of the first students to earn the diploma in probability that Fréchet introduced at the University of Paris in 1931, Ville had been awarded scholarships to study in Berlin in 1933–1934 and in Vienna in 1934–1935. Fréchet had sent Ville to Berlin to get started on a doctoral thesis using analysis, but Ville had been more interested in von Mises's work even then, and he was fascinated by the new mathematics and new applications he encountered in Menger's seminar.

Ville's training in Borel's denumerable probabilities gave him, nonetheless, a perspective very different from von Mises's. Von Mises began with a single infinite imagined sequence and derived a probability distribution from it. Borel began with a probability measure for a denumerable (countably infinite) sequence. Fréchet, along with Kolmogorov and other functional analysts, were finding in the notion of a probability measure a foundation not only for applications but also for new mathematical theory.

The 19th-century law of large numbers said that if s_N is the number of heads among the first N tosses of a coin that comes up heads with probability p, and N is large, then the ratio s_N/N is close to p with high probability. Borel's strong law of large numbers made the sequence infinite, the "close" into convergence, and the high probability into probability one. Other mathematicians had investigated the rate of convergence, culminating in Khinchin's law of the iterated logarithm, which gave a rate of convergence that will happen with probability one.

According to Cournot's principle,[10] which was popular among French probabilists when Ville was a student, probability theory makes contact with the empirical world only by making predictions with probability near or equal to one. The law of large numbers is one such prediction: the frequency of 1s in a sequence of tosses of a fair coin will converge to $1/2$. The law of the iterated logarithm is another: the frequency will oscillate around $1/2$, converging at a certain specified rate. From this point of view, von Mises was too exclusively focused on the convergence of frequen-

[10] For a history of Cournot's principle and examples of statements embracing it by Jacques Hadamard, Paul Lévy, Maurice Fréchet, and Émile Borel, see [74]. The principle was named after Cournot by Fréchet around 1950.

cies. What about the other predictions probability theory makes with probability one? Will collectives in von Mises's sense also satisfy them? Not necessarily, Ville concluded. There are some probability-one predictions that we cannot guarantee a collective to have through our choice of the system of selection rules. Or to put the point positively, there are properties with measure zero that will be possessed by some collective no matter what system of selection rules we adopt.

Ville first made his point in the *Comptes rendus* in July 1936 [85], in a concise note without examples or proofs that considered only the familiar case of collectives consisting of 1s and 0s with limiting frequency $1/2$. Under the Bernoulli measure, the sequences that are not collectives with respect to a given countable system of selection rules have measure zero. But, Ville asserted, not every property of measure zero can be ruled out in this way. He further asserted that this shortcoming of collectives can be corrected by replacing the system of selection rules by a martingale, i.e., a betting strategy.[11] Ville considered strategies satisfying the following conditions:

1 You start with unit capital.
2 At every trial, you bet only a fraction α of your current capital, where $0 \leq \alpha \leq 1$, on 1 or on 0, so that your capital will remain nonnegative no matter how the trial comes out.

There conditions are natural for describing play in a casino, where you must put the money you bet on the table. It is easy to show that the resulting capital will remain bounded with probability one. So there exists a continuum of sequences for which it remains bounded; we may call these collectives with respect to the betting strategy. In the *Comptes rendus* note, Ville asserted without proof that for any property E to which the Bernoulli measure assigns measure zero, there exists a strategy satisfying his conditions for which the capital is unbounded if E happens. Thus we can rule out any property of measure zero for our collectives by properly choosing the strategy.

For those not steeped in the thinking of the French probabilists, or for whom probability could only mean frequency, Ville's results did not seem well motivated. William Feller, in a two-sentence review in *Zentralblatt* (Zbl 0014.16802), summarized what Ville claimed to have proven while making it clear that he could not figure out why Ville should want to prove it.

Ville's ideas received a fuller hearing the following year, when Fréchet presented them to the colloquium at Geneva as part of a wide-ranging argument against collectives and in favor of the axiomatic approach perfected by Andrei Kolmogorov. In Fréchet's contribution to the colloquium's proceedings [27], published in 1938, we see for the first time in print an example of a property of measure zero that cannot be ruled out by a system of selection rules. Probability theory tells us that the frequency of 1s should oscillate above and below $1/2$ as it converges to $1/2$.

[11] For centuries the word *martingale* has referred to the strategy for betting that doubles one's bet after every loss, and it was also used for more complicated strategies. See Roger Mansuy's chapter on the word *martingale* and Glenn Shafer's chapter "Martingales at the casino" in the present volume.

But a collective need not have this property. Its frequency can instead approach the limit from above, for example. It is instructive to point out (although Fréchet did not) that this happens in the construction by Wald that we reviewed in Sect. 3. The sequence constructed there is the result of interleaving many copies of the sequence $101010\ldots$, and because the frequency of 1s in any prefix (finite initial segment) of each copy is always greater than or equal to $1/2$, this must also be true for the whole sequence. This shows that no matter what countable system of selection rules we adopt, there will be a collective in which the frequency converges to $1/2$ from above. We cannot force the frequency to oscillate above and below $1/2$ as required by the law of the iterated logarithm by a clever choice of the selection rules.

Wald stood his ground. At Geneva, he protested that those who had criticized the theory of collectives for excluding some sequences were now criticizing it because it did not exclude enough sequences [27, p. 35]. In his contribution to the proceedings [89], he questioned whether every asymptotic property should be accorded the same significance as the convergence of frequencies.[12] Then, conceding that strengthening the concept of a collective so as to guarantee other asymptotic properties is of some interest, he proposed a way to do this while preserving von Mises's emphasis on frequencies. Call a selection rule *singular* if the sequences of 1s and 0s for which it produces infinite subsequences have total measure zero, he proposed, and call a collective ω with respect to a countable system S of selection rules *strong* if no singular selection rule in S produces an infinite subsequence when applied to ω. There exists a continuum of strong collectives for any countable system of selection rules.[13] For every property A of probability zero, there is a singular selection rule that produces infinite subsequences when applied to sequences in A; so by adding this selection rule to S, we can guarantee that every strong collective avoids the property A. And we can do this for countably many A.

Fréchet expressed his admiration for Wald's ingenuity but objected that the new concepts weakened the simplicity that made von Mises's picture attractive. We might add that they threaten to push frequencies out of the picture. Why not make all the selection rules singular, and why not combine all the asymptotic properties we want, including the frequency properties, into one property A, to be enforced by means of just one singular selection rule? It takes only one more step to get us to Ville's picture: Define the singular selection rule using a strategy whose capital process is unbounded on A. For example, include the next bit ω_i every time the capital hits a new high. To the best of our knowledge, neither Wald nor anyone else ever promoted the concept of a strong collective further. Wald was simply marshaling every argument he could think of against Fréchet's equally broad offensive.

In an obituary of Wald [93, p. 13], his colleague Jacob Wolfowitz ascribed to him "an unusual aversion to all forms of controversy". The unusual vigor of Wald's

[12] An asymptotic property of $\omega_1\omega_2\ldots$ is one that does not depend on any finite prefix. In 1933 [32], Kolmogorov had shown that the probability of an asymptotic property is either zero or one.

[13] In the case of the singular rules, the sequence must be outside the set of probability zero on which the rule produces an infinite subsequence; in the case of the nonsingular rules, it must be outside the set of probability zero on which the rule produces a subsequence that does not converge to $1/2$.

defense of von Mises may be connected with the fact that he was in touch with von Mises about ways he might escape Vienna. He reported on the Geneva conference to von Mises in a letter dated 18 October 1937, and von Mises's response, on 27 October, included assurances that he would do what he could do to help Wald find a way out.[14]

Hitler annexed Austria in 1938. Wald fled from Vienna to Transylvania and then immigrated to the United States in the summer of 1938; most of his family were murdered by the Germans. One of Wald's first publications in the United States was a very positive review, in 1939 [90], of the English version of the second edition of von Mises's *Wahrscheinlichkeit, Statistik, und Wahrheit*. The review only obliquely touched on Wald's own contribution and made no reference to Ville's. Wald's initial employment in the United States was with the Cowles Commission, which had already offered him a position in 1937, but he quickly moved to Columbia University. In 1946, he became head of a newly created Department of Mathematical Statistics at Columbia. By the time of his death in 1950, in an airplane accident in India, he was widely regarded as the leading voice in mathematical statistics in the world.

Von Mises, like Wald, was unconvinced by Fréchet's arguments. He accepted Ville's theorem that there exist asymptotic properties that have probability zero under the theory of denumerable probabilities (this was Borel's name for the extension of classical probability theory to infinite sequences of trials) and that are satisfied by some collectives, no matter what system of selection rules is adopted. But he saw no problem with this—no reason to modify the theory of collectives to avoid it [58, p. 66].

As for Ville's proposal to substitute a martingale for a system of selection rules, it is not clear that anyone aside from Wald understood Frechet's explanation of Ville's idea at Geneva, and Wald seems never to have mentioned Ville's idea after his response to Frechet. Von Mises admitted that he did not understand Ville's theory [58, p. 66]. De Finetti, in his summary of the colloquium, incorrectly stated Ville's definition of a collective relative to a martingale [25, p. 22]. Decades later, in 1964, Lévy wrote to Fréchet that he had never quite understood Ville's definition of a martingale, and that Michel Loève and Aleksandr Khinchin had told him that they had never understood it either [3, p. 292].

Ville might have been better served to speak for himself. But the work on martingales was his thesis. French practice did not permit him to publish his proofs until the thesis was accepted, and this was delayed by Fréchet's insistence that he add enough analysis to make it respectable. He did this during the academic year 1937–1938, using the concept of a martingale to prove new results for stochastic processes in discrete time and trying to extend these results to continuous time in the framework being developed by Doob. Borel and Fréchet finally allowed Ville to defend his thesis only in March 1939, on the eve of World War II. Borel then published it in his series of monographs on probability [86]. This was a prestigious publication, at least in France, and the book was widely distributed, though apparently not widely read.

[14] Papers of Richard von Mises, HUG 4574.5 Box 3, Folder 1936, Harvard University Archives.

As Fréchet had explained at Geneva, but too cryptically, Ville found it convenient to work not with the strategies he initially called martingales but with the capital processes they determine. A strategy tells us how to bet on ω_n after seeing $\omega_1 \ldots \omega_{n-1}$. In the simple case of 1s and 0s with $p = 1/2$, this means that it tells us, as a function of $\omega_1 \ldots \omega_{n-1}$, whether to bet on $\omega_n = 1$ or $\omega_n = 0$ and how much to bet. The strategy together with the initial capital determines, for every finite string x of 1s and 0s, the capital we will have after observing x, say $m(x)$. The condition that the bets be at even odds dictates that

$$m(x) = \frac{m(x0) + m(x1)}{2}. \tag{2}$$

Any function m satisfying (2) for every finite string x is a capital process arising from a strategy and from some initial capital, and uniquely determines that strategy and initial capital. So Ville found it convenient to "define" the martingale by giving its capital process. Consequently, we now use the word *martingale* not to mean strategy but to mean a function on finite strings satisfying (2). In Ville's story, all martingales were nonnegative. They started with a positive initial capital, say $m(\Box) = 1$, where $\Box$ is the empty string, and satisfied $m(x) \geq 0$ for every finite string x. The capital remained nonnegative because the strategy never bet more than it had.

Each of the selection rules considered by Wald and von Mises excluded a property of measure zero. Wald considered countably many selection rules, and the union of a countable number of sets of measure zero still has measure zero. So Wald could exclude certain properties of measure zero. In the case of 1s and 0s, Ville could do better: he could exclude any property of measure zero, and he could exclude a countable number of them with a single nonnegative martinagle. Nothing was gained by using a countable system of nonnegative martingales. This is because whenever $m_1, m_2, \ldots$ are nonnegative martingales starting with 1, we can choose positive real numbers α_i adding to 1 such that $\sum_i \alpha_i m_i$ is also a nonnegative martingale starting with 1.[15] It is obtained by dividing our initial capital among the strategies that produce the m_i: we assign initial capital α_i to the strategy that makes, at each trial, α_i times the bet made by the strategy that produces m_i when you start with 1. The sum $\sum_i \alpha_i m_i$ is unbounded if one of the m_i is unbounded; it therefore excludes (at least) the union of the sets of measure zero excluded individually by the m_i.

Ville's claim that for any event E of measure zero there exists a nonnegative martingale that is unbounded on E is not difficult to prove from our modern perspective. A set E of measure zero can be covered by an open set U_ε of measure at most ε, for arbitrary small values of ε. For every ε, we consider a martingale $m_\varepsilon(x)$ defined as the fraction of elements from U_ε among the continuations of x. The martingale m_ε starts with capital at most ε and reaches 1 on all elements of U_ε. Then we may consider the sum $2^n U_{4^{-n}}$ that reaches 2^n on all elements of $U_{4^{-n}}$ and, therefore, is unbounded on all elements of E.

[15] See [75, Lemma 1.6].

Notice also that if a nonnegative martingale m is unbounded on an event E, then there is another nonnegative martingale that tends to infinity on E. This is because we can stop the strategy at an arbitrarily large value for m, and by taking a weighted sum of versions of m that stop at increasingly large values (stopping at $1/\alpha_i$, for example, where α_i is the corresponding weight in the weighted sum), we obtain a martingale that tends to infinity on E. So instead of ruling out sequences on which a particular nonnegative martingale is unbounded, we can rule out sequences on which a particular nonnegative martingale tends to infinity.

In the end, however, Ville gave up trying to define the notion of a collective. As its title said, his book was only a *critique* of the notion. On p. 93, he wrote,

> …the irregularity condition using a martingale remains relative. It supposes a prior choice of properties (of probability zero) to exclude. If, in a certain sense, it resolves the question of irregularity more completely than Mr. Wald's condition, it does not succeed in giving an arithmetic model of a sequence having *all* the characteristics of a sequence chosen at random. We consider this last problem insoluble, and we acquiesce on this point[16] to the opinion of numerous mathematicians, including E. Borel, Fréchet, P. Lévy.

Why can we not capture all the characteristics of a sequence chosen at random, or at least all the characteristics that can be effectively defined? On p. 134, in a final philosophical chapter not in the thesis, Ville returned to Wald's definitions and discussed the difficulty later clarified, as we have already noted, by Church in 1940 [16]. As Ville explained, you must avoid supposing that the set of effectively defined selection rules is itself effectively defined, for then you can construct a collective that resists them all and then paradoxically construct a selection rule that defeats it. This was just one example, Ville noted, of how paradoxes arise from non-predicative definitions.

Church's later clarification, according to which a collective resisting all computable selection rules exists but is not computable, can be extended to Ville's stronger notion: an infinite sequence of outcomes for which all computable nonnegative martingales remain bounded exists but is not computable. On the other hand, Ville was correct to conclude that the randomness enforced by any single nonnegative martingale could only be "relative". The set of computable nonnegative martingales, while countable, is not itself computable, and hence it is computationally impossible to average its elements so as to obtain a single computable nonnegative martingale that enforces all the properties enforced by its individual elements.

The negative tone of Ville's final conclusion may have contributed to the subsequent neglect of his ideas, and the shortcomings of his exposition must also have been a factor. On the whole, it remained a thesis rather than a more mature exposition. A whole chapter is devoted to an elaborate notation for working with sequences of 1s and 0s, and another is devoted to Popper and Reichenbach. The simple explanation we have given concerning how to construct a collective that approaches $1/2$

[16] In the original: "nous nous soumettons en ce point"

from above is obscured by the apparatus, to the extent that some recent readers have resorted to working out their own constructions [43].

One of von Mises's arguments for his second axiom was that it prevents a gambler from making money by selecting trials on which to bet. Ville argued that this "principle of the excluded gambling system" should apply equally to a strategy that can vary the amount bet, and so his martingale theory of collectives is a natural strengthening of von Mises's and Wald's theory. But whereas Ville's 1936 note in the *Comptes rendus* had positioned his theory as a new and better theory of collectives, his thesis and book were positioned, as their title said, as a critique of collectives. As he said, he was acquiescing in the views of his mentors. For them, probability theory was now an application of functional analysis and measure theory. To the extent that it is independently axiomatized, it should start with an axiomatic system like Kolmogorov's or like Borel's, which differed from Kolmogorov's in that conditional probability was taken as primitive and related to unconditional probability by the axiom $P(A\&B) = P(A)P(B|A)$ [86, p. 10].

The thesis and book were reviewed in half a dozen mathematical journals. Two of the reviews, de Finetti's review of the thesis in *Zentralblatt* (Zbl 0021.14505) and Doob's review of the book in the *Bulletin of the American Mathematical Society* (45(11):824, 1939), mentioned how martingales could replace systems of selection rules in the definition of collectives. Others gave the impression that Ville was merely reviewing the literature on collectives.

It was only through Doob that Ville's work on martingales contributed to mathematical probability in the second half of the twentieth century. Giving Ville full credit for introducing the concept of a martingale, Doob developed the study of martingales within measure-theoretic probability, where they have become increasingly central. (See Paul-André Meyer's chapter in the present volume.)

5 The Status Quo of the 1950s

The 1937 colloquium at Geneva is sometimes seen as a watershed. A substantial mathematical literature had been devoted to collectives during the 1920s and 1930s, but the Geneva colloquium showed that most probabilists favored working in the measure-theoretic framework of Kolmogorov's axioms. By the 1950s, almost all mathematical work in probability and statistics was in Kolmogorov's framework, and little mathematical attention was being paid to collectives.

Von Mises's collectives did remain a topic of discussion among philosophers and philosophically minded mathematicians and statisticians, at least in the West.[17] Most people, including most philosophers and mathematicians, intuitively identified probability with frequency, and the theory of collectives was the simplest way to make that identification into a theory. The notion of irregularity embodied in von Mises's second axiom was sometimes influential, moreover, even when collectives

[17] Concerning early criticism of von Mises by Soviet philosophers, see Siegmund-Schultze [80].

were not mentioned; see for example R. A. Fisher's comments about relevant subsets in his 1956 book on statistical inference [26, pp. 34–35].

Even among philosophers, however, Ville's concept of a collective based on martingales seems to have completely disappeared by the 1950s. Church's 1940 article [16], often regarded as the last word on collectives, had made no mention of Ville's work. The French logician and philosopher Jean Cavaillès wrote about Ville's ideas in 1940 [10], but his example was not followed by philosophers writing in English. (Cavaillès became a leader in the resistance to the German occupation of France and was shot by the German military in 1944.)

Whereas von Mises energetically promoted his theory for decades, Ville, as we have seen, was already diffident about collectives based on martingales in his 1939 thesis, and he then went on to other things. Mobilized in the fall of 1939 along with all other French reservists, Ville spent a year at the front and then a year in a German prison camp before returning to France in June 1941. During the remainder of the war, he worked mainly on statistical problems, returning to martingales only briefly, when he tried to use them to study Brownian motion but then realized that the results he was obtaining had already been found by different methods by Lévy. After the war, Ville worked on Shannon information, signal theory, and mathematical economics.

The degree to which Ville's collectives based on martingales had been forgotten in the 1950s can be measured by the ill informed praise for his thesis when he was appointed to a chair in econometrics in the Paris Faculty of Sciences in 1959. His fellow mathematician Luc Gauthier, in the report on Ville's work that justified the vote to appoint him, recalled that the thesis had earned the highest praise from Fréchet and Borel. The foundations of measure theory were far from clarified at the time, Gauthier added, and Ville's thesis had strongly contributed to their being put in order.[18]

6 The Invention of the Algorithmic Definition of Randomness in the 1960s

The study of random sequences revived in the 1960s, when it became clear that new ideas from mathematical logic and programming could be used to characterize the complexity of sequences. The complexity of a sequence or other finite object can be defined as the length of the shortest program that generates it (this is *description* complexity, as opposed to *computation* complexity, since we ignore the time and other resources needed), and the most complex objects can be considered random.

[18] In French: "La thèse de Monsieur Jean VILLE, intitulée Étude critique de la notion de Collectif, est une étude sur les fondements du calcul des probabilités, qui a eu les plus vifs éloges de Monsieur FRECHET et de Monsieur BOREL. Il est de fait que les assises de la théorie de la mesure étaient loin d'être clarifiées à l'époque où Jean VILLE a fait sa thèse, et que cette dernière a fortement contribué à sa mise au point." (Archives Nationales, Fontainebleau, Cote 19840325, art. 542.)

The association of randomness with complexity was not new in the 1960s. As Warren Weaver wrote in 1953, in a discussion of Claude Shannon's information theory [92],

> In the physical sciences, the entropy associated with a situation is a measure of the degree of randomness, or of "shuffledness" if you will, in the situation…

In the 1940s, as Weaver explained, Shannon had connected this idea with the coding of messages, measuring the complexity of a message by the length of its shortest encoding.

Shannon considered only very specific encodings, but mathematical logicians later began to study compressibility for finite objects more abstractly. One motivation for this came from the study of unsolvable algorithmic problems. As A. A. Markov, Jr. explained in 1967 [45, p. 161], these problems had arisen in many fields—the theory of algorithms itself, mathematical logic, algebra, analysis, topology, and mathematical linguistics. But they had been formulated in a way too general to be practical:

> The essential feature in common to all unsolvable problems is their great generality: we seek an algorithm applicable to every object of some infinite class and leading invariably to a certain desirable result. But such a formulation of algorithmic problems removes them to some extent from the domain of practical application. In practice, all that is usually required of an algorithm is that it should be applicable to every object of a given class, where the class is finite (though perhaps very large).

One aspect of applicability, Markov further explained, is the size of the algorithm:

> An algorithm is a set of instructions, usually formulated in a precise artificial language. Naturally, we must try to prevent the instructions from being too lengthy, since any algorithm must first be invented; it must be the result of creative activity on our part. But such activity is limited in its powers. If the required algorithm, regarded as a word in some alphabet, necessarily turns out to be very lengthy, our powers of thought will simply be unable to grasp the word.

Markov then suggested measuring the complexity of a Boolean function by the size of the simplest algorithm that computes it: "We shall be interested in the problem of constructing the simplest possible normal algorithms computing a given Boolean function". In other words, a short algorithm computing a Boolean function can be considered as its compressed encoding, and execution of this algorithm implements decompression. One then looks for the shortest possible encoding of a given object (Boolean function).

Markov considered some specific kind of algorithms (he called them *normal algorithms*), and one could ask in which extent his choice affects the results. The key step in defining algorithmic complexity was the realization and demonstration that there exist decompression algorithms for finite objects that are universal and recover the objects from descriptions (encodings) that are the shortest possible, in an asymptotic sense that we will review. The shortest description of an object with respect to such a universal algorithm is the object's algorithmic complexity (or Kolmogorov complexity, as we now say). Once this definition is established, it makes sense to take the second step and say that objects with maximal complexity (i.e., longest descriptions) among the objects of some class are random in this class.

Kolmogorov took these two steps in a celebrated article published in 1965 [34]. In this section, we review what is known about how Kolmogorov came to these ideas. We also discuss two other authors who arrived independently at similar ideas at around the same time: Ray Solomonoff and Gregory Chaitin.

6.1 Kolmogorov

Milestones for the evolution of Kolmogorov's thinking about algorithmic complexity and randomness in the early 1960s are provided by the titles of talks that he gave at the Moscow Mathematical Society, which we list here in translation:

1 Data reduction that conserves information, 22 March 1961.
2 What is information?, 4 April 1961.
3 On tables of random numbers, 24 October 1962. This talk probably corresponds to the article Kolmogorov published in *Sankhyā* in 1963 [33].
4 A complexity measure for finite binary strings, 24 April 1963.
5 Computable functions and the foundations of information theory and probability theory, 19 November 1963.
6 Asymptotic behavior of the complexities of finite prefixes of an infinite sequence), 15 December 1964. The title suggests that this talk might have discussed Martin-Löf's results, but Martin-Löf remembers discussing them with Kolmogorov only the following spring (see Sect. 7).

Three later talks about algorithmic complexity, given from 1967 to 1974, have short published abstracts, which are translated in the chapter in the present volume entitled "Andrei Kolmogorov and Leonid Levin on Randomness".

In his obituary for Kolmogorov written in 1988 [64], K. R. Parthasarathy recalled that Kolmogorov had traveled by sea to India in the spring of 1962 to work at the Indian Statistical Institute and receive an honorary degree from the University of Calcutta. When he arrived in Calcutta, he told the students at the institute about his work, while on the ship, "on tables of random numbers, and the measurement of randomness of a sequence of numbers using ideas borrowed from mathematical logic." This may refer to the work that Kolmogorov published in *Sankhyā* in 1963 [33]. The third talk in the list above, on 24 October 1962, would have been given after he returned to Moscow from India.

In the *Sankhyā* article, Kolmogorov does not yet adopt the idea that maximally complex sequences are random. Instead, he offers a finitary version of von Mises's picture, in which random sequences are those whose frequencies are not changed by the simplest selection rules. In the article, Kolmogorov writes as follows:

> I have already expressed the view ... that the basis for the applicability of the results of the mathematical theory of probability to real 'random phenomena' must depend on some form of the *frequency concept of probability*, the unavoidable nature of which has been established by von Mises in a spirited manner. However, for a long time I had the following views[19]:
>
> (1) The frequency concept based on the notion of *limiting frequency* as the number of trials increases to infinity, does not contribute anything to substantiate the applicability of the results of probability theory to real practical problems where we have always to deal with a finite number of trials.
>
> (2) The frequency concept applied to a large but finite number of trials does not admit a rigorous formal exposition within the framework of pure mathematics.
>
> Accordingly I have sometimes put forward the frequency concept which involves the conscious use of certain not rigorously formal ideas about 'practical reliability', 'approximate stability of the frequency in a long series of trials', without the precise definition of the series which are 'sufficiently large'...
>
> I still maintain the first of the two theses mentioned above. As regards the second, however, I have come to realize that the concept of random distribution of a property in a large finite population can have a strict formal mathematical exposition. In fact, we can show that in sufficiently large populations the distribution of the property may be such that the frequency of its occurrence will be almost the same for all sufficiently large sub-populations, when the *law of choosing these is sufficiently simple*. Such a conception in its full development requires the introduction of a measure of the complexity of the algorithm. I propose to discuss this question in another article. In the present article, however, I shall use the fact that there cannot be a very large number of simple algorithms.

Whereas von Mises considered an infinite binary sequence random if the frequency of 1s has a limit and the selection rules we consider do not change this limit, Kolmogorov now considered a finite binary sequence random if the simplest selection rules do not change the frequency of 1s very much. Whereas Wald had relied on the number of constructible selection rules being countable, Kolmogorov relied on the number of simple rules being finite and relatively small. His formalization of the idea of a selection rule also differed from von Mises; for example, it allowed the decision whether to include a particular term to depend on later as well as earlier terms. He did not, however, consider anything like a martingale for testing randomness. We have no evidence that he ever took notice of Ville's work.

The article was received by *Sankhyā* in April 1963. Kolmogorov's hint that he will write another article showing how to measure the complexity of an algorithm suggests that he may have already worked out the difficulties in defining algorithmic complexity when he submitted the article. This is also suggested by the title of the talk he gave at the Moscow Mathematical Society on 24 April 1963. We can be confident, in any case, that he had the definition by the autumn of 1964, because we have Per Martin-Löf's testimony that he learned about it then from Leonid Bassalygo [51]. Bassalygo confirms this (in a private communication to Alexander Shen); he recalls a walk with Kolmogorov in the early spring or late autumn in which Kolmogorov tried to explain the definition, which he found difficult to grasp.

[19] This is corroborated by a letter Kolmogorov wrote to Fréchet in 1939; see the chapter "Andrei Kolmogorov and Leonid Levin on Randomness" in the present volume.

Bassalygo was not the only person to have difficulty understanding Kolmogorov's definition of algorithmic complexity. The problem lies in sorting out and keeping in mind the sense in which the measurement of complexity is invariant when we change from one universal algorithm to another. If we write $K_{\mathscr{A}}(x)$ for the shortest description of a finite string x by a universal algorithm $\mathscr{A}$ and $K_{\mathscr{B}}(x)$ for the shortest description by a second algorithm $\mathscr{B}$, then the universality of $\mathscr{A}$ implies that there exists a constant c such that

$$K_{\mathscr{A}}(x) \leq K_{\mathscr{B}}(x) + c$$

for all x, no matter how long. Because the constant c might be very large, this inequality has only an asymptotic significance: it says that $\mathscr{A}$ does at least nearly as well as $\mathscr{B}$ for very complex x, those for which $K_{\mathscr{A}}(x)$ and $K_{\mathscr{B}}(x)$ are both so large that c is negligible in comparison. If we compare $\mathscr{A}$ to yet another algorithm $\mathscr{C}$ instead of $\mathscr{B}$, the constant c may change. So when we choose $\mathscr{A}$ as our standard for measuring complexity—i.e., set $K(x)$ equal to $K_{\mathscr{A}}(x)$ and call it the algorithmic complexity of x,[20] we must keep in mind that this algorithmic complexity $K(x)$ is meaningful only up to an arbitrary constant that is independent of x. Because of this arbitrary constant, the number $K(x)$ does not have any meaning or use for a particular string x. But we can use the function K to make asymptotic statements about the complexity of strings as they are made longer and longer. These subtleties and limitations have served as a brake on interest in algorithmic complexity. Some people are confused by the definition; others find it too asymptotic for their taste.

Kolmogorov was the first to publish a precise statement of the definition of algorithmic complexity and a proof of the existence of universal algorithms. In the 1965 article in which he first did this [34], he contrasted this new way of measuring information to the familiar idea of Shannon information or entropy. The proposal to consider maximally complex objects random appears only in a single sentence at the end of the article.

There are now many tutorials that provide further explanations concerning the definition of Kolmogorov complexity and the existence of universal algorithms. See, e.g., [42,78].

6.2 Solomonoff

Kolmogorov's invention of algorithmic complexity was anticipated by Ray Solomonoff (1926–2009). Solomonoff issued technical reports explaining the idea in 1960 and 1962, before Kolmogorov had arrived at it, and he also anticipated Kolmogorov in publication, with articles in *Information and Control* in 1964 [82,83].

Solomonoff was interested in inductive inference. He proposed to formalize Occam's razor by basing predictions on the simplest law that fits the data—i.e.,

[20] Many authors now use $C(x)$ instead of $K(x)$.

the simplest program that generates it. He proved the invariance of the length of this program, which is the same as proving the universality of Kolmogorov's measure of complexity. He also defined a universal prior distribution for prediction by averaging all possible laws, giving smaller weights to laws with longer programs required to describe them, and he conditioned this universal prior on what has been observed so far to make predictions.

The shortcoming of this early work, which helps explain its lack of influence, is its lack of rigor. Solomonoff did not do mathematics with the rigor that might be expected for so abstract a topic. He acknowledged this in the reports and articles themselves. A proof of invariance can be extracted from Solomonoff's article [82], but what is being proven is not clearly stated and the reasoning is introduced with an apology: "an outline of the heuristic reasoning behind this statement will give clues as to the meanings of the terms used and the degree of validity to be expected of the statement itself." Elsewhere in the article, he writes, "If Eq. (1) is found to be meaningless, inconsistent or somehow gives results that are intuitively unreasonable, then Eq. (1) should be modified in ways that do not destroy the validity of the methods used in Sects. 4.1 to 4.3." Kolmogorov's student Leonid Levin remembers that when Kolmogorov instructed him to read and cite Solomonoff, he was frustrated by this aspect of the work and soon gave up.

Kolmogorov made a point of acknowledging Solomonoff's priority in publication after he learned about it. In [35] he wrote: "As far as I know, the first paper published on the idea of revising information theory so as to satisfy the above conditions [dealing with individual objects, not random variables] was the article of Solomonoff [82]. I came to similar conclusions, before becoming aware of Solomonoff's work, in 1963–1964, and published my first article on the subject [34] in early 1965". Unlike Kolmogorov, Solomonoff had not used the concept of algorithmic complexity to define randomness; Solomonoff was interested instead in induction.

Solomonoff's 1964 articles also contain other ideas that were developed much later. In Sect. 3.2 (in the first of the two articles), for example, Solomonoff gives a simple formula for predictions in terms of conditional a priori probability, using monotonic machines much before Levin and Schnorr. In 1978, Solomonoff formally proved that this formula works for all computable probability distributions [84].

6.3 Chaitin

Gregory Chaitin was born in the United States in 1947, into a family from Argentina. He recalls that in an essay he wrote as he entered the Bronx High School of Science in 1962, he suggested that a finite binary string is random if it cannot be compressed into a program shorter than itself [15]. He entered City College in 1964, and after his first year there, in the summer of 1965, he wrote "a single paper that is of a size of a small book" [15]. A condensed version was published in two parts in the *Journal of the ACM*. In the first part, published in 1966 [11], he defines the complexity of a binary string in terms of the size of a Turing machine; in the second, submitted in November 1965 but published only in 1969 [12], he defines complexity more generally, in the same way as Kolmogorov did in his 1965 article.

Chaitin and his family returned to Buenos Aires in 1966, and he joined IBM Argentina as a programmer in 1967. His work on algorithmic complexity made a jump forward when he visited IBM's Watson Laboratory in New York for a few months in 1974. He joined this laboratory full-time in 1975 and spent the period from 1976 to 1985 concentrating on IBM's RISC (Reduced Instruction Set Computer) project. He resumed his work on algorithmic information theory in 1985. After 2000, he worked at the University of Auckland in New Zealand and at the Federal University of Rio de Janeiro.

We will discuss some of Chaitin's work in the 1970s in Sect. 10. His most famous discovery, which we will not discuss in this article, is probably his proof of Gödel's incompleteness theorem based on the Berry paradox [13].

7 Martin-Löf's Definition of Randomness

The Swedish mathematician Per Martin-Löf (born 1942) went to Moscow to study with Kolmogorov during 1964–65, after learning Russian during his military service. In an interview with Alexander Shen [51], he explained that he had not previously worked on randomness and did not immediately do so when he arrived. Kolmogorov first gave him a problem in discriminant analysis, which he solved but considered insufficiently challenging. In late autumn 1964, however, Leonid Bassalygo told him about Kolmogorov's new ideas about complexity and randomness, which he found very exciting. He set about learning about recursive function theory and soon obtained interesting results about unavoidable oscillations in complexity in the prefixes of infinite binary sequences, which he discovered when trying to make the complexity of these prefixes as large as possible.

In March 1965, in a train to the Caucasus, Martin-Löf told Kolmogorov about two theorems he had proven on these oscillations. Kolmogorov was so interested that he asked Martin-Löf to present his results as a sequel to a lecture that Kolmogorov gave in Tbilisi, on their way back to Moscow in late March. Martin-Löf wrote two papers in Russian on the oscillations. The first one he wrote was incorporated into an article that appeared in English in 1971 [50]. The second was published in 1966, as the written version of a presentation to the Moscow Mathematical Society on 2 June 1965, again following Kolmogorov; see [48] for the English translation.

Kolmogorov had been interested in finite sequences, but in order to get away from the finitary theory's annoying constants, Martin-Löf investigated instead the question of how to define randomness for an infinite binary sequence. Martin-Löf's first thought was that an infinite binary sequence $\omega_1\omega_2\dots$ might be considered random if the complexity of a prefix $\omega_1\dots\omega_n$ is always maximal up to a constant, i.e.,

$$K(\omega_1\dots\omega_n) = n + O(1). \tag{3}$$

(This means that there exists a constant c such that $n - c \leq K(\omega_1 \ldots \omega_n) \leq n + c$ for all n.) But there are no sequences with this property, Martin-Löf discovered, because of the unavoidable oscillations in complexity.

By the time he left Moscow in July 1965, Martin-Löf was on his way to a definition of randomness for infinite sequences using an approach that mixed logic with measure theory: effectively null sets. In his interview with Alexander Shen [51], Martin-Löf recalls that although he was not familiar with the work of Wald, Church, and Ville, he had absorbed from his reading of Borel the idea that a random sequence should avoid properties with probability zero, or null sets (see, for example, [7]). It is impossible to avoid all null sets; any single sequence itself has probability zero. But it is possible to avoid countably many null sets, and Martin-Löf realized that only countably many can be effectively constructed.

Whereas Wald had constructed null sets by way of selection rules, and Ville had constructed them by way of martingales, Martin-Löf considered how null sets are defined in measure theory. Consider as usual the simple case of the Bernoulli measure with $p = 1/2$. Ever since Borel's 1909 article, mathematicians had understood that this measure is the same as Lebesgue measure on the interval [0, 1] when each real number in [0, 1] is identified with the sequence of 1s and 0s formed by its dyadic expansion. Measure theory says that a subset A of [0, 1] is null (has measure zero or probability zero) if for every $\varepsilon > 0$ there exists a sequence of intervals covering A whose total measure is at most ε. Martin-Löf called A *effectively null* if there exists an algorithm that takes any positive rational ε as input and generates a sequence of intervals that cover A and have total measure at most ε. It is obvious that the union of all effectively null sets is a null set, since there are only countably many algorithms. Sequences that do not belong to any effectively null set therefore exist and form a set with measure one. These are the sequences Martin-Löf considered random. Now they are called *Martin-Löf random* sequences.

Martin-Löf also proved that the union of all effectively null sets is effectively null—in other words, there exists a largest effectively null set. This maximal set consists of all nonrandom sequences. A set A is effectively null if and only if A is a subset of this maximal effectively null set, i.e., A does not contain any random sequence.

Martin-Löf arrived at his definition and results while back in Sweden during the academic year 1965–66. He published them in 1966, in an article that was received by the journal on 1 April 1966 [47]. Later in April, he gave four lectures on his results at the University of Erlangen-Nürnberg, and notes from his lectures [46], in German, were widely distributed, making his and Kolmogorov's work on complexity and randomness relatively well known in Germany.

In his first Erlangen lecture, Martin-Löf contrasted the foundations for probability proposed by von Mises and Kolmogorov. Von Mises, he explained, wanted to base probability on the concept of a collective, whereas Kolmogorov had proposed to begin with the axioms for probability and base applications on two ideas: that frequency approximates probability when an experiment is repeated, and that an event of very small probability can be expected not to happen on a single trial (Cournot's principle). He cited Ville's book, the Geneva colloquium, and other contributions to the literature on collectives and declared that Ville's counterexample, in which the convergence to $1/2$ is from above, had brought discussion of von Mises's Axiom II to an end for the time being.

In his 1966 article and in his Erlangen lectures, Martin-Löf begins how own contribution with the concept of a universal test for the randomness of finite sequences. This is a reformulation of Kolmogorov's definition of randomness for finite sequences by means of a universal algorithm, but Martin-Löf found it could be adapted more readily to infinite sequences. He showed that there exists a universal sequential test for the randomness of infinite sequences, and that this way of defining randomness for infinite sequences is equivalent to the definition in terms of the maximal effectively null set.

Martin-Löf never had an opportunity to discuss his definition of a random sequence with Kolmogorov, but they were mentioned in a detailed survey article [94], published in 1970 by Leonid Levin and Alexander Zvonkin, two of Kolmogorov's students, on Kolmogorov's suggestion; Kolmogorov carefully reviewed this article and suggested many corrections. In addition to Martin-Löf's results, the article covered other results about complexity and randomness obtained by the Kolmogorov school in Moscow.

Martin-Löf later studied the earlier literature on random sequences in more detail and published a review of it in 1969 in English in the Swedish philosophical journal *Theoria* [49]. This was the first survey in the English language of the work by von Mises, Wald, and Ville, and others that we mentioned in Sects. 2, 3, and 4, and in some respects it rescued Ville from obscurity. Whereas the influence of Ville's martingales in measure-theoretic probability was by way of Doob, its influence in algorithmic randomness seems to have been by way of Martin-Löf.

8 Claus-Peter Schnorr's Computable Martingales

Claus-Peter Schnorr (born 1943), who was looking for new research topics after earning a doctoral degree for work in mathematical logic at Saarbrücken in 1967, encountered algorithmic randomness through the notes from Martin-Löf's Erlangen lectures. Building on Martin-Löf's results, Schnorr brought martingales back into the story. His work on algorithmic martingales during the late 1960s culminated, in 1970, in his habilitation and in a series of lectures that appeared as a book in 1971 [70]. (See also [69,71,72].)

According to Schnorr's talk at Dagstuhl [68], he never read Ville's book, having learned about the notion of a martingale indirectly. Schnorr's book is the first publication in which martingales were used in connection with algorithmic randomness.

Schnorr studied *computable* and *lower semicomputable* martingales. A function f (arguments are finite strings of 1s and 0s, values are reals) is called computable if there is an algorithm that computes the values of f with any given precision: given x and positive rational ε, the algorithm computes some rational ε-approximation to $f(x)$. A function is lower semicomputable if there is an algorithm that, given x, generates a sequence of rational numbers that approach $f(x)$ from below. It is easy to see that f is computable if and only if both f and $-f$ are lower semicomputable.

Schnorr characterized Martin-Löf randomness in terms of martingales as follows: an infinite binary sequence is Martin-Löf random if and only if no lower semicomputable nonnegative martingale wins against it (by becoming unbounded). (The initial capital can be noncomputable in this setting.) He also brought the notion of a *supermartingale*, introduced into measure-theoretic probability by Doob in the 1950s, into the theory of algorithmic randomness. A function m on finite strings is a supermartingale if it satisfies the supermartingale inequality,

$$m(x) \geq \frac{m(x0) + m(x1)}{2}.$$

This can be the capital process of a gambler who is allowed to throw money away at each trial. Schnorr proved that lower semicomputable supermartingales characterize Martin-Löf randomness in the same way as lower semicomputable martingales do.

But Schnorr was dissatisfied with this formulation. He proved that there exists a sequence that wins against all computable martingales but is not Martin-Löf random, and he considered computability more appropriate as a condition on martingales than semicomputability. Why should we generate approximations from below but not above? He concluded that semicomputable martingales (or supermartingales) are too broad a class, and that the corresponding class of sequences, the Martin-Löf random sequences, is too narrow.

Trying to find a definition of randomness that better matched his intuition, Schnorr considered a smaller class of effectively null sets, now sometimes called *Schnorr null*. For an effectively null set A there exists an algorithm that given $\varepsilon > 0$ generates a sequence of intervals that cover A and have total measure *at most* ε. For a Schnorr null set, this total measure should *equal* ε. This may sound a bit artificial, but it is equivalent to asking for a computably converging series of lengths of covering intervals. The sequences that are outside all Schnorr null sets he called random ("zufällig" in German; we now call them *Schnorr random*). Schnorr proved that this class of sequences is indeed larger than the class of Martin-Löf random sequences. He also proved that a sequence is Schnorr random if and only if no computable martingale computably wins on it. Here "computably wins" means that there exists a nondecreasing unbounded computable function $h(n)$ such that the player's capital after n steps is greater than $h(n)$ for infinitely many n.

Schnorr also considered a natural intermediate requirement: no computable martingale wins (computably or not) on a sequence, i.e., all computable martingales are bounded on its prefixes. Schnorr proved that this class (now its members are sometimes called *computably random* sequences) is broader than the class of Martin-Löf random sequences; much later Wang [91] showed that it is still smaller than the class of all Schnorr random sequences.

Schnorr's work during this period also contained many other ideas that endured and were developed further much later. For example, he considers how fast a player's capital increases during the game. If a sequence violates the strong law of large numbers, there exists a computable martingale that wins exponentially fast against it, but the violation of more delicate laws may involve slower growth in the player's capital. Since 2000, the growth of lower semicomputable martingales has been connected to notions of effective dimension [44].

One of Schnorr's goals was to develop concepts of pseudorandomness. An object with a short description can be called pseudorandom if the time needed to decompress short descriptions is unreasonably large. So Schnorr considered complexity with bounded resources in his book. He later worked in computational cryptography, where more recent and more practical theories of pseudorandomness are used [30].

9 Leonid Levin's Semimeasures

Leonid Levin was born in 1948. His semimeasures, which are closely related to supermartingales, were introduced in his 1970 article with Zvonkin [94], mentioned earlier.[21]

Let Σ be the set of all finite and infinite binary sequences, and let Σ_x be the set of all extensions (finite and infinite) of a binary string x. Then $\Sigma_x = \Sigma_{x0} \cup \Sigma_{x1} \cup \{x\}$. A *semimeasure* is a measure on Σ. It is convenient to specify a semimeasure in terms of the value it assigns to Σ_x for each x, say $q(x)$. A nonnegative real-valued function q on finite strings defines a semimeasure if and only if

$$q(x) \geq q(x0) + q(x1) \tag{4}$$

for every finite string x. We usually assume also that $q(\square) = 1$ (this says that the measure assigns the value 1 to the whole set Σ; it is a probability measure). The difference between the two sides of the inequality (4) is the measure of the finite string x. A semimeasure is said to be *lower semicomputable* if the function $x \mapsto q(x)$ is lower semicomputable.

As Levin showed in the article with Zvonkin, lower semicomputable semimeasures are output distributions of randomized algorithms. Consider a black box that has a random bit generator inside and, being started, produces a string of 1s and 0s bit by bit (pausing between each bit for an unpredictable amount of time; we do not want to have an a priori bound for the computation time, so we allow the machine to work as long as needed, and it may even happen that the next bit will never appear). This machine can produce both finite (if no bits appear after some moment) and infinite sequences and therefore determines a probability distribution on Σ. This distribution is a lower semicomputable semimeasure and every lower semicomputable semimeasure (that equals 1 on the entire set Σ) can be obtained in this way.

What is the connection between semimeasures and supermartingales? As Ville had explained in 1939 [86, pp. 88–89], a nonnegative martingale m is a ratio of

[21] Zvonkin is listed as the first author of the article; note that Z comes before L in the Cyrillic alphabet.

two probability measures. To see what this means, write $p(x)$ for the probability the Bernoulli measure with parameter $1/2$ assigns to x being a prefix of the infinite binary sequence. Then $p(x) = (1/2)^n$, where n is the length of x. Because $p(x0) = p(x1) = (1/2)p(x)$, Eq. (2) tells us that

$$m(x)p(x) = m(x0)p(x0) + m(x1)p(x1). \tag{5}$$

If m is nonnegative and starts at 1, this implies that $m(x)p(x)$ can be interpreted as the value assigned to Σ_x by a probability measure. Writing $q(x)$ for $m(x)p(x)$, we have $m(x) = q(x)/p(x)$. Every nonnegative martingale $m(x)$ starting at 1 can be represented in this way, and every such ratio is a nonnegative martingale starting at 1. This generalizes to supermartingales and semimeasures. If q is a semimeasure, then the ratio $q(x)/p(x)$ is a nonnegative supermartingale starting at 1, and every nonnegative supermartingale starting at 1 can be obtained in this way. Lower semicomputable semimeasures correspond to lower semicomputable supermartingales.

The article with Zvonkin also included Levin's proof of the existence of a maximal lower semicomputable semimeasure, called the *universal semimeasure* or *a priori probability on a binary tree*.[22] This is a lower semicomputable semimeasure r such that for any other lower semicomputable semimeasure q there exists a constant c such that

$$r(x) \geq \frac{q(x)}{c}$$

for any finite string x.

Semimeasures can be used to define supermartingales with respect to any measure, not only uniform Bernoulli measure. Ville had already shown that the representation of a martingale as a ratio of measures generalizes to the case where p is any measure on $\{0, 1\}^\infty$: a martingale with respect to p is the ratio of some measure q to p. A supermartingale with respect to an arbitrary measure p is similarly the ratio of a semimeasure q to p. This implies that for any computable measure p there exists a maximal lower semicomputable p-supermartingale: it is the ratio of the universal semimeasure r to p. This connects maximal p-supermartingales for different p: when we switch from semimeasures to supermartingales, one object (the universal semimeasure) is transformed into a family of seemingly different objects (maximal lower semicomputable supermartingales with respect to different measures).

Zvonkin and Levin's 1970 article [94] had the ingredients needed to provide a criterion of randomness in terms of semimeasures: a sequence ω is Martin-Löf random with respect to a computable measure p if and only if the ratio $r(x)/p(x)$ is bounded for prefixes x of ω, where $r(x)$ is the universal semimeasure. (This statement is a reformulation of Schnorr's characterization of Martin-Löf randomness in terms of lower semicomputable supermartingales.) However, Levin discovered this result

[22] Later Levin introduced the universal semimeasure on natural numbers; now it is sometimes called *discrete a priori probability*, while the universal semimeasure on a binary tree introduced in [94] is called *continuous a priori probability*.

only later; see Levin's letters to Kolmogorov in the chapter in the present volume entitled "Andrei Kolmogorov and Leonid Levin on Randomness".

10 Characterizing Martin-Löf Randomness Using Complexity

The goal of characterizing the randomness of an infinite sequence in terms of the complexity of its prefixes was finally achieved in the 1970s by Schnorr and Levin. To do this (and this itself was a very important development), they modified the definition of algorithmic complexity. Schnorr and Levin introduced monotone complexity, and Levin and Chaitin introduced prefix complexity.

The history of these discoveries is complicated, because different people, working independently, sometimes used slightly different definitions, and sometimes the results remained unpublished for several years or were published without proofs in a short and sometimes cryptic form. We begin this section with some biographical information about Levin, which explains in part why this happened with some of his results.

10.1 Leonid Levin in the Soviet Union

In an interview [41], Leonid Levin recalled that as a student in a high school for gifted children in Kiev in 1963–64 he had been thinking about the length of the shortest arithmetic formula with one free variable that is provable for one and only one value of the variable. He realized that he did not know how to make this definition invariant—i.e., how to make the complexity independent of the specific formalization of arithmetic. The following year, 1964–65, he was studying in a boarding school for gifted children in Moscow, founded by Kolmogorov, and he posed his question to Alexey Sossinsky, a teacher there. Sossinsky asked Kolmogorov about the question, and Kolmogorov replied that he had answered it in a forthcoming article.

In January 1966, Levin entered Moscow State University, becoming a first-year undergraduate in the middle of the academic year. This was unusual, but it was permitted for students at Kolmogorov's school that year, because the Soviet Union was changing from an 11-year to a 10-year curriculum. Early during his study at the university, he obtained a result on the symmetry of information, which he hoped to use to convince Kolmogorov to be his adviser. But Kolmogorov was always busy, and the appointment to talk with him was postponed several times from February to August 1967. Finally, when Levin called him again, Kolmogorov agreed to see him and mentioned that he would tell him something he had just discovered—that information is symmetric. Levin was surprised: "But, Andrei Nikolaevich, this is exactly what I wanted to tell you."—"But do you know that the symmetry is only up to logarithmic terms?"—"Yes."—"And you can give a specific example?"—"Yes." The results they had discovered independently were published without proof by Kolmogorov in 1968 [35], and the proofs were published in the 1970 article by Zvonkin and Levin [94].[23] Levin continued to work with Kolmogorov during his undergraduate years, but because Kolmogorov did not officially belong to the Mathematical Logic Division of the Mathematics Department, where Levin was enrolled, Vladimir Uspensky, who had been Kolmogorov's student in the 1950s, served as Levin's official advisor.

The typical track for a future mathematician in the Mathematics Department of Moscow State University at that time was 5 years of undergraduate studies plus 3 years of graduate school. Then the student was supposed to defend a thesis, becoming a "kandidat", which is roughly equivalent to having a doctoral degree in the United States. To enter graduate school after finishing 5 years of undergraduate studies, one needed a good academic record *and* a recommendation from the local communist party and komsomol. Komsomol (Communist Union of Young People) was almost obligatory for those from 14 to 28 years of age. Most university students were members, although there were some exceptions and the requirement was never formalized as a law.

Being Jewish, already a handicap at that time, and also a nonconformist, Levin created a lot of trouble for the local university authorities as an undergraduate. He became an elected local komsomol leader but did not follow the instructions given by his Communist Party supervisors. Noisy and arrogant, as he later described himself [76, p. 152], he got away with his behavior because the local authorities did not want to take disciplinary actions that would show higher-ups they were having difficulties, but this tolerance faded after the Prague Spring of 1968, and when Levin finished his undergraduate studies in 1970, his misbehavior was mentioned in his graduation letter of recommendation. Not surprisingly, he was not admitted to the graduate school. But with the help of the university rector, Ivan Petrovsky, Kolmogorov managed to secure a job for him in the university's statistical laboratory, which Kolmogorov headed.

An individual could defend a "kandidat" thesis without having been enrolled in a graduate program. So Levin prepared a thesis, consisting of results he had published in the 1970 article with Zvonkin, along with a few others. It was clearly impossible to defend it in Moscow, but a defense finally took place in Novosibirsk in Siberia in 1971. Very untypically, it was unsuccessful. Though all the reviews were positive, the jury not only rejected the thesis, but they included reference to Levin's "unclear political position" in their report. It effectively barred him from defending any thesis in the Soviet Union.

Levin realized that he might be soon barred from publishing in Soviet journals, and many of his important results, including the definition of prefix complexity, remained unpublished at the time. So he rushed a number of articles into print from 1973 to 1977. These articles were short and cryptic, containing many claims without proofs and many ideas that were understood only much later.

[23] Kolmogorov's 1968 paper did not mention Levin explicitly, but the Zvonkin and Levin's 1970 paper, carefully edited by Kolmogorov, mentioned that Levin and Kolmogorov independently came to the result, and there is no doubt that Kolmogorov agreed with this remark.

Some of Levin's results also appeared in a paper published in 1974 by Peter Gács. While working in Hungary, Gács had read Kolmogorov's 1965 article, Martin-Löf's lecture notes from Erlangen, and Zvonkin and Levin's 1970 article, and he had begun corresponding with Levin. He spent the 1972–73 academic year in Moscow working with Levin.

Levin was eventually given permission to leave the Soviet Union. As he recalls, the KGB made it known to him through Kolmogorov that they thought emigration was his best option. Kolmogorov did not say whether he agreed with their advice. In 1978, Levin immigrated to the United States, where he became well known for work in a number of areas of theoretical computer science, including one-way functions, holographic proofs, and for discovering (independently from Cook and Karp) the phenomenon of NP-completeness (the article [38] appeared while he was still in Russia).

10.2 Monotone Complexity: Levin and Schnorr

By 1971–72, Levin and Schnorr had both realized, independently, that the oscillations in complexity that had stood in the way of Martin-Löf's goal of characterizing randomness by requiring maximal complexity for all prefixes can be eliminated if the algorithms or machines used to define complexity are required to be monotone in a certain sense.

We see the idea of monotone complexity already in a letter from Levin to Kolmogorov, written in January 1971 or earlier (Letter II in the chapter in the present volume entitled "Andrei Kolmogorov and Leonid Levin on Randomness"). There Levin calls an algorithm A is monotone if whenever $A(x)$ is defined and y is a prefix of x, $A(y)$ is also defined and is a prefix of $A(x)$.[24] Let us define monotone complexity as the minimal length of a program that produces x—as before, but with the additional restriction that the programming language we use (its interpreter) has to be a monotone algorithm. Levin formulates the following criterion: a sequence is Martin-Löf random with respect to a computable measure p if and only if the monotone complexity of its prefixes equals $-\log_2 p(x) + O(1)$. For the uniform Bernoulli measure this means that $\omega_1\omega_2\ldots$ is random if and only if the monotone complexity of $\omega_1\ldots\omega_n$ equals $n + O(1)$. Note that the monotone complexity of any string of length n is at most $n + O(1)$, and this criterion characterizes random sequences as sequences whose prefixes have maximal possible complexity.

Schnorr advocated a version of monotone complexity, which he called *process complexity*, in May 1972 at the Fourth ACM Symposium on the Theory of Computing (STOC), in Denver [71]. In the proceedings, he proved that a sequence is Martin-Löf random if and only if its n-bit prefix has monotone complexity $n + O(1)$. This

[24] If understood literally, this definition does not have the properties mentioned by Levin in his letter; one should allow the output of A to appear bit by bit. The correct definition appeared in Levin's paper [37].

seems to be first time this result appears in print, but as Schnorr pointed out, the basic properties of monotone algorithms had already been studied by himself [70] and by Zvonkin and Levin [94].

In an article that appeared in 1973 [37], Levin formulated essentially the same result using a slightly different version of monotone complexity, which Schnorr adopted in a subsequent article [72]. Levin also noted that the same proof works for *a priori complexity*—i.e., minus the binary logarithm of the universal semimeasure on the binary tree. The characterization of randomness in terms of a priori complexity is equivalent to Schnorr's characterization of randomness in terms of semicomputable supermartingales.

10.3 Prefix Complexity

Prefix complexity can be defined in different ways. First, the prefix complexity of a natural number i can be defined as $-\log_2 m_i$ where m_i is the maximal lower semicomputable converging series of non-negative reals. (A series $\sum_i a_i$ is lower semicomputable if the function $i \mapsto a_i$ is lower semicomputable, i.e., for every i one can effectively generate approximations to a_i from below.) The prefix complexity of binary strings is then defined using some computable bijection between strings and natural numbers. (Of course, we need to prove that there exists a maximal converging lower semicomputable series; this can be done in the same way as for universal semimeasures on the binary tree. This maximal converging lower semicomputable series is also called the universal semimeasure on natural numbers, or discrete a priori probability, as mentioned above.)

Another definition explains the name used: the *prefix complexity* of a string x is the length of the shortest program p, considered as a bit string, that produces x, assuming that the programming language used has the following "prefix" (self-delimiting) property: if some program p produces some output, any extension of it produces the same output.

Levin and Gács were the first to publish a definition of prefix complexity. They did so in Russian in 1974. Levin's 1974 article [39] appeared in English translation in 1976, and Gács' 1974 article, which attributed the idea to Levin, appeared in English translation in 1975 [28] (see [29]). The two authors' articles state, without proof, the equivalence of the two definitions mentioned above. Levin's article refers for details to an unpublished paper of his and to Gács' article. The unpublished paper mentioned by Levin appeared only in 1976 [40].

The prefix complexity defined as $-\log_2 m_i$ (but not the other definition) appeared also in Levin's unpublished 1971 thesis. In the 1970 article [94] there is a footnote suggesting consideration of the a priori probability of the string $0^i 1$ (i zeros followed by one); this quantity coincides with m_i. But this idea is not developed further in the article.

Chaitin independently worked out similar ideas during his work at the Watson Laboratory in 1974, and his resulting article, which appeared in 1975 [14], contained similar definitions and a proof that prefix complexity and minus the logarithm of the maximal converging series are equal—the first published proof of this result. This article by Chaitin is also the first publication to state that prefix complexity characterizes Martin-Löf randomness: a sequence $\omega_1\omega_2\ldots$ is Martin-Löf random with respect to the uniform Bernoulli measure if and only if the prefix complexity of $\omega_1\ldots\omega_n$ is at least $n - O(1)$. (For prefix complexity the upper bound $n + O(1)$ is no longer valid, but the lower bound still provides a randomness criterion.) In the article [14], Chaitin attributes this result to Schnorr: Chaitin suggested the requirement "prefix complexity of $\omega_1\ldots\omega_n$ is at least $n - O(1)$" as the definition of randomness (now this is often called "Chaitin randomness") and Schnorr, acting as a referee of the paper, informed Chaitin about the equivalence. In his talk at Dagstuhl [68], Schnorr says, "I knew the first paper of Chaitin that has been published one year later after the Kolmogorov 1965 paper, but the next important paper made Chaitin one of the basic investigators of complexity. This was a paper on self-delimiting or prefix-free descriptions, and this was published in 1975 in the *Journal of the ACM*. In fact I was a referee of this paper, and I think Chaitin knew this because I've sent my personal comments and suggestions to him, and he used them."

Chaitin's definition of prefix complexity was slightly different from Levin's: whereas Levin required that extensions of a program p that produces x should produce x, too, Chaitin required that such extensions always produce nothing. Both restrictions reflect (in different ways) the intuitive idea of a self-delimiting program, which allows the machine to find out the program has ended without the use of an end-marker. The differences are not important; the two definitions lead to the same quantity up a $O(1)$ term and so are equivalent.

The possibility of switching back and forth between two definitions of prefix complexity (in terms of a series and self-delimiting programs) is an important technical advantage. Another advantage of prefix complexity over complexity as originally defined (*plain* complexity) is that it allows an improvement in the result on symmetry of information originally discovered by Kolmogorov and Levin. We can relate the complexity of a pair to the conditional complexities with an $O(1)$ error term instead of the the logarithmic error term obtained by Kolmogorov and Levin. This was discovered independently by Levin and Chaitin; the first proofs were published in Gács' 1974 article [28] and Chaitin's 1975 article [14].

11 After the 1970s

The mathematical theory of randomness and algorithmic information theory have continued to develop since the seminal works of the 1960s and 1970s. They have benefited from advanced techniques of recursion theory and have found many applications in fields ranging from combinatorics and computation complexity to recursion theory and Hausdorff dimension theory. See Christian Calude's *Information and Randomness. An Algorithmic Perspective* [9] and Ming Li and Paul Vitányi's *An Introduction to Kolmogorov Complexity and Its Applications* [42]; both have copious historical notes. Connections with recursion theory are covered in *Computability and Randomness* by André Nies [63] and in *Algorithmic Randomness and Complexity* by Rod Downey and Denis Hirschfeldt [22]. Some other results about Kolmogorov complexity are covered in [79].

Most of the work on algorithmic randomness since the 1970s has been concerned with infinite sequences. But Kolmogorov was always more interested in finite random objects, because only finite objects can be relevant to our experience. Some of his ideas for using the theory of complexity in probability modeling were extended by his student Evgeny Asarin [1,2].

Martingales, which can have a finite or infinite horizon, have also recently been considered as a foundation for probabilistic reasoning independently of the classical axioms [73,75]. Instead of forbidding a nonnegative martingale to diverge to infinity in an infinite number of trials, one forbids it to multiply its initial capital by a large factor in a finite number of trials. Predictions are made and theorems proven by constructing martingales. Tests are conducted by checking whether martingales do multiply their initial capital handsomely. The picture that emerges is a little different from classical probability theory, because the logic does not depend on there being enough bets to define probability distributions.

Acknowledgements In addition to published sources, we have drawn on interviews with Peter Gács, Leonid Levin, and Per Martin-Löf, and on discussions at a meeting at Dagstuhl in late January and early February 2006. Marcus Hutter recorded historical talks at Dagstuhl by Christian Calude, Claus-Peter Schnorr, and Paul Vitányi and posted them at http://www.hut\discretionary-ter1.net\discretionary-/dagstuhl. We have also profited from discussions with Leonid Bassalygo, Reinhard Siegmund-Schultze, Vladimir Uspensky, Vladimir Vovk, Vladimir Vyugin, and others. We are grateful to Reinhard Siegmund-Schultze for calling to our attention the letters between von Mises and Wald in the Harvard Archives.

Preparation of an earlier version of this chapter, which appeared in the *Electronic Journal for History of Probability and Statistics* in June 2009, was supported in part by ANR grant NAFIT-08-EMER-008-01. An earlier preliminary version [5], by Laurent Bienvenu and Alexander Shen, contains additional information about the history of algorithmic information theory; see also [77].

References

1. Asarin, E.A.: Some properties of Kolmogorov δ-random finite sequences. Theory of Probability and its Applications **32**, 507–508 (1987)
2. Asarin, E.A.: On some properties of finite objects random in the algorithmic sense. Soviet Mathematics Doklady **36**, 109–112 (1988)
3. Barbut, M., Locker, B., Mazliak, L.: Paul Lévy and Maurice Fréchet: 50 years of correspondence in 107 letters. Springer (2014)
4. Becchio, G.: Unexplored Dimensions: Karl Menger on Economics and Philosophy (1923-1938). Emerald (2009)
5. Bienvenu, L., Shen, A.: Algorithmic information theory and martingales (2009). Arxiv:0906.2614
6. Borel, E.: Les probabilités dénombrables et leurs applications arithmétiques. Rendiconti del Circolo Matematico de Palermo **27**, 247–270 (1909)
7. Borel, E.: Probabilité et certitude. Presses Universitaires de France, Paris (1950)
8. Bruns, H.: Wahrscheinlichkeitsrechnung und Kollektivmasslehre. Teubner (1906)
9. Calude, C.S.: Information and Randomness. An Algorithmic Perspective. Springer (1994). Second edition 2002
10. Cavaillès, J.: Du collectif au pari. Revue de métaphysique et de morale **47**, 139–163 (1940)

11. Chaitin, G.J.: On the length of programs for computing finite binary sequences. Journal of the ACM **13**, 547–569 (1966)
12. Chaitin, G.J.: On the length of programs for computing finite binary sequences: statistical considerations. Journal of the ACM **16**, 145–159 (1969)
13. Chaitin, G.J.: Computational complexity and Gödel's incompleteness theorem. ACM SIGACT News **9**, 11–12 (1971)
14. Chaitin, G.J.: A theory of program size formally identical to information theory. Journal of the ACM **22**, 329–340 (1975). Received April 1974; revised December 1974
15. Chaitin, G.J.: Algorithmic information theory. Some recollections. In: C.S. Calude (ed.) Randomness and Complexity. From Leibniz to Chaitin, pp. 423–441. World Scientific (2007). ArXiv:math/0701164 [math.HO]
16. Church, A.: On the concept of a random sequence. Bull. Amer. Math. Soc. **46**(2), 130–135 (1940)
17. Coolidge, J.L.: An Introduction to Mathematical Probability. Oxford University Press, London (1925)
18. Copeland Sr., A.H.: Admissible numbers in the theory of probability. American Journal of Mathematics **50**, 535–552 (1928)
19. Davis, M. (ed.): The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions. Raven Press, New York (1965)
20. Dierker, E., Sigmund, K. (eds.): Karl Menger: Ergebnisse eines Mathematischen Kolloquiums. Springer (1998)
21. Doob, J.L.: Note on probability. Annals of Mathematics **37**(2), 363–367 (1936)
22. Downey, R.G., Hirschfeldt, D.R.: Algorithmic Randomness and Complexity. Springer (2010)
23. Du Pasquier, L.G.: Le calcul des probabilités, son évolution mathématique et philosophique. Hermann, Paris (1926)
24. Feller, W.: Über die Existenz von sogenannten Kollektiven. Fundamenta Mathematicae **32**(1), 87–96 (1939)
25. de Finetti, B.: Compte rendu critique du colloque de Genève sur la théorie des probabilités. Hermann, Paris (1939). This is the eighth installment (number 766) of [?].
26. Fisher, R.A.: Statistical Methods and Scientific Inference. Oliver and Boyd, Edinburgh (1956)
27. Fréchet, M.: Exposé et discussion de quelques recherches récentes sur les fondements du calcul des probabilités. pp. 23–55. in Fréchet, M. et Wavre, R., éditeurs. Colloque consacré à la théorie des probabilités. Hermann (1938)
28. Gács, P.: On the symmetry of algorithmic information. Soviet Math. Doklady **15**(5), 1477–1480 (1974)
29. Gács, P.: Review of Algorithmic Information Theory, by Gregory J. Chaitin. The Journal of Symbolic Logic **54**(2), 624–637 (1989)
30. Goldreich, O.: An Introduction to Cryptography, Vol. 1. Cambridge (2001)
31. Hausdorff, F.: Gesammelte Werke. Band V: Astronomie, Optik und Wahrscheinlichkeitstheorie. Springer (2006)
32. Kolmogorov, A.N.: Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer, Berlin (1933). An English translation by Nathan Morrison appeared under the title *Foundations of the Theory of Probability* (Chelsea, New York) in 1950, with a second edition in 1956.
33. Kolmogorov, A.N.: On tables of random numbers. Sankhyā: The Indian Journal of Statistics, Series A **25**(4), 369–376 (1963)
34. Kolmogorov, A.N.: Three approaches to the quantitative definition of information. Problems of Information Transmission **1**, 1–7 (1965)
35. Kolmogorov, A.N.: Logical basis for information theory and probability theory. IEEE Trans. Inform. Theory **14**, 662–664 (1968). "Manuscript received December 13, 1967. The work is based on an invited lecture given at the International Symposium on Information Theory, San Remo, Italy, September, 1967. Translation courtesy of AFOSR, USAF. Edited by A.V.Balakrishnan."
36. van Lambalgen, M.: Von Mises' definition of random sequences reconsidered. The Journal of Symbolic Logic **52**(3), 725–755 (1987)

37. Levin, L.A.: On the notion of a random sequence. Soviet Math. Doklady **14**(5), 1413–1416 (1973)
38. Levin, L.A.: Universal sequential search problems. Problems of Information Transmission **9**(3), 265–266 (1973)
39. Levin, L.A.: Laws of information conservation (nongrowth) and aspects of the foundation of probability theory. Problems of Information Transmission **10**, 206–210 (1974)
40. Levin, L.A.: Various measures of complexity for finite objects (axiomatic description). Soviet Math. Doklady **17**(2), 522–526 (1976)
41. Levin, L.A.: Interview with Alexander Shen (2008)
42. Li, M., Vitányi, P.: An Introduction to Kolmogorov Complexity and Its Applications, 4th edn. Springer (2019)
43. Lieb, E.H., Osherson, D., Weinstein, S.: Elementary proof of a theorem of Jean Ville (2006). Arxiv:cs/0607054
44. Lutz, J.H.: Dimension in complexity classes. SIAM Journal on Computing **32**, 1236–1259 (2003)
45. Markov Jr., A.A.: Normal algorithms connected with the computation of Boolean functions. Mathematics of the USSR. Izvestija **1**, 151–194 (1967)
46. Martin-Löf, P.: Algorithmen und zufällige Folgen. Vier Vorträge von Per Martin-Löf (Stockholm) gehalten am Mathematischen Institut der Universität Erlangen-Nürnberg (1966). This document, dated 16 April 1966, consists of notes taken by K. Jacobs and W. Müller from lectures by Martin-Löf at Erlangen on 5, 6, 14, and 15 April.
47. Martin-Löf, P.: The definition of random sequences. Information and Control **9**(6), 602–619 (1966)
48. Martin-Löf, P.: On the concept of a random sequence. Theory of Probability and Its Applications **11**(1), 177–179 (1966)
49. Martin-Löf, P.: The literature on von Mises' Kollektivs revisited. Theoria **35**(1), 12–37 (1969)
50. Martin-Löf, P.: Complexity oscillations in infinite binary sequences. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **19**, 225–230 (1971)
51. Martin-Löf, P.: (2008). Private communication to Alexander Shen.
52. Menger, K.: The formative years of Abraham Wald and his work in geometry. Annals of Mathematical Statistics **23**, 14–20 (1952)
53. Menger, K.: Reminiscences of the Vienna Circle and the Mathematical Colloquium. Springer (1994)
54. von Mises, R.: Fundamentalsätze der Wahrscheinlichkeitsrechnung. Mathematische Zeitschrift **4**, 1–97 (1919)
55. von Mises, R.: Grundlagen der Wahrscheinlichkeitsrechnung. Mathematische Zeitschrift **5**, 52–99 (1919)
56. von Mises, R.: Wahrscheinlichkeit, Statistik und Wahrheit. Springer, Vienna (1928). 2nd edition 1936
57. von Mises, R.: Wahrscheinlichkeitsrechnung und ihre Anwendungen in der Statistik und der theoretischen Physik. F. Deuticke, Leipzig and Vienna (1931)
58. von Mises, R.: Quelques remarques sur les fondements du calcul des probabilités. pp. 57–66. in Fréchet, M. et Wavre, R., éditeurs. Colloque consacré à la théorie des probabilités. Hermann (1938)
59. von Mises, R.: On the foundations of probability and statistics. Annals of Mathematical Statistics **12**, 191–205 (1941)
60. von Mises, R.: Mathematical Theory of Probability and Statistics. Academic Press, New York and London (1964). Edited and complemented by Hilda Geiringer.
61. von Mises, R., Doob, J.L.: Discussion of papers on probability theory. Annals of Mathematical Statistics **12**, 215–217 (1941)
62. Morgenstern, O.: Abraham Wald, 1902–1950. Econometrica **19**(4), 361–367 (1951)
63. Nies, A.: Computability and Randomness. Oxford (2009)
64. Parthasarathy, K.R.: Obituary of Andrei Kolmogorov. Journal of Applied Probability **25**(1), 445–450 (1988)

65. von Plato, J.: Creating Modern Probability. Cambridge (1994)
66. Popper, K.R.: Logik der Forschung: Zur Erkenntnistheorie der modernen Naturwissenschaft. Springer, Vienna (1935). English translation, The Logic of Scientific Discovery, with extensive new appendices, Hutchinson, London, in 1959.
67. Reichenbach, H.: Axiomatik der Wahrscheinlichkeitsrechnung. Mathematische Zeitschrift **34**, 568–619 (1932)
68. Schnorr, C.P.: A talk during Dagstuhl seminar 06051, 29 January–3 February 2006. Http://www.hutter1.net/dagstuhl/schnorr.mp3
69. Schnorr, C.P.: A unified approach to the definition of random sequences. Mathematical Systems Theory **5**(3), 246–258 (1971)
70. Schnorr, C.P.: Zufälligkeit und Wahrscheinlichkeit. Eine algorithmische Begründung der Wahrscheinlichkeitstheorie. Springer (1971)
71. Schnorr, C.P.: Process complexity and effective random tests. Journal of Computer and System Sciences **7**, 376–388 (1973)
72. Schnorr, C.P.: A survey of the theory of random sequences. In: R. Butts, J. Hintikka (eds.) Basic problems in methodology and linguistics, pp. 193–211. Reidel (1977)
73. Shafer, G., Vovk, V.: Probability and Finance: It's Only a Game! Wiley, New York (2001)
74. Shafer, G., Vovk, V.: The sources of Kolmogorov's Grundbegriffe. Statistical Science **21**(1), 70–98 (2006). A longer version is at www.probabilityandfinance.com as Working Paper 4.
75. Shafer, G., Vovk, V.: Game-Theoretic Foundations for Probability and Finance. Wiley, Hoboken, New Jersey (2019)
76. Shasha, D., Lazere, V.: Out of their Minds: The Lives and Discoveries of 15 Great Computer Scientists. Springer (1998)
77. Shen, A.: Algorithmic information theory and foundations of probability (2009). Arxiv:0906.4411
78. Shen, A.: Algorithmic information theory and Kolmogorov complexity. In: Measures of Complexity. Festschrift for Alexey Chervonenkis. Springer (2015)
79. Shen, A., Uspensky, V., Vereshchagin, N.: Kolmogorov complexity and algorithmic randomness. American Mathematical Society (2017)
80. Siegmund-Schultze, R.: Mathematicians forced to philosophize: An introduction to Khinchin's paper on von Mises' theory of probability. Science in Context **17**(3), 373–390 (2004). A translation of Khinchin's paper into English, by Oscar Sheynin, follows on pp. 391–422.
81. Siegmund-Schultze, R.: Probability in 1919/20: the von Mises-Pólya controversy. Archive for History of Exact Sciences **60**(5), 431–515 (2006)
82. Solomonoff, R.J.: A formal theory of inductive inference, Part I. Information and Control **7**(1), 1–22 (1964)
83. Solomonoff, R.J.: A formal theory of inductive inference, Part II. Information and Control **7**(2), 224–254 (1964)
84. Solomonoff, R.J.: Complexity-based induction systems: Comparisons and convergence theorems. IEEE Transactions on Information Theory **IT-24**(4), 422–432 (1978)
85. Ville, J.: Sur la notion de collectif. Comptes rendus **203**, 26–27 (1936)
86. Ville, J.: Étude critique de la notion de collectif. Gauthier-Villars, Paris (1939)
87. Wald, A.: Sur la notion de collectif dans le calcul des probabilités. Comptes rendus **202**, 180–183 (1936)
88. Wald, A.: Die Wiederspruchsfreiheit des Kollektivbegriffes der Wahrscheinlichkeitsrechnung. Ergebnisse eines Mathematischen Kolloquiums **8**, 38–72 (1937)
89. Wald, A.: Die Widerspruchfreiheit des Kollectivbegriffes. In: Les fondements du calcul des probabilités, pp. 79–99. Hermann (1938). Second installment (number 735) of [?].
90. Wald, A.: Review of Probability, Statistics and Truth, by Richard von Mises. Journal of the American Statistical Association **34**(207), 591–592 (1939)
91. Wang, Y.: A separation of two randomness concepts. Information Processing Letters **69**(3), 115–118 (1999)
92. Weaver, W.: Recent contributions to the mathematical theory of communication. ETC: A Review of General Semantics **10**(4), 261–281 (1953)

93. Wolfowitz, J.: Abraham Wald, 1902–1950. Annals of Mathematical Statistics **23**(1), 1–13 (1952)
94. Zvonkin, A.K., L. A. Levin, L.A.: The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. Russian Math. Surveys **25**(6), 83–124 (1970)

Encounters with Martingales in Statistics and Stochastic Optimization

Tze Leung Lai

Abstract

In this intellectual memoir, Professor Lai describes how martingales came into his world of mathematical statistics, first in sequential tests and confidence intervals, then in time series, stochastic approximation, sequential design of experiments, and stochastic optimization. He sketches the trajectories of many other statisticians that he met along the way. He emphasizes the roles of Harold Hotelling, Abraham Wald, and Herbert Robbins in their creation of the environment for the study of martingales at Columbia University and then his own subsequent work at Stanford University. At Stanford, he came to see stochastic optimization as a unifying theme for the use of martingales in statistical modeling. In conclusion, he describes the multifaceted applications of martingales to statistics and stochastic optimization in the BigData and MultiCloud era.

Keywords

Martingales · Harold Hotelling · Joseph Doob · Abraham Wald · Herbert Robbins · Sequential experimental design · Stochastic approximation · Multi-armed bandits · Hidden Markov models · Particle filters

1 Introduction

In this chapter I describe my encounters with martingales beginning with my years at Columbia University as a graduate student (1968–1971) and then on the faculty before moving to Stanford University in 1987.

In Sect. 2, I recount how the stage was set for the study of martingales at Columbia beginning in 1931, when Harold Hotelling joined the Department of Economics. I

T. L. Lai (✉)
Department of Statistics, Stanford University, California 94305, USA
e-mail: lait@stanford.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*, Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_12

emphasize the roles of Abraham Wald, whom Hotelling recruited to Columbia in 1938, and Herbert Robbins, who came to Columbia in 1946.

Section 3 discusses Wald's work on the sequential probability ratio test during the Second World War and the subsequent development of sequential tests and confidence intervals, which had its origins in Robbins' collaboration with Siegmund, Darling, and others beginning around 1968, and on which I wrote my Ph.D. thesis. Section 4 recalls Robbins' ambitious agenda for sequential design and analysis and discusses work by him and others on bandit problems. Section 5 discusses the theory of stochastic approximation, a variation on sequential design that began with the work of Robbins and Sutton Monro in 1951. Section 6 explains how I learned time series. Section 7 discusses the diverse applications of martingales to statistics and stochastic optimization with which I have been involved since my move to Stanford in 1987. Section 8 offers some concluding remarks.

2 Setting the Stage

The major figure paving the way for the development of martingales at Columbia was Harold Hotelling (1895–1973). Other major figures were Abraham Wald (1902–1950) and Herbert Robbins (1915–2001).

It was Hotelling who secured a Carnegie Corporation Fellowship for Joseph Leo Doob to study and do research in probability and statistics at Columbia in 1933–1934. In 1935, Doob left Columbia to become associate professor of mathematics at the University of Illinois at Urbana-Champagne (UIUC). In 1938, Hotelling invited Abraham Wald to work with him at Columbia under the Cowles Commission for Research in Economics. In 1946, when Hotelling left Columbia to become the founding chair of the Department of Mathematical Statistics at the University of North Carolina (UNC), Wald followed his example, becoming the founding chair of the Department of Mathematical Statistics at Columbia.

Even at UNC, Hotelling influenced the development of statistics and stochastic approximation at Columbia, for he recruited Herbert Robbins to join his new UNC department as associate professor. Robbins invented compound decision theory and empirical Bayes methods, stochastic approximation and multi-armed bandit theory, and also introduced new approaches to sequential analysis [60], all before leaving UNC in 1953 to become the chair at Columbia and rebuild the department there. Wald had died in 1950, and Jacob Wolfowitz had left Columbia for Cornell in 1951.

In 1968, in the midst of student demonstrations against the Vietnam war at Columbia, both Robbins and I arrived at its Department of Mathematical Statistics. I arrived as a new graduate student, and he returned after spending the previous three years at different universities that had tried to lure him away. I wrote my Ph.D. thesis under the supervision of his former Ph.D. student David Siegmund, whom he recruited from Stanford in 1969, and stayed on the faculty after receiving my Ph.D. in 1971. In 1975–1976, I spent a year on leave at the University of Illinois at Urbana-Champagne (UIUC), and after my return I worked with Robbins on sequential analysis and stochastic optimization until 1985, when he retired from Columbia.

2.1 Harold Hotelling

Harold Hotelling received his Ph.D. in mathematics from Princeton University in 1924 at the age of 29, under the supervision of the famous geometer and topologist Oswald Veblen. Before attending Princeton, he studied mathematics at the University of Washington, where he received his B.A. and M.A. degrees and was encouraged by the famous mathematician Eric Temple Bell, author of *Men of Mathematics*, to switch from pure mathematics to mathematical economics. Hotelling followed Bell's advice in his career, taking a position as research associate at the Food Research Institute of Stanford University after leaving Princeton, followed by his professorship in the Economics Department at Columbia in 1931, where he taught future Nobel laureates Kenneth Arrow and Milton Friedman. In the 1920s and 1930s he published a series of seminal papers on a wide range of topics in economics and statistics, from the general theory of depreciation [33] to the use of error theory to analyze economics trends [94] and problems of taxation and railway and utility rates [34].

As a student of Veblen, Hotelling was adept at the mathematical challenges of multivariate statistical problems, and he was an early advocate of R. A. Fisher's statistical methods. He extended Student's t-statistic to the multivariate case as Hotelling's T^2-statistic and introduced canonical correlation analysis. His 1939 paper on tubes and spheres in n-spaces [35], with applications to tests and confidence regions in nonlinear regression, was a tour de force and was immediately followed by a paper on the volume of tubes by the distinguished mathematician Hermann Weyl in the same journal [37,66]. He also initiated the study of stochastic optimization with "experimental attainment of optimal conditions" in polynomial regression models [36].

In 1946, Hotelling accepted an offer from University of North Carolina at Chapel Hill to build a Department of Mathematical Statistics, complementing the Department of Experimental Statistics at North Carolina State in Raleigh, headed by Gertrude Cox. Hotelling continued his work in multivariate statistics at Chapel Hill, advising Ph.D. students who later became leaders in the field, including Seymour Geisser and Ingram Olkin.

2.2 Abraham Wald

Abraham Wald entered the graduate school at the University of Vienna in 1927 before receiving his bachelor's degree in 1928 from King Ferdinand University in Cluj, Romania. In 1931 he received his Ph.D. in mathematics under Karl Menger, son of the famous economist Carl Menger. His interest in mathematics was in set theory, metric spaces and differential geometry, for which he produced numerous papers in the years 1931–1936. As a Jewish refugee, Wald was unable to find appropriate work elsewhere in Vienna, but Menger approached Oskar Morgenstern, director of the Austrian Institute of Business Cycle Research about finding funds to support him. This led to Wald's working at the Institute with Morgenstern, who soon became captivated by his "great ability, gentleness, and the extraordinary strength with which he attacked his problems." Thanks to Rockefeller Foundation support, Wald became a full research associate of the Institute in 1938.

Wald's work in Menger's seminar and Morgenstern's institute had been noticed in the United States. In 1937 he was invited to joint the staff of the Cowles Commission. According to Morgenstern, Wald was reluctant to leave Vienna but was persuaded to accept the invitation because of the political situation. Morgenstern assured him that his future as a statistician in America was certain. Before his scheduled departure, Wald was dismissed from the institute by its new Nazi director. He and his family fled to Romania, from where he went to the United States. While he was at the Cowles Commission, Hotelling invited him to Columbia.[1]

During the Second World War, Wald became a member of the Statistical Research Group (SRG) at Columbia, where he developed the sequential probability ratio test (SPRT) to improve on traditional fixed-sample-size plans [47, Sect. 2.1]. SRG also evaluated damage to returning aircraft with a view to minimizing losses to enemy fire. Wald derived a useful estimate of the damage distribution from all aircraft that flew and those that returned [63]. When Hotelling left Columbia's Economics Department in 1946, Wald was asked to rejoin the Cowles Commission, then at University of Chicago, with the possibility of eventually building a statistics department there. Eager to keep Wald, Columbia offered to create a Department of Mathematical Statistics with him as chair. Wald accepted the offer, worked further in sequential analysis, and developed a new statistical decision theory. He and his wife died in an airplane crash in India in 1950.

2.3 Herbert Robbins

When an undergraduate at Harvard, Robbins was persuaded by the distinguished mathematician Marston Morse to change his major from English and literature to mathematics, and he received his Ph.D. in differential topology from Harvard under the supervision of Hassler Whitney in 1938 at the age of 23. During 1939–1941, he was an instructor in mathematics at New York University, where he coauthored the popularization *What is Mathematics?* with Richard Courant. He enlisted in the Navy during the Second World War and was demobilized as a lieutenant commander in 1945.

Robbins' interest in probability and mathematical statistics began during the war when he overheard a conversation between two senior naval officers about the random scatter of bomb impacts. Although he lacked the security clearance to pursue this problem during his service, his later work on it led to his fundamental papers on the measure of a random set in in 1944 and 1945 [73,74]. This seminal work in geometric probability paved the way for Hotelling to recruit him to teach "measure theory, probability, analytic methods, etc." as an associate professor at UNC. He first

[1] See [65] and the chapter by Laurent Bienvenu, Glenn Shafer, and Alexander Shen in the present volume.

thought that Hotelling had called the wrong person because he "knew nothing about statistics" [68, pp. 8–11].

After joining UNC in the then-nascent field of statistics, Robbins collaborated with his much more experienced colleagues in several emerging areas, created new research directions with his daring originality, and supervised a number of Ph.D. students who later became well known. One was Sutton Monro, who was born in 1914, a year before Robbins, and who had received his B.S. and M.S. from MIT. Monro had enlisted in the Navy during the Second World War, was demobilized in 1946 and then taught mathematics at the University of Maine at Orono until 1948 before entering the Ph.D. program at UNC. He had wanted to work with Hotelling on finding the optimum of a regression function, a problem subsequently advanced by Box and Wilson in 1951 [14]. Because Hotelling was no longer working in this area, Robbins became his advisor, leading to their seminal 1951 paper on stochastic approximation [78], which we will review in Sect. 5.

Another of Robbins' talented Ph.D. students was Gopinath Kallianpur, who wrote his thesis on stochastic processes in 1951 and then advised a number of outstanding Ph.D. students at the University of Minnesota, Michigan State University, and UNC. Among them was Raoul LePage who taught me when I was a first-year Ph.D. student at Columbia. Robbins also worked with the famous probabilist Pao-Lu Hsu (who later returned to China to lead the development of probability and statistics at Tsinghua and Peking Universities) in their development of complete convergence, which later influenced my own research. He also collaborated with Wassily Hoeffding on a central limit theorem for m-dependent random vectors [32]. Hoeffding received his Ph.D. from University of Berlin in 1940 and stayed at UNC from 1946 when Hotelling recruited him until his death in 1991.

In 1951, while on leave from UNC on a Guggenheim Fellowship to visit the Institute for Advanced Study at Princeton, Robbins invented compound decision theory [75]. He supervised Raj Bahadur's 1950 and James Hannan's 1953 Ph.D. theses on the subject.

In 1953 Robbins moved back to New York to serve as chair of Columbia's department after Wald's tragic death. He served in this position for nearly 30 years, from 1953 to 1965 and then again from 1968 to 1985.

During the first period 1953–1965, Columbia quickly regained its prominence in statistics. Robbins supervised a number of outstanding Ph.D. students, including Herbert Wilf, Cyrus Derman, Vernon Johns, Ester Samuel, and David Siegmund. They worked in combinatorics/graph theory/multilayered slab geometry, Markov decision processes, empirical Bayes, and optimal stopping. The department remained small, but the full professors, Robbins, T. W. Anderson and Howard Levene, interacted closely with other departments, and outstanding junior faculty included Lajos Takács, Jerome Sacks, and Ronald Pyke. The department also had adjunct faculty from Bell Labs and IBM and visiting professors, such as Kai Lai Chung, who co-advised Derman's thesis with Robbins [99].

3 Sequential Testing and Confidence Intervals

This section reviews the early history of sequential testing and confidence intervals, from Wald's work during the Second World War, through the subsequent development rooted in Robbins' collaboration with Siegmund, Darling, and others beginning around 1968, to the early termination of the Beta-Blocker Heart Attack Trial in 1982, which spurred widespread applications of time-sequential survival analysis in clinical trials.

3.1 Wald's Seminal Work During the Second World War

Here I briefly summarize Wald's work on the sequential probability ratio test (SPRT) during the Second World War and its early sequels. Wald's seminal 1945 article on the topic is more fully reviewed in [47, Sect. 2.1].

Wald developed the SPRT in order to speed up quality control testing of ammunition. It tests a simple null hypothesis $H_0 : f = f_0$ versus a simple alternative hypothesis $H_1 : f = f_1$ based on independent and identically distributed (i.i.d.) observations $X_1, X_2, \ldots$ having a common density function f (with respect to some measure m). Let $L_n = \prod_{i=1}^{n}(f_1(X_i)/f_0(X_i))$ be the likelihood ratio statistic based on $X_1, \ldots, X_n$. The SPRT stops sampling at stage

$$N = \inf\{n \geq 1 : L_n \notin (A, B)\}, \tag{1}$$

with $A < 1 < B$, and rejects H_0 if $L_N \geq B$. To analyze the error probabilities of the SPRT, Wald introduced the likelihood ratio identities

$$P_0(L_N \geq B) = E_1\left(L_N^{-1} 1_{\{L_N \geq B\}}\right), \quad P_1(L_N \leq A) = E_0\left(L_N^{-1} 1_{\{L_N \leq A\}}\right). \tag{2}$$

It follows from (2) that $P_0(L_N \geq B) \leq B^{-1} P_1(L_N \geq B)$ and $P_1(L_N \leq A) \leq A P_0(L_N \leq A)$, in which $\leq$ can be replaced by $=$ if L_N has to fall on either boundary exactly (i.e., if there is no overshoot). Ignoring overshoots, Wald used (2) to obtain approximations for the error probabilities $P_0(L_N \geq B)$ and $P_1(L_N \leq A)$.

In 1944 [87], Wald used another identity, which he proved for more general i.i.d. random variables than $\log(f_1(X_i)/f_0(X_i))$ that are the summands of $\log L_n$, to analyze the operating characteristics of the SPRT. Although he derived it without martingale theory by using the particular structure (1) of the stopping rule, Blackwell and Girshick (1946, [10]) used the martingale structure to generalize the identity to more general stopping times. Doob [22, Sect. VII.10] extended the result further to the following form, which he called "the fundamental theorem of sequential analysis." Let $Y_1, Y_2, \ldots$ be i.i.d. random variables and let z be a complex number such that $|\psi(z)| \geq 1$, where $\psi(z) = E(\mathrm{e}^{zY_1})$. Let $S_n = Y_1 + \cdots + Y_n$. Then $\{\mathrm{e}^{zS_n}/(\psi(z))^n, n \geq 1\}$ is a martingale with mean 1. Moreover, if N is a stopping time such that $\max_{n \le N} |\mathscr{R}(e^{zS_n})|$ is a bounded random variable, where $\mathscr{R}(\cdot)$ denotes the real part, then

$$E\left(\frac{e^{zS_N}}{(\psi(z))^N}\right) = 1 \tag{3}$$

by the optional stopping theorem for martingales. In the case of real $z \neq 0$ so that $\psi(z)$ is the moment generating function of Y_1, Bahadur [5] subsequently showed that the left-hand side of (3) is equal to $Q(N < \infty)$ if $\psi(z) < \infty$, where Q is the probability measure under which $Y_1, Y_2, \ldots$ are i.i.d. with common density function $e^{zy}/\psi(z)$ with respect to the original probability measure.

Another tool Wald developed in 1945 [88] to analyze the SPRT was his equation

$$E\left(\sum_{i=1}^{N} Y_i\right) = \mu E(N)$$

for any stopping time N and i.i.d. random variables Y_i with mean μ. Doob derived this result in 1953 [22, Sect. VII.10] by applying the optional stopping theorem to the martingale $\{S_n - n\mu, n \ge 1\}$. In 1965, Chow, Robbins and Teicher [20] used martingale theory to analyze the higher moments $E\left(\sum_{i=1}^{N} Y_i\right)^r$ for $r = 2, 3, 4$. Noting that the likelihood ratio statistics $L_n, n \ge 1$, form a martingale with mean 1 under P_0, Doob [22, Sect. VII.9] used the martingale convergence theorem to show that L_n converges a.s. $[P_0]$ (almost surely, or with probability 1, under P_0). This martingale property, and therefore also the martingale convergence theorem, are in fact applicable to dependent X_i, with joint density function f_n for $X_1, \ldots, X_n$, where the likelihood ratio now takes the form

$$L_n = q_n(X_1, \ldots, X_n)/p_n(X_1, \ldots, X_n),$$

where $f_n = q_n$ under H_1 and $f_n = p_n$ under H_0. Doob showed that the a.s. limit of L_n (under H_0) is 0 when the X_i are i.i.d. except for the case $P_0\{f_1(X_1) = f_0(X_1)\} = 1$, or equivalently, $L_n = 1$ a.s. $[P_0]$. Although Wald had developed from scratch tools to analyze the SPRT in 1945, his approach was essentially of "martingale-type". Alternative approaches that were used subsequently include analytic methods based on the strong Markov property and the fluctuation theory of random walks [25,85,93]. Doob had learned about martingales from Jean Ville and Paul Lévy and had already begun laying the foundations for the martingale approach to the analysis of randomly stopped sums in 1940.[2] So he immediately understood the martingale structure in Wald's work and included Wald's results in the chapter on martingales in his monumental 1953 book [22, Chap. VII]. In [46] I survey likelihood ratio identities and related methods in sequential analysis that have been developed on the foundations laid down by Wald and Doob.

[2] See Bernard Locker's chapter in the present volume.

3.2 Sequential Tests with Power 1 and Confidence Sequences

During the period 1968–1974, Robbins and Siegmund focused their research on the development of martingale methods for boundary crossing probabilities in sequential tests with power 1. In the case of simple hypotheses $H_0 : f = f_0$ and $H_1 : f = f_1$, a one-sided SPRT with stopping rule $\tilde{N} = \inf\{n : L_n \geq B\}$ (i.e., letting $A = 0$ in (1) and rejecting H_0 upon stopping) has power 1 and type I error probability α if B is so chosen that $P_0(\tilde{N} = \infty) = \alpha$. On the other hand, for composite hypotheses of the type $H_0 : \theta \leq 0$ versus $H_1 : \theta > 0$ when $f = f_\theta$, how can power-one tests such that $\sup_{\theta \leq 0} P_\theta(\text{Reject} H_0) \leq \alpha$ be constructed?

In the case where $X_1, X_2, \ldots$ are i.i.d. random variables from an exponential family of densities $f_\theta(x) = \mathrm{e}^{\theta x - \psi(\theta)}$ with respect to P_0 such that $E_0 X_1 = 0$, Darling and Robbins [21] used in 1967 the fact that $Z_n(\theta) := \mathrm{e}^{\theta S_n - n\psi(\theta)}$, $n \geq 1$, is a nonnegative martingale with mean 1 under P_0, where $S_n = \sum_{i=1}^n X_i$, to conclude from Ville's inequality that for $c_i > 1$,

$$\begin{aligned} &P_0\{Z_n(\theta_i) \geq c_i \text{ for some } m_i \leq n < m_{i+1}\} \\ &= P_0\{S_n \geq \theta_i^{-1} \log c_i + n\theta_i^{-1}\psi(\theta_i) \text{ for some } m_i \leq n < m_{i+1}\} \\ &\leq 1/c_i. \end{aligned} \tag{4}$$

By choosing m_i, c_i and θ_i suitably, they derived *iterated logarithm inequalities* of the form

$$P_0\{S_n \geq b_n(\varepsilon) \text{ for some } n \geq 1\} \leq \varepsilon$$

for given $\varepsilon > 0$, where

$$b_n(\varepsilon) \sim (E_0 X_1^2)^{1/2}(2n \log\log n)^{1/2} \quad \text{as } n \to \infty. \tag{5}$$

Instead of letting θ and c vary with n as in (4), Robbins and Siegmund integrated $Z_n(\theta)$ with respect to a probability measure on θ, noting that $\int_0^\infty Z_n(\theta)\mathrm{d}F(\theta)$ is also a nonnegative martingale with mean 1 for any probability distribution F on $(0, \infty)$ and that

$$\int_0^\infty Z_n(\theta)\mathrm{d}F(\theta) \geq c \iff S_n \geq \beta_F(n, c)$$

for $c > 0$, where $x = \beta_F(n, c)$ is the unique positive solution of $\int_0^\infty \mathrm{e}^{\theta x - n\psi(\theta)}\mathrm{d}F(\theta) = c$. Therefore Ville's inequality again yields

$$\begin{aligned} &P_0\{S_n \geq \beta_F(n, c) \text{ for some } n \geq 1\} \\ &= P_0\left\{\int_0^\infty Z_n(\theta)\mathrm{d}F(\theta) \geq c \text{ for some } n \geq 1\right\} \leq c^{-1}. \end{aligned} \tag{6}$$

They also showed how F can be chosen so that the boundary $\beta_F(n, c)$ has the iterated logarithm growth in (5).

They applied and refined this idea in a series of papers during the period 1968–1970; see the references cited in Robbins [77], which was based on his 1968 Wald Lectures, and Robbins and Siegmund [79] who proved the following results for continuous martingales: let $\varepsilon > 0$ and let $\{Z_t, \mathscr{F}_t, t \geq a\}$ be a nonnegative martingale that has continuous sample paths on $\{Z_a < \varepsilon\}$ and converges to 0 in probability as t tends to 0 on $\{\sup_{s>a} Z_s < \varepsilon\}$. Then

$$P\left\{\sup_{t>a} Z_t \geq \varepsilon | \mathscr{F}_a\right\} = Z_a/\varepsilon \ a.s. \quad \text{on } \{Z_a < \varepsilon\}.$$

Consequently, $P\{\sup_{t\geq a} Z_t \geq \varepsilon\} = P\{Z_a \geq \varepsilon\} + \varepsilon^{-1} E(Z_a 1_{\{Z_a<\varepsilon\}})$. Applying this result to $Z_t = f(W_t + b, t + h)$, where W_t is Brownian motion and

$$f(x,t) = \int_0^\infty \mathrm{e}^{\theta x - \theta^2 t/2} \mathrm{d}F(\theta),$$

they showed that for any $b \in \mathbb{R}$, $h \geq 0$ and $a > 0$,

$$\begin{aligned} P\{f(W_t + b, t + h) \geq \varepsilon \text{ for some } t \geq a\} &= P\{f(W_a + b, a + h) \geq \varepsilon\} \\ &+ \frac{1}{\varepsilon}\int_0^\infty \exp\left(b\theta - \frac{h}{2}\theta^2\right) \Phi\left(\frac{\beta_F(a+h,\varepsilon) - b}{\sqrt{a}} - \sqrt{a}\theta\right) \mathrm{d}F(\theta), \end{aligned}$$

where Φ is the standard normal distribution function and $\beta_F(t,\varepsilon) = \inf\{x : f(x,t) \geq \varepsilon\}$.

In 1973, Robbins and Siegmund [82] provided a probabilistic proof of the integral representations, introduced in 1944 by Widder [90], of positive solutions of the heat equation. They showed that the following statements are equivalent for any continuous $f : \mathbb{R} \times (0, \infty) \to [0, \infty)$:

$$\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2} = 0 \text{ on } \mathbb{R} \times (0, \infty), \tag{7}$$

$$f(x,t) = \int_{-\infty}^\infty \mathrm{e}^{\theta x - \theta^2 t/2} \mathrm{d}F(\theta) \text{ for all } x \in \mathbb{R}, t > 0 \text{ and some measure } F, \tag{8}$$

$$f(W_t, t), t \geq 0, \text{ is a martingale.} \tag{9}$$

Since Widder [91] had also established in 1953 integral representations of positive solutions of the heat equation on the half-line $x > 0$ (semi-infinite rod), Robbins and Siegmund [82] considered in 1973 extensions of their result to Brownian motion with reflecting barrier at 0 and to the radial part of 3-dimensional Brownian motion (Bessel process), noting that Brownian motion is recurrent in dimensions 1 and 2 but transient in higher dimensions. They showed that in this case, (9) has to be replaced by

$$f(r_{t\wedge T_a}, t \wedge T_a), t \geq 0, \text{ is a martingale for every } a > 0,$$

where r_t is either reflected Brownian motion or the Bessel process and $T_a = \inf\{t : r_t \le a\}$. The integral representation (8) takes a different form here: it is a sum of two integrals with respect to measures F_1 on the time axis $[0, \infty)$ and F_2 on the space axis $(0, \infty)$. Lai [41] and Sawyer [84] extended these results to a general continuous Markov process X_t on an interval I with endpoints r_0 and r_1, where $-\infty \le r_0 < r_1 \le \infty$. Let $\mathscr{A}$ be the infinitesimal generator of X_t. Then $\partial/\partial t + \mathscr{A}$ is the infinitesimal generator of the space-time process (t, X_t). Suppose $f : I \times [0, \infty) \to \mathbb{R}$ satisfies $(\partial/\partial t + \mathscr{A}) f(x, t) = 0$ for $r_0 < x < r_1$ and $t > 0$, which is an extension of (7). Lai studied the analog of (9) and gave conditions on the boundaries r_0 and r_1 under which $f(X_t, t)$, $t \ge 0$, is a martingale. As an analog of (8), Sawyer derived integral representations of nonnegative weak solutions of $(\partial/\partial t + \mathscr{A}) f$, thereby generalizing the Robbins-Siegmund martingale characterization of Widder's results.

As an alternative to mixture likelihood ratios, Robbins and Siegmund [81,83] introduced adaptive likelihood ratio statistics of the form

$$\tilde{L}_n = \prod_{i=1}^{n} \frac{f_{\hat{\theta}_{i-1}}(X_i)}{f_{\theta_0}(X_i)}$$

to construct power-one tests of $H_0 : \theta \le \theta_0$ versus $H_1 : \theta > \theta_0$ for the parameter θ of an exponential family $f_\theta(x) = \mathrm{e}^{\theta x - \psi(\theta)}$, where $\hat{\theta}_{i-1} \ge \theta_0$ is an estimate (e.g., by constrained maximum likelihood) of θ based on $X_1, \ldots, X_{i-1}$. Note that $\hat{\theta}_{i-1}$ is measurable with respect to the σ-field $\mathscr{F}_{i-1}$ generated by $X_1, \ldots, X_{i-1}$ while X_i is independent of $\mathscr{F}_{i-1}$. Hence $\{\tilde{L}_n, n \ge 1\}$ is still a nonnegative martingale under P_{θ_0} and therefore Ville's inequality can be applied as in (6) to ensure that $P_\theta\{N_\alpha < \infty\} \le P_{\theta_0}\{N_\alpha < \infty\} \le \alpha$ for $\theta \le \theta_0$, where $N_\alpha = \inf\{n : \tilde{L}_n \ge \alpha^{-1}\}$. Robbins and Siegmund [83] showed how $\hat{\theta}_{i-1}$ can be chosen so that $E_\theta N_\alpha$ attains Farrell's asymptotic lower bound [24], as $\theta \downarrow \theta_0$, for $E_\theta T$ subject to the constraint $P_{\theta_0}(T < \infty) \le \alpha$. In 1977, I developed a theory of power-one tests [44], for the parameter θ of a one-parameter exponential family, based on the sequence of sample sums S_n which are sufficient statistics for θ. Taking $\theta_0 = 0$ without loss of generality, I used Wald's equation and likelihood ratio identities to show that for $T_b = \inf\{n \ge n_0 : S_n \ge b(n)\}$,

$$\lim_{\theta \downarrow 0} E_\theta T_b / g(\mu_\theta) = P_0(T = \infty),$$

where $b(\cdot)$ is a continuous upper-class boundary satisfying certain regularity conditions, $t = g(\theta)$ is the solution of $\theta t = b(t)$, and $\mu_\theta = E_\theta X_1 = \psi'(\theta)$. In particular, for an upper-class boundary b such that $b(t) \sim (E_0 X_1^2)^{1/2} (2t \log\log t)^{1/2}$ as $t \to \infty$, the stopping rule T_b attains Farrell's lower bound, which I learned when I visited UIUC in 1975–1976.

Robbins' revolutionary idea of terminating a test only when there is enough evidence against the null hypothesis and his theory of power-one tests were praised by Jerzy Neyman as "a remarkable achievement" [67]. Even though practical constraints on time and resources make open-ended tests infeasible in practice, this achievement in statistical theory paved the way for subsequent breakthroughs. In particular, Lorden's seminal work seminal work on the theory of control charts and change-point detection [61] involves the following connection between the stopping time N of a sequential detection rule and an open-ended test τ: Let τ be a stopping time based on i.i.d. random variables $X_1, X_2, \ldots$, such that $P(\tau < \infty) \leq \alpha$. For $k = 1, 2, \ldots$, let N_k denote the stopping time obtained by applying τ to $X_k, X_{k+1}, \ldots$ and let $N = \min_{k \geq 1}(N_k + k - 1)$. Then N is a stopping time and $EN \geq 1/\alpha$. This allows one to derive the properties of a sequential detection rule from those of its associated power-one test, as Lorden did in relating the CUSUM rule to the one-sided SPRT.

Let $X_1, X_2, \ldots$ be i.i.d. random variables whose common distribution depends on an unknown parameter $\theta \in \Theta$. A sequence of confidence sets $\Gamma_n = \Gamma_n(X_1, \ldots, X_n)$ is called a $(1 - \alpha)$-level *confidence sequence* if

$$P_\theta\{\theta \in \Gamma_n \text{ for all } n \geq 1\} \geq 1 - \alpha \text{ for all } \theta \in \Theta. \tag{10}$$

Darling and Robbins introduced this concept in 1967 [21], relating it to the boundary crossing probabilities developed in that paper by using martingale inequalities. In 1976 [43], I showed that for an exponential family with parameter θ, the Robbins-Siegmund method of mixture likelihood ratio martingales leads to a confidence sequence of intervals which have the desirable property of eventually shrinking to θ if the mixing distribution F is so chosen that $F(I) > 0$ for every open interval I contained in the natural parameter space Θ. I also used invariance with respect to transformation groups to handle nuisance parameters, thereby constructing invariant confidence sequences.

Roger Farrell received his Ph.D. from UIUC in 1959 under Burkholder's supervision and then went to Cornell University. As mentioned earlier, I collaborated with Robert Wijsmann in sequential analysis projects in which we had similar interests. He was interested in exponential boundedness of the stopping rules of invariant sequential tests (e.g., sequential t-tests) at that time, and introduced the term "obstructive distributions" for cases when exponential boundedness fails to hold. The starting point of our paper Lai and Wijsman (1979 [57]) was to answer the question of possible obstructive distributions and their characterizations for the sequential F-test and the sequential rank test of Savage and Sethuraman, which Wijsman asked me shortly after my arrival at UIUC. We formulated the problem more generally, in the framework of a random walk S_n or a univariate function of multidimensional random walks, crossing moving boundaries of the form $f(n) \pm cg(n)$, with which I was more familiar and which I could relate to the works of Breiman, Brown, Chow-Robbins-Teicher and Gundy-Siegmund cited in the paper. I had also just finished the paper [42] in 1975 on Chernoff-Savage statistics and sequential rank tests and therefore Wijsman could bring me up to speed quickly to work on the sequential Savage-Sethuraman rank test. I was particularly attracted to the sequential t- and F-statistics because they are studentized (or "self-normalized") and many years later, after I moved to Stanford, I came up with a definitive solution to the exponential boundedness problem related to them. Wijsman was also interested in the duality between sequential testing and sequential confidence intervals [92].

3.3 BHAT and Time-Sequential Survival Analysis

The early termination of the Beta-Blocker Heart Attack Trial (BHAT) in 1982 was a historic event that led to the application of sequential analysis in testing new drugs and treatments not only in early-phase clinical studies but also in late-phase clinical trials for their regulatory approval. BHAT was an NIH-sponsored multi-center trial with a Data and Safety Monitoring Board (DSMB) that met periodically (roughly once every six months) to review patient accrual and adverse outcomes. Its early termination led to FDA's approval of propranolol, the beta-blocker used in the trial, for treatment of myocardial infarction (MI, or heart attack) and contributed to James Black winning the Nobel Prize in Medicine in 1988 for discovery of the use of propranolol to treat MI.

As I explained in a recent interview [62, p. 161], BHAT's early termination also "caught the immediate attention of the pharmaceutical companies in the New York–New Jersey area, which called me for consultation in designing similar trials for regulatory approval of their drugs (since) I was a recognized expert in sequential experimentation and analysis". This led me and my Ph.D. students Minggao Gu, Zhiliang Ying and Zukang Zheng to learn survival analysis quickly and efficiently to meet the demands of the task at hand. It turned out that martingales provided the quickest way and most efficient tools for our task; we learned the key tools quickly from Richard Gill's 1980 monograph [26]. We expanded the martingale tools later after reading many of the publications cited in [1],[3] and started building a theory of time-sequential analysis which came to fruition after I moved to Stanford, where I attended the lectures of Siegmund, whose 1985 monograph [85] has developed time-sequential proportional hazards regression in V.5 and illustrated its applications with a re-analysis of the BHAT data in V.6. I also attended the lectures on the bootstrap and other resampling methods by Efron and Peter Hall, who was visiting Stanford frequently from 1987 to 1989. I synthesized time-sequential censored rank statistics with resampling and fully developed the versatile "hybrid resampling approach" with successive generations of Ph.D. students whose contributions appear in Chaps. 6 and 7 of [7].

4 Martingales in Sequential Design of Experiments and Bandit Problems

In his 1952 paper [76] introducing sequential design of experiments in an American Mathematical Society meeting, Robbins said:

> We are indebted to Wald for his significant contribution to the theory of sequential design. His 1947 book *Sequential Analysis* states the problem in full generality and giving the outline of a general inductive method of solution. The probability problems involved are

[3] A revision of this 2009 article on martingales in survival analysis, by Aalen, Andersen, Borgan, Gill and Keiding, appears as a chapter in the present volume.

> formidable, since dependent probabilities occur in all their complexity, and explicit recipes are not yet available for handling problems of practical interest. Nevertheless, enough is visible to justify a prediction that future results in the theory of sequential design will be of greatest importance in mathematical statistics and to science as a whole.

Robbins mentioned these problems that are “different from those usually met in the statistical literature”:

(i) optimal stopping in testing hypotheses,
(ii) adaptive selection of the k populations with largest mean reward, and
(iii) determining the maximum of a regression function.

Problems (i) and (iii) are discussed in Sects. 3 and 5, respectively. The remainder of this section will consider the developments of problem (ii) from 1952 to 1987, highlighting the role played by martingales in this development.

In the 1952 paper, Robbins introduced the k-armed bandit problem for $k = 2$. He considered sequential sampling with unknown means to maximize the total expected reward $E(y_1 + \cdots + y_n)$, where y_i has mean μ_1 (or μ_2) if it is sampled from population 1 (or 2) and n is the total sample size. Letting $s_n = y_1 + \cdots + y_n$, he applied the law of large numbers to show that $\lim_{n\to\infty} n^{-1} E s_n = \max(\mu_1, \mu_2)$ is attained by the following rule:

> Sample from the population with the larger sample mean except at times belonging to a designated sparse set T_n of times, and sample from the population with the smaller sample size at these designated times.

T_n is “sparse” if $\#(T_n) \to \infty$ but $\#(T_n)/n \to 0$ as $n \to \infty$, where $\#(\cdot)$ denotes the cardinality of a set.

Thirty years later, in 1985, Lai and Robbins [54] developed a definitive solution to the problem of finding the optimal rate of convergence for $n \max(\mu_1, \ldots, \mu_k) - E\left(\sum_{t=1}^n y_t\right)$. Their key idea was to formulate an “adaptive allocation rule” ϕ as a sequence of random variables $\phi_1, \ldots, \phi_n$ with values in the set $\{1, \ldots, k\}$ and such that the event $\{\phi_i = j\}$, $j \in \{1, \ldots, k\}$, belongs to the σ-field $\mathscr{F}_{i-1}$ generated by the previous observations $\phi_1, y_1, \ldots, \phi_{i-1}, y_{i-1}$. Letting $\mu(\theta) = E_\theta y$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$, it follows that

$$E_{\boldsymbol{\theta}}\left(\sum_{t=1}^n y_t\right) = \sum_{t=1}^n \sum_{j=1}^k E_{\boldsymbol{\theta}}\left\{E_{\boldsymbol{\theta}}(y_t I_{\{\phi_t = j\}} | \mathscr{F}_{t-1})\right\} = \sum_{j=1}^k \mu(\theta_j) E_{\boldsymbol{\theta}} \tau_n(j),$$

where $\tau_n(j) = \#\{1 \le t \le n : \phi_t = j\}$ and population j is assumed to have density function $f_{\theta_j}(\cdot)$ from a parametric family of distributions. Hence, maximizing $E_{\boldsymbol{\theta}}\left(\sum_{t=1}^n y_t\right)$ is equivalent to minimizing the regret

$$
\begin{aligned}
R_n(\boldsymbol{\theta}) &= n\mu^*(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}}\left(\sum_{t=1}^{n} y_t\right) \\
&= \sum_{j:\mu(\theta_j)<\mu^*(\boldsymbol{\theta})} \left(\mu^*(\boldsymbol{\theta}) - \mu(\theta_j)\right) E_{\boldsymbol{\theta}}\tau_n(j),
\end{aligned} \tag{11}
$$

where $\mu^*(\boldsymbol{\theta}) = \max_{1\le j\le k} \mu(\theta_j)$. This martingale representation permitted the use of sequential testing theory, with which Lai and Robbins derived the basic lower bound for the regret (11) of uniformly good rules:

$$
R_n(\boldsymbol{\theta}) \ge \left\{ \sum_{j:\mu(\theta_j)<\mu^*(\boldsymbol{\theta})} \frac{\mu^*(\boldsymbol{\theta}) - \mu(\theta_j)}{I(\theta_j, \theta^*)} + o(1) \right\} \log n, \tag{12}
$$

where $\theta^* = \theta_{j(\boldsymbol{\theta})}$, $j(\boldsymbol{\theta}) = \arg\max_{1\le j\le k} \mu(\theta_j)$, and an adaptive allocation rule is called "uniformly good" if $R_n(\boldsymbol{\theta}) = o(n^a)$ for every $a > 0$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}^k$. Making use of the duality between hypothesis testing and confidence intervals, they developed "upper confidence bound" (UCB) rules to attain the asymptotic lower bound of (12).

In 1957, Bellman [9] introduced the dynamic programming approach to the 2-armed adaptive allocation problem considered by Robbins in 1952, generalizing it to k arms and calling it a "k-armed bandit problem". The name derives from an imagined slot machine with k arms such that when an arm is pulled the player wins a random reward. For each arm j, there is an unknown probability distribution Π_j of the reward, hence there is a fundamental dilemma between "exploration" (to generate information about $\Pi_1, \ldots, \Pi_k$ by pulling the individual arms) and "exploitation" (of the information so that inferior arms are pulled minimally). Dynamic programming offers a systematic solution of the dilemma in the Bayesian setting but suffers from the "curse of dimensionality" as k and n increase.

In 1974 and 1979, Gittins and Jones [28] and Gittins [27] considered the discounted version of this problem (thereby circumventing the issue of large horizon n) and showed that the k-dimensional stochastic optimization problem has an "index policy" (which does not have the curse of dimensionality) as its solution: At stage t, pull the arm with the largest "dynamic allocation index" (DAI) that depends only on the posterior distribution of the reward given the observed rewards from that arm up to stage t. The DAI is the solution to a non-standard optimal stopping problem that maximizes the quotient $E_j\left(\sum_{t=0}^{\tau-1} \beta^t Z_t\right) / E_j\left(\sum_{t=0}^{\tau-1} \beta^t\right)$, where E_j denotes expectation under the posterior distribution of Π_j of the reward Z_t from arm j, given the observed rewards from the arm up to the stopping time τ, and $0 < \beta < 1$ is a discount factor. In 1980, Whittle [89] provided an alternative formulation of the DAI, which he called the "Gittins index", in terms of a family (indexed by a retirement reward M) of standard optimal stopping problems (involving E_j but not the quotient) that can be solved by dynamic programming. My paper [45] in 1987 connects UCB to generalized likelihood ratio (GLR) test statistics and to the Gittins index and showed that the UCB rule is uniformly good and attains the asymptotic lower bound (12) for the regret. It is also shown in [45] that the UCB rule asymptotically minimizes the Bayes regret as $n \to \infty$ over a general class of prior distribution H of θ:

$$\int R_n(\boldsymbol{\theta})\mathrm{d}H(\boldsymbol{\theta}) \sim C(\log n)^2,$$

where C depends on the prior density function which is assumed to be positive and continuous over $\theta_j \in (\theta^* - \rho, \theta^* + \rho)$ for $1 \leq j \leq k$, with $\rho > 0$ and $\theta_j^* = \max_{i \neq j} \theta_i$.

5 Stochastic Approximation (SA) and Adaptive SA

The 1951 paper of Robbins and Monro [78] represents a major major departure from the framework of sequential analysis adopted by Wald and his contemporaries, for whom the *sequential* element of the data-generating mechanism (or *experiment*) came from a data-dependent (instead of predetermined) sample size. The sequential experiments in stochastic approximation do not have stopping times; instead they involve choosing the design levels x_i in a regression model sequentially, on the basis of past observations, so that the x_i eventually converge to some desired level. The regression model considered is of the general form

$$y_i = M(x_i) + \varepsilon_i \quad (i = 1, 2, \ldots), \tag{13}$$

where y_i denotes the response at x_i, M is an unknown regression function, and ε_i represents unobservable noise (error). In the deterministic case (where $\varepsilon_i = 0$ for all i), Newton's method for finding the root θ of a smooth function M is a sequential scheme defined by the recursion

$$x_{n+1} = x_n - y_n/M'(x_n). \tag{14}$$

When errors ε_i are present, using Newton's method (14) entails that

$$x_{n+1} = x_n - M(x_n)/M'(x_n) - \varepsilon_n/M'(x_n). \tag{15}$$

Hence, if x_n should converge to θ so that $M(x_n) \to 0$ and $M'(x_n) \to M'(\theta)$, assuming M to be smooth and to have a unique root θ such that $M'(\theta) \neq 0$, then (15) implies that $\varepsilon_n \to 0$, which is not possible for many kinds of random errors ε_i (e.g., when the ε_i are i.i.d. with mean 0 and variance $\sigma^2 > 0$). To dampen the effect of the errors ε_i, Robbins and Monro replaced $1/M'(x_n)$ in (14) by constants that converge to 0. Specifically, assuming that

$$M(\theta) = 0, \qquad \inf_{\varepsilon < |x-\theta| < 1/\varepsilon} (x - \theta)M(x) > 0 \text{ for all } 0 < \varepsilon < 1,$$

$$|M(x)| \leq c(|x - \theta| + 1) \text{ for some } c > 0 \text{ and all } x, \tag{16}$$

the *Robbins-Monro scheme* is defined by the recursion

$$x_{n+1} = x_n - a_n y_n \quad (x_1 = \text{ initial guess of } \theta), \tag{17}$$

where a_n are positive constants such that

$$\sum_1^\infty a_n^2 < \infty, \quad \sum_1^\infty a_n = \infty.$$

To find the maximum of the regression function M in (13) without using M', Kiefer and Wolfowitz [38] used a year later the recursion

$$x_{n+1} = x_n + a_n \Delta(x_n), \tag{18}$$

where at the nth stage observations y_n'' and y_n' are taken at the design levels $x_n'' = x_n + c_n$ and $x_n' = x_n - c_n$, respectively, a_n and c_n are positive constants, and

$$\begin{aligned}\Delta(x_n) &= (y_n'' - y_n')/2c_n \\ &= \frac{M(x_n + c_n) - M(x_n - c_n)}{2c_n} + \frac{\varepsilon_n'' - \varepsilon_n'}{2c_n}.\end{aligned}$$

To dampen the effect of the errors ε_n' and ε_n'', Keifer and Wolfowitz assumed that

$$c_n \to 0, \quad \sum_1^\infty (a_n/c_n)^2 < \infty, \quad \sum_1^\infty a_n c_n < \infty, \quad \text{and} \quad \sum_1^\infty a_n = \infty. \tag{19}$$

By deriving recursions for $E(x_{n+1} - \theta)^2$ from (17) or (18) under the assumption $\sup_i E(\varepsilon_i^2 | x_1, \ldots, x_{i-1}) \le \sigma^2$, Kiefer and Wolfowitz, like Robbins and Monro, proved that their stochastic approximation schemes converge in L_2 , and therefore also in probability, to the value θ satisfying $M(\theta) = 0$ or $M'(\theta) = 0$.

Subsequently, Blum [12] cited a convergence theorem for square-integrable martingales (although he did not use the martingale terminology, then unfamiliar to the statistical community) to prove the a.s. convergence of the Robbins-Monro and Kiefer-Wolfowitz schemes. He was also able to remove the assumption $\sum_1^\infty a_n c_n < \infty$ in (19). Dvoretzky [23] then proved the a.s. and L_2 convergence of a general class of recursive stochastic algorithms; this result is often called *Dvoretzky's approximation theorem*.

Gladyšev [29] gave a simple proof of the a.s. convergence of the Robbins-Monro scheme by an ingenious application of Doob's supermartingale convergence theorem, paving the way for a subsequent generalization of supermartingales by Robbins and Siegmund [80] in 1971. Let $\{\varepsilon_i, \mathscr{F}_i, i \ge 1\}$ be a martingale difference sequence such that

$$\sup_i E(\varepsilon_i^2 | \mathscr{F}_{i-1}) < \infty \; a.s.$$

Putting (13) into the recursion (17) yields a corresponding recursion for $V_n := (x_{n+1} - \theta)^2$. From (16) and the assumption that $E(\varepsilon_i|\mathscr{F}_{i-1}) = 0$, it then follows from this recursion for V_n that

$$E(V_n|\mathscr{F}_{n-1}) \le (1 + 2c^2 a_n^2)V_{n-1} + a_n^2\{2c^2 + E(\varepsilon_n^2|\mathscr{F}_{n-1})\} - 2a_n(x_n - \theta)M(x_n), \tag{20}$$

which can be written in the form

$$E(V_n|\mathscr{F}_{n-1}) \le (1 + \alpha_{n-1})V_{n-1} + \beta_{n-1} - \gamma_{n-1}, \tag{21}$$

in which α_i, β_i and γ_i are nonnegative $\mathscr{F}_i$-measurable random variables. Robbins and Siegmund [80] called V_n that satisfies (21) an *almost supermartingale*, noting that V_n is indeed a supermartingale if $\alpha_{n-1} = \beta_{n-1} = \gamma_{n-1} = 0$. They showed that if V_n is a nonnegative almost supermartingale, then

$$V_n \text{ converges and } \sum_1^\infty \gamma_n < \infty \text{ a.s. on } \left\{\sum_1^\infty \alpha_i < \infty, \sum_i^\infty \beta_i < \infty\right\}. \tag{22}$$

They applied this result to derive the a.s. part of Dvoretzky's approximation theorem and certain convergence results in two-person games and cluster analysis as corollaries. Although V_n satisfying (21) is not a supermartingale, it can be transformed into one via

$$U_n = \frac{V_n}{\prod_{i=1}^{n-1}(1+\alpha_i)} - \sum_{i=1}^{n-1} \frac{\beta_i - \gamma_i}{\prod_{j=1}^{i}(1+\alpha_j)},$$

which is a supermartingale by (21). Let $\beta_i' = \beta_i / \prod_{j=1}^{i}(1+\alpha_j)$. Although U_n need not be nonnegative, it is bounded below on the event $\{\sum_1^\infty \beta_i' \le k\}$ for every $k = 1, 2, \ldots$. Therefore by Doob's supermartingale convergence theorem, U_n converges a.s. on $\left\{\sum_1^\infty \alpha_i < \infty, \sum_i^\infty \beta_i < \infty\right\}$. Robbins and Siegmund [80] made use of this argument to prove (22). Earlier, Gladyšev [29] used a somewhat different argument to transform (20) to a nonnegative supermartingale, to which he applied Doob's supermartingale convergence theorem.

As noted by Lai and Yuan [60, Sect. 2.2], Robbins spent the 1975–1976 academic year as a Guggenheim Fellow at Imperial College in London, where he heard T. W. Anderson lecture on the "multi-period control problem" in econometrics, which is concerned with choosing the inputs $x_1, \ldots, x_N$ sequentially in the linear regression model $Y_i = \alpha + \beta x_i + \varepsilon_i$ (with unknown parameters $\beta \ne 0$ and α and i.i.d. random errors ε_i having mean 0 and variance σ^2) so that the outputs are as close as possible to a target value y^*. Assuming prior knowledge of bounds K_1 and K_2 such that $K_1 < \theta := (y^* - \alpha)/\beta < K_2$, Anderson and Taylor's rule is defined by

$$x_{n+1} = K_1 \vee \{\widehat{\beta}_n^{-1}(y^* - \widehat{\alpha}_n) \wedge K_2\}, \quad n \ge 2,$$

where $\widehat{\alpha}_n$ and $\widehat{\beta}_n$ are the least squares estimates of α and β at stage n. In 1976, based on the results of simulation studies, Anderson and Taylor [3] conjectured that this rule would converge to θ a.s., with $\sqrt{n}(x_n - \theta)$ having a limiting $N(0, \sigma^2/\beta^2)$ distribution. They also raised the question whether $\widehat{\alpha}_n$ and $\widehat{\beta}_n$ are strongly consistent. Clearly, if the x_i should cluster around θ, then there would not be much information for estimating the slope β. There is, therefore, an apparent dilemma between the control objective of setting the design levels as close as possible to θ and the need for an informative design with sufficient dispersion to estimate β.

To resolve this dilemma, Lai and Robbins [52] began by considering the case of known β. Replacing Y_i by $Y_i - y^*$, it can be assumed without loss of generality that $y^* = 0$ so that $Y_i = \beta(x_i - \theta) + \varepsilon_i$. Let $\bar{x}_n = n^{-1}\sum_{i=1}^n x_i$, $\bar{Y}_n = n^{-1}\sum_{i=1}^n Y_i$. With known β, the least squares certainty equivalence rule becomes $x_{n+1} = \bar{x}_n - \bar{Y}_n/\beta$, which turns out to be equivalent to the stochastic approximation recursion $x_{n+1} = x_n - (n\beta)^{-1}Y_n$. Since $\bar{x}_n - \bar{Y}_n/\beta = \theta - \bar{\varepsilon}_n/\beta$, $E(x_{n+1} - \theta)^2 = \sigma^2/(n\beta^2)$ for $n \geq 1$ and, therefore,

$$
\begin{aligned}
E\left(\sum_{n=1}^N Y_n^2\right) &= \sum_{n=1}^N E\{\beta^2(x_n - \theta)^2 + \varepsilon_n^2\} \\
&= \sigma^2(N + \log N + O(1)).
\end{aligned}
$$

Moreover, $\sqrt{N}(x_N - \theta) \Rightarrow N(0, \sigma^2/\beta^2)$ and $\beta^2\sum_{n=1}^N (x_n - \theta)^2$ $\left(\sum_{n=1}^N (Y_n - \varepsilon_n)^2\right)$, called the *regret* (due to ignorance of θ) of the design, is of order $\sigma^2 \log N$. To achieve this when β is unknown, an obvious way to modify the preceding rule for the case of unknown β is to use an estimate b_n to substitute for β either in the recursion $x_{n+1} = \bar{x}_n - \bar{Y}_n/\beta$ or in the equivalent stochastic approximation scheme $x_{n+1} = x_n - Y_n/(n\beta)$. The equivalence between the two recursive schemes, however, no longer holds when β is replaced by b_n. The second recursion, called adaptive stochastic approximation, was treated in Lai and Robbins [52,53] in 1979 and 1981. In particular, Lai and Robbins [52] considered adaptive stochastic approximation schemes of the form $x_{n+1} = x_n - Y_n/(nb_n)$, where b_n is $\mathscr{F}_{n-1}$-measurable and $\lim_{n\to\infty} b_n = b > 0$ a.s. By representing x_n as a weighted sum of the i.i.d. random variables ε_i, they proved limit theorems on $x_N - \theta$ and $\sum_{n=1}^N (x_n - \theta)^2$. Specifically, for $0 < b < 2\beta$, they proved that

$$
\beta^2 \sum_{n=1}^N (x_n - \theta)^2 \sim \sigma^2 g(b/\beta) \log N \ a.s., \quad \sqrt{N}(x_N - \theta) \Rightarrow N(0, (\sigma^2/\beta^2) g(b/\beta)),
$$

where $g(t) = 1/\{t(2-t)\}$ for $0 < t < 2$ and has a minimum value of 1 at $t = 1$. They later showed in [53] that by choosing b_n to be a truncated version of the least squares estimate in the adaptive stochastic approximation scheme, one indeed has $b_n \to \beta$ a.s. and, therefore, the adaptive scheme has the same asymptotic properties as the “oracle” stochastic approximation scheme that assumes known β.

6 Martingales and Biorhythms in Time Series

My entry into time series research was a secondary outcome of my teaching assignments. As the most junior member of the faculty in Columbia's Department of Mathematical Statistics where other faculty members had been my teachers, I was assigned to (learn and) teach new areas that the department would like to offer besides martingale theory, stochastic processes and sequential analysis, in which the department was widely recognized as a world leader. Time series analysis was one such course that I had been teaching since 1973. Fortunately, Box and Jenkins' 1971 book [13] had appeared so that I could use it as a textbook. T.W. Anderson's book [2] had also appeared and was used as a reference because it was too advanced for an introductory course. Later, Peter Bloomfield's book on the Fourier analysis of time series [11] was also included as a reference. My toolkit for deriving results in the time domain was greatly broadened during my visit to UIUC where I learned from Bill Stout the almost sure invariance principles (Philipp and Stout [72]) and martingale methods for time series analysis.

My toolkit for developing results in the frequency domain was also considerably broadened after I returned to Columbia and worked on biorhythms in the Pediatric Pulmonary Division of Columbia's Medical School for an NIH project grant, thanks to the time series course that I resumed teaching from 1976 to 1980. In late 1970s, the Pediatric Pulmonary Group at Columbia applied for a program project of the NIH to study sleep physiology of infants, particularly those at risk of SIDS (sudden infant death syndrome) and asked the director of Biostatistics to suggest someone in his division with expertise in time series to put on the grant application. He replied that his division did not have such experts and suggested me since I had been teaching time series from my assistant professor days to my recent promotion to full professor. This was how I got on the program project when it was funded and had to learn SIDS, cardiorespiratory physiology and some bioengineering besides the relevant time series tools.

The person in the team with whom I interacted most was Gabriel Haddad who had been a postdoctoral fellow and was just appointed assistant professor, leading to a series of papers in the period 1981–1986 summarized in [95, Sect. 3]. Haddad moved to UC San Diego as the chair of the Department of Pediatrics and physician-in-chief and chief scientific officer of Rady's Children's Hospital in San Diego and invited me to visit him ten years ago to discuss potential collaboration. It took a long time to get started because Stanford is about 500 miles away from San Diego, but we finally set up a team and had our first publication with Hua-Tieng Wu, who is at Duke on the east coast (even much farther away) and who was a physician in Taiwan before going to Princeton to study for his Ph.D. in mathematics under Ingrid Daubechies, Haddad and his colleagues Alysson Muotri, and me as coauthors.

7 Martingales in Stochastic Optimization, 1987–2021

I have been at Stanford since January 1987 [62, p. 160]. During the past 34 years (twice as long as my time at Columbia, minus one year at UIUC), the time series and sequential design/analysis work begun at Columbia and UIUC came to fruition just in time for martingales in statistics and stochastic optimization to synthesize with other ideas in these fields and thereby to realize their potential in multifaceted applications in a new millennium marked by scientific innovations, technological advances, and BigData and the MultiCloud.

7.1 Contextual Bandits in Reinforcement Learning and Personalization, Modified Gradient Boosting and SA in AI

Kim, Lai and Xu [39, Sect. 1.3] generalize the definition of regret to contextual k-armed bandits, or k-armed bandits with covariate information, for which the decision maker also observes a covariate vector $\mathbf{x}_t$ that contains information on θ_j if y_t is sampled from arm j (i.e., $\phi_t = j$). First assume that the $\mathbf{x}_t$ are i.i.d. with common distribution H. Let supp$H$ denote the support of H, $f_\theta(y|\mathbf{x})$ denote the density function, depending on a parameter $\theta \in \Theta$ of the reward Y (with respect to some dominating measure ν on the real line) when the covariate vector has value $\mathbf{x}$, $\mu(\theta, \mathbf{x}) = \int y f_\theta(y|\mathbf{x}) d\nu(y)$, and

$$j^*(\mathbf{x}) = \arg\max_{1 \le j \le k} \mu(\theta_j, \mathbf{x}), \quad \theta^*(\mathbf{x}) = \theta_{j^*(\mathbf{x})},$$

where θ_j is the parameter associated with arm j. Letting $\mathscr{F}_{t-1}$ denote the σ-field generated by $\{\mathbf{x}_t\} \cup \{(\mathbf{x}_s, y_s) : s \le t-1\}$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$, the problem of choosing an adaptive allocation rule $(\phi_1, \dots, \phi_n)$ to maximize $E_\theta \left(\sum_{t=1}^n y_t\right)$ is equivalent to minimizing the regret

$$\begin{aligned} R_n(\boldsymbol{\theta}, B) &= n \int_B \mu(\theta^*(\mathbf{x}), \mathbf{x}) \mathrm{d}H(\mathbf{x}) - \sum_{t=1}^n \sum_{j=1}^k E_{\boldsymbol{\theta}} \left\{ E_{\boldsymbol{\theta}}[y_t I_{\{\phi_t = j, \mathbf{x}_t \in B\}} | \mathscr{F}_{t-1}] \right\} \\ &= \sum_{j=1}^k \int_B \left(\mu(\theta^*(\mathbf{x}), \mathbf{x}) - \mu(\theta_j, \mathbf{x}) \right) E_{\boldsymbol{\theta}} \tau_n(j, \mathbf{x}) \mathrm{d}H(\mathbf{x}) \end{aligned} \tag{23}$$

for Borel subsets B of supp$H$, for which $E_\theta \tau_n(j, B) := \sum_{t=1}^n P_{\boldsymbol{\theta}}\{\phi_t = j, \mathbf{x}_t \in B\}$ defines a measure that is absolutely continuous with respect to the common distribution H of the i.i.d. covariate vectors $\mathbf{x}_t$. Hence, the term $E_\theta \tau_n(j, \mathbf{x})$ in (23) is the Radon-Nikodym derivative of the measure $E_{\boldsymbol{\theta}} \tau_n(j, B)$ over Borel subsets B of supp$H$. Using this martingale representation for contextual bandits, Kim, Lai and Xu [39, Sect. 1.3] extend the asymptotic lower bound (12) for the regret $R_n(\boldsymbol{\theta}, B)$: If j^* is constant over B, then

$$R_n(\boldsymbol{\theta}, B) \geq (1 + o(1)) \sum_{j:p_j(\boldsymbol{\theta})=0} \log n \int_B \frac{\mu(\theta^*(\mathbf{x}), \mathbf{x}) - \mu(\theta_j, \mathbf{x})}{I_{\mathbf{x}}(\theta_j, \theta^*(\mathbf{x}))} \mathrm{d}H(\mathbf{x}),$$

where $\sum_j$ over an empty set is interpreted as $O(1)$, $p_j(\boldsymbol{\theta}) = P_{\boldsymbol{\theta}}\{j^*(\mathbf{X}) = j\}$, in which $\mathbf{X} \in \mathbb{R}^p$ has distribution H, and

$$I_{\mathbf{x}}(\theta, \theta') = \inf_{\lambda:\mu(\lambda,\mathbf{x})=\mu(\theta',\mathbf{x})} E_\lambda\{\log\left(f_\theta(Y|\mathbf{x})/f_\lambda(Y|\mathbf{x})\right)\}.$$

If j^* is non-constant over B (i.e., B contains leading arm transitions), then

$$R_n(\boldsymbol{\theta}, B) \geq C(\boldsymbol{\theta})(\log n)^2.$$

Moreover, Kim, Lai and Xu [39, Sect. 2.2] use ε-greedy randomization and likelihood ratio arm elimination schemes to develop adaptive allocation rules that attain these asymptotic lower bounds. Their Sects. 2.1 and 2.3 also extend these results to nonparametric contextual bandits.

Lai, Sklar and Xu [56] develop the ε-greedy randomization and the arm elimination strategies further for nonparametric contextual bandits, from finite to non-denumerable set of arms and from discrete to continuous time. They show how ε-greedy randomization can circumvent the difficulties with defining UCB or index policies in continuous time, and how (self-normalized) Welch t-statistics, which satisfy exponential bounds for large deviation probabilities, can be defined for arm elimination. Their Sect. 3.1 also shows how decoupling inequalities (de la Peña and Lai, 2001 [70]) can be combined with maximal dependence (Lai and Robbins, 1978 [40]) to relax the assumption of i.i.d. arms to identically distributed arms.

Kim, Lai and Xu (2021, [39, abstract, last paragraph of Sect. 2.3]) point out that "the turn of the millennium marked the onset of a "personalization revolution," from personalized medicine to online personalized advertising and recommender systems", and give references on "the importance of nonparametric contextual bandit methodology to precision medicine and drug development" and on " its important role in recommender systems, online experimentation and precision health." Their §3 also describes extensions to high-dimensional covariates in the current BigData and MultiCloud era.

Lai and Yuan [60, Sect. 2.3] describe how SA flourished under multidisciplinary input and development, citing references in signal processing and adaptive control during the period 1971–1990. Their Sect. 3 cites many additional references during the BigData era 2000–2021: high-dimensional sparse linear stochastic regression models in Sect. 3.1, modified gradient boosting for nonlinear stochastic regression models in Sect. 3.2, and SA in particle swarm optimization and artificial (machine) intelligence (AI) in Sect. 3.3. Hongsong Yuan received his Ph.D. under my supervision in 2012.

7.2 Joint State and Parameter Estimation in Hidden Markov Models, with Uncertainty Quantification

My 2021 papers Lai [48] and Wu et al. [95] describe, with extensive references, major breakthroughs in this very important area in time series analysis, image and signal processing, robotics, automatic navigation, bioengineering, and control systems. The starting point of these breakthroughs was the martingale representation of the particle filter (also called sequential Monte Carlo method) in a hidden Markov model (HMM) given by Chan and Lai [17, Lemmas 1 and 4]. The particle filter was introduced in 1993 by Gordon, Salmond and Smith [30]. The assumption of a single, fully specified, HMM is too restrictive in applications since the model parameters are usually unknown and need to be estimated sequentially from the observed data, hence joint state and parameter estimation is the usual estimation task in the aforementioned applications of HMMs. In the Bayesian framework, the unknown parameter θ has a prior distribution so that the posterior joint distribution of θ and the states is the target distribution to be estimated.

The Metropolis-Hastings Markov chain Monte Carlo (MCMC) scheme provides a simulation-based method to estimate a target distribution. In 2010, Andrieu, Doucet and Holenstein [4], and in 2013 Chopin, Jacob and Papaspiliopoulos [19] used this approach in their "particle MCMC" and "SMC2" procedures, respectively, but encountered difficulties in approximating the joint target and state distribution. Together with Hock Peng Chan, who received his Ph.D. under me in 1998, and my current Ph.D. students Huangzhong Xu and Michael Hongyu Zhu, I circumvented this difficulty by developing a new MCMC scheme called MCMC with sequential substitutions (MCMC-SS). In [48, Theorem 2] we provide (i) a comprehensive asymptotic theory of MCMC-SS showing its asymptotic optimality with respect to computational and statistical criteria, and (ii) consistent estimators of standard errors of the Monte Carlo state and parameter estimates. MCMC-SS uses B disjoint blocks $\Theta_{b,k}$ of ν atoms within each block and carries out the following sequential substitution procedure $SS(\Theta_{b,k}, \mathbf{w}_k^b)$ at stage k to update the atom set in $\Theta_{b,k}, b = 1, \ldots, B$:

(a) Let $\{q(\cdot|\gamma) : \gamma \in \Gamma\}$ be a family of proposal densities with respect to some measure m, and sample $\widetilde{\theta}$ from $q(\cdot|\gamma_{b,k-1})$ as the candidate atom.
(b) Let $\theta^b_{\nu+1,k-1} = \widetilde{\theta}$ and compute $\lambda^b_{i,k} = q\left(\theta^b_{i,k-1}|\gamma_{b,k-1}\right)/f\left(\theta^b_{i,k-1}\right)$, $i = 1, \ldots, \nu+1$, in which f is a given function that is proportional to the target density.
(c) Sample J from $\{1, \ldots, \nu+1\}$ with probability $\pi_{i,k} = \lambda^b_{i,k}/\left(\sum_{j=1}^{\nu+1} \lambda^b_{j,k}\right)$ for i.
(d) If $J = \nu+1$, let $\Theta_{b,k} = \Theta_{b,k-1}$. Otherwise let $\Theta_{b,k} = \left(\Theta_{b,k-1} \cup \{\widetilde{\theta}\}\right) \setminus \left\{\theta^b_{\widetilde{J},k-1}\right\}$.
(e) Let $w^b_{i,k} = 1/\pi^b_{i,k}$ for $i = 1, \ldots, \nu$, and $\mathbf{w}^b_k = \left(w^b_{1,k}, \ldots, w^b_{\nu,k}\right)$.

In many applications, the parameter γ in the proposal density $q(\cdot|\gamma)$ is a function $\gamma : \mathscr{P} \to \Gamma$, where $\mathscr{P}$ is the space of probability measures on Θ. Assuming this framework, Lai [48] describes the choice of $\gamma_{b,k-1}$ in the updating algorithm of MCMC-SS. Let κ represent an initial burn-in period that is asymptotically negligible in comparison with the total number K of iterations in the asymptotic theory with $\kappa \to \infty$ such that $\kappa = o(K)$. For $k \le \kappa$, let $\gamma_{b,k-1} = \nu^{-1} \sum_{\theta \in \Theta_{b,k-1}} \gamma(\theta)$, which is the mean of the empirical measure of the atoms in the bth block at the end of stage $k-1$. On the other hand, for $k > \kappa$, pool across blocks by letting $\widetilde{\gamma}_{k-1} = B^{-1} \sum_{b=1}^{B} \gamma_{b,k-1}$, which we use as the modified $\gamma_{b,k-1}$ for all blocks. Therefore, after the burn-in period, carry out the update $SS(\theta_{b,k})$ in the order $b = 1, \ldots, B$, so that if the candidate atom in $SS(\theta_{b,k})$ is not used for block b, it can serve as candidate atom for block $b+1 (\le B)$, which then does not need to generate another random variable from $q(\cdot|\widetilde{\gamma}_{k-1})$, an obvious advantage for high-dimensional complicated states. MCMC-SS estimates $\mu = E_p \psi(\theta)$ for which $E_p \psi^2(\theta) < \infty$ by

$$\widehat{\psi} = \frac{1}{B(K-\kappa)} \sum_{b=1}^{B} \sum_{k=\kappa+1}^{K} \widehat{\psi}_{b,k}, \quad \text{with } \widehat{\psi}_{b,k} = \frac{\sum_{i=1}^{\nu} w_{i,k}^{b} \psi(\theta_{i,k}^{b})}{\sum_{i=1}^{\nu} w_{i,k}^{b}}, \tag{24}$$

and $\sigma^2 := \mathrm{Var}_p\,(\psi(\theta))$ by

$$\hat{\sigma}^2 = \frac{1}{B(K-\kappa)} \sum_{b=1}^{B} \sum_{k=\kappa}^{K} \frac{1}{\nu - 1} \sum_{\theta \in \Theta_{b,k}} (\psi(\theta) - \widehat{\psi}_{b,k})^2. \tag{25}$$

Moreover, $\hat{\sigma}^2$ is a consistent estimate of σ^2 and with probability approaching 1 for large k, the candidate atom $\widetilde{\theta}$ indeed substitutes some existing atom in $\Theta_{b,k-1}$. Hence, similar to the case of known target density p, each random variable generated in the MCMC-SS scheme asymptotically contributes weight $(B(K-\kappa))^{-1}$, to (a) the estimate $\widehat{\psi}$ of μ and (b) the asymptotic variance of $\widehat{\psi}$. Wu et al. (2021) describe applications of MCMC-SS to image reconstruction in [48] and latent variable analysis with uncertainty quantification in infants' brain network development in [95].

In [48, Sect. 3], I describe the recursive variant of MCMC-SS for online system control. Noting that the particle filter of Gordon, Salmond and Smith, under the assumption of known θ, is a recursive procedure, I point out that the recursive feature of the particle filter is lost when θ is unknown and replaced by the posterior distribution of θ given $Y_{1:T} := (Y_1, \ldots, Y_T)$, because $X_{0:t}$ requires θ to be generated from the posterior distribution $\pi(\cdot|Y_{1:t})$ which is different from $\pi(\cdot|Y_{1:t-1})$. I use a group sequential approach as in [7, Chap. 4] that given the posterior distribution $\pi(\cdot|Y_{1:t_j})$ to update the state trajectory $X_{t_j+1:t})$ for $t_j < t \le t_{j+1}$, $j = 1, 2, \ldots$. In this way, the empirical measure of

$$\Omega(t_{j-1}) := \{X_{t_{j-1}}^{b,i,k,m} : 1 \le b \le B, 1 \le i \le \nu, \kappa+1 \le k \le n_1 + \cdots + n_{j-1}, 1 \le m \le M\}$$

is used to generate M particles $\widetilde{X}^{1}_{t_{j-1}:t}(\theta), \ldots, \widetilde{X}^{M}_{t_{j-1}:t}(\theta)$ after initializing, by a random draw of $\widetilde{X}^{m}_{t_{j-1}}(\theta)$ from $\Omega(t_{j-1})$, using $\Omega(t_1)$ and $\Theta_{b,n_1}(t_1)$ to initialize at time

t_1, for $1 \le b \le B$. This recursive variant of MCMC-SS is denoted by recMCMC-SS. In [48] I give illustrative applications to automatic navigation and to high-frequency econometrics of transactions in electronic trading.

Alternative approaches via solving the stochastic partial differential equation (SPDE) defining the target density function have been developed in 1997 by Yau and Yau [96] for automatic navigation, and by Zeng [97,98] in 2003–2004 for high-frequency trading. Yau and Yau have referred to Brockett and Clark [16], Brockett [15], and Mitter [64] who also pointed out that the SPDE approach to nonlinear filtering is not computable and proposed to "use the Lie algebraic method to solve the Duncan-Mortensen-Zakai (DMZ) equation", which is the SPDE pointed out by Yau and Yau [96], who show that the DMZ equation for the Kalman-Bucy and the Benes filtering systems "can be solved explicitly with an arbitrary initial condition by solving a system of ordinary differential equations and a Kolmogorov-type equation", requiring n sufficient statistics if the state space has dimension n. In 2002, Storvik [86] has developed an analogous result for particle filters when $\pi(\cdot|Y_{1:t})$ "depends on some low-dimensional sufficient statistics". Hence, without requiring these restrictive assumptions, recMCMC-SS provides a simulation-based solution to the SPDE defining the target density.

8 Concluding Remarks

My primary appointment at Stanford is in the Department of Statistics of the School of of Humanities and Sciences, but I also hold courtesy appointments in the Department of Biomedical Data Science of the School of Medicine and in the Institute of Computational & Mathematical Engineering of the School of Engineering. I am also director of the Financial and Risk Modeling Institute and co-director of the Center for innovative Study Design, and a faculty affiliate of the Woods Institute for the Environment, the Wu-Tsai Neurosciences Institute, Center for Precision Mental Health and Wellness, Center for Innovation in Global Health. The stochastic optimization and joint state and parameter estimation in hidden Markov models described in Sects. 7.1 and 7.2 have applications in the projects that I am working on with colleagues and students from these departments and centers. In particular, my papers [6,49–51,55] and books [8,18] are related to the research of my group at the Center for Innovative Study Design. Moreover, time series and martingales have also played an important role in my teaching and research as director of the Financial and Risk Modeling Institute, leading to the books [58,59] with my former Ph.D. student Haipeng Xing, and [31] with my collaborator Xin Guo of UC Berkeley and former Ph.D. students Howard Shek and Samuel Po-Shing Wong.

Stanford is only 40 miles from UC Berkeley and it is therefore convenient to collaborate with the faculty there, e.g., with Michael Klass who was the Ph.D. adviser of Victor de la Peña. We have coauthored a number of joint papers, including [69], which led to subsequent papers on self-normalized martingales. Concerning Studentized statistics and self-normalized processes, Qiman Shao visited me and de la Peña in 2007, and that led to our 2009 book on Studentized statistics and self-normalized processes [71], to celebrate the centennial of Student's t-statistic, introduced by Gosset in 1908.

In conclusion, martingales in statistics and stochastic optimization have played a central role in my career, dating back to my graduate student days at Columbia and still continuing with vibrancy. The year 1975–1976 at UIUC, when I learned from Doob and his martingale group, proved to be immensely beneficial. I witnessed the cross-fertilization of martingale theory with other branches of mathematics and the interactions of mathematics with other departments and schools in the university. This experience was very valuable after I returned to Columbia and later when I moved to Stanford, where the multifaceted applications of martingales in statistics and stochastic optimization have been steadily growing.

Acknowledgements This chapter expands my 2009 review of the history martingales in the *Electronic Journal for History of Probability and Statistics* [47] by expanding the scope from "Sequential Analysis and Time Series" to "Statistics and Stochastic Optimization" and expanding the period covered from 1945–1985 to 1933–2021. The research was supported by National Science Foundation grant DMS-1811818.

References

1. Aalen, O., Andersen, P.K., Borgan, Ø., Gill, R., Keiding, N.: History of applications of martingales in survival analysis. Electronic Journal for History of Probability and Statistics **5**(1) (2009). 28 pp.
2. Anderson, T.W.: The Statistical Analysis of Time Series. Wiley (1971)
3. Anderson, T.W., Taylor, J.B.: Some experimental results on the statistical properties of least squares estimates in control problems. Econometrica **44**(6), 1289–1302 (1976)
4. Andrieu, C., Doucet, A., Holenstein, R.: Particle Markov Chain Monte Carlo methods. Journal of the Royal Statistical Society: Series B **72**(3), 269–342 (2010)
5. Bahadur, R.R.: A note on the fundamental identity of sequential analysis. Ann. Math. Statist. **29**, 534–543 (1958)
6. Bartroff, J., Lai, T.L., Narasimhan, B.: A new approach to designing phase I–II cancer trials for cytotoxic chemotherapies. Stat. Med. **33**, 2718–2735 (2014)
7. Bartroff, J., Lai, T.L., Shih, M.C.: Sequential Experimentation in Clinical Trials: Design and Analysis. Springer, New York (2013)
8. Baumgartner, R., Chen, J., Lai, T.L.: Real World Data and Evidence: Applications in Precision Medicine and Healthcare. Chapman & Hall/CRC (2021). Forthcoming
9. Bellman, R.E.: Dynamic Programming. Princeton University Press (1957)
10. Blackwell, D., Girshick, M.A.: On functions of sequences of independent chance vectors with applications to the problem of the "random walk" in k dimensions. Ann. Math. Statist. **17**, 310–317 (1946)
11. Bloomfield, P.: Fourier Analysis of Time Series: An Introduction. Wiley (1976)
12. Blum, J.R.: Multidimensional stochastic approximation methods. Ann. Math. Statist. **25**, 737–744 (1954)
13. Box, G.E.P., Jenkins, G.M.: Time Series Analysis: Forecasting and Control. Holden-Day (1971)
14. Box, G.E.P., Wilson, K.B.: On the experimental attainment of optimum conditions. J. Roy. Statist. Soc. Ser. B **13**(1), 1–38 (1951)
15. Brockett, R.W.: Nonlinear systems and nonlinear estimation theory. In: J. Hazewinkel, J.C. Willems (eds.) Stochastic Systems: The Mathematics of Filtering and Identification, and Applications, pp. 441–477. Springer (1981)

16. Brockett, R.W., Clark, J.M.C.: The geometry of the conditional density functions. In: O.L.R. Jacobs et al. (ed.) Analysis and Optimization of Stochastic Systems, pp. 299–309. Academic Press (1980)
17. Chan, H.P., Lai, T.L.: A general theory of particle filters in hidden Markov models and some applications. Ann. Statist. **41**, 2877–2904 (2013)
18. Chen, J., Heyse, J., Lai, T.L.: Medical Product Safety Evaluation: Biological Models and Statistical Methods. Chapman & Hall/CRC (2018)
19. Chopin, N., Jacob, P.E., Papaspiliopoulos, O.: SMC2: an efficient algorithm for sequential analysis of state space models. Journal of the Royal Statistical Society: Series B **75**(3), 397–426 (2013)
20. Chow, Y.S., Robbins, H., Teicher, H.: Moments of randomly stopped sums. Annals of Mathematical Statistics **36**(3), 789 – 799 (1965)
21. Darling, D.A., Robbins, H.: Iterated logarithm inequalities. Proc. Nat. Acad. Sci. U.S.A. **57**, 1188–1192 (1967)
22. Doob, J.L.: Stochastic Processes. Wiley, New York (1953)
23. Dvoretzky, A.: On stochastic approximation. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955 **I**, 39–55 (1956)
24. Farrell, R.H.: Asymptotic behavior of expected sample size in certain one sided tests. Ann. Math. Statist. **35**, 36–72 (1964)
25. Ghosh, B.K.: Sequential Tests of Statistical Hypotheses. Addison-Wesley (1970)
26. Gill, R.: Censoring and Stochastic Integrals, Mathematical Centre Tracts 124. Mathematisch Centrum, Amsterdam (1980)
27. Gittins, J.C.: Bandit processes and dynamic allocation indices. J. Roy. Statist. Soc. Ser. B **41**, 148–177 (1979)
28. Gittins, J.C., M., J.D.: A dynamic allocation index for the sequential design of experiments. Tech. rep., University of Cambridge, Department of Engineering (1974)
29. Gladyšev, E.G.: On stochastic approximation. Teor. Verojatnost. i Primenen. **10** (1965)
30. Gordon, N.J., Salmond, D.J., Smith, A.F.M.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proc. F (Radar and Signal Processing) **140**, 107–113 (1993)
31. Guo, X., Lai, T., H., S., Wong, S.P.: Quantitative Trading: Algorithms, Analytics, Data, Models, Optimization. Chapman & Hall/CRC (2017)
32. Hoeffding, W., Robbins, H.: The central limit theorem for dependent random variables. Duke Mathematical Journal **15**(3), 773–780 (1948)
33. Hotelling, H.: A general mathematical theory of depreciation. Journal of the American Statistical Association **20**(151), 340–353 (1925)
34. Hotelling, H.: The general welfare in relation to problems of taxation and of railway and utility rates. Econometrica **6**(3), 242–269 (1938)
35. Hotelling, H.: Tubes and spheres in n-spaces, and a class of statistical problems. American Journal of Mathematics **61**(2), 440–460 (1939)
36. Hotelling, H.: Experimental determination of the maximum of a function. Ann. Math, Statist. **12**, 20–45 (1941)
37. Johnstone, I., Siegmund, D.: On Hotelling's formula for the volume of tubes and Naiman's inequality. Ann. Statist. **17**, 184–194 (1989)
38. Kiefer, J., Wolfowitz, J.: Stochastic estimation of the maximum of a regression function. Ann. Math. Statist. **23**, 462–466 (1952)
39. Kim, D.W., Lai, T.L., Xu, H.: Multi-armed bandits with covariates: Theory and applications. Statistica Sinica **31**, in press. https://doi.org/10.5705/ss.202,020.0454(2021)
40. Lai, T., Robbins, H.: A class of dependent random variables and their maxima. Z. Wahrsch. verw. Gebiete **42**(2), 89–111 (1978)
41. Lai, T.L.: Space-time processes, parabolic functions and one-dimensional diffusions. Trans. Amer. Math. Soc. **175**, 409–438 (1973)
42. Lai, T.L.: On Chernoff-Savage statistics and sequential rank tests. Ann. Statist. **3**, 825–845 (1975)
43. Lai, T.L.: On confidence sequences. Ann. Statist. **4**(2), 265–280 (1976)

44. Lai, T.L.: Power-one tests based on sample sums. Ann. Statist. **5**, 866–880 (1977)
45. Lai, T.L.: Adaptive treatment allocation and the multi-armed bandit problem. Ann. Statist. **15**, 1091–1114 (1987)
46. Lai, T.L.: Likelihood ratio identities and their applications to sequential analysis. Sequential Anal. **23**, 467–497 (2004)
47. Lai, T.L.: Martingales in sequential analysis and time series, 1945–1985. Electronic Journal for History of Probability and Statistics **5**(1) (2009). 31 pp.
48. Lai, T.L.: Recursive particle filters for joint state and parameter estimation in hidden markov models with multifaceted applications (2021). Proceedings of International Congress of Chinese Mathematicians, Beijing, 2019, in press. International Press of Boston
49. Lai, T.L., Choi, A., Tsang, K.W.: Statistical science in information technology and precision medicine. Ann. Math. Sci. & Appl. **4**, 413–438 (2019)
50. Lai, T.L., Lavori, P.W., Tsang, K.W.: Adaptive design of confirmatory trials: Advances and challenges. Contemp. Clin. Trials **45,10th Anniversary Special Issue, Part A**, 93–102 (2015)
51. Lai, T.L., Lavori, P.W., Tsang, K.W.: Adaptive enrichment designs for confirmatory trials. Stat. Med. **38**, 613–624 (2019)
52. Lai, T.L., Robbins, H.: Adaptive design and stochastic approximation. Ann. Statist. **7**, 1196–1221 (1979)
53. Lai, T.L., Robbins, H.: Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes. Z. Wahrsch. verw. Gebiete **56**, 329–360 (1981)
54. Lai, T.L., Robbins, H.: Asymptotically efficient adaptive allocation rules. Adv. in Appl. Math. **6**, 4–22 (1985)
55. Lai, T.L., Sklar, M., Weissmueller, N.T.: Novel clinical trial designs and statistical methods in the era of precision medicine. Statistics in Biopharmaceutical Research **13**(2), 133–146 (2021)
56. Lai, T.L., Sklar, M.B., Xu, H.: Bandit and covariate processes with finite or non-denumerable set of arms. Stochastic Processes and Their Applications **to appear**, to appear (2021)
57. Lai, T.L., Wijsman, R.A.: First exit time of a random walk from the bounds $f(n) \pm cg(n)$, with applications. Ann. Probab. **7**, 672–692 (1979)
58. Lai, T.L., Xing, H.: Data Analytics and Risk Management in Finance and Insurance. Chapman & Hall/CRC (2008)
59. Lai, T.L., Xing, H.: Statistical Models and Methods in Financial Markets. Springer (2008)
60. Lai, T.L., Yuan, H.: Stochastic approximation: From statistical origin to big-data, multidisciplinary applications. Statistical Science **36**(2), 291–302 (2021)
61. Lorden, G.: Procedures for reacting to a change in distribution. Ann. Math. Statist. **42**, 1897–1908 (1971)
62. Lu, Y., Small, D., Ying, Z.: A conversation with Tze Leung Lai. Statistical Science **36**(1), 158–167 (2021)
63. Mangel, M., Samaniego, F.: Abraham Wald's work on aircraft survivability. J. Amer. Statist. Assoc. **79**, 259–267 (1984)
64. Mitter, S.K.: On the analogy between mathematical problems of non-linear filtering and quantum physics. Ricerche di Automatica **10**(2), 163–216 (1979)
65. Morgenstern, O.: Abraham Wald, 1902–1950. Econometrica **19**(4), 361–367 (1951)
66. Naiman, D.Q.: Conservative confidence bands in curvilinear regression. Ann. Statist. **14**, 896–906 (1986)
67. Neyman, J.: Foundations of behavioristic statistics. In: Foundations of Statistical Inference, Proc. Sympos., Univ. Waterloo, Waterloo, Ontario, 1970, pp. 1–19. Holt, Toronto (1971)
68. Page, W.: An interview with Herbert Robbins. College Math. J. **15**, 2–24 (1944)
69. de la Peña, V., Klass, M.J., Lai, T.L.: Self-normalized processes: Exponential inequalities, moment bounds and iterated logarithm laws. Annals of Probability **32**(3A), 1902–1933 (2004)
70. de la Peña, V., Lai, T.L.: Theory and applications of decoupling. In: C. Charalambides, M. Koutras, N. Balakrishnan (eds.) Probability and Statistical Models with Applications, pp. 115–145. Chapman & Hall/CRC (2001)
71. de la Peña, V., Lai, T.L., Shao, Q.M.: Self-Normalized Processes: Limit Theory and Statistical Applications. Springer (2009)

72. Philipp, W., Stout, W.: Almost sure invariance principles for partial sums of weakly dependent random variables. Amer. Math. Soc. (1975)
73. Robbins, H.: On the measure of a random set. Annals of Mathematical Statistics **15**(1), 70–74 (1944)
74. Robbins, H.: On the measure of a random set II. Annals of Mathematical Statistics **16**(4), 342–347 (1945)
75. Robbins, H.: Asymptotically subminimax solutions of compound statistical decision problems. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability pp. 131–148 (1951)
76. Robbins, H.: Some aspects of the sequential design of experiments. Bull. Amer. Math. Soc. **58**(5), 527–535 (1952)
77. Robbins, H.: Statistical methods related to the law of the iterated logarithm. Ann. Math. Statist. **41**(5), 1397–1409 (1970)
78. Robbins, H., Monro, S.: A stochastic approximation method. Annals of Mathematical Statistics **22**(3), 400–407 (1951)
79. Robbins, H., Siegmund, D.: Boundary crossing probabilities for the Wiener process and sample sums. Ann. Math. Statist. **41**, 1410–1429 (1970)
80. Robbins, H., Siegmund, D.: A convergence theorem for nonnegative almost supermartingales and some applications. In: Optimizing Methods in Statistics (Proc. Sympos., Ohio State Univ.), pp. 233–257. Academic Press (1971)
81. Robbins, H., Siegmund, D.: A class of stopping rules for testing parametric hypotheses. Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, 1970/1971 **IV: Biology and Health**, 37–41 (1972)
82. Robbins, H., Siegmund, D.: Statistical tests of power one and the integral representation of solutions of certain partial differential equations. Bull. Inst. Math. Acad. Sinica **1**, 93–120 (1973)
83. Robbins, H., Siegmund, D.: The expected sample size of some tests of power one. Ann. Statist. **2**, 415–436 (1974)
84. Sawyer, S.: A Fatou theorem for the general one-dimensional parabolic equation. Indiana Univ. Math. J. **24**, 451–498 (1974/75)
85. Siegmund, D.: Sequential Analysis. Springer (1985)
86. Storvik, G.: Particle filters for state-space models with the presence of unknown static parameters. IEEE Transactions on Signal Processing **50**(2), 281–289 (2002)
87. Wald, A.: On cumulative sums of random variables. Ann. Math. Statist. **15**(3), 283–296 (1944)
88. Wald, A.: Sequential tests of statistical hypothesis. Ann. Math. Statist. **16**(2), 117–186 (1945)
89. Whittle, P.: Multi-armed bandits and the Gittins index. J. Roy. Statist. Soc. Ser. B **42** (1980)
90. Widder, D.V.: Positive temperatures on an infinite rod. Trans. Amer. Math. Soc. **55**, 85–95 (1944)
91. Widder, D.V.: Positive temperatures on a semi-infinite rod. Trans. Amer. Math. Soc. **75**, 510—525 (1953)
92. Wijsman, R.: Sequential confidence intervals with beta protection in one-parameter families. In: J. Van Ryzin (ed.) Adaptive Statistical Procedure and Related Topics, pp. 96–107. Institute of Mathematical Statistics (1985)
93. Woodroofe, M.: Nonlinear Renewal Theory in Sequential Analysis. SIAM (1982). CBMS-NSF Regional Conference Series in Applied Mathematics 39
94. Working, H., Hotelling, H.: Applications of the theory of error to the interpretation of trends. J. Amer. Statist. Assoc. **24**(165A), 73–85 (1929)
95. Wu, H.T., Lai, T.L., Haddad, G.G., Muotri, A.: Oscillatory biomedical signals: Frontiers in mathematical models and statistical analysis. Frontiers in Applied Mathematics and Statistics **7** (2021). 31 pp.
96. Yau, S.T., Yau, S.S.T.: Finite-dimensional filters with nonlinear drift. III: Duncan-Mortensen-Zakai equation with arbitrary initial condition for the linear filtering system and the benes filtering system. IEEE Transactions on Aerospace and Electronic Yystems **33**(4), 1277–1294 (1997)

97. Zeng, Y.: A partially observed model for micromovement of asset prices with bayes estimation via filtering. Mathematical Finance **13**(3), 411–444 (2003)
98. Zeng, Y.: Estimating stochastic volatility via filtering for the micromovement of asset prices. IEEE Transactions on Automatic Control **49**(3), 338–348 (2004)
99. Zheng, T., Ying, Z.: Columbia University statistics. In: A. Agresti, X. Meng (eds.) Strength in Numbers: The Rising of Academic Departments in the U.S. Springer (2013)

Martingales in Survival Analysis

Odd O. Aalen, Per K. Andersen, Ørnulf Borgan, Richard D. Gill and Niels Keiding

Abstract

The paper traces the development of the use of martingale methods in survival analysis from the mid 1970s to the early 1990s. This development was initiated by Aalen's Berkeley Ph.D.-thesis in 1975, progressed in the late 1970s and early 1980s through work on the estimation of Markov transition probabilities, non-parametric tests and Cox's regression model, and was consolidated in the early 1990s with the publication of the monographs by Fleming and Harrington and by Andersen, Borgan, Gill and Keiding. The development was made possible by an unusually fast technology transfer of pure mathematical concepts, primarily from French probability, into practical biostatistical methodology, and we attempt to outline some of the personal relationships that helped this happen. We also point out that survival analysis was ready for this development since the martingale

Niels Keilding (1944–2022) was affiliated to the Department of Biostatistics, University of Copenhagen, Denmark. In March 2022, before the final publication of this book, Niels Keiding passed away. Niels had been a main contributor to the field of survival analysis based on martingales and he took the initiative to establishing the group of authors of the 1993 monograph 'Statistical Models Based on Counting processes'.

O. O. Aalen (✉)
Department of Biostatistics, University of Oslo, Oslo, Norway
e-mail: o.o.aalen@medisin.uio.no

P. K. Andersen · N. Keiding
Department of Biostatistics, University of Copenhagen, Copenhagen, Denmark
e-mail: P.K.Andersen@biostat.ku.dk

Ø. Borgan
Department of Mathematics, University of Oslo, Oslo, Norway
e-mail: borgan@math.uio.no

R. D. Gill
Mathematical Institute, Leiden University, Leiden, Netherlands
e-mail: gill@math.leidenuniv.nl

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*,
Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_13

ideas inherent in the deep understanding of temporal development so intrinsic to the French theory of processes were already quite close to the surface in survival analysis.

1 Introduction

Survival analysis is one of the oldest fields of statistics, going back to the beginning of the development of actuarial science and demography in the 17th century. The first life table was presented by John Graunt [60]. Until well after the Second World War the field was dominated by the classical approaches developed by the early actuaries, like e.g. Wilhelm Lexis [19].

As the name indicates, survival analysis may be about the analysis of actual survival in the true sense of the word, that is death rates, or mortality. However, survival analysis today has a much broader meaning, as the analysis of the time of occurrence of any kind of event one might want to study. A problem with survival data, which does not generally arise with other types of data, is the occurrence of censoring. By this one means that the event to be studied, may not necessarily happen in the time window of observation. So observation of survival data is typically incomplete; the event is observed for some individuals and not for others. This mixture of complete and incomplete data is a major characteristic of survival data, and it is a main reason why special methods have been developed to analyse this type of data.

A major advance in the field of survival analysis took place from the 1950s. The inauguration of this new phase is represented by the paper by Kaplan and Meier [59] where they propose their famous estimator of the survival curve. This is one of the most cited papers in the history of statistics with more than 49,000 citations in the Web of Science (by July 2022). While the classical life table method was based on a coarse division of time into fixed intervals, e.g. one-year or five-year intervals, Kaplan and Meier realized that the method worked quite as well for short intervals, and actually for intervals of infinitesimal length. Hence they proposed what one might call a continuous-time version of the old life table. Their proposal corresponded to the development of a new type of survival data, namely those arising in clinical trials where individual patients were followed on a day to day basis and times of events could be registered precisely. Also, for such clinical research the number of individual subjects was generally much smaller than in the actuarial or demographic studies. So, the development of the Kaplan–Meier method was a response to a new situation creating new types of data.

The 1958 Kaplan–Meier paper opened a new area, but also raised a number of new questions. How, for instance, does one compare survival curves? A literature of tests for differences between survival curves for two or more samples in the 1960s and 1970s, but it was rather confusing. The more general issue of how to adjust for covariates was first resolved by the introduction of the proportional hazards model by David Cox [36]. This was a major advance, and the more than 37,000 citations that Cox's paper has attracted in the Web of Science (by July 2022) is a proof of its huge impact.

However, with this development the theory lagged behind. Why did the Cox model work? How should one understand the plethora of tests? What were the asymptotic properties of the Kaplan–Meier estimator? In order to understand this, one had to take seriously the stochastic process character of the data, and the martingale concept turned out to be very useful in the quest for a general theory. The present authors were involved in pioneering work in this area from the mid-seventies and we shall describe the development of these ideas. It turned out that the martingale concept had an important role to play in statistics. In the 45 years gone by since the start of this development, there is now an elaborate theory, and it has started to penetrate into the general theory of longitudinal data [39]. However, martingales are not really entrenched in statistics in the sense that statistics students are routinely taught about martingales. While almost every statistician will know the concept of a Markov process, far fewer will have a clear understanding of the concept of a martingale. We hope that this historical account will help statisticians, and probabilists, understand why martingales are so valuable in survival analysis.

It should be mentioned that this was of course not the first use of martingales in statistics. For instance, martingales have played an important role in sequential analysis [61].

The introduction of martingales into survival analysis started with the 1975 Berkeley Ph.D. thesis of Odd Aalen [2] and was then followed up by the Copenhagen based cooperation between several of the present authors. The first journal presentation of the theory was given in 1978 by Aalen [7]. General textbook introductions from our group have been given by Andersen, Borgan, Gill and Keiding [17], and by Aalen, Borgan and Gjessing [9]. An earlier textbook was the one by Fleming and Harrington [45].

In a sense, martingales were latent in the survival field prior to the formal introduction. With hindsight there is a lot of martingale intuition in the famous Mantel–Haenszel test [66] and in the fundamental partial likelihood paper by Cox [37], but martingales were not mentioned in these papers. Interestingly, Tarone and Ware [79] use dependent central limit theory which is really of a martingale nature.

The present authors were all strongly involved in the developments we describe here, so our views represent the subjective perspective of active participants.

Below we shall focus on how hazard rates can be naturally understood in a martingale context. In particular, the theory of stochastic integration plays a major role, as well as central limit theory.

2 The Hazard Rate and a Martingale Estimator

In order to understand the events leading to the introduction of martingales in survival analysis, one must take a look at an estimator which is connected to the Kaplan–Meier estimator, and which today is called the Nelson–Aalen estimator. This estimation procedure focuses on the concept of a hazard rate. While the survival curve simply tells us how many have survived up to a certain time, the hazard rate gives us the risk of the event happening as a function of time, conditional on not having happened previously.

Mathematically, let the random variable T denote the survival time of an individual. The survival curve is then given by $S(t) = P(T > t)$. The hazard rate is defined by means of a conditional probability. Assuming that T is absolutely continuous (i.e., has a probability density), one looks at those who have survived up to some time t, and considers the probability of the event happening in a small time interval $[t, t + \mathrm{d}t)$. The hazard rate is defined as the following limit:

$$\alpha(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T < t + \Delta t \,|T \geq t).$$

Notice that, while the survival curve is a function that starts in 1 and then declines (or is partly constant) over time, the hazard function can be essentially any non-negative function.

While it is simple to estimate the survival curve, it is more difficult to estimate the hazard rate as an arbitrary function of time. What, however, is quite easy is to estimate the cumulative hazard rate defined as

$$A(t) = \int_0^t \alpha(s)\mathrm{d}s.$$

A non-parametric estimator of $A(t)$ was first suggested by Wayne Nelson [71,72] as a graphical tool to obtain engineering information on the form of the survival distribution in reliability studies; see also [73]. The same estimator was independently suggested by Altshuler [13] and by Aalen in his 1972 master thesis, which was partly published as a statistical research report from the University of Oslo [1] and later in [3]. The mathematical definition of the estimator is given in (2) below.

In the 1970s there were close connections between Norwegian statisticians and the Department of Statistics at Berkeley, with the Berkeley professors Kjell Doksum (originally Norwegian) and Erich Lehmann playing particularly important roles. Several Norwegian statisticians went to Berkeley in order to take a Ph.D. The main reason for this was to get into a larger setting, which could give more impulses than what could be offered in a small country like Norway. Also, Berkeley offered a regular Ph.D. program that was an alternative to the independent type doctoral dissertation in the old European tradition, which was common in Norway at the time. Odd Aalen also went there with the intention to follow up on his work in his master thesis. The introduction of martingales in survival analysis was first presented in his 1975 Berkeley Ph.D. thesis [2] and was in a sense a continuation of his master thesis. Aalen was influenced by his master thesis supervisor Jan M. Hoem who emphasized the importance of continuous-time Markov chains as a tool in the analysis when several events may occur to each individual (e.g., first the occurrence of an illness, and then maybe death; or the occurrence of several births for a woman). A subset of a state space for such a Markov chain may be illustrated as in Fig. 1. Consider two states i and j in the state space, with $Y(t)$ the number of individuals being in state i at time t, and with $N(t)$ denoting the number of transitions from i to j in the time interval $[0, t]$. The intensity of a new event, i.e., a new transition occurring, is then seen to be $\lambda(t) = \alpha(t)Y(t)$. Censoring is easily incorporated in this setup, and the setup covers the usual survival situation if the two states i and j are the only states in the system with one possible transition, namely the one from i to j.

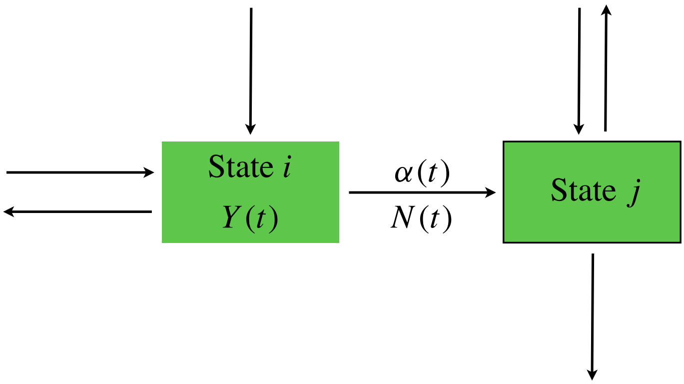


Fig. 1 Transition in a subset of a Markov chain

The idea was to abstract from the above a general model, later termed the multiplicative intensity model; namely where the intensity $\lambda(t)$ of a counting process $N(t)$ can be written as the product of an observed process $Y(t)$ and an unknown intensity function $\alpha(t)$, i.e.

$$\lambda(t) = \alpha(t)Y(t). \tag{1}$$

This gives approximately

$$dN(t) \approx \lambda(t)dt = \alpha(t)Y(t)dt,$$

that is

$$\frac{dN(t)}{Y(t)} \approx \alpha(t)dt,$$

and hence a reasonable estimate of $A(t) = \int_0^t \alpha(s)ds$ would be:

$$\widehat{A}(t) = \int_0^t \frac{dN(s)}{Y(s)}. \tag{2}$$

This is precisely the Nelson–Aalen estimator.

Although a general formulation of this concept can be based within the Markov chain framework as defined above, it is clear that this really has nothing to do with the Markov property. Rather, the correct setting would be a general point process, or counting process, $N(t)$ where the intensity process as a function of past occurrences, $\lambda(t)$, satisfied the property (1).

This was clear to Aalen before entering the Ph.D. study at the University of California at Berkeley in 1973. The trouble was that no mathematical theory for counting processes with intensity processes dependent on the past had been published in the general literature by that time. Hence there was no possibility of formulating general results for the Nelson–Aalen estimator and related quantities. On arrival in Berkeley, Aalen was checking the literature and at one time in 1974 he asked professor David Brillinger at the Department of Statistics whether he knew about any such theory. Brillinger had then recently received the Ph.D. thesis of Pierre Bremaud

[30], who had been a student at the Electronics Research Laboratory in Berkeley, as well as preprints of papers by Boel, Varaiya and Wong [28,29] from the same department. These fundamental papers laid a theoretical probabilistic foundation for counting processes. The papers were not related to statistical analysis, but Aalen noted that the theory they contained was precisely the right tool for giving a proper statistical theory for the Nelson–Aalen estimator. Soon it turned out that the theory led to a much wider reformulation of the mathematical basis of the whole of survival and event history analysis, the latter meaning the extension to transitions between several different possible states.

The mentioned papers by Bremaud and by Boel, Varayia and Wong were apparently the first to give a proper mathematical theory for counting processes with a general intensity process. As explained in this historical account, it turned out that martingale theory was of fundamental importance. With hindsight, it is easy to see why this is so. Let us start with a natural heuristic definition of an intensity process formulated as follows:

$$\lambda(t) = \frac{1}{\mathrm{d}t} P(\mathrm{d}N(t) = 1 \mid \text{past}), \tag{3}$$

where $\mathrm{d}N(t)$ denotes the number of jumps (essentially 0 or 1) in $[t, t + dt)$. We can rewrite the above as

$$\lambda(t) = \frac{1}{dt} E(\mathrm{d}N(t) \mid \text{past}),$$

that is

$$E(\mathrm{d}N(t) - \lambda(t)\mathrm{d}t \mid \text{past}) = 0, \tag{4}$$

where $\lambda(t)$ can be moved inside the conditional expectation since it is a function of the past. Let us now introduce the following process:

$$M(t) = N(t) - \int_0^t \lambda(s)\mathrm{d}s. \tag{5}$$

Note that (4) can be rewritten

$$E(\mathrm{d}M(t) \mid \text{past}) = 0.$$

This is of course a (heuristic) definition of a martingale. Hence the natural intuitive concept of an intensity process (3) is equivalent to asserting that the counting process minus the integrated intensity process is a martingale.

The Nelson–Aalen estimator is now derived as follows. Using the multiplicative intensity model of formula (1) we can write:

$$\mathrm{d}N(t) = \alpha(t)\, Y(t)\mathrm{d}t + \mathrm{d}M(t). \tag{6}$$

For simplicity, we shall assume $Y(t) > 0$ (this may be modified, see e.g. [17]). Dividing over (6) by $Y(t)$ yields

$$\frac{1}{Y(t)}\mathrm{d}N(t) = \alpha(t) + \frac{1}{Y(t)}\mathrm{d}M(t).$$

By integration we get

$$\int_0^t \frac{\mathrm{d}N(s)}{Y(s)} = \int_0^t \alpha(s)\mathrm{d}s + \int_0^t \frac{\mathrm{d}M(s)}{Y(s)}. \tag{7}$$

The right-most integral is recognized as a stochastic integral with respect to a martingale, and is therefore itself a zero-mean martingale. This represents noise in our setting and therefore $\widehat{A}(t)$ is an unbiased estimator of $A(t)$, with the difference $\widehat{A}(t) - A(t)$ being a martingale. Usually there is some probability that $Y(t)$ may become zero, which gives a slight bias.

The focus of the Nelson–Aalen estimator is the hazard $\alpha(t)$, where $\alpha(t)\mathrm{d}t$ is the instantaneous probability that an individual at risk at time t has an event in the next little time interval $[t, t + \mathrm{d}t)$. In the special case of survival analysis we study the distribution function $F(t)$ of a nonnegative random variable, which we for simplicity assume has density $f(t) = F'(t)$, which implies $\alpha(t) = f(t)/(1 - F(t))$, $t > 0$. Rather than studying the hazard $\alpha(t)$, interest is often on the survival function $S(t) = 1 - F(t)$, relevant to calculating the probability of an event happening over some finite time interval $(s, t]$.

To transform the Nelson–Aalen estimator into an estimator of $S(t)$ it is useful to consider the *product-integral* transformation [49,50]:

$$S(t) = \prod_{(0,t]} \{1 - \mathrm{d}A(s)\}.$$

Since $A(t) = \int_0^t \alpha(s)\mathrm{d}s$ is the cumulative intensity corresponding to the hazard function $\alpha(t)$, we have

$$\prod_{(0,t]} \{1 - \mathrm{d}A(s)\} = \exp\left(-\int_0^t \alpha(s)\mathrm{d}s\right),$$

while if $A(t) = \sum_{s_j \le t} h_j$ is the cumulative intensity corresponding to a discrete measure with jump h_j at time s_j ($s_1 < s_2 < \cdots$) then

$$\prod_{(0,t]} \{1 - \mathrm{d}A(s)\} = \prod_{s_j \le t} \left\{1 - h_j\right\}.$$

The plug-in estimator

$$\widehat{S}(t) = \prod_{(0,t]} \left\{1 - \mathrm{d}\widehat{A}(s)\right\} \tag{8}$$

is the Kaplan–Meier estimator [59]. It is a finite product of the factors $1 - 1/Y(t_j)$ for $t_j \leq t$, where $t_1 < t_2 < \cdots$ are the times of the observed events.

A basic martingale representation is available for the Kaplan–Meier estimator as follows. Still assuming $Y(t) > 0$ (see [17] for how to relax this assumption) it may be shown by Duhamel's equation that

$$\frac{\widehat{S}(t)}{S(t)} - 1 = -\int_0^t \frac{\widehat{S}(s-)}{S(s)Y(s)} \mathrm{d}M(s), \tag{9}$$

where the right-hand side is a stochastic integral of a predictable process with respect to a zero-mean martingale, that is, itself a martingale. "Predictable" is a mathematical formulation of the idea that the value is determined by the past, in our context it is sufficient that the process is adapted and has left-continuous sample paths. This representation is very useful for proving properties of the Kaplan–Meier estimator as shown by Gill [48].

3 Stochastic Integration and Statistical Estimation

The discussion in the previous section shows that the martingale property arises naturally in the modelling of counting processes. It is not a modelling assumption imposed from the outside, but is an integral part of an approach where one considers how the past affects the future. This dynamic view of stochastic processes represents what is often termed the French probability school. A central concept is the local characteristic, examples of which are transition intensities of a Markov chain, the intensity process of a counting process, drift and volatility of a diffusion process, and the generator of an Ornstein–Uhlenbeck process. The same concept is valid for discrete time processes, see [39] for a statistical application of discrete time local characteristics.

It is clearly important in this context to have a formal definition of what we mean by the "past". In stochastic process theory the past is formulated as a σ-algebra $\mathscr{F}_t$ of events, that is the family of events that can be decided to have happened or not happened by observing the past. We denote $\mathscr{F}_t$ as the *history at time t*, so that the entire history (or filtration) is represented by the increasing family of σ-algebras $\{\mathscr{F}_t\}$. Unless otherwise specified, processes will be adapted to $\{\mathscr{F}_t\}$, i.e., measurable with respect to $\mathscr{F}_t$ at any time t. The definition of a martingale $M(t)$ in this setting will be that it fulfils the relation:

$$\mathrm{E}(M(t) \mid \mathscr{F}_s) = M(s) \text{ for all} t > s.$$

In the present setting there are certain concepts from martingale theory that are of particular interest. Firstly, Eq. (5) can be rewritten as

$$N(t) = M(t) + \int_0^t \lambda(s)\mathrm{d}s.$$

This is a special case of the *Doob–Meyer decomposition*. This is a very general result, stating under a certain uniform integrability assumption that any submartingale can be decomposed into the sum of a martingale and a predictable process, which is often denoted a *compensator*. The compensator in our case is the stochastic process $\int_0^t \lambda(s)\mathrm{d}s$.

Two important variation processes for martingales are defined, namely the predictable variation process $\langle M \rangle$, and the optional variation process $[M]$. Assume that the time interval $[0, t]$ is divided into n equally long intervals, and define $\Delta M_k = M(kt/n) - M((k-1)t/n)$. Then

$$\langle M \rangle_t = \lim_{n\to\infty} \sum_{k=1}^{n} \mathrm{Var}(\Delta M_k \mid \mathscr{F}_{(k-1)/n}) \quad \text{and} \quad [M]_t = \lim_{n\to\infty} \sum_{k=1}^{n} (\Delta M_k)^2,$$

where the limits are in probability.

A second concept of great importance is stochastic integration. There is a general theory of stochastic integration with respect to martingales. Under certain assumptions, the central results are of the following kind:

1. A stochastic integral $\int_0^t H(s)\mathrm{d}M(s)$ of a predictable process $H(t)$ with respect to a martingale $M(t)$ is itself a martingale.
2. The variation processes satisfy:

$$\left\langle \int H\mathrm{d}M \right\rangle = \int H^2 \mathrm{d}\langle M \rangle \ \text{ and } \ \left[\int H\mathrm{d}M \right] = \int H^2 \mathrm{d}\,[M]\,. \qquad (10)$$

These formulas can be used to immediately derive variance formulas for estimators and tests in survival and event history analysis.

The general mathematical theory of stochastic integration is quite complex. What is needed for our application, however, is relatively simple. Firstly, one should note that the stochastic integral in Eq. (7) (the right-most integral) is simply the difference between an integral with respect to a counting processes and an ordinary Riemann integral. The integral with respect to a counting process is of course just of the sum of the integrand over jump times of the process. Hence, the stochastic integral in our context is really quite simple compared to the more general theory of martingales, where the martingales may have sample paths of infinite total variation on any interval, and where the Itō integral is the relevant theory. Still the above rules 1 and 2 are very useful in organizing and simplifying calculations and proofs.

4 Stopping Times, Unbiasedness and Independent Censoring

The concepts of martingale and stopping time in probability theory are both connected to the notion of a fair game and originate in the work of Ville [82,83]. In fact one of the older (non-mathematical) meanings of martingale is a fair-coin tosses betting system which is supposed to give a guaranteed payoff. The requirement of unbiasedness in statistics can be viewed as essentially the same concept as a fair game. This is particularly relevant in connection with the concept of censoring which pervades survival and event history analysis. As mentioned above, censoring simply means that the observation of an individual process stops at a certain time, and after this time there is no more knowledge about what happened.

In the 1960s and 1970s survival analysis methods were studied within reliability theory and the biostatistical literature assuming specific censoring schemes. The most important of these censoring schemes were the following:

- For *type I censoring*, the survival time T_i for individual i is observed if it is no larger than a fixed censoring time c_i, otherwise we only know that T_i exceeds c_i.
- For *type II censoring*, observation is continued until a given number of events r is observed, and then the remaining units are censored.
- *Random censoring* is similar to type I censoring, but the censoring times c_i are here the observed values of random variables C_i that are independent of the T_i's.

However, by adopting the counting process formulation, Aalen noted in his Ph.D. thesis and later journal publications, e.g. [7], that if censoring takes place at a stopping time, as is the case for the specific censoring schemes mentioned above, then the martingale property will be preserved and no further assumptions on the form of censoring is needed to obtain unbiased estimators and tests.

Aalen's argument assumed a specific form of the history, or filtration, $\{\mathscr{F}_t\}$. Namely that it is given as $\mathscr{F}_t = \mathscr{F}_0 \vee \mathscr{N}_t$, where $\{\mathscr{N}_t\}$ is the filtration generated by the uncensored individual counting processes, and $\mathscr{F}_0$ represents information available to the researcher at the outset of the study. However, censoring may induce additional variation not described by a filtration of the above form, so one may have to consider a larger filtration $\{\mathscr{G}_t\}$ also describing this additional randomness. The fact that we have to consider a larger filtration may have the consequence that the intensity processes of the counting processes may change. However, if this is not the case, so that the intensity processes with respect to $\{\mathscr{G}_t\}$ are the same as the $\{\mathscr{F}_t\}$-intensity processes, censoring is said to be *independent*. Intuitively this means that the additional knowledge of censoring times up to time t does not carry any information on an individual's risk of experiencing an event at time t.

A careful study of independent censoring for marked point process models along these lines was first carried out by Arjas and Hara [22]. The ideas of Arjas and Hara were taken up and further developed by Per Kragh Andersen, Ørnulf Borgan, Richard Gill, and Niels Keiding as part of their work on the monograph *Statistical Models Based on Counting Processes*; cf. Sect. 11 below. Discussions with Martin Jacobsen were also useful in this connection, see also [56]. Their results were published in [15] and later Chap. 3 of their monograph. It should be noted that there is a close connection between drop-outs in longitudinal data and censoring for survival data. In fact, independent censoring in survival analysis is essentially the same as *sequential missingness at random* in longitudinal data analysis, e.g. [52].

In many standard statistical models there is an intrinsic assumption of independence between outcome variables. While, in event history analysis, such an assumption may well be reasonable for the basic, uncensored observations, censoring may destroy this independence. An example is survival data in an industrial setting subject to type II censoring; that is the situation where items are put on test simultaneously and the experiment is terminated at the time of the rth failure (cf. above). However, for such situations martingale properties may be preserved; in fact, for type II censoring $\{\mathscr{G}_t\} = \{\mathscr{F}_t\}$ and censoring is trivially independent according to the definition just given. This suggests that, for event history data, the counting process and martingale framework is, indeed, the natural framework and that the martingale property replaces the traditional independence assumption, also in the sense that it forms the basis of central limit theorems, which will be discussed next.

5 Martingale Central Limit Theorems

As mentioned, the martingale property replaces the common independence assumption. One reason for the ubiquitous assumption of independence in statistics is to get some asymptotic distributional results of use in estimation and testing, and the martingale assumption can fulfil this need as well. After Bernstein's [26] and Levy's [62] early work, study of central limit theorems for discrete-time martingales was taken up in the 1960s and 1970s by Billingsley [27], Ibragimov [54], Brown [34], Dvoretzky [40], and McLeish [68], among others. Of particular importance for our historical account is the paper by McLeish [68] on central limit theorems for triangular arrays of martingale differences. The potential usefulness of this paper was pointed out to Aalen by his Ph.D. supervisor Lucien Le Cam. In fact this happened before the connection had been made to Bremaud's new theory of counting processes, and it was first after the discovery of this theory that the real usefulness of McLeish's paper became apparent. The application of counting processes to survival analysis including the application of McLeish's paper was done by Aalen during 1974–75.

The theory of McLeish was developed for the discrete-time case, and had to be further developed to cover the continuous-time setting of the counting process theory. What presumably was the first central limit theorem for continuous time martingales was published by Aalen [5]. A far more elegant and complete result was given by Rebolledo [75], and this formed the basis for further developments of the statistical theory; see [17] for an overview. A nice early result was also given by Helland [51].

The central limit theorem for martingales is related to the fact that a martingale with continuous sample paths and a deterministic predictable variation process is a Gaussian martingale, i.e., with normal finite-dimensional distributions. Hence one would expect a central limit theorem for counting process associated martingales to depend on two conditions:

(i) the sizes of the jumps go to zero (i.e., approximating continuity of sample paths)
(ii) either the predictable or the optional variation process converges to a deterministic function

In fact, the conditions in the papers by Aalen [5] and Rebolledo [75] are precisely of this nature.

Without giving the precise formulations of these conditions, let us look informally at how they work out for the Nelson–Aalen estimator. We saw in formula (7) that the difference between estimator and estimand of cumulative hazard up to time t could be expressed as $\int_0^t \mathrm{d}M(s)/Y(s)$, the stochastic integral of the process $1/Y$ with respect to the counting process martingale M. Considered as a stochastic process (i.e., indexed by time t), this "estimation-error process" is therefore itself a martingale. Using the rules (10) we can compute its optional variation process to be $\int_0^t \mathrm{d}N(s)/Y(s)^2$ and its predictable variation process to be $\int_0^t \alpha(s)\mathrm{d}s/Y(s)$. The error process only has jumps where N does, and at a jump time s, the size of the jump is $1/Y(s)$.

As a first attempt to get some large sample information about the Nelson–Aalen estimator, let us consider what the martingale central limit theorem could say about the Nelson–Aalen estimation-error process. Clearly we would need the number at risk process Y to get uniformly large, in order for the jumps to get small. In that case, the predictable variation process $\int_0^t \alpha(s)\mathrm{d}s/Y(s)$ is forced to be getting smaller and smaller. Going to the limit, we will have convergence to a continuous Gaussian martingale with zero predictable variation process. But the only such process is the constant process, equal to zero at all times. Thus in fact we obtain a consistency result: if the number at risk process gets uniformly large, in probability, the estimation error converges uniformly to zero, in probability. (Actually there are martingale inequalities of Chebyshev type which allow one to draw this kind of conclusion without going via central limit theory.)

In order to get nondegenerate asymptotic normality results, we should zoom in on the estimation error. A quite natural assumption in many applications is that there is some index n, standing perhaps for sample size, such that for each t, $Y(t)/n$ is roughly constant (non random) when n is large. Taking our cue from classical statistics, let us take a look at $\sqrt{n}$ times the estimation error process $\int_0^t \mathrm{d}M(s)/Y(s)$. This has jumps of size $(1/\sqrt{n})(Y(s)/n)^{-1}$. The predictable variation process of the rescaled estimation error is n times what it was before: it becomes $\int_0^t (Y(s)/n)^{-1}\alpha(s)\mathrm{d}s$. So, the convergence of Y/n to a deterministic function ensures simultaneously that the jumps of the rescaled estimation error process become vanishingly small and that its predictable variation process converges to a deterministic function.

The martingale central limit theorem turns out to be extremely effective in allowing us to guess the kind of results which might be true. Technicalities are reduced to a minimum; results are essentially optimal, i.e., the assumptions are minimal.

Why is that so? In probability theory, the 1960s and 1970s were the heyday of study of martingale central limit theorems. The outcome of all this work was that the martingale central limit theorem was not only a generalization of the classical Lindeberg central limit theorem, but that the proof was the same: it was simply a question of judicious insertion of conditional expectations, and taking expectations by repeated conditioning, so that the same line of proof worked exactly. In other words, the classical Lindeberg proof of the central limit theorem, see e.g. [42], already is the proof of the martingale central limit theorem.

The difficult extension, taking place in the 1970s to the 1980s, was in going from discrete time to continuous time processes. This required a major technical investigation of what are the continuous time processes which we are able to study effectively. This is quite different from research into central limit theorems for other kinds of processes, e.g., for stationary time series. In that field, one splits the process under study into many blocks, and tries to show that the separate blocks are almost independent if the distance between the blocks is large enough. The distance between the blocks should be small enough that one can forget about what goes on between. The central limit theorem comes from looking for approximately independent summands hidden somewhere inside the process of interest. However in the martingale case, one is already studying exactly the kind of process for which the best (sharpest, strongest) proofs are already attuned. No approximations are involved.

At the time martingales made their entry to survival analysis, statisticians were using many different tools to get large sample approximations in statistics. One had different classes of statistics for which special tools had been developed. Each time something was generalized from classical data to survival data, the inventors first showed that the old tools still worked to get some information about large sample properties (e.g. U-statistics, rank tests). Just occasionally, researchers saw a glimmering of martingales behind the scenes, as when Tarone and Ware [79] used the martingale central limit theorem of Dvoretzky [40] in the study of their class of non-parametric tests. Another important example of work where martingale type arguments were used, is the paper of Cox [37] on partial likelihood; cf. Sect. 10.

6 Two-Sample Tests for Counting Processes

During the 1960s and early 1970s a plethora of tests for comparing two or more survival functions were suggested [31,41,47,65,74]. The big challenge was to handle the censoring, and various simplified censoring mechanisms were proposed with different versions of the tests fitted to the particular censoring scheme. The whole setting was rather confusing, with an absence of a theory connecting the various specific cases. The first connection to counting processes was made by Aalen in his Ph.D. thesis when it was shown that a generalized Savage test (which is equivalent to the logrank test) could be given a martingale formulation. In a Copenhagen research report [4], Aalen extended this to a general martingale formulation of two-sample tests which turned out to encompass a number of previous proposals as special cases. The very simple idea was to write the test statistic as a weighted stochastic integral over the difference between two Nelson–Aalen estimators. Let the processes to be compared be indexed by $i = 1, 2$. A class of tests for comparing the two intensity functions $\alpha_1(t)$ and $\alpha_2(t)$ is then defined by

$$X(t) = \int_0^t L(s)\mathrm{d}(\widehat{A}_1(s) - \widehat{A}_2(s)) = \int_0^t L(s)\left(\frac{\mathrm{d}N_1(s)}{Y_1(s)} - \frac{\mathrm{d}N_2(s)}{Y_2(s)}\right).$$

Under the null hypothesis of $\alpha_1(s) \equiv \alpha_2(s)$ it follows that $X(t)$ is a martingale since it is a stochastic integral. An estimator of the variance can be derived from the rules for the variation processes, and the asymptotics is taken care of by the martingale central limit theorem. It was found by Aalen [7] and detailed by Gill [48] that almost all previous proposals for censored two-sample tests in the literature were special cases that could be arrived at by judicious choice of the weight function $L(t)$.

A thorough study of two-sample tests from this point of view was first given by Richard Gill in his Ph.D. thesis from Amsterdam [48]. The inspiration for Gill's work was a talk given by Odd Aalen at the European Meeting of Statisticians in Grenoble in 1976. At that time Gill was about to decide on the topic for his Ph.D. thesis, one option being two sample censored data rank tests. He was very inspired by Aalen's talk and the uniform way to treat all the different two-sample statistics offered by the counting process formulation, so this decided the topic for his thesis work. At that time, Gill had no experience with martingales in continuous time. But by reading Aalen's thesis and other relevant publications, he soon mastered the theory. To that end it also helped him that there was organized a study group in Amsterdam on counting processes with Piet Groeneboom as a key contributor.

7 The Copenhagen Environment

Much of the further development of counting process theory to statistical issues springs out of the statistics group at the University of Copenhagen. After his Ph.D. study in Berkeley, Aalen was invited by his former master thesis supervisor, Jan M. Hoem, to visit the University of Copenhagen, where Hoem had taken a position as professor in actuarial mathematics. Aalen spent 8 months there (November 1975 to June 1976) and his work immediately caught the attention of Niels Keiding, Søren Johansen, and Martin Jacobsen, among others. The Danish statistical tradition at the time had a strong mathematical basis combined with a growing interest in applications. Internationally, this combination was not so common at the time; mostly the theoreticians tended to do only theory while the applied statisticians were less interested in the mathematical aspects. Copenhagen made a fertile soil for the further development of the theory.

It was characteristic that for such a new paradigm, it took time to generate an intuition for what was obvious and what really required detailed study. For example, when Keiding gave graduate lectures on the approach in 1976/77 and 1977/78, he followed Aalen's thesis closely and elaborated on the mathematical prerequisites (stochastic processes in the French way, counting processes [57], square integrable martingales, martingale central limit theorem [68]). This was done in more mathematical generality than became the standard later. For example, he patiently went through the Doob–Meyer decompositions following Meyer's *Probabilités et Potentiel* [69], and he quoted the derivation by Courrège and Priouret [35] of the following result:

If (N_t) is a stochastic process, $\{\mathcal{N}_t\}$ is the family of σ-algebras generated by (N_t), and T is a stopping time (i.e. $\{T \leq t\} \in \mathcal{N}_t$ for all t), then the conventional definition of the σ-algebra $\mathcal{N}_T$ of events happening before T is

$$A \in \mathcal{N}_T \iff \forall t : A \cap \{T \le t\} \in \mathcal{N}_t.$$

A more intuitive way of defining this σ-algebra is

$$\mathcal{N}_T^* = \sigma\{N_{T \wedge u},\ u \ge 0\}.$$

Courrège and Priouret [35] proved that $\mathcal{N}_T = \mathcal{N}_T^*$ through a delicate analysis of the path properties of (N_t).

Keiding quoted the general definition, valid for measures with both discrete and continuous components, of predictability, not satisfying himself with the "essential equivalence to left-continuous sample paths" that we work with nowadays. Keiding had many discussions with his colleague, the probabilist Martin Jacobsen, who had long focused on path properties of stochastic processes. Jacobsen developed his own independent version of the course in 1980 and wrote his lecture notes up in the *Springer Lecture Notes in Statistics* series [55].

Among those who happened to be around in the initial phase was Niels Becker from Melbourne, Australia, already then well established with his work in infectious disease modelling. For many years to come martingale arguments were used as important tools in Becker's further work on statistical models for infectious disease data; see [25] for an overview of this work. A parallel development was the interesting work of Arjas and coauthors on statistical models for marked point processes, see e.g. [21,22].

8 From Kaplan–Meier to the Empirical Transition Matrix

A central effort initiated in Copenhagen in 1976 was the generalization from scalar to matrix values of the Kaplan–Meier estimator. This started out with the estimation of transition probabilities in the competing risks model developed by Aalen [1]; a journal publication of this work first came in [6]. This work was done prior to the introduction of martingale theory, and just like the treatment of the cumulative hazard estimator in [3] it demonstrates the complications that arose before the martingale tools had been introduced. In 1973 Aalen had found a matrix version of the Kaplan–Meier estimator for Markov chains, but did not attempt a mathematical treatment because this seemed too complex. It was the martingale theory that allowed an elegant and compact treatment of these attempts to generalize the Kaplan–Meier estimator, and the breakthrough here was made by Søren Johansen in 1975–76. It turned out that martingale theory could be combined with the product-integral approach to non-homogeneous Markov chains via an application of Duhamel's equality; cf. (12) below. The theory of stochastic integrals could then be used in a simple and elegant way. This was written down in a research report [11] and published by Aalen and Johansen [12].

Independently of this the same estimator was developed by Fleming [43,44] and published just prior to the publication of Aalen and Johansen (and duly acknowledged in their paper). Tom Fleming and David Harrington were Ph.D. students of Grace

Yang at the University of Maryland, and they have later often told us that they learned about Aalen's counting process theory from Grace Yang's contact with her own former Ph.D. advisor, Lucien Le Cam. Fleming also based his work on the martingale counting process approach. He had a more complex presentation of the estimator presenting it as a recursive solution of equations; he did not have the simple matrix product version of the estimator nor the compact presentation through the Duhamel equality which allowed for general censoring and very compact formulas for covariances.

The estimator is named the empirical transition matrix, see e.g. [9]. The compact matrix product version of the estimator presented in [12] is often called the Aalen–Johansen estimator, and we are going to explain the role of martingales in this estimator.

More specifically, consider an inhomogeneous continuous-time Markov process with finite state space $\{1, \dots, k\}$ and transition intensities $\alpha_{hj}(t)$ between states h and j, where in addition we define $\alpha_{hh}(t) = -\sum_{j \neq h} \alpha_{hj}(t)$ and denote the matrix of all $A_{hj}(t) = \int_0^t \alpha_{hj}(s)\mathrm{d}s$ as $\mathbf{A}(t)$. Nelson–Aalen estimators $\widehat{A}_{hj}(t)$ of the cumulative transition intensities $A_{hj}(t)$ may be collected in the matrix $\widehat{\mathbf{A}}(t) = \{\widehat{A}_{hj}(t)\}$. To derive an estimator of the transition probability matrix $\mathbf{P}(s,t) = \{P_{hj}(s,t)\}$ it is useful to represent it as the matrix product-integral

$$\mathbf{P}(s,t) = \prod_{(s,t]} \{\mathbf{I} + \mathrm{d}\mathbf{A}(u)\},$$

which may be defined as

$$\prod_{(s,t]} \{\mathbf{I} + \mathrm{d}\mathbf{A}(u)\} = \lim_{\max|u_i - u_{i-1}| \to 0} \prod_i \{\mathbf{I} + \mathbf{A}(u_i) - \mathbf{A}(u_{i-1})\},$$

where $s = u_0 < u_1 < \dots < u_n = t$ is a partition of $(s,t]$ and the matrix product is taken in its natural order from left to right.

The empirical transition matrix or Aalen–Johansen estimator is the plug-in estimator

$$\widehat{\mathbf{P}}(s,t) = \prod_{(s,t]} \left\{\mathbf{I} + \mathrm{d}\widehat{\mathbf{A}}(u)\right\}, \tag{11}$$

which may be evaluated as a finite matrix product over the times in $(s,t]$ when transitions are observed. Note that (11) is a multivariate version of the Kaplan–Meier estimator (8). A matrix martingale relation may derived from a matrix version of the Duhamel equation (9). For the case where all numbers at risk in the various states, $Y_h(t)$, are positive this reads

$$\widehat{\mathbf{P}}(s,t)\mathbf{P}(s,t)^{-1} - \mathbf{I} = \int_s^t \widehat{\mathbf{P}}(s,u-)\mathrm{d}(\widehat{\mathbf{A}} - \mathbf{A})(u)\mathbf{P}(s,u)^{-1}. \tag{12}$$

This is a stochastic integral representation from which covariances and asymptotic properties can be deduced directly. This particular formulation is from [12].

9 Pustulosis Palmo-Plantaris and k-Sample Tests

One of the projects that were started when Aalen visited the University of Copenhagen was an epidemiological study of the skin disease pustulosis palmo-plantaris with Aalen, Keiding and the medical doctor Jens Thormann as collaborators. Pustulosis palmo-plantaris is mainly a disease among women, and the question was whether the risk of the disease was related to the occurrence of menopause. Consecutive patients from a hospital out-patient clinic were recruited, so the data could be considered a random sample from the prevalent population. At the initiative of Jan M. Hoem, another of his former master students from Oslo, Ørnulf Borgan, was asked to work out the details. Borgan had since 1977 been assistant professor in Copenhagen, and he had learnt the counting process approach to survival analysis from the above mentioned series of lectures by Niels Keiding. The cooperation resulted in the paper [10].

In order to be able to compare patients without menopause with patients with natural menopause and with patients with induced menopause, the statistical analysis required an extension of Aalen's work on two-sample tests to more than two samples. (The work of Richard Gill on two-sample tests was not known in Copenhagen at that time.) The framework for such an extension is k counting processes $N_1, \ldots, N_k$, with intensity processes $\lambda_1, \ldots, \lambda_k$ of the multiplicative form $\lambda_j(t) = \alpha_j(t)Y_j(t)$; $j = 1, 2, \ldots, k$; and where the aim is to test the hypothesis that all the α_j are identical. Such a test may be based on the processes

$$X_j(t) = \int_0^t K_j(s)\mathrm{d}(\widehat{A}_j(s) - \widehat{A}(s)), \qquad j = 1, 2, \ldots, k,$$

where $\widehat{A}_j$ is the Nelson–Aalen estimator based on the jth counting process, and $\widehat{A}$ is the Nelson–Aalen estimator based on the aggregated counting process $N = \sum_{j=1}^k N_j$.

This experience inspired a decision to give a careful presentation of the k-sample tests for counting processes and how they gave a unified formulation of most rank based tests for censored survival data, and Per K. Andersen (who also had followed Keiding's lectures), Ørnulf Borgan, and Niels Keiding embarked on this task in the fall of 1979. During the work on this project, Keiding was (by Terry Speed) made aware of Richard Gill's work on two-sample tests. (Speed, who was then on sabbatical in Copenhagen, was at a visit in Delft where he came across an abstract book for the Dutch statistical association's annual gathering with a talk by Gill about the counting process approach to censored data rank tests.) Gill was invited to spend the fall of 1980 in Copenhagen. There he got a draft manuscript by Andersen, Borgan and Keiding on k-sample tests, and as he made a number of substantial comments to the manuscript, he was invited to co-author the paper [16].

10 The Cox Model

With the development of clinical trials in the 1950s and 1960s the need to analyze censored survival data dramatically increased, and a major breakthrough in this direction was the Cox proportional hazards model published in 1972 [36]. Now, regression analysis of survival data was possible. Specifically, the Cox model describes the hazard rate for a subject, i with covariates $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{ip})^\mathsf{T}$ as

$$\alpha(t \mid \mathbf{Z}_i) = \alpha_0(t) \exp(\boldsymbol{\beta}^\mathsf{T} \mathbf{Z}_i).$$

This is a product of a *baseline* hazard rate $\alpha_0(t)$, common to all subjects, and the exponential function of the linear predictor, $\boldsymbol{\beta}^\mathsf{T} \mathbf{Z}_i = \sum_j \beta_j Z_{ij}$. With this specification, hazard rates for all subjects are proportional and $\exp(\beta_j)$ is the hazard rate ratio associated with an increase of 1 unit for the jth covariate Z_j, that is the ratio

$$\exp(\beta_j) = \frac{\alpha(t \mid Z_1, Z_2, ..., Z_{j-1}, Z_j + 1, Z_{j+1}, ..., Z_p)}{\alpha(t \mid Z_1, Z_2, ..., Z_{j-1}, Z_j, Z_{j+1}, ..., Z_p)},$$

where Z_ℓ for $\ell \neq j$ are the same in numerator and denominator. The model formulation of Cox [36] allowed for covariates to be time-dependent and it was suggested to estimate $\boldsymbol{\beta}$ by the value $\widehat{\boldsymbol{\beta}}$ maximizing the *partial likelihood*

$$L(\boldsymbol{\beta}) = \prod_{i:D_i=1} \frac{\exp(\boldsymbol{\beta}^\mathsf{T} \mathbf{Z}_i(T_i))}{\sum_{j \in R_i} \exp(\boldsymbol{\beta}^\mathsf{T} \mathbf{Z}_j(T_i))}. \tag{13}$$

Here, $D_i = I(i \text{ was observed to fail})$ and R_i is the *risk set*, i.e., the set of subjects still at risk at the time, T_i, of failure for subject i. The cumulative baseline hazard rate $A_0(t) = \int_0^t \alpha_0(u) \mathrm{d}u$ was estimated by the Breslow estimator [32,33]

$$\widehat{A}_0(t) = \sum_{i:T_i \le t} \frac{D_i}{\sum_{j \in R_i} \exp(\widehat{\boldsymbol{\beta}}^\mathsf{T} \mathbf{Z}_j(T_i))}. \tag{14}$$

Cox's work triggered a number of methodological questions concerning inference in the Cox model. In what respect could the partial likelihood (13) be interpreted as a proper likelihood function? How could the large sample properties of the resulting estimators be established? Cox himself used repeated conditional expectations (which essentially was a martingale argument) to show informally that his partial likelihood (13) had similar properties as an ordinary likelihood [37], while Tsiatis [81] used classical methods to provide a thorough treatment of large sample properties of the estimators $(\widehat{\boldsymbol{\beta}}, \widehat{A}_0(t))$ when only time-fixed covariates were considered. The study of large sample properties, however, were particularly intriguing when time-dependent covariates were allowed in the model.

At the Statistical Research Unit in Copenhagen, established in 1978, analysis of survival data was one of the key research areas and several applied medical projects using the Cox model were conducted. One of these projects, initiated in 1978 and published in [20], dealt with recurrent events: admissions to psychiatric hospitals among pregnant women and among women having given birth or having had induced abortion. Here, a model for the intensity of admissions was needed and since previous admissions were strongly predictive for new admissions, time-dependent covariates should be accounted for. Counting processes provided a natural framework in which to study the phenomenon and research activities in this area were already on the agenda, as exemplified above.

It soon became apparent that the Cox model could be immediately applied for the recurrent event intensity, and Johansen's derivation of Cox's partial likelihood as a profile likelihood [58] also generalized quite easily. The individual counting processes, $N_i(t)$, counting admissions for woman i could then be "Doob–Meyer decomposed" as

$$N_i(t) = \int_0^t Y_i(u)\alpha_0(u)\exp(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}_i(u))\mathrm{d}u + M_i(t). \tag{15}$$

Here, $Y_i(t)$ is the at-risk indicator process for woman i (indicating that she is still in the study and out of hospital at time t), $\mathbf{Z}_i(t)$ is the, possibly time-dependent, covariate vector including information on admissions before t, and $\alpha_0(t)$ the unspecified baseline hazard. Finally, $M_i(t)$ is the martingale. We may write the sum over event times in the score $\mathbf{U}(\boldsymbol{\beta}) = \partial \log L(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$, derived from Cox's partial likelihood (13), as the counting process integral

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_i \int_0^\infty \left(\mathbf{Z}_i(u) - \frac{\sum_j Y_j(u)\mathbf{Z}_j(u)\exp(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}_j(u))}{\sum_j Y_j(u)\exp(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}_j(u))}\right)\mathrm{d}N_i(u).$$

Then using (15), the score can be re-written as $U_\infty(\boldsymbol{\beta})$, where

$$\mathbf{U}_t(\boldsymbol{\beta}) = \sum_i \int_0^t \left(\mathbf{Z}_i(u) - \frac{\sum_j Y_j(u)\mathbf{Z}_j(u)\exp(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}_j(u))}{\sum_j Y_j(u)\exp(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}_j(u))}\right)\mathrm{d}M_i(u).$$

Thus, evaluated at the true parameter values, the Cox score, considered as a process in t is a martingale stochastic integral, provided the time-dependent covariates (and $Y_i(t)$) are predictable.

Large sample properties for the score could then be established using the martingale central limit theorem and transformed into a large sample result for $\widehat{\boldsymbol{\beta}}$ by standard Taylor expansions. Also, asymptotic properties of the Breslow estimator (15) could be established using martingale methods. This is because we may write the estimator as $\widehat{A}_0(t) = \widehat{A}_0(t \mid \widehat{\boldsymbol{\beta}})$, where for the true value of $\boldsymbol{\beta}$ we have

$$\widehat{A}_0(t \mid \boldsymbol{\beta}) = \int_0^t \frac{\sum_i \mathrm{d}N_i(u)}{\sum_j Y_j(u)\exp(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}_j(u))} = A_0(t) + \int_0^t \frac{\sum_i \mathrm{d}M_i(u)}{\sum_j Y_j(u)\exp(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}_j(u))}.$$

That is, $\widehat{A}_0(t \mid \boldsymbol{\beta}) - A_0(t)$ is a martingale stochastic integral. These results were obtained by Per Kragh Andersen in 1979–80, but a number of technicalities remained to get proper proofs.

As mentioned above, Richard Gill visited Copenhagen in 1980 and he was able to provide the proof of consistency and work out the detailed verifications of the general conditions for the asymptotic results in Andersen and Gill's *Annals of Statistics* paper [18]. It should be noted that Næs [70], independently, published similar results under somewhat more restrictive conditions using discrete-time martingale results.

Obviously, the results mentioned above also hold for counting processes $N_i(t) = I(T_i \le t, D_i = 1)$ derived from censored survival times and censoring indicators, (T_i, D_i), but historically the result was first derived for the "Andersen-Gill" recurrent events process. Andersen and Borgan [14], see also [17, Chap. VII], extended these results to multivariate counting processes modelling the occurrence of several types of events in the same subjects.

Later, Barlow and Prentice [24] and Therneau, Grambsch and Fleming [80] used the Doob–Meyer decomposition (15) to define *martingale residuals*

$$\widehat{M}_i(t) = N_i(t) - \int_0^t \exp(\widehat{\boldsymbol{\beta}}^{\mathsf{T}} \mathbf{Z}_i(u)) \mathrm{d}\widehat{A}_0(u). \tag{16}$$

Note how $N_i(t)$ plays the role of the observed data while the compensator term estimates the expectation. We are then left with the martingale noise term.

The martingale residuals (16) provide the basis for a number of goodness-of-fit techniques for the Cox model. First, they were used to study whether the functional form of a quantitative covariate was modelled in a sensible way. Later, cumulative sums of martingale residuals have proven useful for examining several features of hazard based models for survival and event history data, including both the Cox model, Aalen's additive hazards model and others, e.g. [63,67]. The additive hazards model was proposed by Aalen [8] as a tool for analyzing survival data with changing effects of covariates. It is also useful for recurrent event data and dynamic path analysis, see e.g. [9].

11 The Monograph *Statistical Models Based on Counting Processes*

As the new approach spread, publishers became interested, and as early as 1982 Martin Jacobsen had published his exposition in the Springer Lecture Notes in Statistics [55]. In 1982 Niels Keiding gave an invited talk "Statistical applications of the theory of martingales on point processes" at the Bernoulli Society conference on Stochastic Processes in Clermont-Ferrand. (One slide showed a graph of a simulated sample function of a compensator, which prompted the leading French probabilist Michel Métivier to exclaim "This is the first time I have seen a compensator".) At that conference Klaus Krickeberg, himself a pioneer in martingale theory and an advisor to the Springer Series in Statistics, invited Keiding to write a monograph on this topic.

Keiding floated this idea in the well-established collaboration with Andersen, Borgan and Gill. Aalen was asked to participate, but had just started to build up a group of medical statistics in Oslo and wanted to give priority to that. So the remaining four embarked upon what became an intense 10-year collaboration resulting in the monograph *Statistical Models Based on Counting Processes* [17]. The monograph combines concrete practical examples, almost all of the authors' own experience, with an exposition of the mathematical background, several detailed chapters on non- and semiparametric models as well as parametric models, and chapters giving preliminary glimpses into topics to come: semiparametric efficiency, frailty models (for more elaborate introductions of frailty models see [53] or [9]) and multiple time-scales. Fleming and Harrington had published their monograph *Counting Processes and Survival Analysis* with Wiley in 1991 [45]. It gives a more textbook-type presentation of the mathematical background and covers survival analysis up to and including the proportional hazards model for survival data.

12 Limitations of Martingales

Martingale tools do not cover all areas where survival and event history analysis may be used. In more complex situations one can see the need to use a variety of tools, alongside of what martingale theory provides. For staggered entry, the Cox frailty model, and in Markov renewal process/semi-Markov models (see e.g. Chaps. IX and X in [17] for references on this work), martingale methods give transparent derivations of mean values and covariances, likelihoods, and maximum likelihood estimators; however to derive large sample theory, one needs input from the theory of empirical processes. Thus in these situations the martingale approach helps at the modelling stage and the stage of constructing promising statistical methodology, but one needs different tools for the asymptotic theory. The reason for this in a number of these examples is that the martingale structure corresponds to the dynamics of the model seen in real (calendar) time, while the principal time scales of statistical interest correspond to time since an event which is repeated many times. In the case of frailty models, the problem is that there is an unobserved covariate associated with each individual; observing that individual at late times gives information about the value of the covariate at earlier times. In all these situations, the natural statistical quantities to study can no longer be directly expressed as sums over contributions from each (calendar) time point, weighted by information only from the (calendar time) past. More complex kinds of missing data (frailty models can be seen as an example of missing data), and biased sampling, lead also to new levels of complexity in which the original dynamical time scale becomes just one feature of the problem at hand, other features which do not mesh well with this time scale become dominating, with regards to the technical investigation of large sample behaviour. A difficulty with the empirical process theory is the return to a basis of independent processes, and so a lot of the niceness of the martingale theory is lost. Martingales allow for very general dependence between processes.

However, the martingale ideas also enter into new fields. Lok [64] used martingale theory to understand the continuous time version of James Robins' theory of causality. This was focused on structural nested models proposed by Robins [76]. Martingales were used to provide a conceptual framework and give asymptotic theory. The fundamental idea is that medical treatments may be started or stopped or changed for patients based on the development of the disease over time. Hence, the intensity process is a natural concept for describing the clinical history and the treatments for patients. And, clearly, treatment decisions have to be adapted to the past developments in accordance with martingale ideas. The fundamental causal question is what happens when one compares two treatments. This comparison has to be understood in a counterfactual sense, that is, it should be correct when considering actual interventions at the two treatment levels. In order for the treatment effect to be estimated correctly from the statistical model, one has to make the "no unmeasured confounding" assumption [76]. This means that the treatment decisions should only depend on the observable information in the model. Lok [64] describes very clearly how counting process martingale ideas can be combined with causal inference ideas in a highly fruitful fashion.

Similarly, Didelez [38] used martingales to understand the modern formulation of local dependence and Granger causality. Connected to this are the work of Arjas and Parner [23] on posterior predictive distributions for marked point process models and the dynamic path analysis of Fosen et al. [46], see also [9]. More recently new developments in causal inference using martingale theory were developed by Røysland [77] and Ryalen et al. [78]. This work use stochastic differential equations and related tools.

Hence, there is a new lease of life for the theory. Fundamentally, the idea of modelling how the past influences the present and the future is inherent to the martingale formulation, and this must with necessity be of importance in understanding causality.

The martingale concepts from the French probability school may be theoretical and difficult to many statisticians. The work of Jacobsen [55] and Helland [51] are nice examples of how the counting process work stimulated probabilists to reappraise the basic probability theory. Both authors succeeded in giving a much more compact and elementary derivation of (different parts of) the basic theory from probability needed for the statistics. This certainly had a big impact at the time, in making the field more accessible to more statisticians. Especially while the fundamental results from probability were still in the course of reaching their definitive forms and were often not published in the most accessible places or languages. Later these results became the material of standard textbooks. In the long run, statisticians tend to use standard results from probability without bothering too much about how one can prove them from scratch. Once the martingale theory became well established people were more confident in just citing the results they needed.

Biostatistical papers have hardly ever cited papers or even books in theoretical probability. However at some point it became almost obligatory to cite Andersen and Gill [18], Andersen et al. [17], and other such works. What was being cited then was worked out examples of applying the counting process approach to various more or less familiar applied statistical tools like the Cox regression model, especially when being used in a little bit non-standard context, e.g., with repeated events. It helped that some software packages also refer to such counting process extensions as the basic biostatistical tool.

The historical overview presented here shows that the elegant theory of martingales has been used fruitfully in statistics. This is another example showing that mathematical theory developed on its own terms may produce very useful practical tools.

Acknowledgements Niels Keiding and Per Kragh Andersen were supported by National Cancer Institute; Grant Number: R01-54706-13 and Danish Natural Science Research Council; Grant Number: 272-06-0442. We are grateful to Judith Lok for comments on the use of martingales in causal inference.

References

1. Aalen, O.O.: Nonparametric inference in connection with multiple decrement models. Statistical Research Report no 6, Department of Mathematics, University of Oslo (1972)
2. Aalen, O.O.: Statistical inference for a family of counting processes. Ph.D. thesis, University of California, Berkeley (1975)
3. Aalen, O.O.: Nonparametric inference in connection with multiple decrement models. Scand. J. Stat. **3**, 15–27 (1976)
4. Aalen, O.O.: On nonparametric tests for comparison of two counting processes. Working paper no. 6, Laboratory of Actuarial Mathematics, University of Copenhagen (1976)
5. Aalen, O.O.: Weak convergence of stochastic integrals related to counting processes. Z. Wahrsch. verw. Geb. **48**, 261–277 (1977)
6. Aalen, O.O.: Nonparametric estimation of partial transition probabilities in multiple decrement models. Ann. Stat. **6**, 534–545 (1978)
7. Aalen, O.O.: Nonparametric inference for a family of counting processes. Ann. Stat. **6**, 701–726 (1978)
8. Aalen, O.O.: A model for non-parametric regression analysis of life times. In: W. Klonecki, A. Kozek, J. Rosinski (eds.) Mathematical Statistics and Probability Theory. Lecture Notes in Statistics, vol. 2, pp. 1–25. Springer-Verlag, New York (1980)
9. Aalen, O.O., Borgan, Ø., Gjessing, H.K.: Survival and Event History Analysis. Springer-Verlag, New York (2008)
10. Aalen, O.O., Borgan, Ø., Keiding, N., Thormann, J.: Interaction between life history events: nonparametric analysis of prospective and retrospective data in the presence of censoring. Scand. J. Stat. **7**, 161–171 (1980)
11. Aalen, O.O., Johansen, S.: An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. Preprint 6/1977, Institute of Mathematical Statistics, University of Copenhagen (1977)
12. Aalen, O.O., Johansen, S.: An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. Scand. J. Stat. **5**, 141–150 (1978)
13. Altshuler, B.: Theory for the measurement of competing risks in animal expreiments. Math. Biosci. **6**, 1–11 (1970)
14. Andersen, P.K., Borgan, Ø.: Counting process models for life history data (with discussion). Scand. J. Stat. **12**, 97–158 (1985)
15. Andersen, P.K., Borgan, Ø., Gill, R., Keiding, N.: Censoring, truncation and filtering in statistical models based on counting processes. Contemp. Math. **80**, 19–60 (1988)

16. Andersen, P.K., Borgan, Ø., Gill, R.D., Keiding, N.: Linear non-parametric tests for comparison of counting processes, with application to censored survival data (with discussion). Int. Stat. Rev. **50**, 219–258 (1982). Amendment 52:225
17. Andersen, P.K., Borgan, Ø., Gill, R.D., Keiding, N.: Statistical Models Based on Counting Processes. Springer-Verlag, New York (1993)
18. Andersen, P.K., Gill, R.D.: Cox's regression model for counting processes: A large sample study. Ann. Stat. **10**, 1100–1120 (1982)
19. Andersen, P.K., Keiding, N.: Survival analysis: Overview. In: P. Armitage, T. Colton (eds.) Encyclopedia of Biostatistics. Volume 6, pp. 4452–4461. Wiley, Chichester (1998)
20. Andersen, P.K., Rasmussen, N.K.: Psychiatric admission and choice of abortion. Stat. Med. **5**, 243–253 (1986)
21. Arjas, E.: Survival models and martingale dynamics (with discussion). Scand. J. Stat. **16**, 177–225 (1989)
22. Arjas, E., Haara, P.: A marked point process approach to censored failure data with complicated covariates. Scand. J. Stat. **11**, 193–209 (1984)
23. Arjas, E., Parner, J.: Causal reasoning from longitudinal data. Scand. J. Stat. **31**, 171–187 (2004)
24. Barlow, W.E., Prentice, R.L.: Residuals for relative risk regression. Biometrika **75**, 65–74 (1988)
25. Becker, N.G.: Martingale methods for the analysis of epidemic data. Stat. Methods Med. Res. **2**, 93–112 (1993)
26. Bernstein, S.: Sur l'extension du théorème limite du calcul des probabilités aux sommes des quantités dépendante. Math. Ann. **97**, 1–59 (1927)
27. Billingsley, P.: The Lindeberg-Lévy theorem for martingales. Proc. Am. Math. Soc. **12**, 788–792 (1961)
28. Boel, R., Varaiya, P., Wong, E.: Martingales on jump processes I: Representation results. SIAM J. Contr. **13**, 999–1021 (1975)
29. Boel, R., Varaiya, P., Wong, E.: Martingales on jump processes II: Applications. SIAM J. Contr. **13**, 1022–1061 (1975)
30. Bremaud, P.: A martingale approach to point processes. Memorandum ERL-M345, Electronics Research Laboratory, University of California, Berkeley (1973)
31. Breslow, N.E.: A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. Biometrika **57**, 579–594 (1970)
32. Breslow, N.E.: Contribution to the discussion of Cox (1972). J. R. Stat. Soc. Series B. Stat. Methodol. **34**, 216–217 (1972)
33. Breslow, N.E.: Covariance analysis of censored survival data. Biometrics **30**, 89–99 (1974)
34. Brown, B.M.: Martingale central limit theorems. Ann. Math. Stat. **42**, 59–66 (1971)
35. Courrège, P., Priouret, P.: Temps d'arrêt d'une fonction aléatoire: relations d'équivalence associées et propriétés de décomposition. Publications de l'Institut de Statistique de l'Université de Paris **14**, 245–274 (1965)
36. Cox, D.R.: Regression models and life-tables (with discussion). J. R. Stat. Soc. Series B. Stat. Methodol. **34**, 187–220 (1972)
37. Cox, D.R.: Partial likelihood. Biometrika **62**, 269–276 (1975)
38. Didelez, V.: Graphical models for composable finite Markov processes. Scand. J. Stat. **34**, 169–185 (2007)
39. Diggle, P., Farewell, D.M., Henderson, R.: Analysis of longitudinal data with drop-out: objectives, assumptions and a proposal. J. R. Stat. Soc. Series C. Appl. Stat. **56**, 499–550 (2007)
40. Dvoretzky, A.: Asymptotic normality for sums of dependent random variables. In: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2, pp. 513–535. University of California Press, Berkeley, California (1972)
41. Efron, B.: The two sample problem with censored data. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 4, pp. 831–853. University of California Press, Berkeley, California (1967)
42. Feller, W.: An Introduction to Probability Theory and Its Applications. Vol. II. Wiley, New York (1967)

43. Fleming, T.R.: Asymptotic distribution results in competing risks estimation. Ann. Stat. **6**, 1071–1079 (1978)
44. Fleming, T.R.: Nonparametric estimation for nonhomogeneous Markov processes in the problem of competing risks. Ann. Stat. **6**, 1057–1070 (1978)
45. Fleming, T.R., Harrington, D.P.: Counting Processes and Survival Analysis. Wiley, New York (1991)
46. Fosen, J., Ferkingstad, E., Borgan, Ø., Aalen, O.O.: Dynamic path analysis – a new approach to analyzing time-dependent covariates. Lifetime Data Anal. **12**, 143–167 (2006)
47. Gehan, E.A.: A generalized Wilcoxon test for comparing arbitrarily singly censored samples. Biometrika **52**, 203–223 (1965)
48. Gill, R.D.: Censoring and stochastic integrals. Mathematical Centre Tracts, vol. 124, Mathematisch Centrum, Amsterdam (1980)
49. Gill, R.D.: Product-integration. In: P. Armitage, T. Colton (eds.) Encyclopedia of Biostatistics. Volume 6, pp. 4246–4250. Wiley, Chichester (2005)
50. Gill, R.D., Johansen, S.: A survey of product-integration with a view towards application in survival analysis. Ann. Stat. **18**, 1501–1555 (1990)
51. Helland, I.S.: Central limit-theorems for martingales with discrete or continuous-time. Scand. J. Stat. **9**, 79–94 (1982)
52. Hogan, J., Roy, J., Korkontzelou, C.: Handling drop-out in longitudinal studies. Stat. Med. **23**, 1455–1497 (2004)
53. Hougaard, P.: Analysis of Multivariate Survival Data. Springer-Verlag, New York (2000)
54. Ibragimov, I.A.: A central limit theorem for a class of dependent variables. Theory Probab. Appl. **8**, 83–89 (1963)
55. Jacobsen, M.: Statistical Analysis of Counting Processes. Lecture Notes in Statistics, Vol. 12. Springer-Verlag, New York (1982)
56. Jacobsen, M.: Right censoring and martingale methods for failure time data. Ann. Stat. **17**, 1133–1156 (1989)
57. Jacod, J.: Multivariate point processes: Predictable projection, Radon-Nikodym derivatives, representation of martingales. Z. Wahrsch. verw. Geb. **31**, 235–253 (1975)
58. Johansen, S.: An extension of Cox's regression model. Int. Stat. Rev. **51**, 165–174 (1983)
59. Kaplan, E.L., Meier, P.: Non-parametric estimation from incomplete observations. J. Am. Stat. Assoc. **53**, 457–481, 562–563 (1958)
60. Kreager, P.: New light on Graunt. Population Studies **42**, 129–149 (1988)
61. Lai, T.L.: Martingales in sequential analysis and time series, 1945–1985. Electron. J. Hist. Probab. Stat. **5** (2009)
62. Lévy, P.: Propriéteés asymptotiques des sommes de variables aléatoires enchainées. Bull. Sci. Math. **59**, 84–96, 109–128 (1935)
63. Lin, D.Y., Wei, L.J., Ying, Z.: Checking the Cox model with cumulative sums of martingale-based residuals. Biometrika **80**, 557–572 (1993)
64. Lok, J.J.: Statistical modelling of causal effects in continuous time. Ann. Stat. **36**, 1464–1507 (2008)
65. Mantel, N.: Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother. Rep. **50**, 163–170 (1966)
66. Mantel, N., Haenszel, W.: Statistical aspects of the analysis of data from retrospective studies of disease. J. Natl. Cancer Inst. **22**, 719–748 (1959)
67. Martinussen, T., Scheike, T.H.: Dynamic Regression Models for Survival Data. Springer-Verlag, New York (2006)
68. McLeish, D.L.: Dependent central limit theorems and invariance principles. Ann. Probab. **2**, 620–628 (1974)
69. Meyer, P.A.: Probabilités et Potentiel. Hermann, Paris (1966)
70. Næs, T.: The asymptotic distribution of the estimator for the regression parameter in Cox's regression model. Scand. J. Stat. **9**, 107–115 (1982)
71. Nelson, W.: Hazard plotting for incomplete failure data. J. Qual. Technol. **1**, 27–52 (1969)

72. Nelson, W.: Theory and applications of hazard plotting for censored failure data. Technometrics **14**, 945–965 (1972)
73. Nelson, W.: Applied Life Data Analysis. Wiley, New York (1982)
74. Peto, R., Peto, J.: Asymptotically efficient rank invariant test procedures (with discussion). J. Roy. Stat. Soc. Series A. General **135**, 185–206 (1972)
75. Rebolledo, R.: Central limit theorems for local martingales. Z. Wahrsch. verw. Geb. **51**, 269–286 (1980)
76. Robins, J.M.: Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. Biometrika **79**, 321–324 (1992)
77. Røysland, K.: A martingale approach to continuous-time marginal structural models. Bernoulli **17**, 895–915 (2011)
78. Ryalen, P.C., Stensrud, M.J., Fosså, S., Røysland, K.: Causal inference in continuous time: an example on prostate cancer therapy. Biostatistics **21**, 172–185 (2020)
79. Tarone, R.E., Ware, J.: On distribution-free tests for equality of survival distributions. Biometrika **64**, 156–160 (1977)
80. Therneau, T.M., Grambsch, P.M., Fleming, T.R.: Martingale-based residuals for survival models. Biometrika **77**, 147–160 (1990)
81. Tsiatis, A.A.: A large sample study of Cox's regression model. Ann. Stat. **9**, 93–108 (1981)
82. Ville, J.A.: Sur la notion de collectif. Comptes rendus des Séances de l'Académie des Sciences **203**, 26–27 (1936)
83. Ville, J.A.: Étude Critique de la Notion de Collectif. Gauthier-Villars, Paris (1939)

Encounters with Martingales in Stochastic Control

Tyrone E. Duncan

Abstract

The martingale approach to stochastic control is very natural and avoids some major mathematical difficulties that arise when Hamilton-Jacobi-Bellman partial differential equations are used to solve optimization problems. The brief history of stochastic control given here commences with the war effort in the US in the 1940s, proceeds through the 1950s and 1960s, and touches on more recent aspects. This history shows that the solution of problems in potential theory related to stochastic control and other stochastic filtering problems is an important mathematical manifestation of martingales.

Keywords

Martingale · Stochastic control

1 Introduction

This paper recounts some of the history of the influence of martingales on stochastic control. Martingales provide a natural dynamic description for solutions of many stochastic control problems and for the closely related problems of estimating a system's stochastic state from noisy observations. This paper is not a complete historical account, but it discusses some major topics and contributors.

Beginning in the 1940s, the study of stochastic control was a response to pressing needs of the time and involved a natural cross fertilization of the related areas of control and optimization. It eventually became clear that martingale theory plays a

---

with an appendix by Laurent Mazliak.

T. E. Duncan (✉)
Department of Mathematics, University of Kansas, Lawrence, Kansas 66045, USA
e-mail: teduncan@ku.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*,
Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_14

fundamental role in problems of stochastic control and filtering. This paper's discussion focuses primarily on the work done in the United States, where some of the initial work on stochastic control was done, but there were important contributions from researchers in other parts of the world.

2 Frequency Domain Methods for Control and Estimation

In the 1940s, particularly with the scientific mobilization in the United States for the war effort, a number of important figures from engineering and mathematics were heavily involved with the study of mathematical models with noise to describe a variety of important engineering problems of control and estimation with military applications.

Three important figures in the United States during the 1940s to study systems with (random) noise were Stephen O. Rice, Claude Shannon, and Norbert Wiener. Steve Rice at Bell Laboratories was especially interested in the modelling of electronic device noise and atmospheric disturbances. He studied level crossings, waiting times and maxima of random curves, though not control or filtering. He typically used methods from frequency domain analysis of random signals [44,45]. Since Claude Shannon's major work was the creation and study of information theory, his contributions will not be described here. Norbert Wiener, who spent his academic career at MIT, had solved a number of important problems of (mathematical) analysis in the 1930s. The Wiener-Hopf equation [48], which he and E. Hopf developed in the 1930s, described some aspects of potential theory using primarily Fourier analysis as opposed to probability theory.

Wiener used this Wiener-Hopf equation during World War II for his solution of a filtering problem [49], often described as Wiener filtering, which was intended to improve the effectiveness of artillery fire both on land and at sea. This filtering problem was to estimate a stochastic signal with additive noise assuming that the infinite past of the observations is available for the estimation solution [49]. Independently, A. N. Kolmogorov in Moscow [36] obtained a solution to a corresponding discrete time linear filtering problem where the signal and noise are indexed by the positive integers. While the time sets are different, the problems are essentially mathematically equivalent. In both cases the noise was often described as "white", that is, the noise random variables at different times or different increments were uncorrelated. These families of noise random variables often formed martingales though they were not described as such. While martingales can be described for both discrete parameter and continuous parameter processes, the discussion here will be primarily focused on continuous time processes, because the modification to discrete time is fairly straightforward.

To fix notation, recall that a real-valued stochastic process $(X(t), t \geq 0)$ is said to be a (continuous parameter) martingale with respect to an increasing family of sub-σ-algebras $(\mathscr{F}(t), t \geq 0)$ on the (complete) probability space $(\Omega, \mathscr{F}, \mathbb{P})$ where Ω is the space, $\mathscr{F}$ is a σ-algebra of subsets of Ω and $\mathbb{P}$ is a probability measure on $\mathscr{F}$, if for all $t > s \geq 0$,

$$\mathbb{E}[X(t)|\mathscr{F}(s)] = X(s) \qquad a.s. \tag{1}$$

where $\mathbb{E}[X(t)|\mathscr{F}(s)]$ denotes conditional expectation of $X(t)$ given the sub-σ-algebra $\mathscr{F}(s)$ and $a.s.$ denotes almost surely, that is, with probability one. Replacing $=$ in (1) by $\leq$ defines a supermartingale and $\geq$ defines a submartingale so the negation of a supermartingale defines a submartingale. This terminology of “super” and “sub” alludes to the relations of these processes to superharmonic and subharmonic functions in potential theory. The notion of a martingale for control can be directly associated with the determination of optimality from the notion of balayage or “sweeping out” a family of measures.

In this paper the control results are primarily focused on the work in the United States though the mathematical foundations, especially in probability, are noted from France, other European countries and the former Soviet Union.

In the 1940s a linear system in engineering was typically described by its impulse response, that is, the response of the linear system to a delta function (Dirac distribution) input or the Fourier transform of its impulse response. This latter description allowed for frequency response interpretations of the systems, which were particularly appealing for applications in engineering. Wiener used his background in Fourier analysis to solve a (linear) filtering problem by transforming the problem to the so-called frequency domain and using some of his prior results from Fourier analysis and particularly potential theory. It was often argued in engineering that the frequency response of a physical system had a much narrower bandwidth than the observed noise frequency response, so the spectral density (the Fourier transform of the correlation function) of the noise was extended as a constant to the whole real line. This extension simplified the mathematical analysis. This noise is called “white” noise because all frequencies contribute the same “power” to the system. The (inverse) Fourier transform of this constant spectral density is a delta function (Dirac distribution) at the origin. While this description of a linear system was often done somewhat mathematically formally, many results that were obtained were mathematically correct. A useful method to make this Gaussian white noise process mathematically rigorous is to take the indefinite integral of it which defines a Brownian motion process, the most important continuous martingale and an even more important process for modeling noise for nonlinear systems and solving nonlinear control problems.

Wiener solved a filtering or estimation problem using the Wiener-Hopf equation from potential theory [48]. One form of a Wiener-Hopf equation for an estimation problem is the following integral equation.

$$u(x) - \int_0^\infty k(x-s)u(s)\mathrm{d}s = f(x) \tag{2}$$

where $x \in \mathbb{R}$. It is necessary to determine the function u given the functions k and f.

Wiener had solved such integral equations using his work on potential theory. Martingales arise directly from a method of solution in potential theory called balayage. It seems that Wiener was never aware of how his work on potential theory and his subsequent work on linear filtering could be naturally described by martingale theory. The word "balayage" means "sweeping out" in French. Balayage is in fact the construction of a (continuous parameter) martingale by sweeping out the support of the measure to the boundary. A fundamental problem for the application of Wiener filtering is the requirement of the infinite past history of the observation signal. For engineering applications, this infinite past history was merely truncated to a finite past history often depending on the physical equipment for filtering. A second important problem with the solution is that it is often difficult practically to implement the solution because typically it requires a spectral factorization and then the realization of the corresponding linear system from this factorization. These problems were subsequently addressed by requiring that the signal and noise model satisfy a linear stochastic equation with a white noise input. Two well known monographs in the 1950s describing the frequency domain approach to linear filtering are [6, 38].

3 Time Domain Methods for Control and Estimation

In 1961 Kalman and Bucy [20] provided a time domain approach to the (continuous time) linear filtering problem (and dually the solution of a linear-quadratic stochastic control problem) by describing the stochastic processes as solutions of linear differential equations driven by white Gaussian noise processes which provided Gaussian process solutions. This duality of problems arises from the natural duality in linear spaces given by vectors and the linear functionals in the space. The white noise processes in the model were the formal derivatives of Brownian motions and for this linear filtering problem it suffices to use these formal derivatives. A white noise process can be described by a delta function (Dirac distribution) covariance function so that the noise at different times is uncorrelated. The formal Fourier transform of this covariance function (or correlation function) is a constant. This approach of white Gaussian noise was primarily limited to linear equations and linear analysis. To extend the models and solutions to nonlinear stochastic equations it was necessary to describe the equations in differential form with the noise being the differential of a Brownian motion. Itô [29–32] had provided the necessary mathematical results, some of which had been anticipated by Wolfgang Doeblin.[1] A particular tool for the analysis of both linear and nonlinear stochastic systems is the so-called Itô formula for change of variables. This stochastic approach of Itô and Doeblin used the ideas

[1] The sealed envelope that Doeblin sent to the Academy of Sciences in Paris in February 1940, shortly before committing suicide as German troops surrounded his military unit, was opened in May 2000. Its contents were analyzed by Bernard Bru and Marc Yor in 2002 [4]. They were published in full along with Doeblin's other mathematical work and extensive commentary, including Bru and Yor's article, in [9].

of martingales in a basic way. Doob [10] had demonstrated in the 1940s the power of the method of martingale analysis for the solutions of some problems of probability and statistics. The development of martingales is described in other chapters in this book.

While the stochastic linear-quadratic control problem, that is, a linear stochastic differential equation and a quadratic cost functional, could be solved for an optimal control by using a white noise model as for the filtering problem or even the algebraic method of completion of squares [18], the analysis of nonlinear systems and in particular the use of stochastic differential equations and martingales is more powerful.

The principle of optimality for control that arose from dynamic programming, popularized by R. E. Bellman in his 1957 monograph [1], is often described by a so-called value function e.g. [24,50], that is, by the remaining cost for a control until the final time, which is a submartingale for an arbitrary control and a martingale for an optimal control. For some stochastic control problems the value function can be shown to satisfy a partial differential equation, called the Hamilton-Jacobi-Bellman equation e.g. [50]. This solution approach has been often used for stochastic control problems such as problems arising in mathematical finance e.g. [21]. The minimum principle for optimization is a local description for optimality and also has a submartingale interpretation from the differential description of a submartingale.

A differential description for a condition of optimality provides a stochastic equation for the optimal control solution. However the coefficients in such a stochastic differential equation typically do not satisfy the Lipschitz continuity conditions usually needed for existence and uniqueness of the (sample path) solutions. Thus other weaker notions of solution have been used, such as a transformation of Wiener measure by absolute continuity of measures, which was used initially by Benes [2] and Duncan and Varaiya [19].

A Radon-Nikodym derivative for Wiener measure as well as other measures can be naturally interpreted in martingale theory. The absolute continuity result for Wiener measure is often called the Girsanov theorem [25]. By conditional expectation it follows directly that this induced family of Radon-Nikodym derivatives indexed by time is a martingale. Since a Radon-Nikodym derivative is a positive function, the natural logarithm of it can be taken which is a concave function, thus providing a supermartingale by Jensen's inequality [40]. The supermartingale decomposition of Meyer [40] provides often an explicit form for the exponent in the exponential description of the Radon-Nikodym derivative (martingale). This Radon-Nikodym derivative can also be used to solve explicitly a variety of stochastic control problems, e.g. when the cost functional is the exponential of a quadratic functional [14]. Such control problems are often called risk sensitive because they can have an interpretation of an investor's propensity for risk by introducing a real-valued parameter in the exponential of the Radon-Nikodym derivative.

The effective use of the stochastic calculus and martingale theory for nonlinear control and filtering developed initially in the mid-1960s. This author (e.g. [11]) at Stanford was one of the initial researchers for nonlinear stochastic filtering and control. At the same time Mortensen [42] was also using stochastic calculus and martingale methods for a nonlinear filtering problem at UC Berkeley. Both of these works were significantly motivated by the martingale work of P. A. Meyer. In the early 1960s Paul-André Meyer [40] completed an extensive study of stochastic methods for potential theory that had its roots in the definition and use of martingales by J. L. Doob in the late 1940s. Meyer refined and extended Doob's results for martingales in the setting of potential theory and also developed a stochastic calculus for a general martingale. Many questions in potential theory can be viewed as optimization problems so it follows naturally that martingales play a fundamental role in optimization and dually optimization problems can be described by notions from potential theory. Thus, while stochastic models in control use martingales, some aspects of stochastic optimization would result in martingales even though they would not appear directly in the problem formulations.

As noted previously the most basic continuous time stochastic control problem is called the stochastic linear-quadratic control problem that is composed of a linear stochastic differential equation and a cost functional for minimization that is typically the integral of a quadratic function in the state and a quadratic function in the control. This problem was initially solved by Kalman [35]. This type of problem is also well known in a special case in physics [41] and it is well known in calculus that a quadratic function has a unique extremum that is determined by the zero of the derivative of the function. Furthermore there is a deterministic control problem that is described by a linear ordinary differential equation, an additive control term and a quadratic cost functional that is the sum of a quadratic functional in the state and another one in the control. Thus this deterministic control problem is very similar to the stochastic linear quadratic control problem. In fact both problems have basically the same solution. This fact results from the linear and quadratic structure of both problems.

For a partially observed stochastic linear system, where the system is observed only as a sum of the stochastic signal and an additional Brownian motion, it is required to perform a filtering operation to obtain an estimate of the state of the system. For the linear filtering problem it had been formally verified in the literature in the early 1960s that there is a "whiting" action on the observations. It was first shown by Duncan [12] in 1968 using measure transformation methods that there is a corresponding result for a nonlinear stochastic differential equation. These results can be viewed as the decomposition of a supermartingale [40]. Furthermore the (Shannon) mutual information between a stochastic signal and the stochastic signal plus white noise is the mean square error for the associated filtering problem (Duncan [13]) so that information theory is naturally related to filtering.

4 Nonlinear Stochastic Control

A so-called principle of optimality that was popularized by Richard Bellman in the 1950s [1] arises from a dynamic programming procedure, that is, an optimal trajectory has to be optimal starting anywhere along the optimal path. This optimality principle can be described as a submartingale inequality which provides a local description of optimality. For Markovian control systems with Markov-type controls another approach is the Hamilton-Jacobi-Bellman nonlinear partial differential equation e.g. [50]. However with this partial differential equation approach it is often difficult to establish existence and uniqueness of the solution of the partial differential equation. The martingale approach to the existence of explicit stochastic optimal controls was initially independently developed by Duncan-Varaiya [19] and Benes [2]. Subsequently martingales with dynamic programming equations were initially used by Davis-Varaiya [8].

A good introduction to these ideas is given by Davis [7]. To see how martingale theory arises naturally, consider the controlled stochastic system

$$\mathrm{d}X(t) = f(t, X(t), U(t)) + g(t, X(t))\mathrm{d}B(t) \tag{3}$$

$$X(0) = x_0 \tag{4}$$

where $X(t) \in \mathbb{R}^n$, $U(t) \in \mathbb{R}^m$, $f : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $g : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$, $(B(t), t \geq 0)$ is a standard $\mathbb{R}^n$-valued Brownian motion and f, g are smooth functions. The cost functional is

$$J(U) = \mathbb{E}[\int_0^T c(t, X(t), U(t))\mathrm{d}t + \Phi(X(T))] \tag{5}$$

The system state is X, the control is U, B is a standard Brownian motion, and $\mathbb{E}$ is expectation. The problem can be scalar or multidimensional. The so-called value function, $V(t, x)$, is the remaining cost for the control problem at time t, that is,

$$V(t, x) = \inf_U \mathbb{E}_{(t,x)}[\int_t^T c(t, X(t), U(t))\mathrm{d}t + \Phi(X(T))] \tag{6}$$

The subscript (t, x) means that $X(t) = x$ and the control U is restricted to the time interval $[t, T]$. The function $M_U(t)$ given by

$$M_U(t) = \int_0^T c(s, X(s), U(s))\mathrm{d}s + V(t, X(t)) \tag{7}$$

Thus for any admissible control $M_U(t)$ is a submartingale and it is a martingale for an optimal control.

Formally applying the principle of optimality to the control problem described above gives the following Hamilton-Jacobi-Bellman (HJB) equation

$$\frac{\partial V}{\partial t} + \Sigma_{i,j} \frac{\partial^2 V}{\partial x_i \partial x_j} + \min_u [\langle DV, f\rangle + c(t, x, u)] = 0 \tag{8}$$

$$V(T, x) = \Phi(x) \tag{9}$$

Hamilton [26] initially developed this equation for the optical path of light and then extended it to mechanics and Jacobi [33] improved on Hamilton's work and made significant applications to mechanics.

For control the equation arises naturally from the change of variables formula and a minimization. This HJB equation is typically a nonlinear partial differential equation. In general it is difficult to establish existence and uniqueness for such partial differential equations. However a few cases can be explicitly solved. The most tractable is the problem of a linear stochastic differential equation with a cost functional that is the sum of a quadratic function of X and quadratic function of U. The stochastic equation is the sum of a linear term in X, a linear term in U and a Brownian motion. This problem is often described as a stochastic linear-quadratic control problem. There is naturally associated a deterministic control problem where the noise term for the linear stochastic differential equation is deleted to obtain an ordinary differential equation. An interesting feature of these two problems is that the optimal control has the same form. The only difference is that for the stochastic problem there is an additional term in the optimal cost from the noise term in the equation. Unfortunately this relation between a deterministic and a stochastic control problem does not carry over to most other control problems.

5 Some Other Related Stochastic Optimization Problems

Some other topics from stochastic systems that are naturally related to stochastic control and martingales are stochastic differential games, mean field games and stochastic adaptive control. These problems are natural generalizations of optimal control problems because the games have two or more players (controllers) and for adaptive control solutions one is required to estimate the system along with controlling it. In stochastic differential games there are two or more players. For two player games, one player desires to minimize the cost functional and the other player desires to maximize the cost functional. For such problems it is necessary to obtain a saddle point, that is, a point that is at least locally optimal for both players. The linear-quadratic control problem can be extended to these two player games by the method described above [15]. For mean field games there are a major player and some minor players e.g. [28]. For adaptive control the system is only partially known and it is desired to control it, for example, for a linear-quadratic control problem e.g. [16]. Furthermore there are control problems that are described by stochastic partial differential equations with the requirement of control e.g. [17]. Each of these types of problems would require significant descriptions to formulate and describe mathematically.

6 Appendix (*by Laurent Mazliak*): Martingale Problems and Stochastic Control of General Processes

Professor Duncan's text is concerned with the control of diffusion processes whose coefficients are sufficiently regular to be treated, with the help of the stochastic calculus, by means of a partial differential equation. This model is in a sense approached in the same way as the deterministic control problem. Its various applications grew in the 1950s and 1960s, not without connection with issues related to aeronautical guidance, especially for rockets, in both civilian and military settings. Its centrality in applied mathematics was established by the 1970s, as illustrated by the now classic volume [24], which appeared in 1975, inaugurating the new collection *Applications of Mathematics* at Springer. On the other hand, the development of the general theory of random processes in the 1960's allowed control problems to be extended, at least theoretically, far beyond the control of diffusions with regular coefficients. The source of this extension is a basic observation: since the criterion to be optimized is in the form of an expectation, it is only the probability law of the controlled process that intervenes in the problem and it must therefore be possible to formulate the problem as an optimization of a set of controlled probability laws.

The first general formulation seems to be due to Charlotte Striebel[2] in 1974 but published ten years later in [46]. A few years later, the most complete presentation of a formulation of stochastic control problems that does not involve the solution of partial differential equations was given by Nicole El Karoui in her course [21] at the summer school of Saint-Flour in France. El Karoui exploited the power of the general theory approach proposed for example in [5], as well as the definition of process laws through martingale problems as exposed in [34] or in [47].

In order to give a more precise idea of the framework I have just mentioned, I will present a situation which does not require the introduction of too many technical elements in addition to those already present in Tyrone Duncan's text, namely the control of diffusions under very general conditions. Still for simplicity, we will limit ourselves to dimension 1 (real-valued diffusions). Moreover, as the aim here is to give general ideas, I will take the liberty of skimming over certain technical details which do not provide decisive elements for the understanding of the method.

6.1 Strong and Weak Solutions of Stochastic Differential Equations. Martingale Problems

It is first necessary to introduce the notion of weak solution. Let us consider a stochastic differential equation

$$\mathrm{d}X_t = b(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t, \tag{10}$$

[2] The reader may have noted the quite sad absence of women mathematicians in the present volume. I (Laurent Mazliak) am therefore very glad to have the opportunity to introduce in this appendix two women who were important actors in stochastic control theory in the 1970s, Charlotte Striebel (1929–2014) and Nicole El Karoui (born in 1944). Striebel was an associate professor at the University of Minnesota for 30 years. Apart from her mathematical duties, she was a strong advocate for equal rights for women, especially in athletics. If El Karoui's achievements are more strictly academic, it must be noted that in the 1990s, she became a world leader in mathematical finance. About the presence of martingales in mathematical finance, the reader may find a lot of information in [3].

where W_t is a Wiener process on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, P)$. A *strong solution* is a process $(X_t)_{t\geq 0}$ on $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, P)$ such that for all $t \geq 0$,

$$X_t = X_0 + \int_0^t b(s, X_s)\mathrm{d}s + \int_0^t \sigma(s, X_s)\mathrm{d}W_s \tag{11}$$

Existence and uniqueness of a strong solution of equation (10) is guaranteed under classical Cauchy-Lipschitz type conditions.

In contrast, there is a *weak solution* of equation (10) when there exists a filtered probability space $(\tilde{\Omega}, \tilde{\mathscr{F}}, (\tilde{\mathscr{F}}_t)_{t\geq 0}, \tilde{P})$, a Wiener process $\tilde{W}$ on this space and a process $\tilde{X}$ such that

$$\tilde{X}_t = X_0 + \int_0^t b(s, \tilde{X}_s)\mathrm{d}s + \int_0^t \sigma(s, \tilde{X}_s)\mathrm{d}\tilde{W}_s.$$

A strong solution is a weak solution, of course, but the existence of weak solutions can be obtained under much more general assumptions, such as the continuity of the coefficients b and σ in equation (10).

A major result in the Stroock and Varadhan's classic 1979 book [47] is that the existence of a weak solution to Eq. (10) can be formulated through a *martingale problem*. To understand how this works, consider the infinitesimal generator of the diffusion process (11), which is a differential operation on the set of $C^{1,2}$-real functions (C^1 in time variable t, C^2 in space variable x) given by

$$L = \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial x^2} + b\frac{\partial}{\partial x}. \tag{12}$$

By Itô's calculus, if X_t satisfies (11) and $f \in C^{1,2}$, then the process

$$f(t, X_t) - \int_0^t \left(\frac{\partial}{\partial t} + L\right) f(s, X_s)\mathrm{d}s$$

is an $(\mathscr{F}_t)$-martingale. In such a case, it is clear that the distribution of the process $(X_t)_{t\geq 0}$ can be seen as a probability distribution over the space of continuous real functions on $\mathbb{R}_+$.

Let us call $\hat{\Omega}$ the (canonical) space of continuous real functions on $\mathbb{R}_+$, $(\hat{\mathscr{F}}_t)_{t\geq 0}$ its natural filtration and $\hat{X}$ the canonical process on $\hat{\Omega}$, meaning that $\hat{X}_t(\varphi) = \varphi(t)$ for $\varphi \in \hat{\Omega}$. A *solution to the martingale problem* associated with generator (12) is a probability measure on $\hat{\Omega}$ such that for all $f \in C^{1,2}$,

$$f(t, \hat{X}_t) - \int_0^t \left(\frac{\partial}{\partial t} + L\right) f(s, \hat{X}_s)\mathrm{d}s$$

is a martingale on $(\hat{\Omega}, (\hat{\mathscr{F}}_t)_{t\geq 0}, \hat{P})$.

Constructing a process by means of a martingale problem is not limited to Itô's diffusions. In 1979 [34] for example, Jacod extended the method to all kinds of semi-martingales for which the stochastic part in (11) may have all kinds of irregularities. For instance, instead of a stochastic integral with respect to a Wiener process, one may consider pure jump processes or processes discontinuous in other ways.

6.2 General Formulation of a Control Problem

Let us now return to Eq. (10) and consider the case where a control parameter enters the equation:

$$\mathrm{d}X_t^{\mathbf{u}} = b(t, X_t^{\mathbf{u}}, u_t)\mathrm{d}t + \sigma(t, X_t^{\mathbf{u}}, u_t)\mathrm{d}W_t, \tag{13}$$

where $\mathbf{u} = (u_t)_{t\geq 0}$ is a control—i.e., a process adapted to the filtration $(\mathscr{F}_t)_{t\geq 0}$, with values in a metric space U with 'reasonable' topological properties (for instance, U can be a compact set, or a convex set). Let us introduce a cost function of the form

$$J(\mathbf{u}) = E\left(\int_0^T h(s, X_s^{\mathbf{u}}, u_s)\mathrm{d}s + g(T, X_T^{\mathbf{u}})\right)$$

The control problem associated with Eq. (13) and the cost function J is to find a control $\mathbf{u}$ that minimizes the cost J.

Itô's calculus allows to formulate a verification theorem as a sufficient condition for optimality. For any $t \geq 0$, consider the set $\mathscr{U}_t$ of controls starting at time t. For $x \in \mathbb{R}$, the *value function* of the problem is given by

$$V(s, x) = \inf_{\mathbf{u}\in\mathscr{U}_\mathbf{s}} E\left(\int_s^T h(t, X_t^{\mathbf{u}}, u_t)\mathrm{d}t + g(T, X_T^{\mathbf{u}})\right),$$

with $X_t^{\mathbf{u}}$ a solution of the equation

$$\mathrm{d}X_t^{\mathbf{u}} = b(t, X_t^{\mathbf{u}}, u_t)\mathrm{d}t + \sigma(t, X_t^{\mathbf{u}}, u_t)\mathrm{d}W_t \tag{14}$$

for all $t \geq s$ and initial condition $X_s^{\mathbf{u}} = x$. Thus $V(s, x)$ is interpreted as the minimal cost that can be obtained starting at time s in position x.

The *verification theorem* stipulates that if a function $F(t, x)$ is a solution to the Hamilton-Jacobi-Bellman equation

$$\frac{\partial}{\partial t}F(t, x) + \min_{\mathbf{u}\in U}(L^{\mathbf{u}}F)(t, x) = 0,$$

where

$$L^{\mathbf{u}}f(t, x) = \frac{1}{2}\sigma^2(t, x, u_t)\frac{\partial^2}{\partial x^2}f(t, x) + b(t, x, u_t)\frac{\partial}{\partial x}f(t, x),$$

then F is the value function of the problem.

There is a probabilistic formulation of the verification theorem through a martingale formulation: a control $\mathbf{u}$ is optimal if and only if the process

$$V_t = V(t, X_t^{\mathbf{u}}) + \int_0^t h(s, X_s^{\mathbf{u}}, u_s)\mathrm{d}s$$

is a $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, P)$-martingale.

Exploiting the notion of weak solutions, a control problem can be formulated in a weak form. Using straightforward notations, a weak control process is given by $(\tilde{\Omega}, \tilde{\mathscr{F}}, (\tilde{\mathscr{F}}_t), \tilde{P}, \tilde{X}_t, \tilde{u}_t)$ such that for all $f \in C^{1,2}$,

$$f(t, X_t^{\mathbf{u}}) - \int_0^t \left(\frac{\partial}{\partial t} + L^{\mathbf{u}}\right) f(s, X_s^{\mathbf{u}})ds$$

is a $(\tilde{\Omega}, \tilde{\mathscr{F}}, (\tilde{\mathscr{F}}_t), \tilde{P})$-martingale.

This general model of control was used in the 1980s to obtain existence results. Compactification methods with measure-valued controls were introduced in 1987 by El Karoui et al. [22]. This technique is also used in [39] in the case of jump processes, while [27] considers convexity hypotheses. In another direction, [37] uses selection theorems to obtain approximate optimal controls satisfying a Markovian property.

Let us finally observe that the martingale property plays also a central role in another important tool for stochastic control introduced in the 1990s by Shige Peng and Etienne Pardoux in [43], the *backward stochastic differential equations*. The interested reader can consult [23], the first textbook proposed on this topic.

Acknowledgements Tyrone Duncan's research was supported by AFOSR grant FA9550-12-1-0384.

References

1. Bellman, R.E.: Dynamic Programming. Princeton University Press, Princeton (1957)
2. Benes, V.E.: Existence of optimal stochastic control laws. SIAM J. Control **9**, 446–472 (1971)
3. Brian, É.: Comment tremble la main invisible. Springer (2009)
4. Bru, B., Yor, M.: Comments on the life and mathematical legacy of Wolfgang Doeblin. Finance and Stochastics **6**(1), 3–47 (2002)
5. C.Dellacherie, Meyer, P.A.: Probabilities and potential. Transl. from the French, vol. 29. Elsevier, Amsterdam (1978)
6. Davenport, W., Root, W.L.: An Introduction to the Theory of Random Signals and Noise. McGraw-Hill (1958)
7. Davis, M.H.A.: Martingale methods in stochastic control. In: Stochastic control theory and stochastic differential systems, pp. 85–117. Springer (1979)
8. Davis, M.H.A., Varaiya, P.P.: Dynamic programming conditions for partially observable stochastic systems. SIAM J. Control **11**, 226–261 (1973)
9. Doeblin, W.: Œuvres complètes – Collected works. Springer (2020). Edited by Marc Yor and Bernard Bru, with forewords by Jean-Michel Bismut and Hans Föllmer and commentaries by Marius Iosifescu, Torgny Lindvall, David Mason, Esa Nummelin and Eugene Seneta

10. Doob, J.L.: Stochastic Processes. Wiley (1953)
11. Duncan, T.E.: Probability densities for diffusion processes with applications to nonlinear filtering and detection theory. Ph.D. thesis, Stanford University (1967). SEL 67-035
12. Duncan, T.E.: Evaluation of likelihood functions. Info. Control **13**, 62–74 (1968)
13. Duncan, T.E.: On the calculation of mutual information. SIAM J. Appl. Math. pp. 215–220 (1970)
14. Duncan, T.E.: Linear-exponential-quadratic Gaussian control. IEEE Trans. Autom. Control **58**, 2910–2911 (2013)
15. Duncan, T.E.: Some partially observed multi-agent linear exponential quadratic stochastic differential games. Evol. Eq. and Control Theory **7**(6), 587–597 (2018)
16. Duncan, T.E., Guo, L., Pasik-Duncan, B.: Adaptive continuous time linear quadratic Gaussian control. IEEE Trans. Autom. Control **44**, 1653–1662 (1999)
17. Duncan, T.E., Maslowski, B., Pasik-Duncan, B.: Adaptive boundary and point control of linear stochastic distributed parameter systems. SIAM J. Control Optim. **32**, 648–672 (1994)
18. Duncan, T.E., Pasik-Duncan, B.: A direct approach to linear-quadratic stochastic control. Opuscula Math. **37**(6) (2017)
19. Duncan, T.E., Varaiya, P.: On the solutions of a stochastic control system. SIAM J. Control **9**, 354–371 (1971)
20. E., K.R., Bucy, R.S.: New results in linear filtering theory. J. Basic Engrg. ASME **83**(D), 95–108 (1961)
21. El Karoui, N.: Les aspects probabilistes du contrôle stochastique. In: École d'été de Probabilités de Saint-Flour IX-1979, pp. 73–238. Springer (1981)
22. El Karoui, N., Du Huu , N., Jeanblanc-Picqué, M.: Compactification methods in the control of degenerate diffusions: Existence of an optimal control. Stochastics **20**, 169–219 (1987). https://doi.org/10.1080/17442508708833443
23. El Karoui, N., Mazliak, L.: Backward stochastic differential equations. Pitman Research Notes in Mathematics Series. 364. Harlow: Longman. 221 p. (1997). (1997)
24. Fleming, W.H., Rishel, R.W.: Deterministic and stochastic optimal control, vol. 1. Springer, New York (1975)
25. Girsanov, I.V.: On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. Theor. Probab. Appl. **5**(3), 285–301 (1960)
26. Hamilton, W.R.: The Mathematical Papers of Sir William Rowan Hamilton. Cambridge (1931)
27. Haussmann, U.G., Suo, W.: Singular optimal stochastic controls I: Existence. SIAM J. Control Optim. **33**(3), 916–936 (1995). https://doi.org/10.1137/S0363012993250256
28. Huang, M.Y., Malhamé, R.P., Caines, P.E.: Large population stochastic dynamic games: Closed loop Mckean-Vlasov systems and the Nash certainty equivalence principle. Comm. Info. Systems **6**(3), 221–252 (2006)
29. Itô, K.: Stochastic integral. Proceedings of the Imperial Academy **20**(8), 519–524 (1944)
30. Itô, K.: On a stochastic integral equation. Proceedings of the Japan Academy **22**(1-4), 32–35 (1946)
31. Itô, K.: Multiple Wiener integral. Journal of the Mathematical Society of Japan **3**(1), 157–169 (1951)
32. Itô, K.: On a formula concerning stochastic differentials. Nagoya Mathematical Journal **3**, 55–65 (1951)
33. Jacobi, C.G.J.: Jacobi's Lectures on Dynamics. Delivered at the University of Königsberg in the winter semester 1842–1843 and according to notes prepared by C. W. Brockart. Hindustan Book Agency, New Dehli (2009)
34. Jacod, J.: Calcul stochastique et problèmes de martingales, vol. 714. Springer, Cham (1979). https://doi.org/10.1007/BFb0064907
35. Kalman, R.E.: Contributions to the theory of optimal control. Bol. Soc. Mat. Mex. **5**, 102–119 (1960)
36. Kolmogorov, A.N.: Interpolation and extrapolation of stationary random sequences (in Russian). Bull. Aca. Sci USSR, Ser. Math. **5** (1941)

37. Krylov, N.V.: Controlled diffusion processes. Translated by A. B. Aries, vol. 14. Berlin: Springer (2009)
38. Laning, J.H., Battin, R.H.: Random Processes in Automatic Control. McGraw-Hill (1956)
39. Mazliak, L.: Mixed control problem under partial observation. Appl. Math. Optim. **27**(1), 57–84 (1993). https://doi.org/10.1007/BF01182598
40. Meyer, P.A.: Probability and Potentials. Blaisdell, Waltham, MA (1966)
41. Mitter, S.K.: On the analogy between mathematical problems of non-linear filtering and quantum physics. Ricerche di Automatica **10**(2), 163–216 (1979)
42. Mortensen, R.E.: Optimal control of continuous-time stochastic systems. Ph.D. thesis, University of California, Berkeley (1966)
43. Pardoux, E., Peng, S.G.: Adapted solution of a backward stochastic differential equation. Syst. Control Lett. **14**(1), 55–61 (1990). https://doi.org/10.1016/0167-6911(90)90082-6
44. Rice, S.O.: Mathematical analysis of random noise. Bell System Technical J. **23**(3), 282–332 (1944)
45. Rice, S.O.: Mathematical analysis of random noise. Bell System Technical J. **24**(1), 46–156 (1945)
46. Striebel, C.: Martingale conditions for the optimal control of continuous time stochastic systems. Stochastic Processes Appl. **18**, 329–347 (1984). https://doi.org/10.1016/0304-4149(84)90304-1
47. Stroock, D.W., Varadhan, S.R.S.: Multidimensional diffusion processes. Berlin: Springer (2006)
48. Wiener, N.: Generalized harmonic analysis. Acta Math. **55**, 117–254 (1930)
49. Wiener, N.: Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications. Wiley (1949). Initially issued as a classified report
50. Yong, J., Zhou, X.Y.: Stochastic Control: Hamiltonian Systems and HJB Equations. Springer (1999)

Part V Documents

$$Z_t = \exp\left\{ \int_0^t \theta(s,\omega)dB(s) - \frac{1}{2}\int_0^t \theta^2(s,\omega)ds \right\}; \qquad 0 \le t \le T$$

Analysis or Probability? Eight Letters Between Børge Jessen and Paul Lévy

Bernard Bru and Salah Eid

Abstract

From September 1934 to August 1935, Børge Jessen and Paul Lévy exchanged eight letters, three from Jessen and five from Lévy, concerning the relation between what we now call Jessen's theorem and Lévy's lemma. Each contended, with some justice, that the other's results could be derived from their own, and we now think of both as versions of the martingale convergence theorem. Jessen's and Lévy's contrasting viewpoints are analyzed in detail in this volume's chapter entitled "The Dawn of Martingale Convergence: Jessen's Theorem and Lévy's Lemma', by Salah Eid. Here we present the letters and drafts that have survived in the Jessen archives at the University of Copenhagen. Jessen's first letter is missing, but we have Lévy's five letters and drafts of two of Jessen's letters. We also include a 1947 letter from Harald Bohr and Jessen to Lévy, which does not to the mathematical discussion but shows the continued relationship between the parties.

Keywords

Børge Jessen · History of probability · Jessen's theorem · Lévy's lemma · Martingale convergence theorem · Lévy · Transfer principle

Introduction and footnotes translated from the French by John Aldrich, University of Southampton.

B. Bru (✉)
Université Paris Descartes, MAP5, Paris, France
e-mail: john.aldrich@soton.ac.uk

S. Eid
Université Paris Diderot, Paris, France
e-mail: salaheid.h@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*,
Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_15

1 Introduction

In the spring of 1934 the brand new Institute of Mathematics was inaugurated in Copenhagen under the direction of Harald Bohr. It was built with a donation from the Carlsberg Foundation on the model of the Institute of Theoretical Physics, which the Rockefeller Foundation had financed for Niels Bohr, the Institute Henri Poincaré in Paris, and the Institute of Mathematics at Göttingen [65,70]. There were funds for visitors and Harald Bohr could invite foreign lecturers of reputation to speak in Copenhagen. Among them was Paul Lévy, who seems to have had good relations with Harald Bohr. Lévy was always more esteemed abroad than in Paris, and there was nothing surprising about this invitation, which came at the beginning of April 1934.

Lévy tells us that he presented his brand new theory of integrals whose elements are independent random variables before the Mathematical Society of Denmark on 9 April 1934. He had presented his first note on the theory, [51], to the Paris Academy of Sciences only a few months before, on 26 February, and he had given the same lecture in Hadamard's seminar on March 16 [53, p. 337, note 1].

One can imagine that the audience was rather surprised by Lévy's performance, but Harald Bohr, a man with a cool head and perfect manners, could at least grasp that it was a matter of integrating functions with an infinity of variables and indicate to his guest that his student Børge Jessen, then in Princeton, had done much work on the question. In any case, we can assume that Bohr gave Lévy Jessen's articles in German and Danish on the question, because §1 of Lévy's 1935 [55] quotes Jessen's [33] and seems to be a kind of review or commentary on Jessen's theory. When Jessen returned to Copenhagen in September 1934, he wrote to Lévy, undoubtedly on Bohr's advice. We have not found this letter but it must have been accompanied by a reprint of his great article in *Acta Mathematica* [37], which he had just received. This letter is the starting point of the correspondence that we now present.

It is well known that Lévy, who worked alone, liked to correspond. There is hardly any difference in style between his letters and his publications, long monologues delivered in a single breath as though he were reporting a film that was rolling in front of him without ever stopping. This is why one can learn much about the mathematical universe he inhabited from reading his correspondence. To be convinced, consider his letters to Fréchet, preserved in the Archives of the Académie des Sciences and published with remarkable commentary in [4].

Lévy's files were destroyed during the war but Jessen kept all his correspondence, and it was deposited after his death in the Archives of the Institute for Mathematical Science at the University of Copenhagen. This correspondence is published here with commentary, after an introduction to the two protagonists and their relevant work. Børge Jessen, born in 1907 and a student of Harald Bohr, was professor of geometry at the Polytechnic School of Copenhagen in 1934 and was already making a name for himself among analysts. Paul Lévy, born in 1886, a student of Hadamard and Borel and professor of analysis at the Paris École Polytechnique, had been developing the modern theory of probability after his own fashion for the previous fifteen years.

The letters presented here, in chronological order, come from the Jessen Archive at the Institute of Mathematics at Copenhagen. Lévy's series appears to be complete. Nothing of Jessen's first letter survived but some copies, or drafts, of other letters did.

Lévy responded promptly to Jessen's letter of September 1934, but the core of the correspondence is Lévy's second letter, dated 4 April 1935, and the response Jessen drafted a few days later. These are followed by two shorter letters from Lévy, on 24 April and 3 May, and then by an exchange in August. The final letter in the collection, from Bohr and Jessen to Lévy, concerns only the lack of funding for a proposed invitation for another visit by Lévy to Copenhagen.

As the letters reveal, Jessen and Lévy brought very different viewpoints to their conversation, and they were often talking past each other. Jessen saw his work as pure analysis, whereas Lévy found his motivation in the intuition and language of probability. Their different viewpoints and the ultimate influence of their interaction are discussed in detail in the present volume's chapter "The Dawn of Martingale Convergence: Jessen's Theorem and Lévy's Lemma', by Salah Eid.

2 Lévy to Jessen. Paris, 27 September 1934

Cher Monsieur, Je vous remercie de votre aimable communication ; dès que j'aurai terminé de rédiger quelques résultats que j'ai obtenus depuis déjà plusieurs mois, je ne manquerai pas de rechercher vos travaux que vous m'indiquez. Malheureusement je suis un peu découragé par la difficulté des recherches sur $\zeta(s)$; je n'ai obtenu personnellement aucun résultat important sur cette question.[1] J'ai été plus heureux pour les fonctions d'une infinité de variables.

M. le professeur Bohr m'a parlé de vos travaux, et ce qu'il m'en a dit m'a beaucoup intéressé. A vrai dire votre théorie de la mesure et celle de H. Steinhaus m'étaient familières depuis longtemps, peut-être depuis 1920.[2] Ce sont pour moi des notions simples que je précise quand j'en ai besoin. Mais par les applications que vous en avez faites, vous avez beaucoup dépassé ce que je savais. J'ai signalé votre communication au Congrès d'Oslo[3] à M. Denjoy, qui vient de nouveau de découvrir la théorie de la mesure (Note du 6 juin 1933 à l'Académie des Sciences),[4] et je la cite dans un Mémoire dont je viens de terminer la rédaction et qui paraîtra dans le Bulletin des

[1] Lévy had discussed the Riemann zeta function at the Zurich Congress [50], and Jessen may have mentioned it in his lost first letter. Jessen had also discussed the topic in Zurich [34] and worked on it with Bohr from 1928–1929 [9]. See also [60, p. 41] and [47, p. 143].

[2] An allusion to Lévy's lectures at the Collège de France in 1919, taken up in [43–46].

[3] Jessen [33].

[4] Denjoy's note, [18], followed Cantelli's lectures at the Institut Henri Poincaré in 1933 [15]. Denjoy used the principle of correspondence to construct infinite sequences of independent random variables, which he called "variables pondérées multipliables" and the Polish School called "fonctions indépendantes", e. g. [39,62].

Sciences Mathématiques.[5] C'est dommage que je ne sache pas le danois et ne puisse lire le Mémoire plus développé que M. Bohr m'a donné.[6]

Je vais présenter à l'Académie une Note qui résume mon nouveau Mémoire.[7] Elle complète, et rectifie le résumé que j'ai exposé à la Société Mathématique de France le 23 mai 1934, et que j'ai communiqué à M. Bohr et à M. Lublin.[8] Je vous signale aussi que, dans mon mémoire de Studia Mathematica, les th. XI et XII sont vrais, non pour $\sum x_n$, mais pour $\sum(x_n - a_n)$, a_n étant, pour chaque n, une constante convenablement déterminée qui peut être le terme d'une série semi-convergente.[9] Je ne me suis aperçu de cette erreur qu'en rentrant du Danemark, de sorte qu'elle n'est pas corrigée sur les exemplaires que j'ai laissés à l'Institut de Copenhague.

[5] Lévy [55]. From this letter, it is safe to conclude that Lévy's article on dependent variables was written after his visit to Copenhagen in April 1934 and before the end of the 1934 summer holidays. As noted above, Lévy added to the proofs a note, on p. 89, where he indicated that he had since been informed of Jessen's article in *Acta Math*. In his *Notice* [54, p. 84, note 1], Lévy also states that he had submitted [55] in September 1934. See the next letter and [4, p. 156, note 111].

[6] Jessen [32]. It is therefore clear that §1 of [55] (or §39 of [58]), that is to say, the "foundations" of Lévy's theory of denumerable probabilities, came from a close reading of [33] and memories, real or imagined, of his first works on integration in infinitely many dimensions.

[7] Lévy [52], presented on October 1, five days after this letter. This note corrects the statement of the central limit theorem for martingales in his communication to the Société Mathématique de France, which we will not examine here. We may recall, however, that the first general central theorem limit for dependent variables is in Bernstein's [5], from which Lévy began. However, Lévy's theorem exploited a novel idea, a random change of time; Lévy imagines that the nth game, whose result is X_n, assumed to have zero mean conditional on the past, has a random duration equal to the conditional variance $\sigma_n^2 = E_{n-1}(X_n)$. If one cumulates the profits of n games by relating them not to the square root of time n or to $\sqrt{\sum_1^n E(X_k^2)}$, but to the new time equal to the sum of the n random durations thus defined, one obtains asymptotic normality [58, p. 242].

This idea, remarkable in every way but consistent with the transformation of prior probabilities into posterior probabilities, was taken up by W. Doeblin in the framework of continuous time martingales with continuous trajectories. Locally such martingales behave like Lévy's martingales at infinity, so that by changing their time in the natural way (by replacing the sum by an integral), one obtains a Brownian motion. This theorem of Doeblin [19, p. 1068, lemma IX], the first known statement of the Dubins-Schwartz theorem [22], enables him to solve in masterly fashion the Bernstein-Kolmogrov problem of finding probabilistic solutions of Kolmogorov's equation. On this topic, see Marc Yor's introduction to [19, pp. 1033–1035]. But Lévy did not have the idea of a theory of continuous martingales, which was treated by Ville [76], Doob [20], and Doeblin [19], without Lévy's ever knowing.

[8] Mogens Lublin was a young Danish mathematician and a contemporary of Jessen's. He submitted his magister thesis, written under the direction of N. Nielsen and N. E. Nørlund, in 1930. In the years 1930–1940 Lublin published actuarial works, and he had a brilliant career as an actuary in Copenhagen; he died in 1972.

[9] Lévy [48, §15, pp. 148–149].

Je voudrais vous charger de mes souvenirs pour tous vos collègues que j'ai vus à Copenhague, M.M. Norlund, Bohr, Steffensen, Petersen, Bonnessen, Mollerup, …[10] mais ils sont trop nombreux et je ne puis vous le demander. J'ai conservé un excellent souvenir des journées que j'ai passé à Copenhague, et regretté de ne pas y faire votre connaissance. Mais j'espère vous voir un jour à Paris.

En attendant croyez à mes sentiments dévoués.

P. Lévy

3 Lévy to Jessen. Paris, 4 April 1935

Cher Monsieur, Je pense vous intéresser en vous envoyant les épreuves d'un mémoire qui va bientôt paraître, et qui est en relation avec vos travaux. Il a été rédigé l'été dernier, et remis à la rédaction du Bulletin des Sciences Mathématiques avant que

---

[10] Niels Erik Nørlund (1885–1981) was a very well known mathematician. He was a professor at the University of Lund and then at the University of Copenhagen. For 55 years he was editor of *Acta Mathematica*. His sister Margrethe Nørlund married Niels Bohr in 1912. See [26].

Johan Frederik Steffensen (1873–1961) was an important Danish statistician, a professor of actuarial science at the University of Copenhagen and correspondent of Fréchet.

Richard Petersen (1894–1968), a pupil of Bohr, was assistant in mathematics at the University of Copenhagen, then professor at the city's Polytechnic School. This institution, founded in 1829 on the model of the Paris École Polytechnique, is now called the Technical University of Denmark; thanks to Christian Berg for this information. Richard Petersen was a pioneer of automatic calculation by computer in Denmark.

Tommy Bonnesen (1873–1935) was professor of descriptive geometry at the Polytechnic School of Copenhagen, which followed Monge's curriculum like its Parisian model. Bonnesen worked in particular on the isoperimetric inequalities, convex bodies, etc. When Bonnesen died in 1935, Jessen succeeded him—he had to learn the geometry of engineers and stone masons.

Johannes Mollerup (1872–1937) was professor of analysis at the Polytechnic School of Copenhagen. Christian Berg provides these details:

> Concerning Mollerup, I mention that Bohr and Mollerup initiated the writing (around 1915) of a 4 volume treatise of mathematical analysis for the Polytechnical School (I suppose inspired by Jordan's Cours d'analyse which Bohr had studied himself). It became a legend for Danish engineers, used in many new editions up to around 1970 and just called Bohr-Mollerup. There is a famous theorem called the Bohr-Mollerup theorem, namely the characterization of the Gamma function as the only log-convex function satisfying the functional equation and being normalized to 1 at 1. It appeared in a slightly disguised version in the 1922 edition of Bohr-Mollerup and was later made known by Artin in his small book about Gamma (with due credit to Bohr and Mollerup). Bohr and Mollerup never published the result in a journal.

On the Danish mathematical community in the 1930s, see Ramskov [64] and Schött [65], who estimates that there were ten mathematical positions in Denmark, the number of chairs having been much reduced. At the University of Copenhagen, for example, there were only three professors of pure mathematics, N. E. Nørlund, H. Bohr, and Johannes Hjelmslev (1873–1950); Jessen would eventually succeed Hjelmslev.

vous m'ayez envoyé votre mémoire des Acta Mathematica. Je n'ai pu que signaler un des points communs dans une note rajoutée après coup.[11] Je m'occupe d'ailleurs depuis quelques jours d'étudier plus complètement votre mémoire, dont je dois faire l'analyse en présence de M. Hadamard, et je vois que les points communs entre vos idées et les miennes sont encore plus nombreux que je ne le pensais.[12]

Mon lemme I est au fond la même chose que votre théorème du §14 (Representation of a function as a limit of an integral). Seulement je l'établis directement.[13] Votre ≪ important lemma ≫ du §11 est alors un cas particulier du mien ; et on arrive aussi assez facilement à votre théorème du §13 en partant de mon lemme I.

Je pense d'autre part que vous ne connaissez pas une conférence que j'ai faite au séminaire de M. Hadamard en janvier 1924; elle a paru dans la Revue de Métaphysique et de Morale, et je l'ai reproduite dans mon Calcul des Probabilités (pp. 325–345).[14] En la relisant récemment, j'ai trouvé que, outre une erreur à la page 330 (l. 12 à 19), qui m'avait été signalée par M. Steinhaus,[15] il y en a une assez fâcheuse,

[11] See footnote 5 above.

[12] This passage shows that until April 1935, when he was preparing his presentation to the Hadamard seminar, Lévy did not realise the similarity between the theorem in §14 of Jessen and his own lemma I, though he had received Jessen's article in September 1934.

[13] At this point, Lévy added in the margin: "[see the note attached to this letter]'.

[14] References [45,46]].

[15] In [45, p. 330], Lévy maintained the possibility of a countably additive (though not invariant and therefore, according to him, "very arbitrary") extension of Lebesgue measure to all the subsets of the interval [0, 1]. On this point there is very interesting information in [4, p. 153, note 103].

Lebesgue posed the problem of extending Lebesgue measure to all the subsets of the real line in his first Peccot course at the Collège de France in 1903, [40, p. 102]. As soon as the following year Vitali [77] showed that an invariant extension is impossible if one accepts the axiom of choice. Vitali's nonmeasurable sets, so contrary to Lebesgue's geometric intuition, led Lebesgue to contest and then to reject the axiom of the choice in mathematics. Saving the intuition was paramount; the axioms must conform or be dismissed. Lévy, not of this opinion, tolerated the axiom of the choice and the transfinite within the much broader limits of his own intuition [58, §39, pp. 124–125]. Lévy returned to the problem of measure in 1961 [59].

In 1914 [30, pp. 469ff], Hausdorff posed the problem of additive invariant extensions of Lebesgue measure and concluded that it was impossible in spaces of three or more dimensions. In 1923 Banach [3] proved the converse, that such an extension (additive and invariant) exists for the line and the plane. But in 1924, when Lévy wrote his note on "the laws of probability in abstract sets", the problem of countably additive noninvariant extensions of Lebesgue measure on the line was open, and Lévy thought it self-evident that such extensions exist using "Mr. Zerlmelo's method". In 1929 however, Banach and Kuratowski, using the continuum hypothesis, showed that it was not possible (also [75]). Steinhaus had to inform Lévy about this result in the course of their correspondence in 1930–1931. At the same time Steinhaus probably also communicated to Lévy his principle of correspondence [1930b], without the latter noticing; we do not know the fate of the Lévy-Steinhaus correspondence. Lévy returned to the point in [58, §9]. In his intellectual autobiography [60, p. 67], Lévy recalls his very great surprise at learning the negative result of Banach-Kuratowski. He adds:

> I had to face the facts; my intuition had misled me, and I am still sometimes astonished that my intuitive idea is false; I remain tempted by the same error.

à la p. 332 (l. 1à 5).[16] Malgré cela j'ai introduit dès cette époque des idées que M. Steinhaus et vous avez développées et précisées, sans vous douter que certaines se trouvaient déjà dans mon article de 1924 — et même dans un cours que j'ai fait en 1919.[17] Ce que j'appelle une partition correspond à ce que vous appelez ≪ construction of nets ≫.[18] Je l'indique pour les ensembles abstraits, et ensuite (p. 334, l. 6 à 13) j'indique la manière de le réaliser pour le cube à une infinité de dimensions;[19] c'est bien ce que vous faîtes. Quant à votre ≪ transferring principle ≫, je ne l'ai pas dans cet article indiqué très explicitement, mais quand j'ai écrit p. 332 (vers le bas) qu'on peut réaliser l'image de la partition sur un segment de droite, c'est

---

This suggests that Lévy's intuition is comfortable with the axiom of the choice but not with the continuum hypothesis, something that we know to be true. To look deeper into this issue, we would need to return to Cantor and his debates with the Paris school, but that would take us too far from our subject. See the very interesting books by Decaillot [23] and by Graham and Kantor [27], and, naturally, Guilbaud [29].

[16] Lévy gives a characterization of summability in an abstract set that is valid only for bounded functions, an error he acknowledged in [57, p. 157, note 2].

[17] This persistent claim appears in all or nearly all of Lévy's writings from 1935 on, for instance in [57, pp. 157–158, 169], [58, p. XII], [4, p. 163]. Steinhaus also reports the claim in a note in his communication to the 1937 Geneva Colloquium, [74, p. 65, note 14]:

> Mr. Paul Lévy has just informed me in private correspondence that the solution of the problem of measure on the [infinite] cube had come to him in the course of general considerations that the reader will find in a note at the end of his Calcul des Probabilités, without his judging it necessary to go into details. This Note of 1925 was followed by two articles by Mr. Paul Lévy [[48,49]] …My article [[71]] appears to have escaped Mr. Lévy's attention.

In note II of the second edition of Lévy's treatise of 1937, written when he was correcting the proofs and therefore around 1954, Lévy finally writes (p. 370, note 1):

> I had hoped to bring the mean in the sense of Gateaux closer to the concept of integral in the sense of Fréchet. It was an attempt doomed to fail. However this overly difficult problem had led me to pay too little attention to simpler questions that seemed trivial to me.

This may explain the obstinacy with which Lévy claimed things he had not written down but which he knew or certainly would have known if he had gone further in this direction. The Gateaux means are not Fréchet integrals (with respect to a probability measure), and everything else is trivial.

[18] Lévy introduced the concept of partition in 1925 [45, p. 331], indicating that the concept came from Norbert Wiener. Lévy developed the idea in [54, §9], [57, Chap. I, §3], and [58, Chap. II, §10], but forgot the correspondence with Jessen and his correspondent's article [37], which were most probably the direct cause of these later developments, given that the 1925 note had a different purpose and the contents of the 1919 Peccot course are unknown.

Jessen introduced the concept of "nets" in §6, "The Construction of Nets", in [37], in his theses [31], and in [32], following La Vallée Poussin.

[19] As Christian Berg points out to us, Jessen underlined this last part of the sentence and added in the margin, in Danish: "This is done by Wiener: Ann. of Math. 22 (1920–21) p. 66–72. Daniell: Bull. Amer. Math. Soc. 26 (1919–20) p. 448 below. Also Rice Inst. Pamphlet 8 (1921) p. 60–61."

In 1925 [45, p. 334, l. 6 to 13], Lévy writes in connection with the construction of partitions of the cube:

bien ce principe que je pensais.[20] Une partition n'est pas en effet simplement une subdivision indéfinie de l'ensemble étudié, mais une subdivision où chaque cellule a un poids et qui conduit à une définition de la probabilité (ou, si vous préférez, de la mesure). Je reconnais que j'aurais dû le dire plus explicitement. Peut-être l'avais-je fait à la conférence ; après 11 ans je n'en suis plus sûr.[21] Ce dont je suis sûr, c'est que c'est un résultat que je connaissais, et qui m'avait paru si évident qu'il suffisait

---

A partition of this cube can be obtained, for example, as follows: at the n-th stage, the interval of variation of each of the n coordinates $a_1, a_2, \ldots, a_n$ will be divided into 2^n equal intervals, making in all 2^{n^3} partial volumes; but these volumes are small in n directions only, and large in all others. For n infinite, the coordinates being fixed one after the other, one arrives at this result that each e contains one point and only one; yet a uniformly continuous functional will not in general be summable.

At this point in his note, Lévy's goal, in effect, is to "define a law of partition such that any uniformly continuous functional is summable" (p. 333, §VIII, first paragraph). So we must acknowledge that Lévy uses substantially the same "construction of nets" as Jessen but uses it to show that a theory of strong integration is impossible in Q_ω (which is perfectly correct), while Jessen uses it to construct a weak integral in Q_ω, having concluded that it is not necessary to consider strong integration (which would integrate all uniformly continuous functionals) for the applications he envisages. This is also true for applications in probability, as Lévy ends up realizing (though not before 1934, as he acknowledged later).

Lévy is thus at once right and wrong in this matter. That, at least, is the conclusion we have arrived at, without advancing it is right or definitive. The whole business is singularly obscure, for Lévy reconstructed the entire history in the 1930s [54,57,58], reinterpreting his earlier work in 1925 [46]—or even in 1919 or in 1918 in the military hospital—in the light of what he had learned since.

Lévy is not the only mathematician to reconstruct in this way, far from it. But we must give him credit for his lack of deceit or malice. He acts in good faith equally in truth and in error, which makes him charming and formidable at the same time. In such situations, all too common, we know that historians are careful about rejecting testimony, even the most erroneous. They put their energy into subjecting these accounts to a benevolent and determined external criticism, trying to find the truth in the error, without always succeeding. Let us hope that we too do not err too much. All the more so that the historian of mathematics faces an even more difficult task. Mathematical understanding is always an act of creation, even if what is created was already created very well by others, before or afterwards, independently or not, so that attributions, with rare exceptions, remain random and partial, and the reasons for them, no matter how firmly presented, remain heterogeneous, fugitive and misleading.

[20] Again Lévy is right and wrong. He uses a principle of transfer, not to establish a correspondence between the measure on Q_ω (of which he seems only to have an implicit idea) and Lebesgue measure on [0, 1], but with a very different aim: to show that one cannot define a probability law that distinguishes points in a set with a cardinality exceeding that of the continuum (Note §VII), a result rediscovered by Ulam in 1930 [75] and taken up by de Finetti in 1936 [24, p. 280].

[21] Lévy reproaches himself in the same way in [54, p. 29, note 1], [57, p. 158], and [58, §10, p. 21, note 1]. He should have noted at this point that the correspondence (which he sees and constructs very well) preserves measure. He even adds in a note in [57, p. 158] that he surely spoke about it at the time of his "lecture of 1924", a presentation to the Hadamard seminar in January 1924 that was published in [46] and included as the final note of [45]. One may doubt this "memory of Lévy", insofar as nowhere in the note does he refer to a "measure" in Q_ω, a measure which now (in 1935) seems so obvious to him that he believes he remembers it, a Socratic recollection as was his way.

de l'indiquer d'un mot. Par contre ce n'est que très récemment, notamment à la suite de la lecture du mémoire de Steinhaus (Studia t. II) et de votre communication au congrès d'Oslo de 1929,[22] que j'ai vu le parti que l'on pouvait tirer de ce principe pour les questions de probabilités dénombrables.

En tout cas il résulte nettement de mon article que les principes de la théorie de la mesure dans un ensemble quelconque sont ceux de M. Lebesgue.

Bien entendu les indications très brèves de mon article ne sont pas toujours suffisantes, et votre étude très complète restait nécessaire. Elle m'a appris d'ailleurs beaucoup de choses que je ne savais pas (notamment le §3 ; avant de l'avoir lu je n'avais pas pensé qu'il y ait intérêt à préciser si l'on considérait des intervalles ouverts ou fermés ; et je n'avais jamais étudié la représentation de la mesure dans Q_ω par le symbole d'une intégrale d'ordre infini[23] ; et je ne parle pas de l'application aux séries de Fourier dont je viens seulement de commencer l'étude).

Excusez cette lettre un peu longue. Comme Steinhaus, qui pourtant connaissait mon mémoire de 1924, ne semble pas s'être douté de ce qu'il contenait, je pense qu'il ne doit pas paraître toujours très clair, et qu'il y a intérêt à ce que je vous indique explicitement les points qui sont en relation avec vos travaux.

Croyez, cher Monsieur, à mes sentiments les plus cordiaux.

P. Lévy

Veuillez me rappeler au souvenir de M. Harald Bohr.

Yet one can maintain that Lévy did indeed mention in his lecture that the correspondence preserves measure and defines it at the same time. There is support for this position in note 2 on p. 44 of the second edition of Lebesgue's *Intégration*, and Lebesgue was probably present at the seminar in 1924. But between these two hypotheses and others that we have not yet imagined, it is best not to take sides and to leave the question to the more learned.

In any event, Lévy readily recognizes in 1935 that Jessen's article was the occasion of a re-awakening (or awakening) of the *correspondence principle* and its "applications to the theory of denumerable probabilities" [54, p. 29, note 1]. So for now on, without hesitation, we may place Jessen in the pantheon of Lévy's masters, alongside Hadamard, Borel, Wiener, Fréchet, Cantelli, Mlle Mezzanotte, Steinhaus, Khinchin, Marcinkiewicz, etc, all of whom had succeeded in awakening or re-awakening Lévy's brain. We could set up a whole hierarchy of Lévy pantheons. Borel and Hadamard are tutelary deities but sometimes also awakeners. Then there is the unattainable paradise where are installed those, like Doeblin or Kolmogorov, who fire so quickly that Lévy's brain does not have time to awaken to find its bearings. On Kolmogorov, see [17, 69]. On Wolfgang Doeblin, see [19], in which Doeblin shows, among many other things, that a continuous martingale is a Brownian motion with a change of time. This was in 1940, twenty or thirty years before the leading specialists of the time realized that this was an important property (see footnote 7). The relations between Doeblin and Lévy are almost Oedipal, like those between Doeblin and his father Alfred Döblin. For this subject see M. Petit's beautiful book [63]. Recall that Doeblin's first publication, jointly signed with Lévy, was for Doeblin an "easy" demonstration of a conjecture of Lévy's.

[22] References [32, 33, 72].

[23] Jessen [37], §9 and especially §§13 and 14, where we find Jessen's theorems, the finest part of the paper. Here perhaps is the very origin of Lévy's "awakening." Obviously Jessen's integral in $\mathbb{Q}_\omega$ is not an integral in Cauchy's sense, for the continuous functions are generally not integrable, but as it is in "correspondence" with the Lebesgue integral, it acquires enough properties from it to be applicable to the theory of denumerable probabilities, in particular the Fubini property, which is also at the heart of Lévy's lemma without Lévy realizing it before he read Jessen.

Attached note: Démonstration d'un théorème de M. Jessen (Acta, vol. 63, p. 273) en partant de mon lemme I (Bull. Sc. Math. 1935).[24]

Soit $f(x)$ mesurable dans Q_ω. On peut supposer $0 < f(x) < 1$. (Autrement on raisonnerait sur $g(x) = \frac{1}{2} + \frac{1}{\pi}\operatorname{Arctg} f(x)$).

Appliquons le lemme I, en appelant E l'inégalité $\frac{h}{2^p} \leq f(x) < \frac{h+1}{2^p}$. A tout $\varepsilon_p > 0$, on peut faire correspondre N_p tel que si E est vérifié et sauf dans des cas de probabilité $< \frac{1}{2^p}\varepsilon_p$, on ait pour tout $n > N_p$,

$$P_n\left(\frac{h}{2^p} \leq f(x) < \frac{h+1}{2^p}\right) > 1 - \varepsilon,$$

et par suite

$$\frac{h-1}{2^p} \leq f_n(x) < \frac{h+2}{2^p}$$

en posant

$$f_n(x) = E_n(f(x)) = \int_{Q_{n,\omega}} f(x) \mathrm{d}w_{n,\omega}$$

et enfin

$$|f(x) - f_n(x)| < \frac{\varepsilon}{2^p}$$

En appliquant ce résultat pour $h = 0, 1, \ldots, 2^p - 1$, on voit que l'inégalité précédente est vérifiée, sauf dans des cas de probabilité $< \varepsilon_p$, pour tout $n > N_p$.

Faisons $p = 1, 2, \ldots; \varepsilon_p = \frac{\varepsilon}{2^p}$. On voit que, sauf dans des cas de probabilité $< \sum \varepsilon_p = \varepsilon$, on a

$$|f(x) - f_n(x)| < \frac{\varepsilon}{2^p}, \text{ pour tout } p, \text{ et } n > N_p.$$

Il y a bien convergence presque partout de $f_n(x)$ vers $f(x)$, c.q.f.d.

---

[24] As the reader can see, Lévy's proof is perfectly correct and very simple, but it assumes that the function f is bounded without correctly indicating how to pass to the case of an unbounded integrable function. On the other hand, the proof of the corollary is quite invalid, as Jessen naturally saw right away.

Basically, the step from Lévy's lemma to Jessen's theorem is natural and quite clear, as Doob grasped at once [20,21]. Lévy gives the reason in [57], which can be seen as a supplementary letter to Jessen. On pp. 177–178, in fact, Lévy notices that, by transfer, his lemma is nothing other than Lebesgue's density theorem, while Jessen's theorem corresponds to Lebesgue's differentiation theorem. Moreover, he states that Lévy's lemma can be proved in this way, although this turns out to be much more complicated than his "direct" method [57, p. 178]. Now we know that Lebesgue first obtained his density theorem from his differentiation theorem [41] and then proved the differentiation theorem from the density theorem [42, §33]; this second method was recommended by La Vallée Poussin in his contemporary work. There is no reason to be surprised that the same situation is found in this new framework.

Corollaire. De

$$|f(x) - f_n(x)| < \frac{\varepsilon}{2^p} \text{ pour } n > N_p, \text{ sauf dans un ensemble de mesure } < \varepsilon_p,$$
$$|f(x) - f_n(x)| < 1 \text{ toujours,}$$

on déduit

$$\left|\int_{Q_\omega} f(x)\mathrm{d}w - \int_{Q_n} f_n(x)\mathrm{d}w_n\right| \leq \int_{Q_\omega} |f(x) - f_n(x)|\mathrm{d}w < \frac{2}{2^p} + \varepsilon_p, \text{ pour } n > N_p.$$

Comme $\frac{2}{2^p} + \varepsilon_p$ est arbitrairement petit, c'est le théorème du §13 de M. Jessen.

Le 4/4/35

P. Lévy

4 Jessen to Lévy. Undated Draft, About 8 April 1935

Dear Professor Lévy.[25]

I have your letter of April 4 and the proofs of your memoir, which is to appear in Bulletin des Sciences Mathématiques, and I am very thankful for both.

From your letter I learn that the notion of measure in infinitely many dimensions as well as the transferring principle is indicated already in your paper from 1924 and reproduced in your Calcul des Probabilités and that your ideas on this subject partly go as far back as 1919. I am sorry not to have known this, as I have done my best to give the complete references in my memoir in Acta mathematica. It seems that the notion of measures in infinitely many dimensions has had the rather curious fate to be discovered and rediscovered at least five times. The priority clearly belongs to Daniell who has given a complete treatment of the notion and not only indications already in 1919 using his theory of general integrals (cf. the references in my memoir). That Daniell has been quite clear about what he did is seen not only from these but also from other papers of his from this period; it might interest you that he indicates a subdivision of the infinite-dimensional cube in Bulletin of the American Math. Soc. 26 (1919–20) p. 448 and explicitly points out the importance of the problem for the calculus of probability in The Rice Institute Pamphlet 8 (1921) pp. 60–61.[26] Wiener did not rediscover the theory as he knew Daniell's papers, but he made several applications of the theory (cf the references in my memoir); he

[25] At the top of this letter Jessen wrote in pencil, "Sendt I noget anden Form," which Christian Berg has kindly deciphered and translated as, "Sent in a somewhat different form." We have made minor corrections that Jessen would surely have made himself before sending the letter.

[26] John Aldrich informs us that the pamphlet by Daniell that Jessen mentions to Lévy was never cited by anyone except by Jessen in this letter to Lévy (see also footnote 19 above). It is an extremely interesting review paper and would merit detailed study. The pages 60 and 61 quoted by Jessen show explicitly how the measure that Jessen constructed in 1929 on Q_ω can be defined starting from a Daniell integral, adding:

emphasized more than Daniell the usefulness of the subdivision, see for instance his paper in Proceedings of the London Math. Soc. (2) 22 (1924) pp. 454–467, as far as I remember, the Transferring principle does not occur explicitly in his early papers where he always need Daniells integrals, but he told me that it was quite familiar to him and in later papers he uses it,[27] in order that he may work with Lebesgue integrals instead of Daniell integrals; see for instance his memoir in Acta mathematica 55 (1930) §13. As I now learn you have had similar ideas without knowing that the problem was already treated by Daniell. The same has been the case with Steinhaus and myself, who found the theory independently of each other and at the same time; neither of us knew Daniell's work. Finally I learned from you that Denjoy had recently rediscovered the theory.[28]

In my memoir in Acta mathematica I did not go too much into the history of the subject which is complicated by the fact that the ideas in question have developed gradually, so that it is now hard to say who has the priority in each case. I hope, however, that the first sentences in §1 have made it clear, that my program was "to study in greater detail than has been done before" the theory in question.

Page 2 [29] :

That is: Q_ω is the product of an infinite number of abstracts spaces $C_1, C_2, \ldots$ in each of which a measure has been defined, so that the measure of the space itself is 1.

I am writing on a paper in which I intend to develop the theory for abstract spaces. (Cf the references in my memoir at the top of p. 251)

In this paper I shall make due reference to your paper from 1924 and to your book. [30]

> This type of integral might possibly be useful in connection with probability of sets of functions defined by means of Fourier constants or by the coefficients of a series expression.

This prophetic sentence seems to anticipate, and in any case announces Steinhaus's works from 1924 to 1930 and those of Paley, Wiener, and Zygmund in 1930 on random Fourier series. For more information, see Aldrich [1], to whom we owe the essence of this note and whom we thank very warmly.

[27] Jessen met Wiener at the end of 1933. In his talk at the conference of the American Mathematical Society in December 1933, in which Jessen must have also participated, Wiener constructed his measure by using the correspondence principle; [79], also [80]. This principle is also the basis for his great article in 1930 [78] quoted by Jessen, and for his work with Paley in 1930. Presumably he learned it from Steinhaus or from Paley. As already briefly indicated, the Daniell integral is not very useful for calculations, it is only there to ensure their coherence. To calculate it is to better use the Lebesgue integral by transfer, or the Gateaux means, or the approximation by the game of heads or tails, or changes of variables and suitable symmetries, etc.

[28] Denjoy [18].

[29] There is clearly a passage missing here, in which Jessen must have told Lévy of his work in progress on the extension of his theory to an abstract framework.

[30] We have not found any reference by Jessen to Lévy's [46] or [45]. He and Sparre Andersen cite Lévy's 1937 book [58] in [2].

Your remark in your letter that my theorem from §§13 and 14 will follow from your lemme I interested me very much as I have tried hard to find simple proofs for these theorems. I do not think however that the proofs which you sent me are sufficient to give my theorems in full generality, for the following reasons:

1°. Is it sufficient to prove the theorems for bounded functions? I do not think that you can deduce them for an arbitrary integrable $f(x)$ from their validity for $g(x) = \frac{1}{2} + \frac{1}{\pi}\mathrm{Arctg} f(x)$.

2°. Is it really possible to deduce the theorem of §13 from that of §14? This is as far as I can see what you wish to do. As it stands your proof is not valid, since the function

$$f_{n,\omega} = \int_{Q_n} f(x) \mathrm{d}w_n$$

does not appear at all. (The term $\int_{Q_\omega} f(x)\mathrm{d}w_\omega - \int_{Q_n} f_n(x)\mathrm{d}w_n$ on the left in your estimation is simply A – A = 0). It might be possible to argue as follows (and this, I believe, is what you have in mind)[31]

$$h_n(x) = f(x) - f_n(x) \to 0 \quad \text{p.p.}$$

This implies

$$\int_{Q_n} h_n(x)\mathrm{d}w_n = f_{n,\omega}(x) - A \to 0 \quad \text{p.p.}$$

The theorem "$h_n(x) \to 0$ p. p. implies $\int_{Q_n} |h_n(x)|\mathrm{d}w_n \to 0$" is actually true for bounded functions but I do not know how to prove it [?] just my theorem of §13. I do not [?] is true for integrable functions.[32]

In the case of abstract spaces the proofs of the theorem at §§11, 13 and 14 must be rearranged (cf my memoir, footnote 2 on page 251), the reason being that the notion of a net can be applied only in special cases of abstract spaces. I intend first to give a new and direct proof of the theorem of §14 ; the lemma of §11 (which is of course only the 0- and 1—law of the calculus of probability in a general form)[33]

---

[31] Jessen's two objections are well founded.

[32] This part is difficult to decipher. Jessen wrote over the original text with blacker ink, obscuring the original formulas.

[33] This parenthesis, which appears in the margin of the text and which was thus added afterwards, seems to be obvious, but it appears neither in Jessen's initial article [37], where §11 mentions only an application of the "important lemma" to a result of Steinhaus [73], nor in Jessen's later article with Wintner [38]. So Jessen learned of the existence of probability's 0–1 law between the publication of his article and this letter of April 1935, perhaps by reading Kolmogorov's *Grundbegriffe* more attentively.

For his part, as we have noted, Lévy initially saw Jessen's important lemma as a consequence of his lemma I in [55], also without indicating that it was a version of Kolmogorov's 0-1 law. Moreover, a short note by Laurent Schwartz [66], which undoubtedly resulted from a conversation between Schwartz and Lévy at a Sunday meal in the spring of 1935, also attributed the 0-1 law in Kolmogorov's probabilistic formulation to Jessen's [37], obviously unaware that Kolmogorov had follows then from this theorem, and the proof of the theorem of §13 may then be left unaltered. The theorem of §14 I prove by generalizing F. Riesz's proof of the differentiation theorem for monotone functions as follows:

Let $f(x)$ be integrable in Q_ω and $f_n(x) = \int_{Q_{n,\omega}} f(x)\mathrm{d}w_{n,\omega}$. In order to prove that $f_n(x) \to f(x)$ p.p., I first prove that lim $f_n(x)$ exists p. p. It is sufficient to consider the case where $f(x) \geq 0$. Put $\varphi(x) = \liminf f_n(x)$, $\psi(x) = \limsup f_n(x)$. It is sufficient to prove that if $0 < \alpha < \beta < \infty$ then the set $D_{\alpha\beta} = [\varphi(x) < \alpha, \psi(x) < \beta]$ is a null set. For $0 \leq m \leq n$ let

$$A_{mn} = [f_{m+1}(x) > \alpha, \ldots, f_{n-1}(x) > \alpha, f_n(x) \leq \alpha]$$
$$B_{mn} = [f_{m+1}(x) < \beta, \ldots, f_{n-1}(x) < \beta, f_n(x) \geq \beta]$$

A_{mn} and B_{mn} are cylinders with basis in Q_n. We shall make use repeatedly of the remark that if C is a cylinder with base in Q_n then $\int_C f(x)\mathrm{d}w_\omega = \int_C f_n(x)\mathrm{d}w_\omega$ so that if $f_n(x) \leq \alpha$ or $\geq \beta$ in C we have $\int_C f(x)\mathrm{d}w_\omega \leq \alpha mC$ or $\geq \beta mC$ respectively. Suppose now that $0 \leq m < n < \beta$ and that C is a cylinder with base in Q_m (if $m = 0$ we take $C = Q_\omega$) — We consider the set $CA_{mn}B_{np}$ which is a cylinder with base in Q_p. Hence since $f_p(x) \geq \beta$ in this set we have $\beta mCA_{mn}B_{np} \leq \int_{CA_{mn}B_{np}} f(x)\mathrm{d}w_\omega$.

formulated it in his *Grundbegriffe* in 1933. The mathematical conversations between Lévy and his future son-in-law are discussed by Scwhartz in [68, pp. 93–95]. Schwartz did not continue in this line at the time but took it up again at the end of the 60s [67,81].

On the other hand, Lévy devoted a paragraph of his 1936 survey article [57, p. 179], to the "Lemma of MM. Kolmogorov et Jessen" (the 0–1 law), quoting the *Grundbegriffe* of 1933 in a note. Thus Lévy finally took note of Kolmogorov's text between May and December 1935; perhaps it was at the time of Jessen's letter of 11 August 11 1935, although we cannot be sure. See also a letter to Fréchet of 29 January 1936, in [4, pp. 95–96], and that book's very enlightening note 108. The entire book is indispensable for anyone studying Lévy's work.

In [57, p. 179], Lévy asked how the 0–1 law of denumerable probabilities can be transferred to the interval [0, 1] endowed with Lebesgue measure using the correspondence principle. In a note he writes,

> There is reason to note that, taking into account the principle of linear representation mentioned in §4, this lemma can be assimilated to a theorem on linear sets established in 1916 by Mr. Burstin, which is a corollary of Lebesgue's theorem, just as Mr. Jessen's lemma is a corollary of our theorem of §10 [Lévy's lemma].

We do not know how Lévy learned of Burstin's result. Did they meet in Bologna? Celestyn Leonovitch Burstin was born on 28 January 1888, in Ternopol in Ukraine. He studied at the University of Vienna where he published interesting work on analysis, in particular [14], which Lévy cited, and on Riemannian geometry; for his work on the latter see [28]. Unable to find a position in Vienna, being Jewish and a member of the Communist party, he emigrated to Belarus, where he was a professor at Minsk and a member of the Academy of Science of Belarus. During the Stalin purges, he was arrested at the end of 1937 and died in prison on 1 October 1938. See also [12]. For the purges in Belarus, see the collection published by the Academy of Science of Minsk and also [61]. This last reference was provided by J.-M. Kantor.

Summing this for all $p > n$ for fixed m, n we get

$$\beta m \sum_p CA_{mn}B_{np} \leq \int_{\sum_p CA_{mn}B_{np}} f(x)\mathrm{d}w_\omega \leq \int_{CA_{mn}} f(x)\mathrm{d}w_\omega \leq \alpha m CA_{mn}$$

since CA_{mn} is a cylinder with base in Q_n in which $f_n(x) \leq \alpha$. Summing now for all $n > m$ we get $\beta m \sum_{np} CA_{mn}B_{np} \leq \alpha m \sum_n CA_{mn} \leq \alpha m C$. We now take first $m = 0$ and $C = Q_\omega$; observing that $D_{\alpha\beta} \subseteq \sum_{np} A_{on}B_{np}$ we get $mD_{\alpha\beta} \leq m \sum_{np} CA_{on}Bnp \leq \frac{\alpha}{\beta} m Q_\omega = \frac{\alpha}{\beta}$. Next we take $C = A_{on}B_{np}$ for a fixed n and p; then we get $m \sum_{qr} A_{on}B_{np}A_{pq}B_{qr} \leq \frac{\alpha}{\beta} m A_{on}B_{np}$ the indices q and r being restricted by $p < q < r$. Summing afterwards over n and p and observing that $D_{\alpha\beta} \subseteq \sum_{npqr} A_{on}B_{np}A_{pq}B_{qr}$ we get $mD_{\alpha\beta} \leq m \sum_{npqr} A_{on}B_{np}A_{pq}B_{qr} \leq \alpha\alpha\beta m \sum_{np} A_{on}B_{np} \leq (\frac{\alpha}{\beta})^2$. Proceeding in this manner we get $mD_{\alpha\beta} \leq (\frac{\alpha}{\beta})^n$ for every n, hence $mD_{\alpha\beta} = 0$.

Page 3 :

It remains to prove that $\lim f_n(x) = f(x)$ p. p.. This may be proved as follows. From the definition of measure in Q_ω one readily deduces the following approximation theorem : If $f(x)$ is integrable in Q_ω and $\varepsilon > 0$ is given then there exists an $m = m(\varepsilon)$ and an integrable function $g(x)$ depending only of $x_1, \ldots, x_m$ so that $\int_{Q_\omega} |f(x) - g(x)|\mathrm{d}w_\omega < \varepsilon$. This implies $\int_{Q_\omega} |f_n(x) - g_n(x)|\mathrm{d}w_\omega < \varepsilon$ for all n. Now $g_n(x) = g(x)$ for $n \geq m$. Hence $\int_{Q_\omega} |f_n(x) - g(x)|\mathrm{d}w_\omega < \varepsilon$ for $n \geq m$ and consequently $\int_{Q_\omega} |f(x) - f_n(x)|\mathrm{d}w_\omega < 2\varepsilon$ for $n \geq m$. Hence $\int_{Q_\omega} |f(x) - f_n(x)|\mathrm{d}w_\omega \to 0$ as $n \to \infty$ and this, together with the existence of $\lim f_n(x)$ p. p. proves that $\lim f_n(x) = f(x)$ p. p.[34]

This is the simplest proof I know of the theorem of §14. For bounded function and more generally for functions of the class L^p $(p > 1)$ it is possible to give very short proofs of the theorems of §§13 and 14 just mentioned and the majorization theorem of §16, but in this way I could not prove the theorem for arbitrary integrable functions.[35]

Excuse me this long digression; I thought it might interest you to know this other proof.

---

[34] This new proof of the theorem of §14 that Jessen is giving to Lévy, where one feels the influence of Lévy's note, is the first direct proof of Jessen's theorem. It makes no appeal to the differentiation theorem or to the transfer principle. It is so close to the analogous theorem in [35, part 4] and in his 1946 article with Sparre Andersen [2] that one cannot see what prevented Jessen from publishing the result ten years earlier, aside from the absence of an abstract framework that suited him, and which Lévy also lacked, so that the correspondence could not come to a resolution. The appropriate framework was almost brought out in the version of Lévy's lemma in his 1937 treatise [58], but Jessen did not see it, and neither did Doob in his first article of 1940. Lévy never saw it, though he was persuaded of the contrary. Simple mathematical ideas are always the most hidden, as Laplace, who could calculate everything, already complained.

[35] Here Jessen uses the same polemic technique as Lévy, for which we cannot reproach him. Drown Lévy's proof for the bounded case, as brief as it is elegant, in a fog of vague commentary that diminishes it value: the bounded case is treated very easily; I knew all that for a long time…

With kind regards also from Prof. Bohr
Sincerely yours
Borge Jessen

[Written over the calculations on page 2 with a darker ink:]

I hope that your memoir in Bulletin des Sciences mathématiques will have appeared when I finish my paper so that there will be nothing to prevent me from using these new proofs (of course with due reference to your work; my paper will at any rate be mainly expository).[36]

5 Lévy to Jessen. Hennequeville, 24 April 1935

Cher Monsieur Jessen,[37]

J'ai bien reçu votre lettre du 8 avril et vos mémoires, et vous remercie.

Ce que vous dîtes de la priorité de Daniell m'intéresse naturellement beaucoup. Je connaissais l'existence de ses travaux ; mais à cette époque je lisais difficilement l'anglais. Quoique j'aie fait des progrès, je ne le lis pas encore facilement. J'avais lu un résumé de quelques résultats de Daniell au début d'un mémoire de Wiener, et pensais connaître ainsi ses résultats les plus importants. J'apprends seulement par votre lettre qu'il connaissait bien avant moi le principe de correspondance.[38]

Je n'ai donc plus aucune raison de demander que mon mémoire de 1924 soit mentionné pour l'histoire de ce principe. Cela ne me surprend pas beaucoup. Je crois bien me rappeler que si je n'ai pas indiqué plus explicitement ce principe de correspondance, c'est qu'il me paraissait probable qu'un principe si simple devait être connu. C'est seulement en le trouvant redécouvert par vous et par Steinhaus que j'avais regretté de ne pas l'avoir exprimé plus explicitement.

Je suis bien d'accord avec ce que vous dîtes au sujet de mon lemme I. En effet les raisonnements par lesquels je pensais en déduire vos §13 et 14 n'étaient pas corrects. J'avais bien remarqué que votre §14 contenait mon lemme I comme cas particulier ; mais, toujours parce que je lis l'anglais très lentement, j'ai dû donner le bon à tirer de mon mémoire avant d'être assez avancé dans la lecture du vôtre, que pour cette raison je n'ai cité qu'en partie. Je m'en excuse.[39]

---

[36] This refers to the article announced in 1935 in [38] that will appear only in 1939 in the 4th installment of [35]. Lévy's name does not appear there. Lévy's 1937 book is nevertheless cited in the general bibliography in the the 10th and final installment in 1947 and in the 1946 article with Sparre Andersen [2].

[37] Hennequeville is a district of Trouville in Calvados where Lévy was probably spending the Easter holidays. Easter was on April 21 in 1935.

[38] Lévy had not understood or not managed to read Daniell. He thought that Daniell used the principle of correspondence to construct his integral. He reconsiders this point in the next letter.

[39] See [55, p. 89 note 1], where Lévy cites only part of Jessen's memoir, §11 "An important lemma". Lévy corrects this in his *Notice* [54, pp. 43–44], where he writes, "This theorem [Lévy's lemma] was obtained independently of me by Mr. Jessen, or at least it appears to be a special case of a theorem of Mr. Jessen, who in addition indicates a more special case of great importance." The

M. Bohr m'avait déjà donné votre thèse. Mais je n'avais pas réussi à la comprendre. C'est seulement maintenant en la rapprochant de votre mémoire écrit en anglais que je vois qu'elle contenait déjà plusieurs résultats importants de ce mémoire.

Comme j'ai maintenant deux exemplaires, je pense bien faire d'en donner un à la Bibliothèque de l'Institut Henri Poincaré.

Je vous enverrai mon mémoire dès que j'aurai les tirages à part ; mais vous aurez sans doute pu le voir plus tôt dans le Bulletin des Sciences Mathématiques. J'en ai rédigé, aussi en 1934, un autre qui doit paraître dans le Journal de Mathématiques, et où je donne la condition nécessaire et suffisante pour que la somme d'un grand nombre de variables aléatoires indépendantes dépende asymptotiquement de la loi de Gauss. On savait déjà que cette condition était suffisante ; le résultat nouveau est qu'elle est nécessaire.[40]

J'ajoute enfin que je rédige une Notice résumant mes travaux. L'impression était commencée avant que j'ai reçu votre lettre. Comme je l'ai fait pour mon Mémoire, je vais ajouter des Notes au bas de la page pour mentionner les priorités nouvellement venues à ma connaissance, c'est-à-dire cette fois celles de Daniell.[41]

Croyez, cher Monsieur, à mes sentiments les plus dévoués.

P. Lévy

6 Lévy to Jessen. Paris, 3 May 1935

Mon cher Collègue,

Je vous écris, après avoir regardé les mémoires de Daniell, et relu votre lettre. Je me suis aperçu qu'il y avait eu un malentendu. Je ne sais pas pourquoi j'avais cru que vous me parliez de la priorité de Daniell dans le principe de correspondance ; je vois que vous ne parliez que de la mesure dans l'espace à une infinité de dimensions.

Or, pour la mesure dans les ensembles abstraits, la priorité appartient à Fréchet (Bulletin de la Société Mathématique de France, 1915, pp. 248–265). Il n'a pas étudié spécialement le cas de l'espace à une infinité de dimensions. Il n'en est pas moins celui qui a donné le premier les éléments essentiels de cette théorie.[42]

reference is to Jessen's "important lemma", i.e. Kolmogorov's 0-1 law in Jessen's framework. We have already commented on the note on page 45 where Lévy states that his article [55] was "sent to the editors of the *Bulletin des Sciences Mathématiques* in September 1934". Again in 1936 [57, p. 179], Lévy states that Mr. Jessen had established "several theorems on integration in Q_ω which include and exceed" Lévy's lemma. Lévy did not include similar statements in his 1937 treatise [58] or later writings.

[40] Lévy [56].

[41] Lévy [54, p. 28, note 1].

[42] In 1915 Fréchet [25] constructed a theory of integration associated with an abstract measure, on the model of the Radon-Stieltjès-Lebesgue integral, but he did not construct a (non-trivial) measure in a space of infinitely many dimensions, contrary to what Lévy seems to think and would continue to suggest in later writings. In his reply Jessen insists on this point, quite correctly but without much effect. Measure in infinite dimensions pre-exists and exhibits itself so clearly (albeit tardily) in Lévy's mind that it has no need for an explicit construction, which is in any case perfectly obvious a

Je ne connaissais pas son travail, paru pendant la guerre. J'ai, comme Daniell, retrouvé ces résultats en 1919. J'ai entendu parler de Daniell pour la première fois en 1922 par N. Wiener, qui m'a indiqué ses principaux résultats, et ce n'est, je crois, que en 1924 ou 1925 que je me suis aperçu de la priorité de Fréchet, de sorte que vous pouvez trouver dans mes travaux l'expression incorrecte d'intégrale de Daniell.[43]

Quant au principe de correspondance, je pense de nouveau que la question se présente bien comme je le pensais d'abord ; d'abord l'indication peu précise de mon mémoire de 1924–1925 ; puis vos travaux et ceux de Steinhaus.

Croyez, mon cher Collègue, à mes sentiments dévoués.

P. Lévy

P. S. – Le N° de mars 1935 du Bulletin des Sciences Mathématiques, contenant le début de mon Mémoire, a paru. Mon Mémoire commence à la p. 84. J'ai compté 28 pages, je pense donc que cela fera p. 84 à 111.

– J'ai remarqué que Daniell citait bien Fréchet dans un de ses premiers mémoires. J'avoue n'avoir pas vu clairement ce qu'il a ajouté d'essentiel aux idées de Fréchet, sauf quelques précisions pour l'espace à une infinité de dimensions.

P. L.

7 Jessen to Lévy. Copenhagen, 11 August 1935

Dear Professor Levy,

I am sorry to have been so long in answering your letter of May 3 and thanking you for kindly sending me your "Notice sur les travaux".

Regarding the question of priority for the transferring principle for the infinite dimensional cube it is so, that, as far as I know the literature, it occurs first in your paper from 1924–1925 and later by Steinhaus and myself in 1929–1930. But it is difficult to separate the transferring principle from the underlying construction of a net in the infinite dimensional cube, and this construction occurs already in a paper of Daniell from 1919–1920 (Bulletin of the American Mathematical Society, Vol. 26, p. 448 below) and is reproduced by Wiener in 1920–1921 (Annals of Mathematics, 2. Series, Vol. 22, pp. 66–72, Example 3). These authors had no reason to use the transferring principle, since they had the general Daniell integral, which is much

posteriori. For Jessen on the contrary, what matters is the measure's construction, in all its rigor and complexity, a construction that we can thus date and attribute as precisely as possible. To Daniell in the first place, but also to Jessen who rediscovered it only a few years later, to Steinhaus and others, and perhaps even to Lévy, though Jessen is too polite to write exactly what he thinks on this point. A dialogue without resolution.

[43] Lévy does not quote Daniell in his note of 1924–1925, but does so in [44] (see his next letter). He then cites him regularly beginning in 1934, in particular in the important article [53], in the *Notice* [54, p. 28, note 1], then in [57, §7, p. 167], [58, §10, pp. 17–18, note 2], where Lévy again merges his own work in 1918–1919 with Daniell's in 1918–1919 and with Fréchet's earlier work of 1915. Yet Fréchet, for his part, never laid any claim on Daniel's integral. Is the reader, even a relatively well-disposed one, convinced? And has Lévy, by constantly repeating the amalgamation, convinced himself?

more satisfactory. It is only a pity that the general integral was not a larger success, and that therefore later writers (including myself) have preferred by means of the transferring principle to reduce everything to ordinary Lebesgue integrals.[44]

I do not think that you estimate Daniell's papers sufficiently since you can say, that you do not see, what he has added of essential to the ideas of Frechet. There is the following essential difference : Frechet (in his paper in Bulletin de la Societe math. De France 1915) starts from a completely additive set function, so that what he has generalized is the definition of the Lebesgue integral when the Lebesgue measure is already known. This he has done in a very elegant way. What Daniell has done is to generalize the definition (due to Young) of the Lebesgue integral based on the properties of the Riemann integral, that is on an object, which is much more elementary than the Lebesgue measure. Personally I prefer the definition of the integral based on a measure to the direct definitions (though the latter have also their great importance); to Daniells work corresponds in the theory of measure a generalization of the definition of the Lebesgue measure based on the properties of the Jordan measure. Of this important problem (which is treated e.g. in Kolmogoroff's "Grundbegriffe der Wahrscheinlichkeitsrechnung") I find nothing in Frechets paper.[45]

With kind regards, I am

Very sincerely.

8 Lévy to Jessen. S. Cristina, 23 August 1935

Mon cher Collègue,[46]

---

[44] Without detracting in any way from Daniell's fundamental contribution, which gave Kolmogorov's foundations their generality and Wiener's and Lévy's measures their first mathematical existence, one can undoubtedly make the opposite case, following Steinhaus [74]. The fact that the Daniell integral was little known in continental Europe was also an opportunity. In place of the Daniell integral, analysts (Danish and Polish especially) created and applied the principle of correspondence, finding new theorems that Daniell's integral did not permit to be seen. These included, for example, Jessen's theorem in §14, the image in infinite dimensions of the differentiation theorem in one dimension, and, in the opposite direction, the following theorem, which Lebesgue seems not to have seen but which Borel certainly anticipated [11, Chap. II]: Riemann sums of a function integrable in Lebesgue's sense converge to the Lebesgue integral of this function for evenly spaced and increasingly fine partitions anchored at almost any point in the interval of definition. This is the image in one dimension of the theorem in Jessen's §13, as Jessen showed in [36]. And, as is well known, the theorem of Borel-Jessen is a way of seeing the strong law of large numbers, as Doob would make clear at the Lyon conference in 1948, placing it at the same time in Kolmogorov's axiomatization; see Bernard Locker's chapter in the present volume.

[45] Jessen's comments are to the point. Yet Fréchet's article does contain a theorem on the extension of a measure on an algebra to the generated σ-algebra, independently of Carathéodory's theorem [16]; see e. g. Bogachev [8, Vol. 1, p. 419]. So one might follow Lévy here and conclude that Fréchet could already have had measure in infinite dimensional space, if only he had thought of looking for it. Bogachev has further comments.

[46] The Lévy family spent their holidays in San Cristina in the Dolomites. Bernard Locker has very kindly given us information on this subject:

J'ai reçu il y a quelques jours votre lettre du 11 août. Je n'ai pas la possibilité de revoir ici les travaux de Daniell ; mais je suis persuadé que vous avez raison.

Je sais très bien que j'ai le défaut d'être absorbé par mes propres travaux au point qu'il m'est toujours très difficile de lire complètement ceux des autres. Je connaissais vaguement ceux de Daniell, et lorsque vous m'en avez reparlé, étant trop occupé, je les ai lus trop rapidement, voulant surtout voir si j'y trouvais l'énoncé du principe de correspondance et si je pouvais maintenir ce que j'avais écrit dans ma notice, dont je devais à ce moment donner le bon à tirer. Aussi ne suis-je pas surpris de votre réponse.

J'ai d'ailleurs mal exprimé ma pensée en vous écrivant. J'aurais dû écrire ≪ je ne vois pas encore bien ce que Daniell a ajouté au mémoire de Fréchet ≫. Votre lettre m'aide à le mieux voir maintenant, et je vous en remercie.

J'avais d'ailleurs déjà cité Daniell, que je connaissais un peu grâce à N. Wiener ; voyez mon fascicule 5 du Mémorial des Sciences Mathématiques. J'ai ensuite retrouvé le Mémoire de Fréchet de 1915, que je n'avais pas connu ou que j'avais oublié ; et il m'a semblé que je n'étais pas le seul à avoir oublié de le citer.

Bien cordialement à vous.

P. Lévy

9 Bohr and Jessen to Lévy. Copenhagen, 14 July 1947

Dear Professor Levy,

First of all we wish to thank you and Mrs. Levy heartily for the most agreeable evening spent with you and your family in Paris,[47] and for all your kindness during the interesting days in Nancy, on which we look back with great pleasure.[48] We also thank you very much for your kind letter. It would have been such a great pleasure

Denise Lévy-Piron often spoke to me about the pleasure her father took in spending time in the mountains, even telling me that "my father was an exceptional mountaineer", which I doubted, attributing the adjective "exceptional" to filial piety…San Cristina is in the valley of Val Gardena in Italy…. and I know from Mr. and Mme Piron that Lévy adored the mountain and Italy….

[47] Lévy liked to entertain foreign mathematicians passing through Paris. See [4, p. ix], where K. L. Chung recalls a dinner at avenue Théophile Gautier, at the end of which Lévy served Port-du-Salut cheese. In those days, this cheese was made by the monks at the Port-du-Salut Abbey, unlike the cheese now sold under the name Port Salut. We do not know which members of the Lévy family were present at the Danish dinner.

[48] The reference is to the conference on harmonic analysis chaired by S. Mandelbrojt and held in Nancy from 15 to 22 June 1947. This prestigious conference was financed by CNRS and the Rockefeller Foundation. It brought together the great names in a field then being transformed. Among the invited lecturers was, of course, Lévy, who presented a paper on the harmonic analysis of stationary random functions, which contained a very beautiful result of Blanc-Lapierre and Fortet [6,7], whether being recalled or rediscovered. This hardly pleased those two authors [13, p. 31]. Bohr and Jessen, for their part, presented a paper on almost periodic functions [10]. The conference proceedings were published by the CNRS in 1949.

to us all to have seen you in Copenhagen already this autumn. However, we are very sorry to say that, as you also felt yourself, in these difficult times it does not seem possible to obtain a sufficient grant to cover the expenses of your contemplated stay here. In earlier times this would have been easy, but at present the funds are rather hard up and have already disposed of their means for the nearest future. We hope, however, that within long it will be possible to make arrangements for your coming here to give some lectures which would be a great pleasure to all the Danish mathematicians.

As you may have heard, Mr. Schwartz has been invited to visit Copenhagen in September to give some lectures on his extraordinary theory of distributions.[49] That this invitation has been possible is due to the interest of this theory also among all the applied mathematicians, which has made a grant available that otherwise would have not been obtainable for mathematical lectures.

With kind regards to yourself and your family.

Yours sincerely

HB (Bohr) BJ (Jessen)

Acknowledgements We are very grateful to Christian Berg, who obtained these letters for us from the Jessen archives at the Institute for Mathematical Sciences of the University of Copenhagen.

References

1. Aldrich, J.: "But you have to remember P. J. Daniell of Sheffield". Electronic Journal for History of Probability and Statistics **3**(2) (2007)
2. Andersen, E.S., Jessen, B.: Some limits theorems on integrals in an abstract set. D. Kgl. Danske Vidensk Selskab Mat. -fys. Medd. **22**(14) (1946)
3. Banach, S.: Sur le problème de la mesure. Fund. Math. **4**, 7–33 (1923)
4. Barbut, M., Locker, B., Mazliak, L.: Paul Lévy and Maurice Fréchet. 50 years of correspondence in 107 letters. Springer (2014)
5. Bernstein, S.N.: Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. Math. Ann. **97**, 1–59 (1926)
6. Blanc-Lapierre, A., Fortet, R.: Sur la structure des fonctions aléatoires strictement stationnaires à spectre totalement discontinu. Comptes rendus Acad. Sci. Paris **222**, 1155–1157 (1946)
7. Blanc-Lapierre, A., Fortet, R.: Théorie des fonctions aléatoires. Applications à divers phénomènes de fluctuation, avec un chapitre sur la mécanique des fluides par J. Kampé de Fériet. Masson, Paris (1953)
8. Bogachev, V.I.: Measure Theory. Springer, New York (2006). 2 vol.
9. Bohr, H., Jessen, B.: Über die Werteverteilung des Riemannschen Zetafunktion I, II. Acta Math. **54**, 1–35 (1930). 58:1–55, 1932

[49] Schwartz's exposition of Fourier analysis for distributions at the Nancy conference greatly impressed Harald Bohr. It was Bohr who would present Schwartz's work to the International Congress of Mathematics in Cambridge, USA, in 1950, helping establish its international prominence. Schwartz describes his visit to Copenhagen in October 1947 at Bohr's invitation in his memoir [68, Chap. VIII, p. 309].

10. Bohr, H., Jessen, B.: Mean motions and almost periodic functions. In: Colloques internationaux du CNRS, 15, Analyse harmonique, Nancy 15–22 juin 1947, pp. 75–84. Gauthier-Villars, Paris (1949)
11. Borel, É.: Le calcul des intégrales définies. J. Math. Pures Appl. **(6) 8**, 159–210 (1912). Oeuvres 2, pp. 827–878
12. Borodin, A.I., Bugai, A.S.: Biographical dictionary of mathematicians (in Russian). Kiev (1987)
13. Brissaud, M. (ed.): Écrits sur les processus aléatoires, mélanges en hommage à Robert Fortet. Hermes, Lavoisier, Paris (2002)
14. Burstin, C.: Über eine spezielle Klasse reeller periodischer Funktionen. Monat. f. Math. und Phys. **26**(1), 229–262 (1915). Correction 27(1):163–165, 1916
15. Cantelli, P.: Considérations sur la convergence dans le calcul des probabilités. Ann. Inst. H. Poincaré **5**, 3–50 (1935)
16. Carathéodory, C.: Vorlesungen über reelle Funktionen. Teubner, Leipzig (1918). 2d edition 1927, 3rd edition 1968
17. Chaumont, L., Mazliak, L., Yor, M.: Some aspects of the probabilistic works. In: É. Charpentier, A. Lesne, N.K. Nikolski (eds.) Kolmogorov's Heritage in Mathematics, pp. 41–66. Springer (2007)
18. Denjoy, A.: Sur les variables pondérées multipliables de M. Cantelli. Comptes rendus Acad. Sci. Paris **196**, 1712–1714 (1933)
19. Doeblin, W.: Sur l'équation de Kolmogoroff. Comptes rendus Acad. Sci. Paris **331**, 1031–1187 (2000). Sealed envelope, accepted by the Académie des Sciences on 26 February 1940, opened 18 May 2000
20. Doob, J.L.: Regularity properties of certain families of chance variables. Trans. Amer. Math. Soc. **47**, 455–486 (1940)
21. Doob, J.L.: Stochastic Processes. Wiley, New York (1953)
22. Dubins, L., Schwarz, G.: On continuous martingales. Proc. Nat. Acad. Sci. **53**, 913–916 (1965)
23. Décaillot, A.M.: Cantor et la France. Kimé, Paris (2008)
24. de Finetti, B.: Les probabilités nulles. Bull. Sci. Math. **(2) 60**, 275–288 (1936)
25. Fréchet, M.: Sur l'intégrale d'une fonctionnelle étendue à un espace abstrait. Bull. Soc. Math. France **43**, 248–265 (1915)
26. Garding, L.: Mathematics in Sweden, History of Mathematics Vol. 13. Amer. Math. Soc., London Math. Soc. (1998)
27. Graham, L., Kantor, J.M.: Naming Infinity: A True Story of Religious Mysticism and Mathematical Creativity. Belknap Press, Cambridge (2009)
28. Gromov, M., Rokhlin, V.: Embeddings and immersions in Riemannian geometry. Russian Math. Surveys **25**(5), 1–57 (1970)
29. Guilbaud, G.T.: Lumen in Manibus : Essai sur la connaissance intermittente. Unpublished (2008). St. Germain-en-Laye
30. Hausdorff, F.: Grundzüge der Mengenlehre. Veit, Leipzig (1914). 2d ed. 1927
31. Jessen, B.: Hovedsætningerne indenfor de næstenperiodiske Funktioners Teori og deres indbyrdes Sammenhæng. In: Magisterkonferens i Matematik, København, 23/4–21/5 1929 (1929)
32. Jessen, B.: Bidrag til Integralteorien for Funktioner af uendelig mange Variable, (doctoral thesis). G. E. C. Gads Forlag, København (1930)
33. Jessen, B.: Über eine Lebesguesche Integrationstheorie für Funktionen unendlich vieler Veränderlichen. In: Comptes rendus du septième Congrès des Mathématiciens scandinaves tenu à Oslo, 19-22 août 1929, pp. 127–138. A. W. Broggers Boktrykkeri, Oslo (1930)
34. Jessen, B.: Eine Integrationstheorie für Funktionen unendlich vieler Veränderlichen, mit Anwendung auf das Werteveilungsproblem für fastperiodische Funktionen, insbesondere für die Riemannsche Zetafunktion. In: Verhandlungen des Internationalen Mathematikerkongresses, Zürich, September 1932, vol. 2, pp. 135–136 (1932). Also Mat. Tidsskr. B (1932), pp. 59–65
35. Jessen, B.: Abstrakt maal- og integralteori, 1–10. Mat. Tidsskr. B pp. 73–84 (1934). Also 1935 pp. 60–74, 1938 pp. 13–26, 1939 pp. 7–21, 1942 p.p 43–53, 1944 pp. 28–34, 35–37, 1947 pp. 1–20, 21–26, 27–36; collected in one volume with the same title, København, Matematisk Forening, 1947

36. Jessen, B.: On the approximation of Lebesgue integrals by Riemann sums. Ann. Math. **35**, 248–251 (1934)
37. Jessen, B.: The theory of integration in a space of an infinite number of dimensions. Acta Math. **63**, 249–323 (1934)
38. Jessen, B., Wintner, A.: Distribution functions and the Riemann Zeta function. Trans. Amer. Math. Soc. **38**(1), 48–88 (1935)
39. Kac, M.: Sur les fonctions indépendantes (I). Studia Math. **6**, 46–58 (1936)
40. Lebesgue, H.: Leçons sur l'intégration et la recherche des fonctions primitives. Gauthier-Villars, Paris (1904). 2nd ed. 1928
41. Lebesgue, H.: Sur les fonctions représentables analytiquement. J. Math. Pures Appl. **(6), 1**, 139–216 (1905)
42. Lebesgue, H.: Sur l'intégration des fonctions discontinues. Ann. Sci. ENS **(3) 27**, 361–450 (1910). Œuvres II, pp. 185–274
43. Lévy, P.: Leçons d'analyse fonctionnelle. Gauthier-Villars, Paris (1922). 2nd ed. 1951
44. Lévy, P.: Analyse fonctionnelle. Gauthier-Villars, Paris (1925). Mémorial des sciences mathématiques, n° 5, 2nd ed. 1951
45. Lévy, P.: Calcul des probabilités. Gauthier-Villars, Paris (1925)
46. Lévy, P.: Les lois de probabilité dans les ensembles abstraits. Revue de Métaphysique et de Morale pp. 149–175 (1925)
47. Lévy, P.: Sur la croissance des fonctions entières. Bulletin de la S. M. F. **58**, 29–59, 127–149 (1930)
48. Lévy, P.: Sur les séries dont les termes sont des variables éventuelles indépendantes. Studia Math. **3**, 119–155 (1931)
49. Lévy, P.: Sur quelques questions de calcul des prebabilités. Prace Mat.-Fiz. **39**(1), 19–28 (1931)
50. Lévy, P.: Sur les méthodes de M. Norbert Wiener et la fonction $\zeta(s)$ (1932). Lecture at the International Mathematical Congress in Zürich, September 1932
51. Lévy, P.: Généralisation de l'espace différentiel de M. Wiener. Comptes rendus Acad. Sci. Paris **198**, 786–788 (1934)
52. Lévy, P.: Propriétés asymptotiques des sommes de variables aléatoires enchaînées. Comptes rendus Acad. Sci. Paris **199**, 627–629 (1934)
53. Lévy, P.: Sur les intégrales dont les éléments sont des variables aléatoires indépendantes. Ann. R. Sc. Norm. Sup. Pisa **(2) 3**, 337–366 (1934). Observations sur le mémoire précédent, ibid., 4 (1935), pp. 217–218
54. Lévy, P.: Notice sur les travaux scientifiques de M. Paul Lévy. Hermann, Paris (1935)
55. Lévy, P.: Propriétés asymptotiques des sommes de variables aléatoires enchaînées. Bull. Sci. Math. **(2) 59**, 84–96, 109–128 (1935)
56. Lévy, P.: Propriétés asymptotiques des sommes de variables aléatoires indépendantes ou enchaînées. J. Math. Pures Appl. **14**, 347–402 (1935)
57. Lévy, P.: Sur quelques points de la théorie des probabilités dénombrables. Ann. Inst. H. Poincaré **6**, 153–184 (1936)
58. Lévy, P.: Théorie de l'addition des variables aléatoires. Gauthier-Villars, Paris (1937). Page numbers refer to the 2nd edition, 1954
59. Lévy, P.: An extension of the Lebesgue measure of linear sets. Proceedings of the Fourth Berkeley Symposium on Math. Stat. and Prob., June–July 1960 **II**, 273–287 (1961)
60. Lévy, P.: Quelques aspects de la pensée d'un mathématicien. Blanchard, Paris (1970)
61. Marakou, L.: Repressed writers, scientists, educators, public and cultural figures of Belarus (in Russian). Smolensk, Minsk (2003–2005). 4 vol.
62. Marcinkiewicz, J.: Sur les fonctions indépendantes (I), (II). Fund. Math. **30**, 202–214, 349–364 (1938)
63. Petit, M.: L'équation de Kolmogoroff. Vie et mort de Wolfgang Doeblin, un génie dans la tourmente. Ramsay, Paris (2003). 2nd ed., Paris, Gallimard (Folio 4240) 2005
64. Ramskov, K.: The Danish Mathematical Society through 125 years. Historia Mathematica **27**, 223–242 (2000)

65. Schött, T.: Fundamental research in a small country: Mathematics in Denmark 1928–1977. Minerva **18**, 243–283 (1980)
66. Schwartz, L.: Sur une question de calcul des probabilités. Bull. Sci. Math. **2, 60**(1), 101–102 (1936)
67. Schwartz, L.: Semi-martingales sur des variétés, et martingales conformes sur des variétés analytiques complexes. Springer (1980)
68. Schwartz, L.: Un mathématicien aux prises avec le siècle. Odile Jacob, Paris (1997)
69. Shafer, G., Vovk, V.: The sources of Kolmogorov's *Grundbegriffe*. Statistical Science **21**, 70–98 (2006). Unabridged version: arXiv:1802.06071 [math.HO], The origins and legacy of Kolmogorov's *Grundbegriffe*
70. Siegmund-Schultze, R.: Rockefeller and the Internationalization of Mathematics Between the Two World Wars. Birkhäuser, Basel (2001)
71. Steinhaus, H.: Les probabilités dénombrables et leurs rapports avec la théorie de la mesure. Fund. Math. **4**, 286–310 (1923)
72. Steinhaus, H.: Sur la probabilité de la convergence des séries. Studia Math. **2**, 21–39 (1930)
73. Steinhaus, H.: Über die Wahrscheinlichkeit dafür, dass des Konvergenzkreiss einer Potenzreihe ihre natürlich Grenze ist. Math. Zeitschrift **31**, 408–416 (1930)
74. Steinhaus, H.: La théorie et les applications des fonctions indépendantes au sens stochastique. In: Colloque consacré à la théorie des probabilités et présidé par M. Maurice Fréchet, Genève 11 au 16 octobre 1937, cinquième partie, (Actualités Sci. Ind. 738), pp. 58–73. Hermann, Paris (1938)
75. Ulam, S.M.: Zur Masstheorie in der allgemeinen Mengenlehre. Fund. Math. **16**, 140–150 (1930)
76. Ville, J.: Étude critique de la notion de collectif. Gauthier-Villars, Paris (1939)
77. Vitali, G.: Sul problema della misura dei gruppi di punti di une retta. Œuvres pp. 231–235 (1905)
78. Wiener, N.: Generalized harmonic analysis. Acta Math. **55**, 117–258 (1930)
79. Wiener, N.: Random functions. J. of Math. Phys. **24**(1), 17–23 (1935)
80. Wiener, N., Zygmund, A.: Notes on random functions. Math. Zeitschrift **37**, 647–668 (1933)
81. Yor, M.: Deux maîtres ès probabilités. Gazette des mathématiciens, Edition spéciale Laurent Schwartz (1915–2002) (2003)

Counterexamples to Abstract Probability: Ten Letters by Jessen, Doob and Dieudonné

Bernard Bru and Salah Eid

Abstract

These ten letters are mainly about two erroneous theorems published by Joseph Doob in 1938. Both concerned probability measures on abstract spaces. The first concerned the construction of measures on infinite-dimensional product spaces from their finite-dimensional margins (the Daniell-Kolmogorov construction). The second concerned the existence of regular conditional probabilities. Ten years after the theorems were published and used by Doob's students, counterexamples were independently discovered by Jean Dieudonné and by Erik Sparre Andersen in collaboration with Børge Jessen. The correspondence begins when Jessen writes to Doob about the counterexample he and Sparre Andersen had discovered. It reveals that Doob had not encountered Jessen's theorem until then, and it suggests that Doob's return to martingales, which he had left aside after his initial work on them in 1940, was inspired by his learning about Jessen's work on the topic.

Keywords

Børge Jessen · Daniell-Kolmogorov theorem · Disintegration of probability measures · Erik Sparre Andersen · History of probability · Jean Dieudonné · Joseph Doob · Martingales

---

Introduction and footnotes translated from the French by John Aldrich, University of Southampton
e-mail: john.aldrich@soton.ac.uk

B. Bru
Université Paris Descartes, MAP5, Paris, France

S. Eid (✉)
Université Paris Diderot, Paris, France
e-mail: salaheid.h@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*,
Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_16

1 Introduction

In 1948, the Danish mathematician Børge Jessen and his younger colleague Erik Sparre Andersen discovered a counterexample to two theorems, published ten years earlier by Joseph Doob, concerning probability measures in infinite-dimensional spaces. Doob's first theorem attempted to generalize the Daniell-Kolmogorov construction to abstract spaces. The second attempted to show that a probability measure on an infinite-dimensional abstract space can always be "disintegrated" into conditional probabilities. Jessen wrote to Doob when he noticed that his and Sparre Andersen's example refuted the first theorem. Doob promptly acknowledged his errors and mentioned to Jessen that he had already known that the second theorem was false, because of an counterexample discovered by Jean Dieudonné. So Jessen also wrote to Dieudonné.[1]

The letters reproduced here, from the Jessen Archive at the Institute of Mathematics at Copenhagen, were written in the course of a year, from the spring of 1948 to the spring of 1949, and we present them chronologically. We have five letters from Jessen to Doob, with two responses, and two letters from Jessen to Dieudonné, with one response.

As recounted in the chapter in the present volume by Bernard Bru and Salah Eid, Børge Jessen (1907–1993) played an important role in the history of martingale theory. His inspiration came not from the language or intuition of probability but from Lebesgue's theory of integration and its transfer principle. When he and Erik Sparre Andersen (1919–2003) discovered their counterexample, they were struggling with the shortcomings of Sparre Andersen's attempt to generalize the Daniell-Kolmogorov construction even further than Doob had attempted.

Joseph Leo Doob (1910–2004), professor at the University of Illinois, Urbana-Champaign from 1935, was one of the most productive analysts of the 20th century. He is clearly the central figure of the modern theory of martingales, which owes him everything, beginning with its name, which he borrowed from Jean Ville; see the chapter by Bernard Locker in the present volume. The two theorems that Doob had believed he had proven in 1938 were not about martingales, but they were so natural, so necessarily true, that they were accepted for ten years without anyone thinking of questioning them.

There is no shortage of literature on Jean Dieudonné (1906–1992), cofounder and principal writer of Bourbaki; see in particular [20]. In 1948 Dieudonné was a professor in the Faculty of Sciences at Nancy, having spent the previous academic year, from May 1946 to December 1947, in Brazil, in Rio de Janeiro and São Paulo, where he visited his friend and teacher André Weil who was professor there. Dieudonné would be with Weil in Chicago in the 50s before being appointed to the Institut des Hautes Études Scientifiques in 1959 and then to the new University of Nice in 1964.

[1] We follow Jessen and other contemporaries in referring to "Sparre Andersen" as if this were his last name. Other sources (the yearbooks of the Danish Academy of Sciences, for example) give his name is given as "Andersen, E. S.", and we follow this practice in our list of references.

Jessen's first letter, which sounds the death-knell for the Daniell-Kolmogoroff-Doob theorem, was a bitter pill for a mathematician of Doob's strength. It has a place in the history of martingales because it may have occasioned Doob's return to his own theory of martingales. It is not easy to identify the moment when Doob truly recognized the importance of this class of random variables, which he had first considered in 1940 [14], following Ville. It was definitely not in 1940 and must have been before 1953, which leaves some margin. We know from the correspondence that by 1948 Doob had read the new probabilistic version of Jessen's theorem, which Sparre Andersen and Jessen had presented in 1946 [3], and which Doob finally adopted in his 1953 book [17]. By the time Doob delivered his lecture at the Lyon conference in late June 1948 [16],[2] one senses that something had changed. Martingales now have their name, and this is a sign that does not mislead. Doob's student Laurie Snell defended his thesis on "martingale systems" in 1951.

Was Doob's reading Jessen cause or consequence of the emergence of the theory? We do not know, but we can assert that the invention occurred suddenly in the spring of 1948 somewhere in Doob's mind or in the surrounding countryside, and that among his personal antecedents and his precursors were Jessen, Lévy, Ville, Doob and a few dozen others. Of course, the Jessen-Lévy-Doob theorem was only a part of Doob's theory of martingales, which draws some of its richness, and not the least part, from stopping properties. A suitably stopped martingale remains a martingale. These properties may have their origins in Doob's conversations with Feller and Chung about heads and tails and in the abundant work of Ville, Doeblin, Lévy and of all the inventors of probability theory over three centuries. This is a matter of the very nature of things. The universe is not obliged to be beautiful, but it is beautiful.

It is always difficult to determine the date and circumstances of the birth of a theory. For example, Laurent Schwartz tells us in his memoirs, [30, p. 223], that the theory of distributions was born "suddenly in only one night" in November 1944 on the ground floor of 11 rue Monticelli in the 14th arrondissement of Paris, where he was living at the time. But he admits that he is unable to understand what triggered this discovery, which would change dramatically his career as a mathematician, nor, moreover, who were "his precursors and his personal antecedents", which leaves the field free for historians to find him his place [25,27].

2 Jessen to Doob, 11 May 1948

Dear Professor Doob

As you will have noticed Mr Sparre Andersen and I have raised doubt as to the validity of the proof of your abstract generalization of the theorem on the introduction of measures in a real Cartesian space of an infinite number of dimensions.[3] The problem had interested us very much, and actually Sparre Andersen had given a proof of an

[2] See the chapter by Bernard Locker in the present volume.

[3] This refers to the first theorem in Doob's 1938 article [13]. The article in which Sparre Andersen and Jessen questioned it, [3], was submitted on 5 October 1945 (before postal service between America

even more general theorem. This proof we had discovered to be wrong before noticing that the theorem was in your paper from 1938. On reading your proof we found, however, that it used an argument, which we had also attempted to apply but could not carry through, and your proof therefore seemed incomplete. I owe you an apology for not having written to you before publishing our paper, my only excuse is that when the paper was written the mail service to America had not yet been opened.

When I write to you now, it is because we believe to have found a counter-example of the theorem, and this we would like to show you before publishing it.

…[4]

I would very grateful to hear your opinion of this example.

I remain very sincerely yours
Børge Jessen

3 Doob to Jessen, 17 May 1948

Dear Professor Jessen:

Thank you for your letter, although I can hardly say its news was welcome. I had already realized that the theorem following the one to which you give a counterexample was false, but I had not realized that the first was false.[5] These two theorems are very closely related, and counterexamples to the two are of essentially the same type.

---

and Denmark was restored) and printed on 1 April 1946. Sparre Andersen and Jessen wrote (p. 22, Sect. 24, note 1):

> An analogous theorem on arbitrary measures in product sets has been given by Doob, but his proof seems incomplete (it is not seen how the sets on p. 92 are chosen). The proof by Sparre Andersen of a more general theorem is incomplete…

Sparre Andersen had published his more general erroneous theorem in 1944 [2].

It is possible that Doob objected to the 1946 criticism, and that Jessen wanted to be apologize, but we have found no proof of this. Doob himself was not sparing in writing notes of this kind (see e.g. the notes on pp. 91 and 135 of [13]), and Doeblin (who was at least as tough as Doob) complained to him about it; see Doeblin's correspondence with Doob in [11]. In any case, the letter's main point is the counterexample discovered in the spring of 1948 by Sparre Andersen and Jessen, which left in no doubt the irreparable inaccuracy of Doob's theorem.

[4] Here we omit a passage that reproduces word for word the description of the counterexample as it appears in the Sparre Andersen's and Jessen's 1948 article [4, Sect. 4]. The article was submitted on 28 June 1948, after Doob's reply.

[5] The "first theorem" in question here and in the following letter is Theorem 1.1 in Doob's 1938 article [13, p. 90], which extends to a general abstract framework the Daniell-Kolmogorov theorem for the real case or for the abstract case for product probabilities (the case of independent variables). The "following theorem" (or the "second theorem" lower down), is Theorem 3.1 on page 96, which asserts the existence of abstract disintegrations (or the existence of regular conditional probabilities), and this is equally erroneous. Jessen's counterexample also works for this theorem, which Doob already knows is false from an example of Dieudonné's that Halmos had communicated to him (see the letter in Sec. 5 below). But until Jessen's first letter arrived, Doob still thought that his Daniell theorem was correct.

Perhaps you have seen Kakutani's proof in the case of independent finite dimensional measures the extension to the infinite dimensional case is correct.[6] His proof is quite simple, of the same type as the treatment of two dimensional measure by Hopf in the latter's Ergodentheorie in the Ergebnisse series. Kakutani's proof can be used word for word in the general case, with the following hypothesis, which saves the theorem for the purposes of probability.

Let $P(E)$ be a probability measure in x_i space.

For $n > 1$, let $P(x_1, \ldots, x_{n-1}; E)$ be for fixed $x_1, \ldots, x_{n-1}$ a probability measure in x_n sets E. Then these probability measures can be used to define finite dimensional measures in the usual way; $P(E)$ is the x_1 probability measure and the other function is the conditional probability of x_n sets if the preceding x_js are known. Under the hypothesis that there are such conditional probabilities to define the finite dimensional measures, it follows that the extension to infinitely many dimensions can be accomplished following Kakutani in the independent case, in which case the conditional measures do not actually depend on the conditioning variables. Conversely, if finite dimensional measures are given, they determine conditional probabilities as described above in very general cases but not always (my second theorem is also false). Your example is also a counterexample to the second theorem.[7]

I think that this way of looking at it is helpful; to find out when the extension to infinitely many dimensions is possible, one may find out when the finite dimensional measures are determined by conditional probability measures. This is true for example if the coordinate spaces are themselves finite dimensional Borel sets, as can be seen by the principle of my proof; the essential property is that the given field can be mapped on Borel sets, of course the conditions I describe are not necessary, but I suspect that they are pretty close to it.[8]

Do you intend to visit this country in the next few years? Our work has many common points and it would be interesting to discuss these and other matters in detail.[9]

Sincerely,
Doob

---

[6] See Kakatuni's [24, I], which has a very simple treatment of the independent case. S. Kakutani (1911–2004) was at the Princeton Institute for Advanced Study between 1940 and 1942, and it was there that he learned Doob's erroneous general theorem.

[7] The formulation Doob proposes here is a version of the very general result, without topological assumptions, that was published in 1949 by C. Ionescu Tulcea. Neveu [29, V-1] has a presentation of this theorem, which is adequate for the general theory of Markov chains, and of which Jessen said in the following letter that he had become convinced "in the course of my attempts to prove this theorem [the theorem of Daniell-Doob]".

[8] The search for the minimal topological assumptions ensuring the validity of Doob's disintegration theorem produced an abundance of literature in the 50s and later. See the very numerous references in [8, Vol. II, p. 462].

[9] From this last sentence it can safely be concluded that by 17 May 1948 Doob had read Sparre Andersen and Jessen's 1946 article [3] and undoubtedly also Jessen's articles from the 30s, and had understood that they contained a satisfactory version of his first theory of 1940 [14], which he was not yet calling the theory of martingales.

4 Jessen to Doob, 29 May 1948

Dear Professor Doob

Thank you for your letter. That a counter example to your first theorem would imply that conditional probability fields need not exist, I knew, having realized in the course of my attempts to prove this theorem, that the proof succeeds along the same lines as the proof in case of product measures when conditional probability fields exist. Kakutani's proof for product measures I do not know, since we have not yet received the Japanese journals from the war. But it can hardly be simpler than my proof which was announced in Wintner's and my paper in Trans. Amer. Math. Soc. 88 (1935) (Sect. 15) and which appeared (in danish) in Mat. Tidsskr. B 1939. This proof (which I believe to be the first that has been published) is reproduced in Sparre Andersen's and my article in Dansk Vid. Selsk. Mat.-fys. Medd. 22, Sect. 19 (1948) (Sect. 23).

The introduction of a measure in an infinite product by means of conditional probability fields I have not worked out in detail. It hardly seemed worth while as long as the validity of your first theorem was undecided. Now it seems to me that it should be done. What would you think if we joined in a little article giving this result which, as you mention, is sufficient for the probability applications? Further results might possibly be included. Sparre Andersen and I might then in our article put in some words to the effect that according to our example the introduction of measure in infinite products intended to cover the case of dependent variables must be done in a different manner and that you and I would treat this question in a forthcoming paper.[10]

I expect to spend the major part of 1949 (from February) in America and hope very much to see you. We might write the paper then or perhaps we might do it by correspondence, though that, of course, is not so convenient.

[10] Such a note is indeed in the final paragraph of Sect. 3 of Sparre Andersen and Jessen's [4]. Yet as we will see, the project of a joint Doob-Jessen article did not materialize (see the letter in 11 below). This consolation prize was obviously not much motivation for Doob, and Jessen rather quickly realized that it held no interest to him, except as a way of expressing his sympathy for a colleague he had put into difficulty.

In any event, Doob preferred to work alone at home and published very little with others. In his interesting conversation with Snell [31], he says:

> I corresponded with many mathematicians but never had detailed interplay with any but Kai Lai Chung and P.- A. Meyer in probability and Brelot in potential theory. My instincts were to work alone and even to collect enough books and reprints so that I could do all my work at home.

Sparre Andersen will be in America this winter spending most of his time with Prof. Feller at Cornell, who, as you may know, is a good friend of the Copenhagen mathematicians.[11]

Sincerely yours
Børge Jessen

5 Doob to Jessen, 4 June 1948

Dear Professor Jessen:

I think it would be a good idea to write a joint paper clearing up this whole subject. I first heard about my error through Halmos who sent me a counter example to my second wrong theorem that had been sent him by Dieudonné. My errors were rather unfortunate, among other reasons because that second theorem was used very essentially in papers by Halmos, Kakutani and Ambrose.[12]

I have just been trying to read your proof of the existence of measure in infinitely many dimensions in the independence case, and as far as I can understand the language it seem to be the same as that of Kakutani which I mentioned to you. Of course yours is much earlier.[13] I have a vague recollection that von Neumann may have also proved the theorem in a course of lectures at Princeton, and that it appeared in a mimeographed edition of his lectures, but we do not have the volume in our library.[14]

I do not think that there is any hurry in our publication. We might as well wait until you are here in this country. Perhaps you will be able to visit Urbana for a while. Please write me your plans when they are definite. I shall be in Europe to attend the Lyon conference on probability and statistics, but shall return immediately after it, leaving July 9, from Cherbourg.[15]

Our work has had many points of contact. The war has confused my records, and I do not know which of my reprints I have sent you. Have I sent you the one Amer.

[11] William Feller (1906–1970) fled Nazi Germany in 1933 for Denmark and Sweden, before emigrating in the United States in 1939. See [7] for Feller's links with Danish mathematicians.

[12] Paul Halmos (1916–2006) was Doob's first doctoral student in Urbana, and he defended his Ph.D. thesis in 1938. His 1941 article [23] was based on Doob's theorem of disintegration; it is referred to below in the correspondence with Dieudonné. See also Halmos's autobiography [22] and Burkholder and Protter's obituary of Doob [9].

Warren Ambrose (1914–1996) completed his Ph.D. under Doob's direction in 1939. He used Doob's theorem in [1].

For Kakutani, see footnote 6 above. He used Doob's theorem in [24, II].

The brilliant careers of these three mathematicians do not seem to have especially suffered from this unfortunate mistake.

[13] Jessen's proof, probably dating back to 1934–1935, was published in Danish in 1939 and in English in 1946.

[14] Reference [28].

[15] The Lyon conference was held from June 28 to July 3. Doob's lecture there, [16], is reprinted in [26]; see also the chapter by Bernard Locker in the present volume.

Math. soc. Trans. 1940 in which I derive theorems which are essentially yours in Kgl. Danske Vi. Sels. 1946?[16] Of course our terminologies and points of view are quite different. I do everything from the point of view of functions, you from the point of view of set functions.[17] The theorems involved are very important in probability theory, and I am going to discuss various applications at Lyon.

If Andersen will be at Cornell, I shall see him. Feller and I visit each other frequently.[18]

Best wishes,
Doob

6 Jessen to Dieudonné, 17 June 1948

Dear Professor Dieudonné.

Together with Mr. Sparre Andersen I have recently found an example showing that the Daniell-Kolmogoroff theorem (Kolmogoroff : Grundbegriffe der Wahrscheinlichkeitsrechnung, p. 27) on the introduction of measure in an infinite Cartesian product by means of consistent measures in the finite sub-products cannot be extended to abstract sets in the case where the coordinates are dependent (in the probability sense). The example will be published in Danske Vid. Selsk. Mat.-fys. Medd.[19]

Professor Doob, who like Sparre Andersen has attempted to prove the extension, and whom I have communicated our example, has informed me, that you have given a counterexample of the related theorem about the existence of conditional probability measures. Naturally, our example is also a counterexample of this theorem, since the extension of the Daniell-Kolmogoroff theorem to abstract sets may be carried

[16] References [3,14]. It is clear that by this time Doob had made the connection between his theory and Jessen's. But Jessen is not quoted in Doob's lecture at Lyon, where only Ville's name appears.
[17] Jessen reconsiders this point in his 1948 article with Sparre Andersen [5, Sect. 1]. He explains there that he was unaware of Doob's 1940 work [14] in 1946, and that while he had preferred to adopt the viewpoint of set functions, this had been only for convenience of exposition, the results being just as valid for point functions. The 1948 article makes this explicit, making the two theorems perfectly symmetrical so that they include Doob's 1940 results. The 1948 article was received by the journal on 16 August 1948, and published on 23 October. It is discussed in the letter in Sect. 9 below.
[18] Doob and Feller met for the first time at the meeting of the American Mathematical Society at Darmouth in 1940 [31]. He was, according to Doob "the first mathematical probabilist I ever met." Doob and Feller had very different visions of mathematics. To be convinced, simply compare Doob's 1953 book and Feller's 1950 book [17,21]. Doob did not like calculations and looked for the most general possible results and concepts. Feller loved only formulas and the rare and precious flowers that appear only after complicated calculations and particularly meticulous investigations. That did not prevent them from getting together to try to convince American mathematicians that the theory of probability was a branch of mathematics like any other, contrary to general belief. See also [15,18,19].
[19] Reference [4].

through in the same manner as for product measures when conditional probability measures are supposed to exist.

I would be very grateful if you would let me know whether your example has been published in order that we may then quote it. If it has not we shall restrict ourselves to mention that you have given such an example.

I allow myself to send you (under separate cover) some of my papers relating of functions of infinitely many variables.

Please give my kind regards to your colleagues. I regret not to have met you when I was at Nancy last year.[20]

Believe me, very sincerely yours
Børge Jessen

7 Dieudonné to Jessen, Nancy, 28 June 1948

Dear Professor Jessen

I have received your letter of June 17 and your reprints on Integration, for which I thank you most heartily. The example Professor Doob refers to was found by me last September, while working on Prof. Halmos's paper "The decomposition of measures".[21] My paper is due to appear in the next few weeks in the "Annales de Grenoble" under the title "Sur le théorème de Lebesgue-Nikodym (III)", p. 25–53; the example which interests you is given in p. 42. As soon as I have reprints of this paper, I shall have great pleasure in sending one to you, together with some of my older papers on Integration and Banach spaces.

Hoping to have the pleasure of meeting you some day, I am

Very sincerely yours
J. Dieudonné, 2, Rue de la Craffe, Nancy.

8 Jessen to Dieudonné, 13 September 1948

Dear Professor Dieudonné,

Thank you very much for your kind letter and for the reprints which I received some time ago.

[20] This is a reference to the Nancy Conference of June 1947 (see Jessen's letter to Paul Lévy, dated 13 September 1948, in the collection of correspondence between the two of them in the present volume). Dieudonné was in Brazil at the time.

[21] Reference [23]. Dieudonné's counterexample [12, p. 42] shows that the fundamental theorem in Halmos's paper (Theorem 1, p. 390) is not true in general. Halmos's argument is correct, but it relies on Doob's erroneous result in [13, Theorem 3.1, p. 96], as Dieudonné points out (p. 42, note 1).

The paper of Sparre Andersen and myself has now appeared and I send you enclosed a copy. As yours, our example is based on non-measurable sets. In order to disprove the existence of conditional probability measures it is, of course, sufficient to work in a product of two sets. An example of this type we found long ago, but it was not until recently that we noticed that by the same idea the extension of the Daniell-Kolmogoroff theorem to abstract sets may be disproved.

With best regards, I am
Very sincerely yours
Børge Jessen

9 Jessen to Doob, 13 September 1948

Dear Professor Doob,

Thank you very much for your letter of June 4. I have postponed the answer until I could send you the paper of Sparre Andersen and myself containing our example. I thank you for the references to Dieudonné (whose paper has just appeared) and von Neumann.

I am looking forward very much to write a joint paper with you on our common results. I except to come to the United States about middle of January and intend to spend the first months at eastern universities. In the second quarter (from March 28 to June 18) I will be in Chicago lecturing, and in the fall (from the middle of September) in Princeton at the Institute for Advanced Study. It will be a great pleasure for me to come to Urbana for a while either before or after the visit in Chicago. Please write me when it would be most convenient for you. (If it is not too hot in June it would perhaps be more convenient to arrange our collaboration after June 18 when I shall not have the lectures to think of.)

Your paper from Trans. Amer. Math. Soc. 1940 I had not received from you, and Sparre Andersen and I were not aware of it when writing our first paper. Your results are, as ours, closely related to my old results on integrals in infinitely many dimensions, though the connection is not so apparent in your exposition. You will notice that in our paper there is a little unsymmetry between the two limit theorems, the first dealing with set-functions which may have a singular part, whereas in the second the set-function is supposed to be continuous with respect to the measure considered. In a note, which will appear in the Danske Vid. Selsk. Mat.-fys. Medd., we give easy generalizations of the two theorems which are completely symmetrical. Here we use the opportunity to quote your paper from 1940. Actually the generalization makes the proofs more conspicuous.

With best wishes, Sincerely yours
Børge Jessen

10 Jessen to Doob, 17 May 1949

Dear Doob[22] :

Just a few lines to thank you and your wife for all your hospitality during my stay in Urbana. I enjoyed very much being with you and talking with you. On the matters we discussed I have had time to think them [?]

I had my promise not to read English detective stories canceled for a night and read one of Steve's novels.[23] Please tell him that it was most exciting as was also Gene Autry,[24] whom I did not know before. It might interest Steve and Peter that we (or rather Hochschild) found a live turtle at one of the creeks in Turkey Run.[25]

11 Jessen to Doob, 23 June 1949

Dear Doob:

I took longer time than I had expected to get the car, but now it is all in order, and with Mr Calderon from Argentina as chief pilot.[26] I expect to leave for Urbana tomorrow about noon (top speed 30 m/h). If the car does not make trouble we shall continue on Monday for New York to meet my wife. As passenger we bring Mr. Nachbin from Brazil,[27] who is also living here. I am ashamed to make the visit such an invasion. I think I did not like to ask one and not the other. He will return to Chicago by train probably on Sunday. If somebody would put him up (I think his main interest is general topology) it would be most welcome but he is prepared to stay in an hotel.

---

[22] This letter and the following one are dated from Chicago. These are incomplete drafts and were doubtless altered at the time of sending. From these letters, it seems that the Jessen-Doob meeting took place in Urbana in the first fortnight of May 1949, without much scientific result, and that the two men met again at the end of June 1949, after Jessen had given his course in Chicago.

[23] This refers no doubt to Doob's oldest son Steve. Burkholder and Protter [9] describe Doob's life in Urbana.

[24] Gene Autry (1907–1998), "the singing cowboy", was a famous American actor and singer.

[25] Turkey Run is a national park in Indiana.

Gerhard Hochschild was born in Berlin in 1915. An important algebraist, he was a student of Chevalley at Princeton. In 1949 he was professor at Urbana, before being appointed to Berkeley.

The draft ends in a badly written, crossed out sentence: "Please give my kind regards to the Cairns, Landen?" Stewart S. Cairns (1904–1982), a student of Marston Morse at Harvard, was mathematics professor at Urbana from 1948 to 1972.

[26] Alberto Calderón (1920–1998), a mathematician of Argentinian origin, was one of the most important analysts of the 20th century. Discovered by A. Zygmund at a conference in Buenos Aires in 1949, he followed Zygmund to the USA, where he spent his whole career. In Chicago, where he was appointed in 1959, he and Zygmund established an important school of analysis. See [10] for a glimpse of his work and his influence.

[27] Leopoldo Nachbin (1922–1993), a brilliant Brazilian mathematician, was a student of Dieudonné and Weil when they taught in São Paulo. He was a professor in Rio de Janeiro and then in Rochester. In 1949 he was a Guggenheim foundation fellow. See [6].

Please do not be unhappy about the proposed collaboration.[28] It has been a great pleasure to discuss the subject with you, if we do not arrive to anything worth while a publication there is the possibility to leave the subject to some …[29]

Acknowledgements We are very grateful to Christian Berg, who obtained these letters for us from the Jessen archives at the Institute for Mathematical Sciences of the University of Copenhagen.

References

1. Ambrose, W.: Some properties of measurable stochastic processes. Trans. Amer. Math. Soc. **47**, pp. 66–79 (1940)
2. Andersen, E.S.: Indhold og Maal i Produktmaengder. Mat. Tidsskrift, B pp. 19–23 (1944)
3. Andersen, E.S., Jessen, B.: Some limits theorems on integrals in an abstract set. D. Kgl. Danske Vidensk Selskab Mat. -fys. Medd. **22**(14) (1946)
4. Andersen, E.S., Jessen, B.: On the introduction of measures in infinite product sets. D. Kgl. Danske Vidensk Selskab Mat. -fys. Medd. **25**(4) (1948)
5. Andersen, E.S., Jessen, B.: Some limit theorems of set-functions. D. Kgl. Danske Vidensk Selskab Mat. -fys. Medd. **25**(5) (1948)
6. Barroso, J.A. (ed.): Aspects of mathematics and its applications, dedicated to Leopoldo Nachbin. North-Holland, Amsterdam (1986)
7. Berg, C.: Børge Jessen, 19.6.1907–20.3.1993. Electronic Journal for History of Probability and Statistics **5**(1) (2009)
8. Bogachev, V.I.: Measure Theory. Springer, New York (2006). 2 vol.
9. Burkholder, D., Protter, P.: Joseph Leo Doob, 1910–2004. Stochastic Processes and their Applications **115**, 1061–1072 (2005)
10. Christ, M., Kenig, C.E., Sadosky, C. (eds.): Harmonic Analysis and Partial Differential Equations: Essays in Honor of Alberto P. Calderón. University of Chicago Press (2001)
11. Cohn, H. (ed.): Doeblin and modern probability. American Mathematical Society (1993)
12. Dieudonné, J.: Sur le théorème de Lebesgue-Nikodym (III). Ann. Univ. Grenoble **23**, 25–53 (1948)
13. Doob, J.L.: Stochastic processes with an integral-valued parameter. Trans. Amer. Math. Soc. **44**, 87–150 (1938)
14. Doob, J.L.: Regularity properties of certain families of chance variables. Trans. Amer. Math. Soc. **47**, 455–486 (1940)
15. Doob, J.L.: Probability as measure. Ann. Math. Stat. **12**, 206–214 (1941)
16. Doob, J.L.: Application of the theory of martingales. In: Actes du Colloque International Le Calcul des Probabilités et ses applications (Lyon, 28 juin au 3 juillet 1948), pp. 23–27. CNRS, Paris (1949)
17. Doob, J.L.: Stochastic Processes. Wiley, New York (1953)
18. Doob, J.L.: William Feller and twentieth century probability. Proceedings of the Sixth Berkeley Symposium on Math. Stat. and Prob. **2**, xv–xx (1972)
19. Doob, J.L.: The development of rigor in mathematical probability, (1900–1950). In: J.P. Pier (ed.) Development of Mathematics 1900–1950, pp. 157–169. Birkhäuser, Basel (1994)
20. Dugac, P.: Jean Dieudonné mathématicien complet. Jacques Gabay, Paris (1995)
21. Feller, W.: An Introduction to Probability Theory and its Applications, Vol. 1. Wiley, New York (1950). 3rd ed. 1968
22. Halmos, P.: I want to be a mathematician, an automathography. Springer (1978)

[28] Here there is a fragment of a crossed out phrase: "If we decide to drop the matter I…"
[29] Thus ended the collaboration of Doob and Jessen.

23. Halmos, P.R.: The decomposition of measures. Duke Math. J. **8**(2), 386–392 (1941)
24. Kakutani, S.: Notes on infinite product measure spaces, i, ii. Proceedings of the Imperial Academy, Japan **19**(3), 148–151, 184–188 (1943)
25. Kantor, J.M.: Mathematics east and west, theory and practice: the example of distributions. Mathematical Intelligencer **26**(1), 39–50 (2004)
26. Locker, B.: Doob at Lyon. Electronic Journal for History of Probability and Statistics **5**(1) (2009)
27. Lützen, J.: The prehistory of the theory of distributions. Springer (1982)
28. von Neumann, J.: Functional Operators, Vol. I: Measures and Integrals. Princeton (1950). Notes on lectures given at the Institute for Advanced Study, Princeton (1933–1934)
29. Neveu, J.: Bases mathématiques du calcul des probabilités. Masson, Paris (1964)
30. Schwartz, L.: Un mathématicien aux prises avec le siècle. Odile Jacob, Paris (1997)
31. Snell, J.M.: A conversation with Joe Doob. Statistical Science **12**(4), 301–311 (1997)

Jean Ville Remembers Martingales

Pierre Crépel

Abstract

In 1984, Pierre Crépel contacted Jean Ville, who had retired from the University of Paris in 1978, to ask him about the sources of his thinking about martingales. This document provides an English translation of their correspondence and a narrative based on Crépel's notes from a face-to-face interview. Ville recounts not only his work on martingales in the 1930s but also his perceptions of mathematical teaching and research in France during the period; his experience with Maurice Fréchet and Émile Borel in Paris and with Karl Menger in Vienna, and his own subsequent mathematical career.

Keywords

History of probability theory · Jean Ville · Martingale

1 Introduction

As a mathematician doing research in the 1970s, I was particularly interested in the connections between random walks on groups and limit theorems for dependent variables. At the beginning of the 1980s, Jean-Luc Verley, sadly deceased in 2007 but then in charge of mathematics for the *Encyclopaedia Universalis*, asked me to write the encyclopedia's article on martingales. One usually begins this kind of article with a short historical introduction, and my inability to do this well pushed me to look

Retired from Université de Lyon. Glenn Shafer of Rutgers University (e-mail: gshafer@business.rutgers.edu) has prepared this translation from the French and has added the footnotes.

P. Crépel (✉)
Université de Lyon, Lyon, France
e-mail: crepel@math.univ-lyon1.fr

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*,
Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_17

more closely at the beginnings of martingale theory, which the mathematical folklore attributed to Joseph Leonard Doob.

In fact, though Doob did indeed develop the concept of a martingale and reorganize whole branches of the probability calculus using it, he never pretended to be the first to introduce the concept or the name into modern, post-Kolmogorov, probability theory. Moreover, his pioneering article of 1940, "Regularity properties of certain families of chance variables,"[1] referred to Jean Ville's thesis, "Étude critique de la notion de collectif," defended and published in 1939.

No doubt the word and/or the thing, or a least closely related ideas, had been around for a long time, at least since the 18th century, in what was called the probability calculus. In the 20th century, certainly, Louis Bachelier, Serge Bernstein, and Paul Lévy had proven interesting results on convergence that we recognize today as theorems about martingales, but it appears that Chap. V of Ville's thesis is the first place where we find all three elements explicit and together: the definition, the name, and an almost sure limit theorem.

Bernstein died in 1968, Lévy in 1971. So when I decided to work on the history of martingales, I was especially interested in getting in touch with Ville and Doob.

Ville had been retired for only a few years, and many people at the Faculty of Sciences in Paris had known him, but he no longer had any relations at all with the probability department, and I was told he was "dead." No one being able to give me a date or show me an obituary, I was skeptical, and I ended up finding him quite alive at Langon, a village in the Loir-et-Cher in the center of France, where he had retired.

I then got in touch with Michel Vacher, a retired physics professor from the University of Rennes whom I knew, who had also retired to Langon. He confirmed that Jean Ville and his wife lived in a hamlet in the district and were known there. So I simply telephoned Ville. He was very friendly and gave me an appointment to see him in the summer. Then on 22 August 1984, I wrote a letter telling him the sort of questions that I wanted to ask, and on the 27th I appeared as he had invited me to do.

I should emphasize that there were no university centers for the history of science in France at the beginning of the 1980s. History of mathematics was practiced, for the most part, either by philosophers or on an amateur basis by mathematicians, who might or might not still be doing mathematics. I had absolutely no training as a historian, which partly explains the unquestionably naive character of my questions.

At the end of 1984, I recorded what I had learned by reading the mathematical literature (mainly works by Bernstein, Lévy, Ville, and Doob), together with miscellaneous information I had gathered on the history of martingales up to 1940, in an article that remained unpublished, part of the "gray literature," namely, "Quelques matériaux pour l'histoire des martingales," Séminaires de probabilités de l'Université de Rennes 1, 1984, 65 pp.[2]

[1] *Transactions of the American Mathematical Society* 47:455–486.
[2] Now published online at Numdam.

I had planned to continue my work in at least two directions: first the connections with analysis and harmonic functions (I had already began this work by corresponding with Doob and talking with Paul-André Meyer), and second the use of martingales and related concepts in mathematical statistics and its applications. This is mentioned at the end of my article. But after 1985, as a good many of the historians of science in France were progressively caught up in the bicentennial of the French Revolution, Roshdi Rashed pushed me to study probability in Condorcet's work. This I did, abandoning the history of martingales.

2 Letter from Crépel to Ville, 22 August 1984

Dear Mr. Ville,

Let me begin by thanking you for agreeing to see me. The time we agreed on, Monday, August 27, around 2:00 pm, is perfectly convenient for me.

I have already explained a bit the purpose of my visit. I am a researcher at the CNRS,[3] a probabilist by education. In 1977 I defended a thesis on random walks with values in locally compact groups. Now I am working on a topic in the history of mathematics: the theory of martingales in the 1930s and 1940s. I have a wide variety of questions, because in my view, the history of science should not neglect any angle of attack. In addition to looking at a topic from inside the particular discipline, we should also look at its scientific, philosophical, socio-economic, and cultural aspects, and so on. Here are a few questions; I am sure our discussion will produce more.

- What did "martingale" mean to probabilists of the 1930s? Were there treatises devoted especially to games of chance? At what point did the word come into mathematical usage?
- What topics were discussed in Karl Menger's colloquium? What was the atmosphere? What questions were considered important? What importance did Vienna have in the world of ideas in general and for probability in particular? Did people like Popper participate?
- It seems to me that Abraham Wald played a rather central role in the evolution of probability and statistics in the 1930s and 1940s. What do you think? Weren't sequential analysis, stopping times, and martingales all connected? To what extent was this clear in people's minds at the time?
- Why was the notion of a collective so important in the 1930s? Why were probabilists writing so much philosophy? How was Kolmogorov's contribution, especially his 1933 book, received and experienced?

[3] The CNRS (Centre National de la Recherche Scientifique) is a branch of the French government. It engages in research in all scientific disciplines, employing tens of thousands of scientists full-time in laboratories spread across France, many of them attached to universities. Crépel was located at the University of Rennes at this time.

- Why and how did the center of theoretical probability shift towards the United States between 1930 and 1950 (emigration, result of the war,...)? Were you well acquainted with Doob, Feller, ...?
- More generally, how would you explain the particularities of the "national schools" (if this expression is appropriate) of probability and mathematical statistics: French, English, American, Russian, German, Austrian...?
- Did people talk about the schools of economists (for example, Keynes, Tinbergen, Koopmans, Morgenstern, Friedman, Wallis...) and their connections with probability and statistics? How and why did operations research evolve in the 1930s and 1940s, and how was this evolution related to probability and statistics?
- In your May 1955 summary of your scientific work, you say this about martingales: "I was calling mathematicians' attention to an idea that existed already, because the word existed, but had not been considered important." Looking backwards, one has the feeling that the need for this notion was felt to a greater or lesser extent in many different branches of analysis, but that mathematicians had not succeeded in bringing it to the surface. I am thinking of Jessen, of Marcinkiewicz.... So how were analysts and probabilists connected at the time (is this even a well posed question)?
- From a philosophical point of view, isn't the emergence of the idea of a martingale in probability theory linked indirectly to a new step in the conception of time?

I will stop here, because I could go on and on; there are so many questions. Impatient to meet you, I offer my best greetings.

Pierre Crépel

3 Crépel's Interview of Ville, 27 August 1984

When I arrived at Jean Ville's home, he was dressed more or less like a gardener, and he received me under a little arbor, in front of a pond where a great number of geese and ducks were splashing about. The homestead was fairly large and disordered. J. Ville told me that he would not ask me to come in the house, where the disorder was even worse and his wife was rather indisposed. He talked with me very straightforwardly, without guile or pomposity. He gave me several documents that he had found for me in his papers, including a report on his articles and work.[4] The conversation lasted about two hours. I took fairly detailed notes, but not knowing shorthand, I was obviously unable to take it all down. The translator and I have composed the following text from these notes. We have put it in the third person to avoid leaving the impression that Ville's exact words are being translated.

[4] Reproduced, in English translation as *Summary of the scientific work of Mr. Jean Ville (May 1955)*, in *Electronic Journal for History of Probability and Statistics* **5**(1), June 2009.

3.1 Mathematics in France in the 1930s

Looking back 50 years, Ville saw France in the 1930s as a black hole for the mathematical topics he found most interesting. There were incredible gaps. Boolean algebra, for example, was almost never taught. Neither probability theory nor set theory were taken seriously. David Hilbert's work in logic was hardly known. Ville learned about existence of propositional functions only from a Romanian student, Barbalat, who wanted to do a thesis on the topic.[5]

There was work on geometry and complex analysis in Paris. Ville did not like complex analysis; he saw it as a bit miraculous but not at all stable. Fréchet was the grand master of topology, of course. But the rigorous exposition of topology was yet to come.

Fréchet wanted Ville to do a thesis on what could be done with two abstract spaces, perhaps a space of sequences and a Hilbert space. Ville found nothing.

To make probability dynamic, you had to discretize time, as in Borel's denumerable probability and Cantelli's strong law of large numbers, topics Fréchet had in the curriculum for the Diplôme d'études supérieures in probability theory.[6] Markov chains were unknown to the Paris faculty.[7] Bachelier was completely unrecognized in France, though Czuber had cited him; his work did not appeal to Fréchet, who took offense at probabilities being treated so lightly.

Fréchet did not believe that there is only a single mathematical foundation for probability.[8] He was interested in von Mises but also enthusiastic about Kolmogorov. He distinguished events that are probabilizable from those that are not.[9] This was

[5] The translator thanks Marius Iosifescu and Solomon Marcus, of the University of Bucharest and the Romanian Academy of Sciences, for information on Ion Barbalat. Barbalat was born on 20 January 1907 in Barlad, Romania. He became a student at the University of Paris in 1926. He and Ville were among the six students who took the examination for the Diplôme d'études supérieures in probability theory in March 1931. He returned to Romania later that year. After completing his military service, he worked for an insurance company and taught at the secondary and university level. His earliest publications, following World War II, were in foundational topics, but he is best known for Barbalat's lemma, about the stability of dynamic systems. He became a full professor in 1963 and finally earned a doctoral degree in Romania in the 1970s. See *George St. Andonie. Istoria matematicii în Romania*, III:274–277, 1967.

[6] Ville earned this degree in 1931, after two years at the Ecole Normale and the University of Paris.

[7] This sentence is not accurate. Fréchet was already enthusiastic about "événements en chaîne" in the early 1930s (see Bernard Bru's "Souvenirs de Bologne", *Journal de la Société Française de Statistique* 144:134–226, 2003). The topic was even treated in the course Ville took from Fréchet.

[8] In a lecture he delivered in 1925 and published in 1955 (*Les mathématiques et le concret*, Presses Universitaires de France, Paris, pp. 1–10), Fréchet argued that geometry and other topics in mathematics, including probability, should be "de-axiomatized".

[9] Fréchet was still using "probabilizable" instead of "measurable" after World War II. See "On two new chapters in the theory of probability," *Mathematics Magazine* 22:1–12, 1948.

interesting only for mathematical technicians. Borel was interested in the imitation of chance—pseudo random events.[10]

The graduates of the École Normale Supérieure were platoon leaders in the infantry in World War I. The school's annual directory shows what massacres befell the classes who had entered in 1912, 1913, 1914, and 1915. It was not as bad later. At first the high command had no idea what they were doing. It took them at least two years to figure it out.[11]

3.2 Vienna and Karl Menger

In 1933 Ville went to Berlin, to pursue his thesis topic in analysis and also to study von Mises's definition of probability. He did not find anyone there working on either topic,[12] so he studied the work of Carathéodory, which he found unpretentious and very clear, with privatdocents. He quarreled by letter with Fréchet, who was angry that he had not produced anything that could be published in the *Comptes rendus* of the Academy of Sciences.

Out of desperation, not knowing where to go after Berlin, Ville went to Vienna. Vienna had great liberty of thought. It was in full decadence in the 1930s.

Karl Menger, who had done dimension theory, was in Vienna. So was Carathéodory. Karl Menger was the son of the great Menger—the distinguished economist Carl Menger who studied marginal value. There were also other famous Mengers.

At one point Karl Menger hosted Ville at a small inn in Kitzbühel,[13] where they talked a good deal. There Menger explained to Ville the main fault of French education: its extraordinary pedanticism. The student is so regimented by the instruction in the last preparatory year that he doesn't dare do anything.[14]

Menger also told Ville about a theorem of Sierpinski's: two unequal balls, say one a millimeter in diameter and the other a kilometer in diameter, can be decomposed into

[10] One of Borel's many notes on this topic appeared as an appendix to the book on games of chance that Ville wrote up from Borel's lectures. Another appeared in 1937 in the *Comptes rendus* 204:203–205.

[11] Ville may be trying to explain the lack of energy in mathematics in Paris in the 1930s. Top students like Ville are often mentored by mathematicians ten to twenty years their senior. But in Ville's time in France, this generation was missing. Borel and Fréchet were more than thirty years older than Ville. See Laurent Mazliak's "The ghosts of the École Normale", *Statistical Science* 30(3):391–412, 2015.

[12] Von Mises fled from Berlin to Istanbul shortly after Hitler came to power early in 1933.

[13] Kitzbühel is over 350 kilometers from Vienna.

[14] Instead of going directly to the university, elite French students continue their secondary education for two more years, to prepare for entrance examinations for schools such as the École Normale Supérieure. The mathematics in the second of these two preparatory years (mathématiques spéciales) is advanced but still taught in the style of secondary education.

a finite number of congruent three-dimensional sets.[15] He interested Ville in formal logic by telling him about Gödel. Menger had studied with Brouwer in Holland. Ville had already read a bit of Lord Russell. Two entire volumes of symbols to show that there are infinitely many prime numbers!

Menger's seminar, the Mathematisches Kolloquium, met once a week, with a dozen people coming on the most crowded days. In the seminar, they called Ville "Mr. Student," or the "King of Counterexamples," because he found counterexamples to the supposed theorems of an American named Blumenthal who was in Vienna at the time.[16]

Tarski was in Vienna, working on Lukasiewicz's three-valued logic. Gödel was also there. Reichenbach was a professor in Istanbul. Menger thought parentheses were useless. He gave a lecture explaining Lukasiewicz's "Polish notation," which eliminates them. One can write the operator in front of its arguments: $\times ab$. In general, it is better to reverse this: $abc \times \times$. All this comes back in machine computation, as Ville realized much later: $A + B$ becomes $\mathrm{S}AB$, $A + B + C$ becomes $\mathrm{SS}ABC$, etc. The seminar also learned about Bergmann's axioms of incidence for n-dimensional geometry.[17]

Wald did a bit of everything in the seminar. He lived with his brother, who repaired radios in the Marienfeldstrasse. He was from Transylvania, which had been part of Hungary before it was invaded by Romania.[18]

Economics probably interested Wald most. Gérard Debreu later had a stunning success with his theory of economic equilibrium, showing how a price system is determined by utility functions as a fixed point, where no one finds any advantage

---

[15] This is proven by Stefen Banach and Alfred Tarski in "Sur le décomposition des ensembles de points en parties respectivement congruentes," *Fundamenta Mathematicae* 6:244–277, 1924.

[16] Ville's first publication, in Menger's *Ergebnisse*, was a note intitled, "Sur une proposition de M. L. M. Blumenthal." Leonard M. Blumenthal (1901–1987) continued to work on distance geometry after he returned to the United States. He was chairman of the Department of Mathematics at the University of Missouri for many years. Ville could not remember Blumenthal's name; he thought it might have been Rosenblum or Rosenblatt.

[17] This sentence is our interpretation of Ville's passing reference to "connecteurs de Rosen…." Gustav Bergmann (1906–1987), a member of the Vienna Circle, later became a professor of philosophy at the University of Iowa. Menger and Alt discussed Bergmann's work in Menger's seminar in January and February of 1935. The French "connecteur" was sometimes used to translate Hilbert's "Verknüpfungsbeziehungen," which names an operation that produces a point from two lines or a line from two points, etc.

[18] Wald's hometown Cluj was in a portion of Transylvania occupied by the allies in World War I and transferred from Hungary to Romania by the Treaty of Trianon in 1920. According to Menger (*Annals of Mathematical Statistics* 23:14–20, 1952), Wald had carried out most of his studies at the elementary and secondary school level under the direction of his older brother Martin, a capable electrical engineer. According to Jacob Wolfowitz (*Annals of Mathematical Statistics* 23:1–13, 1952), Martin had many inventions to his credit. He and most of the rest of the family perished under the Nazis.

in trading. Wald had already done this in a more classical framework. Ville had read the articles on economics in the old mathematical encyclopedia.[19]

Georges Alexits, ten years older than Ville, wanted to construct torsion with respect to Wald's total curvature in dimensions greater than three, but this did not work out.[20]

None of the young people in the seminar had a dime. Ville was able to be there because of an Arconati-Visconti scholarship.[21]

3.3 Random Sequences and Martingales

The idea of a random sequence interested people. Von Mises was one of many who had his own definition.

Karl Popper proposed "nachwirkungsfreie Folgen", sequences with no aftereffect. For him, probability was a sequence of frequencies; he wanted sequences in which 0 had probability $1/2$, while 00, 01, 10, and 11 had probability $1/4$, and so on. The work on this idea did not get very far.[22]

For Wald, a random sequence was one in which the selection of a subsequence does not change the frequency. This led quickly to logic. According to Wald, a selection rule had to be defined by a finite number of symbols, and this depends on the algebraic formalization. Given a particular formalization, Wald could form a collective, making von Mises's theory rigorous. "I can tell Fréchet it is finished", Wald said. "Too bad for me," Ville thought to himself.

Ville spent hours and hours constructing sequences with regular patterns of zeros and ones satisfying Popper's conditions: 00011101, etc. Someone named Posthumus later showed that this could be done using mere arithmetic.[23] Ville obtained a sequence that respected Wald's condition but gives preference to 1, approaching the

[19] The *Encyclopédie des sciences mathématiques pures et appliquées*, published beginning in 1906, was an expanded French version of the earlier German *Encyklopädie der mathematischen Wissenschaften*.

[20] Georges Alexits, "La torsion des espaces distanciés", *Compositio Mathematica* 6:471–477, 1939.

[21] Crépel's notes indicate that Ville said "Asconati." His 1955 report on his work, already cited, also makes this error.

[22] After hearing a semi-technical exposition of Popper's ideas in Moritz Schlick's seminar (the "Vienna Circle"), Menger asked Popper to present them in detail in his seminar, and Wald became greatly interested.

[23] Consider the sequence 00011011, arranged in a circle (the first 0 following the last 1) to form a cycle. Each of the eight possible triplets (000, 001, etc.) occurs exactly once in the cycle. There are exactly 16 cycles of length eight with this property. Around 1944, the well known Dutch engineer Klaus Posthumus (1902–1990) conjectured that in general the number of cycles of length 2^n that have each of the 2^n possible sequences of length n occurring exactly once is $2^{2^{n-1}-n}$. In 1946, N. G. de Bruijn proved the conjecture in "A Combinatorial Problem," *Koninklijke Nederlandse Akademie v. Wetenschappen* 49:758–764, 1946. In 1950, N. M. Korobov showed that the result can be used to construct sequences satisfying Popper's conditions (Izvestiya Akad. Nauk SSSR 14:215–238 and Uspehi Matem. Nauk (N. S.) 5(3)(37):135–137). De Bruijn subsequently discovered that his

frequency $1/2$ from above. This event has probability zero under classical probability theory, and so Wald's conditions are not sufficient to represent the classical theory.

It was at this point that Ville tried the notion of a martingale. He thought "martingale" might be an Italian name. The martingale is the gambling system that tells you to double your bet every time you lose. Or rather, any system that permits the player, if and when he finally wins, to regain all the money previously bet. The word was associated with a classical argument for all gambling systems being illusory. Time would have to be infinite. Governments engage in such illusions nowadays.

A martingale must tell how much of your money to risk and how much to keep on each round, as a function of what has happened before. The expectation should be the same with or without such a system of play. Ville looked for simple martingales that can prove complicated theorems.

Ville found a simple martingale for the game of heads and tails. A player starting with unit capital with this martingale has

$$\frac{\alpha!\beta!}{(\alpha+\beta+1)!}p^{-\alpha}q^{-\beta}$$

after α tails and β heads. This being bounded implies that the frequency of heads tends to p at a logarithmic rate.[24] The people in Vienna liked this, but none of them were professional probabilists.

Schnorr's Lecture Notes no. 218[25] should be read. It talks about Ville's collective. The more complicated a probability law, the longer it takes to describe the martingale that would make it happen. See Kolmogorov.

People did not appreciate sequential analysis until they did quality control.

3.4 Probability Back in France

Fréchet had a hard time understanding Ville's results. Borel said to Ville: When are you going to decide to do analysis?

Paul Lévy did not read his thesis. "I don't read," he told Ville.[26] Aside from his aversion to reading other mathematicians, Lévy was displeased that Ville's thesis had been printed by the *Rendiconti del Circolo Matematico di Palermo*. "You had your thesis printed by the fascists," he objected. "I didn't have any money," Ville responded.

---

result had been obtained by others much earlier. He discussed the history of the problem in 1975 in T.H.-Report 75-WSK-06, Department of Mathematics, Technological University Eindhoven.

[24] See Sect. 3 of Chap. V of Ville's *Étude critique*. The system of play is to risk the fraction $(\alpha+1)/(\alpha+\beta+2)$ of your capital on tails on the next round when you have α tails and β heads so far; the rest on heads.

[25] Claus-Peter Schnorr, *Zufälligkeit und Wahrscheinlichkeit*, Lecture Notes in Mathematics, Vol. 218, Springer, 1971.

[26] Lévy's explained his unwillingness and even inability to read the work of other mathematicians in his autobiography, *Quelques aspects de la pensée d'un mathématicien*, Blanchard, Paris, 1970.

Fellow students included Wolfgang Doeblin, Michel Loève, Félix Rosenfeld, and Robert Fortet. Doeblin was not at all a pedant. In 1938–39,[27] Ville and he began a seminar on probability, which was soon anschlussed by Borel.[28] Loève was from Egypt, from a Russian family who had taken refuge there after the Russian Revolution. Rosenfeld was a statistician. Fortet was excused from military service. He worked on the theory of heat. He thought of probability as a fluid; when an event happens it gives probabilities to other events.

In Poitiers, where Ville taught during the war, there was no probability. Not much was going on. It was much more hierarchical then. People were occupied with their everyday business. There was not much scholarly research in France during the war, except for the Germans.

3.5 Other Aspects of Probability

Conditional probability is essential for writing down a martingale. What is a conditional probability? It is in the rule $P(AB) = P(A)P(B|A)$. A martingale is the quotient of a false conditional probability by the true one:

$$\frac{P(x \text{ if } y)}{P_0(x \text{ if } y)}.$$

It's very simple.

There was no talk about R. A. Fisher in Vienna. We should expect not an outcome with nearly maximum likelihood, as Fisher suggested, but an outcome with likelihood close to average.[29] But the calculations are impossible.

Ville became upset with his colleagues at Lyon in 1947, when he was passed over for a vacant post; he was ranked second, after the person appointed. So he quit. He worked in the electrical industry and threw himself into working on quality control and signal theory at the Société Alsacienne de Constructions Mécaniques.

Shannon information came along. Shannon had not been in contact with Vienna, but his ideas were close to those explored in Menger's seminar. Ville worked on random coding, which is easier than algebraic coding, and on Monte Carlo methods. He studied the problem of removing mines. You have to look at the probability that they are laid and the probability that the adversary passes through. This leads to game theory.

The Wiener lattice and the probability calculus: Write the matrix equation $A^2 = A$ instead of the equation $x^2 = x$, which has only 0 and 1 as solutions.

[27] The seminar actually took place in 1937–38. Doeblin did his military service in 1938–39, and documents in the Doeblin archives at Marbach indicate that Doeblin made presentations in the seminar in 1937.

[28] Here Ville used the invented French verb *anschlusser*. After Hitler's *Anschluss* of Austria in March 1938, the word was used as a verb in French and other languages.

[29] Average in the sense of the geometric mean. Ville spelled this idea out on pp. 92–95 of his "Leçons sur quelques aspects nouveaux de la théorie des probabilités," *Annales de l'Institut Henri Poincaré* 14:5–143, 1954–1955.

The Americans took the lead in operations research. Claude Berge did graphs and networks, but these topics came late to France; things like that were not done in France before the war.[30]

People had long believed that the first bullets in a burst of gunfire are the most dangerous; the end of the burst is wasted. By numbering the bullets, operations researchers found that the opposite is true.

The Americans had a more open conception of mathematics. The success of modern algebra is due to them. The flowering of Bourbaki also owes a lot to them. They published many books on such topics.

French mathematics lost its originality during the war. It was hard just to find a place to live. It was hard to pass between the occupied and unoccupied zones.

Ville knew Doob and Feller only by their books. He had no relation with them or with von Neumann. Feller was both a probabilist and an analyst. He was interested in properties of independent normal random variables.

3.6 Economics

Von Neumann was the great man in economics. It gave Borel fits not to have proven the fundamental minimax theorem.[31]

Ville had never taught probability in his life before doing so in econometrics.[32] In econometrics, he studied matrices of positive numbers—weighted sum of permutation matrices. He was also interested in preference orderings. Price theory can be substituted for value theory. Dual prices are important.

An Arab economist wrote a book on prices in the Soviet Union, based on the mathematics of the Russian school. Ville wrote an article about this around 1976 in the newspaper *Le Monde*. But people told him: The Russians only pretend to do mathematical economics.

Leontief's technology matrices are still not properly understood.

---

[30] Berge's doctoral thesis, "Sur une théorie ensembliste des jeux alternatifs," *J. Math. pures et appl.* 32:129–184, 1953, was cited by Ville in the article cited in the preceding footnote.

[31] On Borel's relation with von Neumann, see Mazliak, L.: The games of Borel and chance. Some comments on Borel's role in the theory of games. In: M. Voorneveld, al (eds.) One Hundred Years of Game Theory: A Nobel Symposium. Cambridge University Press (2022).

[32] This statement overlooks his having taught probability as a prisoner of war in the second world war.

3.7 Computing at the University of Paris

The French computer company Bull, under Philippe Dreyfus, loaned a computer to the University of Paris. Georges Darmois said to himself, perhaps we should get that franc-tireur Ville to work on it.[33] They were infernal, those computers with lamps.

Ville knew the pioneers of computer science in France. The Société Alsacienne de Constructions Mécaniques (SACM) was located on the street named for Admiral Mouchez.

How do you decide whether to reject a cable? You are playing a game against the computer's breakdown.

Ville also worked with the Compagnie Générale d'Électricité (CGE), which merged with part of SACM and evolved into Alcatel. He did not know what exactly Alcatel had become.

4 Letter from Crépel to Ville, 21 January 1985

Dear Mr. Ville,

To begin, I send my best wishes for 1985, wishing good health for you and all those dear to you.

For the past few months I have been working on the history of martingales before the war, and I have completed writing up a presentation I made on the subject for the probability seminar at Rennes. Before having the secretary type it, I am sending you the part that concerns your work. If you find errors or omissions, don't hesitate to let me know, so that I can correct them. Of course I will also send you the whole thing once it is typed.

While thinking about the topic, I have thought of further questions to ask you:

- What effect did Kolmogorov's book, the *Grundbegriffe*, have on people in Vienna in 1934–35, and on Wald, Popper, and yourself in particular? Was it discussed in Karl Menger's seminar, and if so how?
- At the 1937 Geneva a colloquium (at least in the written version of his remarks, dated 1938), well before your thesis appeared, Maurice Fréchet explained its results in some detail. I conjecture that you were not in Geneva in 1937. Were you invited? Do you know whether Fréchet actually talked about your work there

[33] For ten years, beginning after he left his position at Lyon in 1947, Ville worked largely as a consultant. The term "franc-tireur" appears in English as well as French dictionaries, but it is no longer common in English. A literal translation is "free shooter." The term was used in the Franco-Prussian War of 1870, when French soldiers acting as snipers operated without being attached to specific military units. It was revived in World War II as a name for partisans who carried out sabotage and other acts of terrorism against the German occupiers.

(or did so only in his written contribution), and if so, what was the reaction of the participants?

- A related question: What was your relationship with Wald at Vienna and after you returned to France? Reading the summary of the Geneva colloquium, I got the impression that he did not agree with your objection to von Mises.[34]
- When did you notice the connections between your ideas and those of Paul Lévy on "chains of variables"? Were you immediately aware of the notes he published in the *Comptes rendus* of the Academy of Sciences while you were in Vienna (volume 199 in 1934 and volume 201 in 1935)?
- I found a 1936 article by Doob entitled "Note on probability" (*Annals of Mathematics* 37, 1936, pp. 363–367), where he demonstrated mathematically the impossibility of a gambling system (one that merely chooses trials on which to bet, without varying the stakes) in the case of independent trials. Did you read this article, and what did you think of it? The same question for Paul Halmos's "Invariants of certain stochastic transformations in the mathematical theory of gambling systems" (*Duke Mathematical Journal* 5, 1939, pp. 461–478), which extends the demonstration to the martingale case.
- Finally, I noticed that Doob and yourself each wrote an article for the colloquium at Lyon (28 June–3 July 1948) on the probability calculus and its applications. Were you actually there together, and did you encounter each other?

Of course I will also be grateful for any other information you think might be useful to me.

I offer my best respects and renew my wishes for a good new year.

Pierre Crépel

5 Letter from Ville to Crépel, 2 February 1985

There were three notes in an envelope postmarked on this date.

5.1 First Note

I am late responding to your letter. Like many others, we were taken by surprise by the extreme cold. Then it got worse. A large portion of the water pipes here gave out, and an equally large portion of the electric lines. There will be a big bill.

[34] The summary, written by Bruno de Finetti (*Compte rendu critique du colloque de Genève sur la théorie des probabilités*, Hermann, Paris, 1939), quotes Wald's response to various criticisms, including Fréchet's report that his student Ville had demonstrated that not all properties with probability one can be represented by subsequence selection rules. Without specifically mentioning Ville or martingales, Wald reiterated that his approach allows the construction of a sequence satisfying any countable set of properties of probability one (p. 15). This is true whether the properties are specified by subsequence selection rules, by martingales, or in some other way.

Catastrophes of this sort, and a good many other kinds of bad luck (such as these interminable lawsuits, rotten from the outset because of the state of our legal system), are never taken into account when pension payments are set.

More exactly, pensions are set so that someone else has to take responsibility for these things. This is just what was explained to me by Mr. Vessiot, director at the École Normale Supérieure when I was there.

5.2 Second Note

Dear Mr. Crépel,

Thank you for the documents you sent. I attach a more detailed letter, to avoid prolonging this one indefinitely.

You make me remember a time when we pondered the meaning of probability, sensing its impending importance.

Fréchet favored: "Probability = physical quantity, measured empirically by the frequency of outcomes in a sequence of trials." P. Lévy favored a more ethereal definition. The students did not have a preconceived opinion. Fréchet would have been persuaded by von Mises, but like everyone raised a trivial objection against him. One should not consider *all* selection rules. This is where the disagreement was in the end. Logic was beginning to be fashionable, and "Bourbakism" had been launched. People were fussing about the meaning of "it exists." The logicians thought they were addressing the problem by writing ∃. People talked about "constructivity." Algorithms were not popular. But this is really the idea we have in our heads, as developed by Claus-Peter Schnorr (*Zufälligkeit und Wahrscheinlichkeit*. Lecture Notes in Mathematics. Edited by A. Dold, Heidelberg, and B. Eckmann, Zürich, Springer[35]). We sensed the germination of today's theories based on algorithms. But this was all muddled for people nourished on Cauchy, Riemann, and Poincaré.

So the hiatus does not surprise me.

At that time I could not manage to make people understand what I wanted to say. Fréchet was openly doubtful; he didn't like the topic. I only managed to move on with the help of Émile Borel. Borel told Fréchet that it was time to finish with the thing and had it published.

At the time, I had given a proof of the minimax theorem for a zero-sum matrix game with two players, showing that it was only a question of convexity. Fréchet did not believe it. It was only accepted when J. von Neumann acknowledged it in *Econometrica*. (Just as Doob's acknowledgement gave me credit for martingales.)

Much later I worked on stochastic processes, using the Wiener-Lévy process. I had already been aware of it, completely by chance, from Bachelier. Now of course all this is so well known that it seems like part of the nursery school curriculum.

Renewing all my thanks to you, I ask you, dear Mr. Crépel, to accept the affirmation of my best wishes.

Jean Ville

[35] 1971.

5.3 Third Note

I and Mrs. Ville thank you for your good wishes. Your care in sifting through the history of the theory of martingales is meritorious.

Martingales were not the subject of any systematic study in France. People were mainly interested in showing that they were ineffective. But I had an acquaintance, a relative of the woman who became my wife, who claimed to make a (modest) living by gambling, which he pursued like a drudge, "working" for hours recording and counting the outcomes of boule or roulette spins, and then betting according to a calculation that he kept secret. His name was Mr. Parcot. I claimed that it was impossible for him to win. The probability calculus showed that for simple martingales, everything ended up a loss. Because the calculus was applied to martingales one by one, the layman was left with the impression that one could find a crack in the armor and slip though. Mr. Parcot claimed to have found a crack. I did not try to convince him; I don't even remember now if I had an opportunity to do so. I knew that there was a general refutation of the possibility of winning for sure, but I went no farther. Mr. Parcot's continuing profits simply made me think. Why not? I knew that a certain role was played by confusions between infinitely small and zero, and between actual and virtual infinity, nothing more. I did not doubt Mr. Parcot's good faith, and there was something that pointed out a path.

You mentioned it in your letter, by the way. I studied the probability calculus in Laplace, and I found there a way to win in heads and tails if you know the coin is asymmetric without knowing which side is favored. From this, I concluded that Parcot had perhaps discovered and taken advantage of a flaw in the roulette wheel. Taking advantage of a *known* flaw in a roulette wheel is child's play, but taking advantage of the fact, for example, that the spins are *not* independent, without knowing *exactly* how they are dependent, is another matter. So this is where I was, say in 1932.

There was a lot of talk about the foundations of probability. There was a fight between the supporters of probability on a single trial (a sequence of trials being the combination of trials whose individual nature is known and which we *join* together) and the supporters of frequency (there is no probability without repetition, each individual trial being "extracted" from the sequence). Then there were positions in between.

People mentioned Kolmogorov, nothing more. Admittedly, the *Grundbegriffe* says nothing about probability.[36] Fréchet spoke about "probabilized" events. As for me, I tried to apply myself to general topology, which was experiencing a painful birth.

In Vienna, people talked about the *Grundbegriffe* no more than in Paris.

I do not remember talking about my thesis, or hearing it talked about, before defending it.[37] The thesis was ready in 1936, but I had difficulties getting it printed

[36] Nothing, that is to say, about the meaning of the word.

[37] In the introduction to the thesis, however, Ville thanks Borel for the opportunity to present his ideas in the seminar on probability theory that Borel was conducting at the Institut Henri Poincaré, presumably the same seminar that he and Doeblin had begun.

and worked on other subjects. So I cannot tell you anything about the Geneva colloquium in 1937.[38]

I did not have any personal contact with Wald after returning to France. If he did not agree with my objection, my best recollection would be that it was because he thought von Mises's axioms, as refined by himself, were more than sufficient in practice. In any case, I was not working regularly on probability. I was occupied with electricity, signal theory, Shannon's information theory, coding and the detection of errors, etc.

Coming back to martingales, I did not hear Paul Lévy's work mentioned in this connection until I was giving a course on signal theory at Toulon,[39] where one of my listeners spoke to me about Lévy and martingales. He told me Lévy had invented them. I corrected him, explaining that Lévy had worked on dependent variables. In fact, I had talked with Lévy. I had defended the principle of compound probability, $\Pr\{a\&b\} = \Pr\{a\} \times \Pr\{b \text{ if } a\}$, while he defended the definition of conditional probability, which defines this probability starting with $\Pr\{a\}$ and $\Pr\{a\&b\}$ (Fubini and company). Of course, if one talks about martingales, it is the first of the two approaches that matters.

Going back even farther, I may point out that probabilities were very classically linked to games of chance. Think about B. de Finetti, who *defined* probability as the *inverse* of the payment, if the event happens, for staking 1 on the event now.

The article by Doob in 1936 did not hold my attention. The impossibility of winning by choosing trials was far too close to von Mises's axiom of selection. As for Halmos's article, I did not read it. The year 1939 was very much a year of crisis. By the end of 1938, I had already been mobilized for several months. I had left my scholarship to take a job teaching the last year of preparatory mathematics at Nantes in 1938. I realized that the defense of my thesis was taking forever to happen, and I had decided to take a regular position. Fréchet did not approve: "A future member of the higher education profession does not go into secondary teaching."

Again about Doob. By 1948, because of the attitude of the Lyon faculty towards me, I had resigned from my position there for personal reasons. I did not receive any invitation from them; they would not have had the nerve.[40] This is why I never met Doob. Some months later, someone at the Paris faculty called the matter to my attention, saying, "An American came to give a lecture at Lyon."

Again on the topic of the impossibility of a winning martingale, I still insist on a point that appears obscure to me on your page V 9: "One will notice the reversal in viewpoint."[41] What is important is not to give a name to the martingale, but to say

[38] In a letter to Fréchet at the time, Doeblin said that Ville would remain in Paris instead of attending the Geneva colloquium because he was busy finishing Borel's book on games of chance. But Ville cited the published proceedings of the colloquium in his thesis.

[39] Toulon is a major base for the French navy, and Ville consulted for the navy after World War II.

[40] Ville did contribute an article to the proceedings of the Lyon conference. Fréchet was the editor.

[41] Ville is quoting the draft report on the history of martingales Crépel had sent him. In the sentence quoted, Crépel contrasts Ville's definition of a martingale $s_1, s_2, \ldots$ (p. 99 of his book) with Doob's later definition of a martingale $Z_1, Z_2, \ldots$. Whereas s_n is a function of random variables $X_1, \ldots, X_n$, Z_n is a random variable measurable with respect to the σ-algebra $\mathscr{F}_n$.

how it is defined from what is given. For me, the X_n are given, and we are to find $s_1(X_1)$, $s_2(X_1, X_2)$, etc. Nowadays one takes as given the $\mathscr{F}_n$ and the Z_n. There has been an emphasis on the fact that Z_n cannot become infinite. My goal, given the X_n and a set to which the sequence of X_n belongs with probability zero, was to define s_n so that they tend to infinity on that set. I insist on the point because it took me so long to make this way of proceeding understood.

Seven Letters from Paul Lévy to Maurice Fréchet

Laurent Mazliak

Abstract

This chapter reproduces and comments on recently rediscovered letters from Paul Lévy to Maurice Fréchet. These letters complement letters from Lévy to Fréchet previously published. They cover many topics, but in particular they extend what we know about Lévy's vision of martingales and his negative view of Jean Ville.

Keywords

Paul Lévy · Maurice Fréchet

1 Introduction

In 2019, during a reorganization of the departments of statistics and probability in Sorbonne University, Paris, France, some forgotten archival documents were rediscovered, including a set of letters sent by Lévy to Fréchet overlooked when the original French version of [1] was prepared in 2003. They were in a file along with a collection of lecture notes by Fréchet and others and had been probably collected at the end of the 1950s when the just retired Fréchet left his office at the Institut Henri Poincaré (IHP). Following an interesting Brownian path, they accompanied the various motions of the mathematical departments for 50 years: first from the IHP to Jussieu campus when the new site of the faculty of science was opened after 1968; then to the temporary address rue du Chevaleret in 1999 when asbestos began to

L. Mazliak (✉)
Sorbonne Université, LPSM, Paris, France
e-mail: laurent.mazliak@upmc.fr

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*, Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_18

be removed from the Jussieu campus, then back to Jussieu when the mathematical departments came back in 2010, and finally to its present location inside the newly created *Laboratoire de Probabilités, Statistique et Modélisation*!

As mentioned in Laurent Mazliak's chapter in the present volume, several of the letters deal with the relation between Lévy and Ville. They deal with other topics as well. Here we present the whole collection, so that they will be in a single place as a complement to the book [1]. The numbering for the letters corresponds to the numbering used there. As the first two letters of the set belong to a series of four from March 1931, we also reproduce letters 26 and 27 from [1].

Letter 26

Paris, 16/3/31–38 rue Théophile Gautier

Mon cher Collègue,

Je vais passer à la Sorbonne déposer quelques tirages à part,[1] et je joins un mot comme suite à notre conversation téléphonique.

La probabilité de convergence de $\sum x_n$, les x_n étant indépendants, est toujours 0 ou 1 et je donne dans une Note[2] présentée aujourd'hui les conditions nécessaires et suffisantes pour qu'on soit dans l'un ou l'autre cas.

Je viens de m'apercevoir qu'en outre, dans le cas de convergence, bien qu'il y ait en général semi-convergence, la loi de probabilité dont dépend la somme est indépendante de l'ordre des termes. On en déduit aisément que

$1 \pm \frac{1}{2} \pm \frac{1}{3} \cdots \pm \frac{1}{n} \pm \ldots$ en tenant compte de $1 \pm \frac{1}{2} \pm \frac{1}{4} \pm \cdots \sim 4x$

et $4(x_1 + \frac{x_2}{2} + \cdots + \frac{x_n}{n} \pm \ldots)$

(x_n étant une variable choisie au hasard entre $-\frac{1}{2}$ et $+\frac{1}{2}$, avec la densité de probabilité unité) dépendent de la même loi de probabilité. Il y a d'autres cas particuliers amusants à étudier.

La série $\sum \pm \frac{1}{n}$ a déjà été étudiée par Norbert Wiener.[3] Je vois qu'à ce sujet il cite Steinhaus, je vais rechercher dans le mémoire cité.

---

[1] Probably Lévy, Paul. 1931. Sur le gain maximum au cours d'une partie de pile ou face. CRAS.192:258-259. At the same time, Lévy published another article on this topic: Lévy, Paul. 1931. Nuove formule relative al giuoco di testa e croce. Giornale Istituto Italiano degli Attuari. 6: 3–36. A shortened version appeared in 1931. Journal de l'Ecole Polytechnique. 3–23.

[2] Lévy, Paul. 1931. Quelques théorèmes sur les probabilités dénombrables. CRAS. 192:658–659, which Lévy presented the same day at the Academy. As the next letter shows, Lévy was unaware of the results of Khinchin and Kolmogorov, despite their having been published five years before in their unique joint paper [3] in which they study the conditions for convergence of series of independent variables.

[3] In 1922, at their first meeting, Wiener showed Lévy some results about this series.

Il est entendu que si, après avoir lu ma Note, vous trouvez une priorité à me signaler, je vous en serai reconnaissant.

Bien cordialement

P.Lévy.

Letter 27

Paris 38 rue Théophile Gautier 18/3/31

Mon cher Collègue,

Je vous remercie bien vivement. Kolmogoroff m'avait laissé 5 ou 6 mémoires que je n'avais pas encore pu étudier ; j'ai pu au reçu de votre lettre, hier soir, me reporter à la source citée, constater qu'il s'agissait bien de mon théorème, qu'il était attribué à Khintchine et Kolmogoroff, (Moscou 1925), et ce matin citer cette référence en addition à ma Note. Il était temps !

J'avais lundi précisément trouvé le Mémoire de Steinhaus dans Studia Math.[4] ; il traite un cas très particulier du problème en question, et cela m'avait encore plus convaincu que la solution générale n'était pas connue. Ainsi ai-je été très surpris en recevant votre lettre.

Dans ma Note, qui va donc paraître avec l'addition faite ce matin, le seul résultat important, qui me reste et que je vous signale, est le théorème IV.

J'ai écrit à Steinhaus que le résultat annoncé comme probable à la fin de son dernier Mémoire de Studia est faux.[5] Il faut enlever le 2 sous le radical. Dans le cas particulier, l'hypoténuse n'est pas plus grande que les deux composantes, les grandes valeurs des deux composantes n'étant pas réalisées simultanément. Cela est d'ailleurs bien évident, car si vous divisez le cercle en p parties (p arbitrairement grand mais fixe) les grandes valeurs de $\sum a_n e^{i\varphi_n}$ ont autant de chances d'être réalisées avec un argument qu'avec un autre ; on peut donc supposer leur argument compris entre $-\frac{\pi}{p}$ et $+\frac{\pi}{p}$, et, avec une erreur relative très petite, assimiler cette somme à sa partie réelle.

Je réponds maintenant à vos objections sur mon Calcul des Probabilités.

1° Je crois que vous avez été troublé par le mot variable; $\lambda(x)$ est en l'espèce quelque chose d'aussi bien déterminé que $\sin x$, et quand je définis la valeur probable par $\sum \alpha_i \lambda\{x_i\}$, α_i est la probabilité de $x = x_i$; elle ne change pas si on l'ajoute à d'autres termes .

Je crois que c'est de Finetti ou en tout cas un italien qui a fait observer que la valeur probable n'est pas exactement la même chose si l'on étudie la valeur probable de λ en f[onction] de x ou en partant de la loi de probabilité de λ; mais la distinction

---

[4] In his article *Les probabilités dénombrables et leur rapport à la théorie de la mesure* (Fundamenta Mathematicae, 4, 1923. 286–310), Hugo Steinhaus studies (p. 295) the convergence of an elementary random series as a consequence of the Rademacher theorem on orthogonal functions.

[5] Lévy's observations concerns Steinhaus' paper *Sur la probabilité de la convergence de séries (Première communication)* (Studia Mathematica, 2, 1930. 21–39). On the last page (39), the author gives a generalization of Khinchin's iterated logarithm law for head and tails. In fact, as seen in the next letter, Lévy was wrong as he misunderstood Steinhaus' formulation which was exact.

n'intervient que dans des cas exceptionnels. Mon énoncé général, ne précisant pas les hypothèses faites sur $\lambda(x)$, est sans doute critiquable ; mais l'application au cas où λ est une fonction à variation bornée ne l'est pas.

Vous demandez ensuite si "la probabilité de E, pour x donné, étant α , il en est de même si l'on ne sait rien sur x."

Evidemment oui; par application des axiomes fondamentaux, et je vois bien ce qui vous arrête, à moins que ce ne soit le souvenir de difficultés rencontrées dans des cas analogues, mais non identiques ; je puis vous en citer un exemple que j'ai rencontré récemment.

Soit x_n le gain après n coups de pile ou face, y_n le plus grand des nombres $x_1, x_2, \ldots, x_n$. Il s'agissait d'avoir la loi de probabilités de y_n, connaissant x_n. Si l'on sait que $y_n = x_\nu$, ν étant connu, le problème était facile ; mais était-il correct d'affirmer que $\mathscr{P}\{y_n > N\}$ est inférieur par exemple à la plus grande des probabilités calculées en faisant successivement $\nu = 1, 2, \ldots, n$. Evidemment non, parce que pour certaines parties le maximum y_n est atteint pour plusieurs valeurs de ν , et dans le compte des cas possibles ces cas sont comptés plusieurs fois. Toute difficulté disparaît si je précise que ν est le plus petit entier pour lequel $y_n = x_\nu$; alors chaque cas est bien compté une fois et une seule.

Je suis persuadé que vous arriverez à cette conclusion qu'il n'y pas de difficulté, moyennant des hypothèses sur $\lambda(x)$, que j'aurais dû préciser , mais qui sont bien vérifiées dans l'application que j'avais en vue.

Quant à l'exemple de la sphère, c'est un exemple du type classique où la probabilité n'est pas bien définie. Mais dans la composition des probabilités indépendantes, si x et y ont des lois déterminées, il en est de même dans le plan dont dépend le point x, y, et par suite de celle dont dépend $x + y$; la déterminer c'est de l'analyse pure.

2° Il faut être timide avec le transfini, sans doute. Toutefois en reprenant le raisonnement de la page 330, il me semble correct. Il faudrait que je voie le Théorème de Vitali dont vous me parlez. Si vous voulez poursuivre cette discussion (pour ma part je serai content de tirer la chose au clair) voulez-vous m'indiquer la référence.[6]

Pour le 1°, si vous n'êtes pas convaincu, je crois qu'il vaudrait mieux en parler de vive-voix.

- Pour vous prouver que je ne suis pas infaillible, et à toutes fins utiles, je vous signale quelques errata

[6] The last section of [4] (which includes page 330) is entitled *Note sur les lois de probabilités dans les espaces abstraits* and is a reproduction of a text published in Revue de Métaphysique et de Morale in 1929. Lévy suggested extending the Lebesgue measure to all the subsets of [0,1]. He noted that an application of Zermelo's theorem allows the assignment of an arbitrary value to the non-measurable sets but that this is of no practical interest. In the second (1957) edition of [5], he mentions on page 370 that the application of Zermelo's theorem was in fact not justified in this case, as had been pointed out to him by Steinhaus. Fréchet no doubt had his reservations about this abstruse remark by Lévy. Fréchet probably mentioned Vitali's construction of a non-measurable set of the real line as contradictory to Lévy's hopes. In letter 27c below, Lévy acknowledged that his reasoning was erratic.

Calcul des Probabilités - p. 202. Tout ce passage est à revoir; j'ai perdu de vue en l'écrivant que $F(x)$, fonction monotone croissante de 0 à 1, a une limite croissant de $x_0 \geq 0$ à $x_1 \leq 1$ mais qu'on n'a pas le droit d'affirmer que $x_0 = 0$ et $x_1 = 1$.[7]

Calcul des Probabilités- p. 161, lignes 1 à 5. C'est faux, cela devient correct en remplaçant l'exposant 2 dans I_n par l'exposant n (résultat qui m'a été communiqué par [??][8])

Sulla legge forte dei grandi numeri , p.7 du tirage à part, remarque 2°. Le résultat énoncé est exact mais le raisonnement défectueux[9] ; même remarque pour le bas de la p.18 du même mémoire.

Bien cordialement

P. Lévy.

Letter 27b

Paris, le 24 mars 1931
Mon cher Collègue,

Je me trouve avoir chez moi le t.IV des Fundamenta, et un coup d'œil rapide m'a fait tomber à la p.30 sur le Th. I de Banach, qui coïncide exactement avec ce que je dis dans mon livre, p. 330, l. 22–24 "on peut au contraire s'arranger pour vérifier ce nouveau principe en abandonnant le principe b".[10] Cette remarque ne répond pas d'ailleurs à votre question; mais je commence à me dire que j'aurai du mal à rédiger mes travaux en cours avant la période d'examens, et si le travail de Vitali ne me paraît pas dès l'abord très clair, j'en remettrai l'étude à plus tard.

Pour votre autre question, je crois que j'ai maintenant bien compris ce que vous voulez dire. Il faudrait en effet remanier assez sérieusement la p.188, mais la formule (56) est bien exacte.

Au point de vue axiomatique, on peut d'abord se demander s'il n'y a pas lieu de compléter l'axiome b de la p.329 par un nouvel axiome, qui serait un principe d'addition pour une infinité non dénombrable d'éléments. Cela ne me paraît pas possible, vu que dans le cas d'une loi continue à une variable, un tel principe devrait

---

[7] As is well known, there is a possibility of evanescence of the total mass, and a condition of conservation of the total mass 1 must be added to the weak convergence conditions of the probability measures tu guarantee the convergence in distribution. This is another instance of the kind of simplification that was brought later to the theory through topological properties on measures.

[8] The name is badly written.

[9] This remark and the following seem to indicate that the offprints in question at the start of letter 1 are from Lévy, Paul. 1931. Sulla legge forte dei grandi numeri. Giornale Istituto Italiano degli Attuari. 6: 3–23.

[10] Lévy refers to Banach's paper *Sur le problème de la mesure* (Fundamenta Mathematicae. 4, 1923. 7–33). However, Lévy made a mistake : on page 30, Banach states the existence of a finitely additive measure extending Lebesgue measure on any subset, but not a real (σ-additive) measure. See the next letter. About Banach's works in Fundamenta Mathematicae, see also [2].

donner $\sum 0 = 1$. Il faut donc s'inspirer de cette idée (indiquée au bas de la page 330) que, à l'opposé des idées de Kronecker, le point n'est pas autre chose que la limite d'un intervalle très petit. Une loi de probabilité à une variable ne peut donc être donnée que si l'on donne la probabilité liée à tout intervalle très petit, et le point de probabilité positive n'est qu'un concept mathématique résultant d'un passage à la limite.

La valeur probable $\int \lambda(x)dF(x)$ est alors bien définie si $\lambda(x)$ est continu, ou dans certains cas un peu plus généraux (par ex. points de discontinuité de 1ère espèce ne coïncidant pas avec ceux de $F(x)$). Mais elle n'a pas de sens pour une fonction absolument quelconque; ainsi, si $\lambda(x)$ n'est pas mesurable, et si $F(x)$ est absolument continu, $\int \lambda(x)dF(x)$ est a priori dépourvu de signification (peut-être pourrait-on lui en donner par des conventions arbitraires, conformément à la remarque finale de la p. 330)

Si maintenant je reviens au problème de la composition des probabilités indépendantes, chaque petit rectangle $dxdy$ a une probabilité déterminée; cela sert de définition à une fonctionnelle linéaire, et ce n'est qu'un problème d'analyse pure de chercher la valeur de cette fonctionnelle pour $z < x + y \leq z + dz$. Il ne peut y avoir de difficulté au point de vue des axiomes; mais je reconnais que mon exposé laisse à désirer.

J'ajoute, ayant précisément besoin en ce moment de cette extension, que le principe de la dispersion croissante des masses représentant les lois de probabilité s'étend au cas de probabilités non indépendantes. En supposant $G_x(y)$ continue en x, on peut écrire

$$H(z_1) - H(z_0) = \int_{-\infty}^{+\infty} [G_x(z_1 - x) - G_x(z_0 - x)]dF(x),$$

et par suite, si $G_x(z_1) - G_x(z_0) \leq \varphi(\ell)$ pour $z_1 - z_0 = \ell$, x quelconque on a a fortiori

$H(z_1) - H(z_0) \leq \varphi(\ell)$, (et en général $<$).

Bien cordialement,

Paul Lévy

Letter 27c

Paris - 38 r. Th. Gautier - 28/3/31

Mon cher collègue,

Vous aviez parfaitement raison de mettre en doute mon énoncé de la p. 330. Steinhaus vient de m'envoyer deux mémoires de Fundamenta, celui de Banach et

Kuratowski, t.XIV,[11] et un plus récent, de Ulam,[12] qui démontre explicitement le contraire de ce que je pensais. Un coup d'œil sur le premier m'avait d'ailleurs tout de suite fait voir le point qui m'avait échappé.

Peut-être y a-t-il quelque intérêt à ce que je précise mon raisonnement et la faute que j'avais faite.

Mon principe de choix successifs impliquait 2 axiomes:

Axiome 1: Si jusqu'à un certain moment on a pu faire des choix exempts de contradiction, cela est possible pour le choix suivant.

Axiome 2 : Si l'on fait indéfiniment et transfiniment des choix exempts de contradiction, aucune contradiction n'apparaît à la limite.

J'avais bien vu la nécessité de l'axiome 1, et je crois bien qu'il est exact. Mais je ne m'étais pas du tout aperçu de celle de l'axiome 2. Or il est manifestement faux, dans le cas d'ensembles intérieurs les uns aux autres (chacun intérieur au précédent) et sans partie commune, si on leur attribue des mesures décroissantes et ne tendant pas vers zéro.

Le principe de Zermelo n'est donc pas en cause. Je m'étais trompé, non dans le transfini, mais déjà dans le dénombrable, ce qui est plus humiliant.

J'en profite pour rectifier 2 points d'une lettre antérieure:

1° - J'avais cru à une erreur dans mon mémoire italien, p.7 des tirages à part; le raisonnement en question est bien exact.

2° - A la fin du mémoire de Steinhaus, dans Studia t.II, j'avais lu = 1 quand il avait écrit = const. Comme en réalité c'est $= \frac{1}{\sqrt{2}}$, j'ai eu tort de dire qu'il fallait répondre négativement à sa question.

Bien cordialement

Paul Lévy

Letter 99b

Paris - 38 Av. Théophile Gautier

Mon cher collègue,

Je vous envoie séparément 3 notices sur mes travaux; une de 1951; une de septembre 1963 rédigée au moment où (sur le conseil de Léauté)[13] j'avais pensé à me présenter à la succession de Ramon[14] ; je pense qu'elle aurait suffi pour ma candidature; mais en voyant que j'avais le temps, je me suis dit qu'il pouvait y avoir intérêt

[11] Lévy refers to Stefan Banach and Kazimierz Kuratowski's paper *Sur une généralisation du problème de la mesure* (Fundamenta Mathematicae, 14. 1929. 127–131) in which both authors prove the impossibility of defining a σ-additive measure to any subset of $\mathbb{R}$ with the condition that any single point has measure 0.

[12] Stanisław Ulam. Zur Masstheorie in der allgemeinen Mengenlehre. Fundamenta Mathematicae, 16, 1930. 140–150.

[13] Pierre Léauté (1882–1966) was professor of physics and Lévy's colleague at the École Polytechnique between 1936 and 1952, and a member of the Academy of Sciences. Ho worked on the improvement of various electronic devices, in particular for navigation control.

à rédiger un exposé d'ensemble de mes travaux sur les probabilités, et je viens de la recevoir.

Je ne sais pas si ce que j'ai dit des mes autres travaux dans ma notice de 1951 ou à la fin de celle de 1964 vous suffira. Sinon, et si vous ne retrouvez pas ma notice de 1935, je vous en prêterai un autre exemplaire. L'élection n'étant pas imminente (je ne sais pas si ce sera en mars ou en avril), nous nous reverrons d'ici là.

Bien cordialement,

Paul Lévy

Letter 99c

Paris - 38 Av. Théophile Gautier

19-1-1964

Mon cher collègue,

Je suis un peu embarrassé par votre lettre. Je vous envoie une liste dans laquelle j'ai marqué surtout 1° les travaux qui ont été le plus nettement approuvés par M.Hadamard (notamment les n$^{\circ s}$ 1 et 9) 2° ceux qui ont été à l'origine du plus grand nombre de travaux d'autres savants notamment les n$^{\circ s}$4 à 8 de ma liste 3° mes premiers travaux sur le calcul des probabilités qui ont été un achèvement de travaux antérieurs et ont par suite suscité peu de travaux nouveaux, sauf en ce qui concerne les lois stables étudiées par plusieurs savants depuis mes travaux de 1922 et 1934.

Vous pourriez peut-être dire qu'en dehors de ces travaux, qui forment chacun un ensemble important, j'ai publié un grand nombre de travaux ayant un caractère différent; ce sont des théorèmes isolés dans mon œuvre, quoique certains aient suscité des prolongements d'autres savants. Ainsi

le théorème cité de géométrie a été suivi par un travail de Hopf (Heinz Hopf),

Un théorème sur le rapport d'une série entière et de son plus grand terme - complété par Valiron

Un théorème sur la convergence absolue des séries de Fourier - complété par J.P.Kahane

Une remarque sur le théorème de Picard, utilisée par P.Bernays

Un théorème sur une équation intégrale d'Émile Picard, complété par Feldheim

D'ailleurs l'ensemble de mes travaux sur divers types d'équations intégrales mérite sans doute d'être mentionné, sans qu'il y ait lieu d'insister sur un type particulier.

Je pense que les notices que je vous ai envoyées contiennent les renseignements suffisants. Il reste entendu qu'en cas de besoin je vous communiquerai celle de 1935.

Bien cordialement,

Paul Lévy

---

[14] Gaston Ramon (1886–1963) was a physician and biologist, who worked on the use of anatoxin for the prevention of diphteria. He was a member of the Academy in the category of *académiciens libres*.

P.S. - Je me suis adressé à vous parce que je pense que dans la Section de géométrie vous êtes le plus qualifié par vos travaux d'analyse fonctionnelle et de calcul des probabilités. Mais si vous voulez partager le travail avec d'autres, je n'y verrai aucun inconvénient.

Si vous désirez d'autres explications, le mieux serait peut-être que je vienne un jour vous voir

---

Letter 99d

Paris - 38 Av. Théophile Gautier

21 janvier 1964

Mon cher collègue,

L'adresse de Pollaczek est 54 rue du Point du Jour à Boulogne (Seine).[15]

J'ai vu Madame Gauja en quittant l'Académie.[16] La date de l'élection est le 20 avril (avec possibilité de changement ultérieur).

Je vous renvoie séparément un exemplaire de ma notice, avec quelques corrections et additions manuscrites qui n'étaient peut-être pas sur le premier exemplaire envoyé. J'en ai assez pour qu'il n'y ait pas d'inconvénient à ce que vous en ayez deux.

Bien que je n'aie pas tout dit, je pense que vous trouverez là de quoi parler suffisamment de mes travaux sur le calcul des probabilités, sur quelques problèmes de géométrie et sur quelques équations intégrales.

Pour l'analyse fonctionnelle, je vais vous faire un résumé, qui aura peut-être 10 pages, et qui vous dispensera de lire les 50 pages de ma notice de 1935.

Quant aux questions diverses d'analyse dont je me suis occupé, je ne sais pas bien sur lesquelles je dois attirer votre attention, il faut que j'y réfléchisse encore. Il est d'ailleurs bien entendu que vous aurez trop d'éléments, et qu'il vous appartiendra de choisir.

J'ajoute encore un mot sur les martingales, que j'ai introduites dans un mémoire de 1935, je crois, puis dans mon livre de 1937 et que Ville a baptisées. Doob y a attaché assez d'importance pour leur consacrer dans ses Stochastic processes tout un chapitre, entre celui des Processus markoviens et celui des Processus additifs.

[15] An Austrian born in Vienna, Felix Pollaczek (1892–1981) was an expert in queuing theory. After finishing his studies he was mobilized in World War I. In 1920 he obtained a doctorate in mathematics (in number theory) at the University of Berlin. He remained in Berlin until 1933, primarily doing research for the post office. With the rise of the Nazis, he emigrated to Paris, then went to Czechoslovakia. He returned to Paris in 1938, managing to escape persecution during the Occupation. At the Liberation he obtained the position, however precarious, of Maître de Recherches in the CNRS. He became a French citizen in 1947.

[16] Madame Gauja, a secretary at the Académie was the wife of the main secretary-archivist of Paris Academy of Sciences for many years, Pierre Gauja, who published several books based on the archival material of the Academy collection.

Il a créé ensuite la théorie des sous-martingales. Il y a maintenant aussi des super martingales; je l'ai vu par les programmes de cours de la Sorbonne, et ne sais même pas ce que c'est ; toutefois je crois le deviner.

Bien cordialement

Paul Lévy

Letter 99e

Paris - 38 Av. Théophile Gautier

30 janvier 1964

Mon cher collègue,

Je viens de réfléchir que Denjoy avait préparé l'année dernière un exposé de mes titres, et avait été de mauvaise humeur parce qu'il avait fait ce travail pour rien. Je ne pense pas qu'il désire l'utiliser maintenant, mais avant de vous mettre au travail il faudrait peut-être vous assurer que vous n'êtes pas en compétition avec lui.

Il reste entendu que préférerais que vous soyez chargé du rapport, tant parce que vous m'avez tout de suite promis votre appui sans réserve qu'à cause de votre compétence en analyse fonctionnelle et en calcul des probabilités.

Naturellement, si, après le rapport, un ou deux autres membres de la Section prennent la parole pour confirmer brièvement leur accord, cela ne peut pas nuire.

Quoi qu'il en soit, je vous envoie ci-joint un résumé en 4 pages de mes travaux d'analyse fonctionnelle. Je crois qu'il donne l'essentiel, et vous dispensera de lire les 27 pages consacrées à ce sujet dans ma notice de 1935 (dans ma nouvelle notice, en bas de la p.22, il faut lire "environ 30 pages"; c'est par erreur qu'on a mis 50).

Pour mes autres travaux (géométrie ou analyse pure), je reste embarrassé. Les travaux les plus importants sont peut-être ceux auxquels j'ai été conduit par les applications au calcul fonctionnel ou au calcul des probabilités (formule d'inversion de $\varphi(z) = \int_{-\infty}^{+\infty} \mathbf{e}^{izx} dF(x)$, étude des fonctions de Green de certaines équations intégrales, des exponentielles de polynômes, etc.). Peut-être aussi un théorème sur les séries de Fourier qui a servi de point de départ à une étude de J.P.Kahane.[17]

[17] Lévy probably alludes to Jean-Pierre Kahane's extension of his and Wiener's result about the absolute convergence of Fourier series of $F \circ f$ when f is a real-valued function whose Fourier series absolutely converges and F an analytical function. Kahane's paper was published as the note to the Academy of Science under the title *Sur un théorème de Wiener-Lévy* (CRAS 246, 1958. 1949–1951).

Je vous en envoie un résumé ci-joint; et je précise aussi mon théorème sur les $\mathbf{e}^{P(z)}$.

Bien cordialement à vous,

Paul Lévy

Letter 99f

Paris - 38 Av. Théophile Gautier

2-4-1964

Mon cher collègue,

Repensant à notre conversation d'hier, je viens de regarder le livre de Doob. Il consacre aux martingales un chapitre de 99 pages, sur 622 pages de texte.

Il y a ensuite un appendice historique, dans lequel le chapitre des martingales commence par la phrase suivante:

"Martingales have been studied by many authors, referred to below. See particularly Lévy (Théorie de l'addition des variables aléatoires, 1937), Ville (Etude critique de la notion de collectif), J.L.Doob (Regularity properties of certain families of chance variables, 1940)

———

Dans les 6 pages qui suivent et donnent plus de détails, je ne trouve le nom de Ville qu'une fois ; je suis cité 6 fois, et Doob lui-même 11 fois. Les autres auteurs cités sont Andersen, Jessen, Zygmund, Marcinkiewicz, de Possel.

Même en admettant que Doob ait développé avec complaisance un chapitre qui doit tant à ses travaux personnels (remarque qui peut être faite aussi pour les processus markoviens), cela montre bien l'importance des martingales ; et je ne crois pas qu'on puisse discuter ma priorité. Mon livre de 1937 avait été précédé par 2 mémoires de 1935 qui contiennent les idées reprises dans ce livre. La thèse de Ville (1939) avait dû être précédée par 1 ou 2 notes, sans doute en 1938, en tout cas après mon livre.

Bien cordialement

Paul Lévy

J'ajoute que la bibliographie de Doob comprend 8 de mes livres ou mémoires; ce nombre n'est dépassé que par Doob lui-même (13) et Kolmogrorov (12).

References

1. Barbut, M., Locker, B., Mazliak, L.: Paul Lévy and Maurice Fréchet: 50 years of correspondence in 107 letters. Springer (2014)
2. Jaëck, F.: From *Fundamenta Mathematicae* to *Studia Mathematica*: the renaissance of Polish mathematics in the light of Banach's publications 1919–1940. In: L. Mazliak, T. R. (eds.) Mathematical communities in the aftermath of the Great War. Birkhäuser-Science (2020). Trends in the History of Science
3. Khinchin, A.Y., Kolmogorov, A.N.: Über Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden. Mat. Sb. **32**, 668–677 (1925)
4. Lévy, P.: Calcul des probabilités. Gauthier-Villars (1925)
5. Lévy, P.: Théorie de l'addition des variables aléatoires. Gauthier-Villars (1937)

Andrei Kolmogorov and Leonid Levin on Randomness

Laurent Bienvenu, Glenn Shafer and Alexander Shen

Abstract

This chapter provides a translation of a letter written in 1939, in which Andrei Kolmogorov's explains his views on the connection between probability theory and its applications. It also provides some historical sources on the early history of Kolmogorov complexity in Russia (then the USSR): abstracts of talks by Kolmogorov and letters from Leonid Levin to Kolmogorov. The relation between Levin's measures and semimeasures on the one hand and martingales and supermartingales on the other is explained in the present volume's chapter "Martingales in the Study of Randomness".

Keywords

Andrei Kolmogorov · Complexity · Leonid Levin · Randomness · Semimeasure

1 Introduction

This chapter reproduces in translation several documents related to the history of martingales in randomness as recounted in the present volume by Laurent Bienvenu, Glenn Shafer, and Alexander Shen in the chapter "Martingales in the Study of Randomness".

L. Bienvenu
LaBRI (Laboratoire Bordelais de Recherche en Informatique), CNRS, Bordeaux, France
e-mail: Laurent.Bienvenu@computability.fr

G. Shafer (✉)
Rutgers Business School, 1 Washington Park, Newark, NJ, USA
e-mail: gshafer@business.rutgers.edu

A. Shen
LIRMM, University of Montpellier, CNRS, Montpellier, France
e-mail: alexander.shen@lirmm.fr

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*,
Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9_19

Section 2 translates from the French a 1939 letter from Andrei Kolmogorov to Maurice Fréchet, in which Kolmogorov agrees with Richard von Mises that only a theory of very large finite collectives can reflect truthfully the use of probability in practice, but adds that such a theory cannot be formalized mathematically.

Section 3 translates from the Russian the abstracts of three talks on algorithmic randomness that Kolmogorov gave in Moscow in the period 1967–1971. In the talk given 31 October 1967, Kolmogorov presented the results later announced in [1], including the Kolmogorov–Levin formula for the complexity of pairs and the existence of Church random sequences with logarithmic complexity of prefixes. In the 23 November 1971 talk Kolmogorov introduces the notion of resource-bounded complexity and discusses some related results. Finally, on 16 April1974, Kolmogorov introduces a notion related to the algorithmic statistics, now called Kolmogorov's structure function.

Section 4 translates from the Russian three letters from Leonid Levin to Kolmogorov, undated but written during the period from August 1970 to January 1971. In the first letter Levin notes that Martin-Löf random sequences (with respect to a computable measure P) can be characterized as sequences that have a bounded ratio of continuous a priori probability and P for its prefixes. The second letter introduces monotone complexity (though some details of this definition were later corrected by Levin in [2]) and formulates the criterion of randomness in terms of monotone complexity. The third letter is rather cryptic; probably it contains some initial version of Levin's observation that non-stochastic objects have large mutual information with the halting problem, see [3, Sect. 4.6].

2 Letter from Kolmogorov to Fréchet, 1939

The Fréchet papers in the archives of the Academy of Sciences in Paris include a letter in French to Fréchet, in which Kolmogorov elaborates briefly on his philosophy of probability. This translation is published with permission from the Academy.

Moscow 6, Staropimenovsky per. 8, flat 5
3 August 1939

Dear Mr. Fréchet,

I thank you sincerely for sending the proceedings of the Geneva Colloquium, which arrived during my absence from Moscow in July.

The conclusions you express on pp. 51–54 are in full agreement with what I said in the introduction to my book:

> In the pertinent mathematical circles it has been common for some time to construct probability theory in accordance with this general point of view. But a complete presentation of the whole system, free from superfluous complications, has been missing…

You are also right to attribute to me (on p. 42) the opinion that the formal axiomatization should be accompanied by an analysis of its real meaning. Such an analysis is given, perhaps too briefly, in the section "The relation to the world of experience" in my book. Here I insist on the view, expressed by Mr. von Mises himself (Wahrscheinlickeitsrechnung 1931, pp. 21–26), that "collectives" are finite (though very large) in real practice.

One can therefore imagine three theories:

A A theory based on the notions of "very large" finite "collectives", "approximate" stability of frequencies, etc. This theory uses ideas that cannot be defined in a purely formal (i.e., mathematical) manner, but it is the only one to reflect experience truthfully.

B A theory based on infinite collectives and limits of frequencies. After Mr. Wald's work we know that this theory can be developed in a purely formal way without contradictions. But in this case its relation to experience cannot have any different nature than for any other axiomatic theory. So in agreement with Mr. von Mises, we should regard theory B as a certain "mathematical idealization" of theory A.

C An axiomatic theory of the sort proposed in my book. Its practical value can be deduced directly from the "approximate" theory A without appealing to theory B. This is the procedure that seems simplest to me.

Yours cordially,
A. Kolmogoroff

3 Abstracts of Three Talks by Kolmogorov, 1967–1974

Abstracts of some of the talks at the meetings of Moscow Mathematical Society were published in the journal *Uspekhi matematicheckikh nauk*. Here we reproduce translations of the abstracts for three talks by Kolmogorov, in 1967, 1971, and 1974, on algorithmic information theory. The translations are by Leonid Levin; we have edited them slightly.

3.1 31 October 1967

A. N. Kolmogorov, "Several theorems about algorithmic entropy and algorithmic amount of information", from Volume 23, no. 2, 1968.

The algorithmic approach to the foundations of information theory and probability theory was not developed far for several years after its appearance, because some questions raised at the very start remained unanswered. Now the situation has changed somewhat. In particular, it is ascertained that the decomposition of entropy $H(x, y) \sim H(x) + H(y|x)$ and the formula $J(x|y) \sim J(y|x)$ hold for the algorithmic concept only with accuracy $O([\log H(x, y)])$ (Levin, Kolmogorov).

The fundamental difference between the algorithmic definition of a Bernoulli sequence (a simplest collective) and the definition of Mises-Church, stated earlier, is concretized in the form of a theorem: there exist Bernoulli (in the sense of Mises-Church) sequences $x = (x_1, x_2, ...)$ with density of ones $p = \frac{1}{2}$, with initial segments of entropy ("complexity") $H(x^n) = H(x_1, x_2, ..., x_n) = O(\log n)$ (Kolmogorov).

For understanding of the talk an intuitive, not formal, familiarity with the concept of a computable function suffices.

3.2 23 November 1971

A. N. Kolmogorov, "Complexity of specifying and complexity of constructing mathematical objects", from Volume 27, no. 2, 1972.

1. Organizing machine computations requires dealing with evaluation of (a) complexity of programs, (b) size of memory used, (c) duration of computation. The talk describes a group of works that consider similar concepts in a more abstract manner.
2. It was noticed in 1964–1965 that the minimal length $K(x)$ of the binary representation of a program specifying the construction of an object x can be defined invariantly up to an additive constant (Solomonoff, A. N. Kolmogorov). This permitted using the concept of *definition complexity* $K(x)$ of constructive mathematical objects as the basis for a new approach to the foundations of information theory (A. N. Kolmogorov, Levin) and probability theory (A. N. Kolmogorov, Martin-Löf, Schnorr, Levin).
3. Such characteristics as "required memory volume," or "required duration of work" are harder to free of technical peculiarities of special machine types. But some results may already be extracted from the axiomatic "machine-independent" theory of a broad class of similar characteristics (Blum, 1967). Let $\Pi(p)$ be a characteristic of "construction complexity" of the object $x = A(p)$ by a program p, and let $\Lambda(p)$ be the length of the program p. The formula $K^n\Pi(x) = \inf(\Lambda(p) : x = A(p), \Pi(p) = n)$ defines the "n-complexity of definition" of object x (when the condition is unsatisfiable, the inf is considered infinite).
4. Barzdin's Theorem on the complexity $K(M_\alpha)$ of prefixes M_α of an enumerable set of natural numbers (1968) and results of Barzdin, Kanovich, and Petri on corresponding complexities $K^n\Pi(M_\alpha)$, are of general mathematical interest, as they shed some new light on the role of extending previously used formalizations in the development of mathematics. The survey of the state of this circle of problems was given in the form free from any cumbersome technical apparatus.

3.3 16 April 1974

A. N. Kolmogorov, "Complexity of algorithms and objective definition of randomness", from Volume 29, no. 4, 1974.

To each constructive object corresponds a function $\Phi_x(k)$ of a natural number k – the log of minimal cardinality of x-containing sets that allow definitions of complexity at most k. If the element x itself allows a simple definition, then the function Φ drops to 1 even for small k. Lacking such a definition, the element is "random" in a negative sense. But it is positively "probabilistically random" only when the function Φ, having taken the value Φ_0 at a relatively small $k = k_0$, then changes approximately as $\Phi(k) = \Phi_0 - (k - k_0)$.

4 Three Letters from Levin to Kolmogorov 1970–1971

These letters are not dated but were written after the submission of [4] in August 1970 and before Kolmogorov left (in January 1971) for an oceanographic expedition on the ship *Dmitri Mendeleev*. Copies (the typescript for the first two letters and the handwritten manuscript for the third one) provided by Leonid Levin and translated by Alexander Shen. The third letter has no salutation. Levin recalls that he often gave notes like this to Kolmogorov, who rarely had much time to hear lengthy explanations and preferred something written in any case.

4.1 Letter I

Dear Andrei Nikolaevich! A few days ago I obtained a result I like a lot. Maybe it could be useful to you if you work on these topics while travelling on the ship.

This result gives a formulation for the foundations of probability theory different from Martin-Löf. I think it is closer to your initial idea about the relation between complexity and randomness and is much clearer from the philosophical point of view (as, e.g., [Yury Tikhonovich] Medvedev says).

Martin-Löf considered (for an arbitrary computable measure P) an algorithm that studies a given sequence and finds more and more deviation from the P-randomness hypothesis. Such an algorithm should be P-consistent, i.e., find deviations of size m only for sequences in a set that has measure at most 2^{-m}. It is evident that a number m produced by such an algorithm on input string x should be between 0 and $-\log_2 P(x)$. Let us consider the complementary value $(\log_2 P(x)) - m$ and call it the "complementary test" (the consistency requirement can be easily reformulated for complementary tests).

Theorem. *The logarithm of a priori probability [on the binary tree] $-\log_2 R(x)$ is a P-consistent complementary test for every measure P and has the usual algorithmic properties.*

Let me remind you that by a priori probability I mean the universal semicomputable measure introduced in our article with Zvonkin. [See [4].] It is shown there that it [minus its logarithm] is numerically close to complexity.

Let us consider a specific computable measure P. Compared to the universal Martin-Löf test f (specific to a given measure P) our test is not optimal up to an additive constant, but is asymptotically optimal. Namely, if the universal Martin-

Löf test finds a deviation m, our test finds a deviation at least $m - 2\log_2 m - c$. Therefore, the class of random infinite binary sequences remains the same.

Now look how nicely it fits the philosophy. We say that a hypothesis "x appeared randomly according to measure P" can be rejected with certainty m if the measure P is much less consistent with the appearance of x than a priori probability (this means simply that $P(x) < R(x)/2^m$. This gives a law of probability theory that is violated with probability at most 2^{-m}. Its violation can be established effectively since R is semicomputable [enumerable from below]. But if this law holds, all other laws of probability theory [i.e., all Martin-Löf tests] hold, too. The drawback is that it gives a bit smaller value of randomness deficiency (only $m - 2\log_2 m - c$ instead of m), but this is a price for the universality (arbitrary probability distribution). The connection with complexity is provided because $-\log_2 R(x)$ almost coincides with the complexity of x. Now this connection does not depend on the measure.

It is worth noting that the universal semicomputable measure has many interesting applications besides the above mentioned. You know its application to the analysis of randomized algorithms. Also it is often useful in proofs (e.g., in the proof of J. T. Schwartz' hypothesis regarding the complexity of almost all trajectories of dynamic systems). Once I used this measure to construct a definition of intuitionistic validity. All this shows that it is a rather natural quantity.

L.

4.2 Letter II

Dear Andrei Nikolaevich!

I would like to show that plain complexity does not work if we want to provide an *exact* definition of randomness, even *for a finite case*. For the uniform distribution on strings of fixed length n the randomness deficiency is defined as n minus the complexity. For a non-uniform distribution length is replaced by minus the logarithm of the probability.

It turns out that even for a distribution on a finite set the randomness deficiency could be high on a set of large measure.

Example. Let

$$P(x) = \begin{cases} 2^{-(l(x)+100)} & \text{if } l(x) \leq 2^{100} \\ 0 & \text{if } l(x) > 2^{100}. \end{cases}$$

Then $|\log_2 P(x)| - K(x)$ exceeds 100 *for all* strings x.

A similar example can be constructed for strings of some fixed length (by adding zero prefixes). The violation could be of logarithmic order.

Let me show you how to sharpen the definition of complexity to get an exact result (both for finite and infinite sequences).

Definitions. Let A be a monotone algorithm, i.e., for every x and every y that is a prefix of x, if $A(x)$ is defined, then $A(y)$ is defined too and $A(y)$ is a prefix of $A(x)$.

Let us define

$$KM_A(x) = \begin{cases} \min \ l(p)\colon x \text{ is a prefix of } A(p) \\ \infty \text{ if there is no such } p \end{cases}$$

The complexity with respect to an optimal algorithm is denoted by $KM(x)$.

Let $P(x)$ be a computable distribution on the Cantor space Ω, i.e., $P(x)$ is the measure of the set Γ_x of all infinite extensions of x.

Theorem 1

$$KM(x) \le |\log_2 P(x)| + O(1);$$

Theorem 2

$$KM((\omega)_n) = |\log_2 P((\omega)_n)| + O(1)$$

for P-almost all ω; here $(\omega)_n$ stands for n-bit prefix of ω. Moreover, the probability that the randomness deficiency exceeds m for some prefix is bounded by 2^{-m}.

Theorem 3 *The sequences ω such that*

$$KM((\omega)_n) = |\log_2 P((\omega)_n)| + O(1);$$

satisfy all laws of probability theory (all Martin-Löf tests).

Let me use this occasion to tell you the results from my talk in the laboratory [of statistical methods in Moscow State University]: why one can omit non-computable tests (i.e., tests not definable without a strong language).

For this we need to improve the definition of complexity once more. The plain complexity $K(x)$ has the following property:

Remark. Let A_i be an effectively given sequence of algorithms such that

$$K_{A_{i+1}}(x) \le K_{A_i(x)}$$

for all i and x. Then there exists an algorithm A_0 such that

$$K_{A_0}(x) = 1 + \min_i K_{A_i}(x).$$

Unfortunately, it seems that $KM(x)$ does not have this property. This can be corrected easily. Let A_i be an effective sequence of monotone algorithms with finite domain (provided as tables) such that

$$KM_{A_{i+1}}(x) \le KM_{A_i(x)}$$

for all i and x. Let us define then

$$\overline{KM}_{A_i}(x) = \min_i KM_{A_i}(x).$$

Among all sequences A_i there exists an optimal one, and the complexity with respect to this optimal sequence is denoted by $\overline{KM}(x)$. This complexity coincides with the logarithm of a universal semicomputable semimeasure [=a priori probability on the binary tree].

Theorem 4 *$\overline{KM}(x)$ is a minimal semicomputable [from above] function that makes Theorem 2 true.*

Therefore no further improvements of $\overline{KM}$ are possible.

Now consider the language [=set] of all functions computable with a fixed non-computable sequence [oracle] α. Assume that α is complicated enough, so this set contains the characteristic function of a universal enumerable set [$\mathbf{0}'$].

We can define then a relativized complexity $\overline{KM}_\alpha(x)$ replacing algorithms by algorithms with oracle α, i.e., functions from this language.

Definition. A sequence ω is called *normal* if

$$\overline{KM}((\omega)_n) = \overline{KM}_\alpha((\omega)_n) + O(1).$$

For a finite sequence ω_n we define the "normality deficiency" as

$$\overline{KM}(\omega_n) - \overline{KM}_\alpha(\omega_n).$$

Theorem 5 *A sequence obtained by an algorithm from a normal sequence is normal itself.*

Theorem 6 *Let P be a probability distribution that is defined (in a natural encoding) by a normal sequence. Then P-almost every sequence is normal.*

This theorem exhibits a law of probability theory that says that a random process cannot produce a non-normal sequence unless the probability distribution itself is not normal. This is a much more general law than standard laws of probability theory since it does not depend on the distribution. Moreover, Theorem 5 shows that this law is not restricted to probability theory and can be considered as a universal law of nature:

Thesis. Every sequence that appears in reality (finite or infinite) has normality deficiency that does not exceed the complexity of the description (in a natural language) of how it is physically produced, or location etc.

It turns out that this normality law (that can be regarded as not confined to probability theory) and the law corresponding to the universal computable test together imply any law of probability theory (not necessary computable) that can be described in the language. Namely, the following result holds:

Theorem 7 *Let P be a computable probability distribution. If a sequence ω is normal and passes the universal computable P-test, then ω passes any test defined in our language (i.e., every test computable with oracle α).*[1]

Let us give one more interesting result that shows that all normal sequences have similar structure.

Theorem 8 *Every normal sequence can be obtained by an algorithm from a sequence that is random with respect to the uniform distribution.*

4.3 Letter III

We use a sequence α that provides a "dense" coding of a universal [recursively] enumerable set. For example, let α be the binary representation of [here the text "the sum of the a priori probabilities of all natural numbers" is crossed out and replaced by the following:] the real number

$$\sum_{p \in A} \frac{1}{p \cdot \log^2 p}$$

where A is the domain of the optimal algorithm.

A binary string p is a "good" code for x if the optimal algorithm converts the pair $(p, K(x))$ into a list of strings that contains x, and the logarithm of the cardinality of this list does not exceed $K(x) + 3 \log K(x) - l(p)$. (The existence of such a code means that x is "random" when $n \geq l(p)$.)

We say that a binary string p is a canonical code for x if every prefix of p either is a "good" code for x or is a prefix of α, and $l(p) = K(x) + 2 \log K(x)$.

Theorem 1 *Every x (with finitely many exceptions) has a canonical code p, and p and x can be effectively transformed into each other if $K(x)$ is given.*

Therefore, the "non-randomness" in x can appear only due to some very special information (a prefix of α) contained in x. I cannot imagine how such an x can be observed in (extracted from) the real world since α is not computable. And the task "to study the prefixes of a specific sequence α" seems to be very special.

[1] In a footnote in the letter, Levin adds, "Note that for every set of measure 0 there exists a test (not necessary computable) that rejects all its elements."

Acknowledgements This material appeared earlier in "On the history of martingales in the study of randomness", in the *Electronic Journal for History of Probability and Statistics* for June 2009. Preparation for that earlier article was supported in part by ANR grant NAFIT-08-EMER-008-01.

References

1. Kolmogorov, A.N.: Logical basis for information theory and probability theory. IEEE Trans. Inform. Theory **14**, 662–664 (1968)
2. Levin, L.A.: On the notion of a random sequence. Soviet Math. Doklady **14**(5), 1413–1416 (1973)
3. Vereshchagin, N.K., Shen, A.: Algorithmic statistics: forty years later. In: Computability and Complexity. Essays Dedicated to Rodney G. Downey on the Occasion of His 60th Birthday. Springer, 2017. Lecture Notes in Computer Science, v. 10010, pp. 669–737. Springer (2017)
4. Zvonkin, A.K., L. A. Levin, L.A.: The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. Russian Math. Surveys **25**(6), 83–124 (1970)

Index

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
L. Mazliak and G. Shafer (eds.), *The Splendors and Miseries of Martingales*, Trends in the History of Science,
https://doi.org/10.1007/978-3-031-05988-9