Clases 9-10: El proceso de Wiener y los paseos al azar: el teorema de Donsker *

6 de noviembre de 2017

Índice

1. Introducción

El teorema de Donsker establece la convergencia débil de un paseo al azar convenientemente normalizado al proceso de Wiener. Es una generalización del teorema central del límite al caso de procesos estocásticos, por lo que se lo llama *teorema central del límite funcional* Es entonces el sustento teórico de la simulación. En estas clases introducimos los paseos al azar y demostramos el teorema de Donsker en dos casos particulares: para el paseo al azar con variables gaussianas y para el paso al azar con variables simétricas de Bernoulli.

2. Paseos al azar

Un paso al azar es un proceso estocástico de tiempo discreto construido a partir de una sucesión de variables aleatorias independientes y con distribución común

$$X_1, X_2, \dots$$

*Notas preparadas por E. Mordecki para el curso de Simulación en procesos estocásticos 2017.

de la siguiente forma:

$$S_0 = 0, \quad S_n = X_1 + \cdots + X_n \ (n \geq 1).$$

Para obtener un límite en primer lugar suponemos que las variables están centradas y tienen varianza uno, es decir $\mathbf{E}X_1 = 0$, $\mathrm{var}X_1 = 1$. De aquí resulta que $\mathbf{E}S_n = 0$, mientras que $\mathrm{var}S_n = n$.

En segundo lugar escalamos el proceso. La idea es que queremos ver a este paseo al azar con una cantidad creciente de sumandos en un intervalo fijo, y parametrizado como el moviemiento Browniano, digamos $t \in [0, 1]$.

Esto se obtiene de considerar $S_{[nt]}$, donde $[nt]$ es la parte entera de nt. Para $t = 1$ obtenemos

$$\mathrm{var}S_{[n1]} = \mathrm{var}S_n = n.$$

Para obtener un límite tenemos que tener una varianza constante. Por eso consideramos

$$X_n(t) = \frac{1}{\sqrt{n}} S_{[nt]}, \quad t \in [0, 1]. \tag{1}$$

El teorema de Donsker establece que X_n converge a W en un sentido que no vamos a precisar[1], pero vamos a demostrarlo en dos casos particulares.

3. Paseo al azar con variables gaussianas

Supongamos ahora que

$$X_1, X_2, \ldots$$

son variables aleatorias independientes gaussianas estándar, y construyamos como antes

$$X_n(t) = \frac{1}{\sqrt{n}} S_{[nt]}, \quad t \in [0, 1].$$

Consideremos un proceso de Wiener $W = \{W(t) \colon 0 \leq t \leq 1\}$, y la función "escalera" $x_n \colon [0, 1] \to [0, 1]$, definida como

$$x_n(t) = \frac{[nt]}{n}, \tag{2}$$

Si $x(t) = t$, observemos que $x_n \rightrightarrows x$, donde la convergencia es uniforme en $[0, 1]$. Recordemos el siguiente resultado.

Proposición 1. *Sea $f \colon [0, 1] \to \mathbf{R}$ una función continua, y una sucesión de funciones (que actúan como cambios de tiempo) $x_n \rightrightarrows x$, donde la convergencia es uniforme. Entonces $f(x_n(t))$ converge uniformemente a $f(x(t))$, es decir*

$$f \circ x_n \rightrightarrows f \circ x, \qquad t \in [0, 1],$$

donde $\circ$ representa la composición de funciones.

[1] El lector interesado puede consultar el libro *Convergence of Probability Measures*, 2nd Edition, por P. Billingsley

Demostración. Sea $\varepsilon > 0$. La función f es uniformemente contina, luego existe $\delta > 0$ tal que $|f(x) - f(y)| < \varepsilon$ siempre que $|x - y| < \delta$. Para ese δ existe n_0 tal que $|x_n(t) - x(t)| < \delta$ si $n \geq n_0$. Entonces, para esos mismos n, tenemos $|f(x_n(t)) - f(x(t))| < \varepsilon$, que es la tesis. □

Consideremos ahora el proceso estocástico

$$Y_n(t) = W(x_n(t)),$$

donde x_n es la función escalera definida en (2). No es difícil verificar que las leyes (finito-dimensionales) de los procesos X_n definido como en (1) con variables gaussianas estándar e Y_n coinciden. Además, como las trayectorias de W son continuas tenemos, casi seguramente

$$Y_n \rightrightarrows W.$$

En conclusión, el teorema de Donsker toma la siguiente forma en este caso.

Teorema 1. *Dada una sucesión de variables aleatorias independientes normales estándar $\{X_k\}$, el proceso estocástico $\{X_n(t) \colon 0 \leq t \leq 1\}$ en* (1) *tiene una copia Y_n (es decir, otro proceso con la misma distribución) tal que*

$$\mathbf{P}(Y_n(t) \rightrightarrows W(t),\ t \in [0,1]) = 1.$$

4. Paseo al azar de Bernoulli simétrico

La segunda versión del teorema de Donsker es más interesante, y no es tan directa como la anterior. Comenzamos enunciado el siguiente resultado sobre las transformaciones del proceso de Wiener.

Proposición 2. *Sea $\{W(t) \colon t \geq 0\}$ un proceso de Wiener. Entonces, son procesos de Wiener los siguientes procesos:*

(1) *El proceso* simétrico*: $\{W_1(t) = -W(t) \colon t \geq 0\}$.*

(2) *El proceso* trasladado*: $\{W_2(t) = W(t + a) - W(a) \colon t \geq 0\}$, cualquiera sea $a > 0$.*

(3) *El proceso* escalado*: $\{W_3(t) = \sqrt{c}W(t/c) \colon t \geq 0\}$, cualquiera sea $c > 0$.*

Demostración. La demostración de esta proposición se obtiene aplicando la definición de proceso de Wiener en cada uno de los casos. □

Las variables aleatorias independientes $\{X_k\}$ ahora tienen distribución común de Bernoulli simétrica, es decir

$$\mathbf{P}(X_k = 1) = \mathbf{P}(X_k = -1) = \frac{1}{2}.$$

De aquí $\mathbf{E}X_k = 0$ y $\mathrm{var}X_k = 1$, y construímos X_n como en (1). La técnica de demostración de la aproximación es la misma: dado un proceso de Wiener

$W = \{W(t)\colon 0 \leq t \leq 1\}$ construiremos un paseo al azar simétrico Y_n con igual ley que X_n, que converge a W (en un sentido a precisar).

Para eso introducimos una nueva variable aleatoria (denominada tiempos de parada) que indica el primer arribo al del proceso a los niveles ± 1:

$$T_1 = \text{ínf}\{t \geq 0\colon |W(t)| = 1\}. \tag{3}$$

Veamos que T_1 es casi seguramente finito. En efecto

$$\begin{aligned}\mathbf{P}(T_1 \geq t) = \mathbf{P}\left(\max_{0\leq s\leq t}|W(s)| \leq 1\right) &\leq \mathbf{P}(|W(t)| \leq 1)\\ &= 2\int_0^{1/\sqrt{t}} \varphi(x)dx \leq \frac{1}{\sqrt{2\pi t}} \to 0 \ (t \to \infty).\end{aligned}$$

Aquí $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ es la densidad normal estándar. De la propiedad (1) (proceso simétrico) obtenemos que el proceso detenido al llegar a ± 1, es decir, $W(T_1)$ es una variable aleatoria simétrica, es decir

$$\mathbf{P}(W(T_1) = 1) = \mathbf{P}(W(T_1) = -1) = \frac{1}{2}.$$

Así construímos $Y_1 = W(T_1)$, la primera variable de Bernoulli simétrica. La construcción de las siguientes requiere algunos resultados auxiliares (que tienen importancia por sí mismos).

Teorema 2. *Dados un proceso de Wiener* $\{W(t)\colon t \geq 0\}$ *y la variable aleatoria* T_1 *construída como en* (3), *el proceso* W_2 *dado por*

$$W_2(t) = W(t + T_1) - W(T_1), \quad t \geq 0,$$

es un nuevo proceso de Wiener, y es independiente del proceso $\{W(t \wedge T_1)\colon t \geq 0\}$.

Demostración. Como $W_2(0) = 0$ y $\{W_2(t)\colon t \geq 0\}$ es continuo casi seguramente, nos resta demostrar las dos últimas condiciones de la definición de movimiento Browniano y la independencia. Esto es equivalente a verificar que, dados $n \geq 1$, $0 = t_0 < t_1 < \cdots t_\ell$, la distribución del vector

$$(W_2(t_1) - W_2(t_0), \dots, W_2(t_\ell) - W_2(t_{\ell-1})) \sim \mathcal{N}\left(0, \mathbf{diag}(t_j - t_{j-1})_{1\leq j\leq \ell}\right),$$

(donde $\mathbf{diag}(x)$ es la matriz con las entradas de x en la diagonal) y que este vector es independiente de sucesos que dependan de las coordenadas de $W_2(t \wedge T_1)$. Si B uno de estos sucesos, tenemos que demostrar que

$$\mathbf{E}\left(\mathbf{1}_B \exp\left(i\sum_{j=1}^{\ell}\lambda_j(W_2(t_j) - W_2(t_{j-1}))\right)\right) = \mathbf{P}(B)\exp\left(-\frac{1}{2}\sum_{j=1}^{\ell}\lambda_j^2(t_j - t_{j-1})\right).$$

Para demostrar esta afirmación, aproximamos nuestro tiempo de parada T_1 mediante

$$T_n = \frac{[2^n T_1]+1}{2^n} = \frac{k+1}{2^n}, \text{ cuando } \frac{k}{2^n} \le T_1 < \frac{k+1}{2^n}.$$

Tenemos que $T_n \searrow T_1$ c.s. Como el integrando es acotado,

$$\begin{aligned}
&\mathbf{E}\left(\mathbf{1}_B \exp\left(i\sum_{j=1}^{\ell}\lambda_j(W_2(t_j)-W_2(t_{j-1}))\right)\right)\\
&\qquad= \mathbf{E}\left(\mathbf{1}_B \exp\left(i\sum_{j=1}^{\ell}\lambda_j(W(T_1+t_j)-W(T_1+t_{j-1}))\right)\right)\\
&\qquad= \lim_{n\to\infty}\mathbf{E}\left(\mathbf{1}_B \exp\left(i\sum_{j=1}^{\ell}\lambda_j(W(T_n+t_j)-W(T_n+t_{j-1}))\right)\right)\\
&= \lim_{n\to\infty}\mathbf{E}\left(\sum_{k=1}^{\infty}\mathbf{1}_{B\cap\{T_n=k2^{-n}\}} \exp\left(i\sum_{j=1}^{\ell}\lambda_j(W(T_n+t_j)-W(T_n+t_{j-1}))\right)\right)\\
&= \lim_{n\to\infty}\mathbf{E}\left(\sum_{k=1}^{\infty}\mathbf{1}_{B\cap\{T_n=k2^{-n}\}} \exp\left(i\sum_{j=1}^{\ell}\lambda_j(W(k2^{-n}+t_j)-W(k2^{-n}+t_{j-1}))\right)\right)\\
&\overset{key}{=} \lim_{n\to\infty}\sum_{k=1}^{\infty}\mathbf{P}(B\cap\{T_n=k2^{-n}\})\mathbf{E}\left(\exp\left(i\sum_{j=1}^{\ell}\lambda_j(W(k2^{-n}+t_j)-W(k2^{-n}+t_{j-1}))\right)\right)\\
&\qquad= \lim_{n\to\infty}\sum_{k=1}^{\infty}\mathbf{P}(B\cap\{T_n=k2^{-n}\})\exp\left(-\frac{1}{2}\sum_{j=1}^{\ell}\lambda_j^2(t_j-t_{j-1})\right)\\
&\qquad\qquad= \mathbf{P}(B)\exp\left(-\frac{1}{2}\sum_{j=1}^{\ell}\lambda_j^2(t_j-t_{j-1})\right).
\end{aligned}$$

La igualdad clave ($\overset{key}{=}$) se basa en la independencia de los incrementos futuros a $k2^{-n}$ y el suceso B y la variable T_1, que dependen de tiempos anteriores a $k2^{-n}$. Esto concluye la demostración. □

Continuamos con la construcción entonces. Definimos

$$T_2 = \inf\{t \ge 0\colon |W_2(t)| = 1\}.$$

La repetición del argumento anterior, nos lleva a obtener que $Y_2 = W_2(T_2)$ es una variable de Bernoulli simétrica, independiente de $W(T_1)$. Observemos que $Y_1 + Y_2 = W_2(T_2) + W(T_1) = W(T_1 + T_2)$. Es claro que este proceso se puede repetir, y en cada paso el esquema es el mismo. Obtenemos entonces una sucesión de variables aleatorias simétricas de Bernoulli $\{Y_k\}$ tales que

$$\sum_{k=1}^{[nt]} Y_k = W\left(\sum_{k=1}^{[nt]} T_k\right). \tag{4}$$

Veamos ahora que, para cualquier $\delta > 0$,

$$\mathbf{P}\left(\sup_{0\leq t\leq 1}\left|\frac{1}{n}\sum_{k=1}^{[nt]} T_k - t\right| \geq \delta\right) \to 0, \quad (n\to\infty),$$

o, en otros términos

$$x_n(t) = \frac{1}{n}\sum_{k=1}^{[nt]} T_k \rightrightarrows t, \quad \text{en probabilidad.} \tag{5}$$

Veamos que probar (5) es equivalente a probar que

$$\frac{1}{n}\sum_{k=1}^{[nt]}(T_k - 1) \rightrightarrows 0, \quad \text{en probabilidad.} \tag{6}$$

Esto es porque

$$\frac{1}{n}\sum_{k=1}^{[nt]} 1 = \frac{[nt]}{n} \rightrightarrows t.$$

Para ver (6) utilizamos el siguiente resultado.

Teorema 3 (Desigualdad de Kolmogorov)**.** *Sean $X_1, \dots, X_n$ variables independientes y centradas, con segundo momento finito, y sea $S_k = X_1 + \cdots + X_k$ ($k = 1, \dots, n$). Entonces, para $\varepsilon > 0$, se verifica*

$$\mathbf{P}\left(\max_{1\leq k\leq n}|S_k| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\mathbf{E}(S_n^2).$$

Demostración. Consideremos el conjunt $A = \{\max_{1\leq k\leq n}|S_k| \geq \varepsilon\}$, y los conjuntos disjuntos

$$A_k = \{|S_1| < \varepsilon, \dots, |S_{k-1}| < \varepsilon, |S_k| \geq \varepsilon\}.$$

que verifican $A = \cup_{k=1}^n A_k$. Tenemos

$$\begin{aligned}\mathbf{E}(S_n^2) \geq \mathbf{E}(S_n^2 \mathbf{1}_A) = \sum_{k=1}^n \mathbf{E}(S_n^2 \mathbf{1}_{A_k}) &= \sum_{k=1}^n \mathbf{E}((S_n - S_k + S_k)^2 \mathbf{1}_{A_k}) \\ &= \sum_{k=1}^n \mathbf{E}([(S_n - S_k)^2 + (S_k)^2 + 2S_k(S_n - S_k)]\mathbf{1}_{A_k}) \\ &= \sum_{k=1}^n \mathbf{E}([(S_n - S_k)^2 + (S_k)^2 + 2S_k(S_n - S_k)]\mathbf{1}_{A_k}) \\ &= \sum_{k=1}^n \mathbf{E}([(S_n - S_k)^2 + (S_k)^2]\mathbf{1}_{A_k}) \\ &= \sum_{k=1}^n \mathbf{E}((S_k)^2 \mathbf{1}_{A_k}) \geq \varepsilon^2 \sum_{k=1}^n \mathbf{E}(\mathbf{1}_{A_k}) = \varepsilon^2 \mathbf{P}(A),\end{aligned}$$

lo que prueba el resultado. Aquí nos basamos en la independencia entre por un lado A_k y S_k y por otro $S_n - S_k$, de la que resulta

$$\mathbf{E}([S_k(S_n - S_k)]\mathbf{1}_{A_k}) = \mathbf{E}(S_n - S_k)\mathbf{E}(S_k \mathbf{1}_{A_k}) = 0$$

□

Para demostrar (6), precisamos saber que $\mathbf{E}\, T_1 = 1$ y que $\mathbf{E}(T_1^2) < \infty$. Con esos datos las variables aleatorias $X_k = T_k - 1$ son independientes, centradas y con varianza finita. Entonces

$$\mathbf{P}\left(\max_{1 \leq k \leq n} \left| \sum_{k=1}^n (T_k - 1) \right| \geq n\varepsilon \right) \leq \frac{1}{n\varepsilon^2} \mathbf{E}(T_1 - 1)^2 \to 0,$$

obteniendo (6).

Ahora aplicamos la transformación del cambio de escala (3) de la proposición 2, donde definimos

$$\widehat{W}(t) = \sqrt{n} W(t/n)$$

que es un Movimiento Browniano. Escribamos la igualdad (4) en términos de $\widehat{W}$. Tenemos

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} Y_k = \widehat{W}\left(\frac{1}{n} \sum_{k=1}^{[nt]} T_k \right) \rightrightarrows \widehat{W}(t), \text{ en probabilidad.} \tag{7}$$

Hemos demostrado el siguiente resultado.

Teorema 4 (Teorema de Donsker para el paseo al azar simétrico).

Existe una sucesión de variables aleatorias de Bernoulli simétricas e independientes $\{Y_k\}$ *y un movimiento Browniano* $\widehat{W}$ *tales que vale* (7).

A. Sobre los momentos de T_1

Para completar la demostración anterior, precisamos establecer que

$$\mathbf{E}\ T_1 = 1, \qquad \mathbf{E}(T_1^2) < \infty. \tag{8}$$

Esto se puede hacer utilizando el siguiente hecho:

Si $H_n(x)$ es el polinomio de Hermite de orden n y $\{W(t)\}$ es un movimiento Browniano, entonces

$$M_n(t) = t^{n/2} H_n(W(t)/\sqrt{t}),$$

es una martingala. Los primeros polinomios de Hermite son

$$H_1(x) = 1, \quad H_2(x) = x^2 - 1, \quad H_3(x) = x^3 - 3x, \quad H_4(x) = x^4 - 6x + 3.$$

Utilizaremos esta identidad para $n = 2$ para la esperanza y $n = 4$ para el segundo momento, aplicando el teorema opcional de Doob para martingalas. Lo haremos directamente, la prueba formal exige considerar los tiempos de parada $T_1 \wedge n$ y tomar límite cuando $n \to \infty$.

Como $M_2(t) = W(t)^2 - t$ es una martingala, la aplicación del Teorema de Doob nos da

$$0 = \mathbf{E}W(T_1^2) - \mathbf{E}\ T_1.$$

Como $W(T_1)^2 = 1$, obtenemos $\mathbf{E}\ T_1 = 1$, la primer afirmación de (8).

Para la segunda utilizamos

$$M_4(t) = W(t)^4 - 6W^2(t)t + 3t^2.$$

Aquí nuevamente $W(T_1)^4 = W(T_1)^2 = 1$, por lo que la aplicación del teorema de Doob nos da

$$0 = 1 - 6\mathbf{E}\ T_1 + 3\mathbf{E}(T_1^2),$$

de donde

$$\mathbf{E}(T_1^2) = 5/3,$$

completando (8).