Clases 7-8: Métodos de reducción de varianza en Monte Carlo *

19 de septiembre de 2017

Índice

1. Variance reduction

As we have seen, a critical issue in MC method is the quality of estimation. The question we face is: can we devise a method that produces, with the same number n of variates, a more precise estimation? The answer is yes, and the

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

general idea is the following: If we want to estimate $\mu = \mathbf{E}X$, to find Y such that

$$\mu = \mathbf{E}X = \mathbf{E}Y, \qquad \mathrm{var}Y < \mathrm{var}X.$$

The way to produce a *good* Y usually departs from the knowledge that we can have about X. There are several methods to reduce variance, however there does not to exist a general method that always produce gain in the variance, the case is that each problem has its own good method.

2. Antithetic variates

The method is simple, and consists in using a symmetrized variate in the cases this is possible. For instance, if we want to compute $\mu = \int_0^1 f(x)dx$, we would have, with U uniform in $[0,1]$,

$$X = f(U), \quad Y = \frac{1}{2}\left(f(U) + f(1-U)\right).$$

We have

$$\mathrm{var}Y = \frac{1}{2}(\mathrm{var}X + \mathbf{cov}(f(U), f(1-U))) \leq \mathrm{var}X.$$

If $\mathbf{cov}(f(U), f(1-U)) < \mathrm{var}(X)$ we have variance reduction.

Proposition 1 *If f is not symmetric (i.e. if $f(x) \neq f(1-x)$ for some x in case of continuity) then*

$$\mathrm{var}Y < \mathrm{var}X$$

Proof. As we have seen, $\mathrm{var}X < \mathrm{var}Y$ iff $\mathbf{cov}(f(U), f(1-U)) < \mathrm{var}f(U)$, and this holds iff

$$\int_0^1 f(x)f(1-x) < \int_0^1 f(x)^2 dx,$$

that is equivalent to

$$2\int_0^1 f(x)f(1-x) < \int_0^1 f(x)^2 dx + \int_0^1 f(1-x)^2 dx,$$

that is equivalent to

$$\int_0^1 (f(x) - f(1-x))^2 dx > 0.$$

This holds iff f is not symmetric.

Remark. For any r.v. with symmetric (w.r.t. μ) distribution, a similar argument holds: the method reduces the variance in the case the function is not symmetric w.r.t. μ.

2.1. Example: Uniform random variables

We compute

$$\pi = 4\int_0^1 \sqrt{1-x^2}dx,$$

with $n = 10^6$ variates. Our results

	Estimate	Variance
Classical Estimate	3.141379	0.000553
Antithetic Variates	3.141536	0.000205
True value	3.141593	

2.2. Example: Tail probabilities of normal random variables

We want to compute the probability that a standard normal variable is larger than 3:

$$\mu = \mathbf{P}(Z > 3), \quad \widehat{\mu} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}\{Z_k > 3\},$$

$$\widehat{\mu}_A = \frac{1}{2n}\sum_{k=1}^{n}(\mathbf{1}\{Z_k > 3\} + \mathbf{1}\{-Z_k > 3\}).$$

Our results with $n = 10^4$ variates:

	Estimate	Variance
Classical Estimate	0.00121	0.00022
Antithetic Variates	0.00137	0.00016
True value	0.0013499	

3. Importance sampling

The importance sampling method consists in changing the underlying distribution of the variable used to simulate. It is specially suited for the estimation of small probabilities (*rare events*). Assuming that $X \sim f$ and $Y \sim g$, it is based in the following identity

$$\mu = \mathbf{E}h(X) = \int h(x)f(x)dx = \int \frac{h(x)f(x)}{g(x)}g(x)dx = \mathbf{E}H(Y),$$

where we define $H(x) = \frac{h(x)f(x)}{g(x)}$. The main idea is to achieve that Y points towards the set where h takes large values. If not correctly applied, the method can enlarge the variance.

3.1. Computation of the variance

The variance of the method is

$$\mathrm{var}H(Y) = \mathbf{E}(H(Y)^2) - (\mathbf{E}H(Y))^2 = \mathbf{E}(H(Y)^2) - \mu^2.$$

As μ is fixed, we should minimize

$$\mathbf{E}(H(Y)^2) = \int \frac{(h(x)f(x))^2}{g(x)} dx.$$

3.2. Example: Tail probabilities of normal random variables

$$\begin{aligned}\mu = \mathbf{P}(Z > 3) &= \int_3^\infty \frac{e^{-x^2/2}}{\sqrt{2\pi}} \frac{e^{-(x-3)^2/2}}{e^{-(x-3)^2/2}} dx \\ &= \int_3^\infty e^{-3x+9/2} \frac{e^{-(x-3)^2/2}}{\sqrt{2\pi}} dx = \mathbf{E} e^{-3Y+9/2} \mathbf{1}_{\{Y>3\}},\end{aligned}$$

where $Y \sim \mathbf{N}(3,1)$. Our results with $n = 10^4$ variates:

	Estimate	Variance
Classical Estimate	0.00121	2.2e-04
Antithetic Variates	0.00137	1.6e-04
Importance sampling	0.001340	1.5e-05
True value	0.0013499	

4. Control variates

Given the problem of simulating $\mu = \mathbf{E}h(X)$ the idea is to "control" the function h through a function g, close as posible to h, and such that *we know* $\beta = \mathbf{E}g(Y)$. We can add a constant c to better adjustment. More concretely, the equation is

$$\begin{aligned}\mu = \mathbf{E}h(X) &= \mathbf{E}h(X) - c(\mathbf{E}g(X) - \beta) \\ &= \mathbf{E}(h(X) - cg(X)) + c\beta.\end{aligned}$$

The coefficient c can be chosen in order to minimize the variance:

$$\mathrm{var}(h(X) - cg(X)) = \mathrm{var}h(X) + c^2 \mathrm{var}g(X) - 2c\,\mathbf{cov}(h(X), g(X)).$$

This gives a minimum when

$$c^* = \frac{\mathbf{cov}(h(X), g(X))}{\mathrm{var}(g(X))}.$$

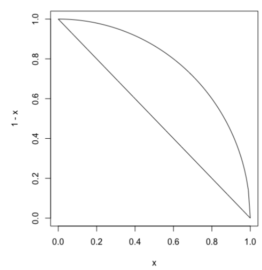
As this quantities are usually unknown, we can first run a MC to estimate c^*. obtaining the following variance:

$$\begin{aligned}\mathrm{var}(h(X) - cg(X)) &= \mathrm{var}(h(X)) - \frac{\mathbf{cov}(h(X), g(X))^2}{\mathrm{var}(g(X))} \\ &= (1 - \rho(h(X), g(X))^2)\mathrm{var}(h(X))\end{aligned}$$

As $\rho(h(X), g(X)) \leq 1$, we usually obtain a variance reduction.

4.1. Example: The computation of π

We choose $g(x) = 1 - x$, that is close to $\sqrt{1 - x^2}$



We first estimate c. In this case we know $\beta = \mathbf{E}(1 - U) = 1/2$ and $\mathrm{var}(1 - U) = 1/12$. After simulation we obtiain

$$\widehat{c^*} \sim 0{,}7$$

So we estimate

$$\pi = 4\mathbf{E}(\sqrt{1 - U^2} - 0{,}7(1 - U - 1/2)).$$

Our results with $n = 10^6$:

	Estimate	Variance
Classical Estimate	3.141379	0.000553
Antithetic Variates	3.141536	0.000205
Control variates	3.141517	0.000215
True value	3.141593	

5. Stratified sampling

The idea[1] to reduce the variance that this method proposes is to produce a partition of the probability space Ω, and distribute the effort of sampling in

[1] Adapted from *Simulation* 5ed.S. M. Ross, (2013) Elsevier

each set of the partition. Suppose we want to estimate

$$\mu = \mathbf{E}(X),$$

and suppose there is some discrete random variable Y, with possible values $y_1, \ldots, y_k$, such that, for each $i = 1, \ldots, k$:

(a) the probability $p_i = \mathbf{P}(Y = y_i)$, is known;

(b) we can simulate the value of X conditional on $Y = y_i$.

The proposal is to estimate

$$\mathbf{E}(X) = \sum_{i=1}^{k} \mathbf{E}(X|Y = y_i)p_i,$$

by estimating the k quantities $\mathbf{E}(X|Y = y_i)$, $i = 1, \ldots, k$. So, rather than generating n independent replications of X, we do np_i of the simulations conditional on the event that $Y = y_i$ for each $i = 1, \ldots, k$. If we let $\widetilde{X}_i$ be the average of the np_i observed values of $X|Y = y_i$, then we would have the unbiased estimator

$$\widehat{\mu} = \sum_{i=1}^{k} \widetilde{X}_i p_i$$

that is called a *stratified sampling* estimator of $\mathbf{E}(X)$.

To compute the variance, we first have

$$\mathrm{var}(\widetilde{X}_i) = \frac{\mathrm{var}(X|Y = y_i)}{np_i}$$

Consequently, using the preceding and that the $\widetilde{X}_i$, are independent, we see that

$$\begin{aligned}\mathrm{var}(\widehat{\mu}) &= \sum_{i=1}^{k} p_i^2 \mathrm{var}(\widetilde{X}_i) \\ &= \frac{1}{n}\sum_{i=1}^{k} p_i \mathrm{var}(X|Y = y_i) = \frac{1}{n}\mathbf{E}(\mathrm{var}(X|Y)).\end{aligned}$$

Because the variance of the classical estimator is $\frac{1}{n}\mathrm{var}(X)$, and $\mathrm{var}(\widehat{\mu}) = \frac{1}{n}\mathbf{E}(\mathrm{var}(X|Y))$, we see from the conditional variance formula

$$\mathrm{var}(X) = \mathbf{E}(\mathrm{var}(X|Y)) + \mathrm{var}(\mathbf{E}(X|Y)),$$

that the variance reduction is

$$\frac{1}{n}\mathrm{var}(X) - \frac{1}{n}\mathbf{E}(\mathrm{var}(X|Y)) = \frac{1}{n}\mathrm{var}\mathbf{E}(X|Y),$$

That is, the variance savings per run is $\mathrm{var}\mathbf{E}(X|Y)$ which can be substantial when the value of Y strongly affects the conditional expectation of X. On the contrary, if X and Y are independent, $\mathbf{E}(X|Y) = \mathbf{E}X$ and $\mathrm{var}\mathbf{E}(X|Y) = 0$.

Observe that the variance of the stratified sampling estimator can be estimated by

$$\widehat{\mathrm{var}(\widehat{\mu})} = \frac{1}{n}\sum_{i=1}^{k} p_i^2 s_i,$$

if s_i is the usual estimator of the sample of $X|Y = y_i$.

Remark: The simulation of np_i variates for each i is called the *proportional sampling*. Alternatively, one can choose $n_1, \ldots, n_k$ s.t. $n_1 + \cdots + n_k = n$ that *minimize* the variance.

5.1. Example: Integrals in $[0, 1]$

Suppose that we want to estimate

$$\mu = \mathbf{E}(h(U)) = \int_0^1 h(x)dx.$$

We put

$$Y = j, \text{ if } \frac{j-1}{n} \leq U < \frac{j}{n}, \text{ for } j = 1, \ldots, n.$$

We have

$$\mu = \mathbf{E}\mathbf{E}(h(U)|Y) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{E}(h(U_{(j)})),$$

where $U_{(j)}$ is uniform in $j - 1 \leq U < j$. In this example we have $k = n$, and we use $n_i = 1$ variates for each value of Y. As

$$U_{(j)} \sim \frac{U + j - 1}{n},$$

the resulting estimator is

$$\widehat{\mu} = \frac{1}{n}\sum_{j=1}^{n} h\left(\frac{U_j + j - 1}{n}\right).$$

To compute the variance, we have

$$\begin{aligned}
\mathrm{var}(\widehat{\mu}) &= \frac{1}{n^2}\sum_{j=1}^{n} \mathrm{var} h\left(\frac{U + j - 1}{n}\right) \\
&= \frac{1}{n}\sum_{j=1}^{n} \int_{\frac{j-1}{n}}^{\frac{j}{n}} (h(x) - \mu_j)^2 dx,
\end{aligned}$$

where $\mu_j = \int_{\frac{j-1}{n}}^{\frac{j}{n}} h(x)dx$.

The reduction is obtained because μ_j is closer to h than μ:

$$\operatorname{var}(\widehat{\mu}_C) = \frac{1}{n}\int_0^1 (h(x)-\mu)^2 dx,$$

where $\widehat{\mu}_C$ stands for the classic MC estimator.

5.2. Example: computation of π

We return to

$$\pi = 4\int_0^1 \sqrt{1-x^2}dx.$$

Observing that

$$\frac{j-U}{n} \sim \frac{U+j-1}{n} \sim U_{(j)},$$

we combine stratified and antithetic sampling:

$$\widehat{\mu} = \frac{2}{n}\sum_{j=1}^{n}\left(\sqrt{1-\left(\frac{U_j+j-1}{n}\right)^2} + \sqrt{1-\left(\frac{j-U_j}{n}\right)^2}\right)$$

For $n = 10^5$ we obtain an estimation

$$\widehat{\mu} = 3{,}14159265\,37 \quad \pi = 3{,}14159265\,358979$$

with 10 correct digits.

6. Conditional sampling

Remember the telescopic (or "tower") property of the conditional expectation:

$$\mathbf{E}(X) = \mathbf{E}(\mathbf{E}(X|\theta)).$$

where θ is an auxiliar random variable. In case we are able to simulate

$$Y = \mathbf{E}(X|\theta),$$

we have the following variance reduction:

$$\operatorname{var}(Y) = \operatorname{var}(\mathbf{E}(X|\theta)) = \operatorname{var}(X) - \mathbf{E}(\operatorname{var}(X|\theta)).$$

6.1. Example: Computing an expectation

Let U be uniform in $[0,1]$ and $Z \sim \mathbf{N}(0,1)$. We want to compute

$$\mu = \mathbf{E}(e^{UZ}).$$

We first compute

$$\mathbf{E}(e^{UZ}|U=u) = \int_{\mathbf{R}} e^{ux}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = e^{u^2/2},$$

so $Y = \mathbf{E}(e^{UZ}|U) = e^{U^2/2}$, and $\mathbf{E}Y = \int_0^1 e^{u^2/2}du$.

Our results for two size samples.

	$n=10^3$	$n=10^6$
Classical	$1{,}2145 \pm 0{,}020$	$1{,}1951 \pm 0{,}00060$
Conditional	$1{,}1962 \pm 0{,}004$	$1{,}1949 \pm 0{,}00012$
True	1.194958	

Note that the classical method requires $2n$ samples.

7. Exercises

When not specified confidence level is $1-\alpha = 0{,}95$ and sample size $n = 10^6$.

The general purpose is to estimate using different methods of variance reduction the following quantities

$$\begin{aligned}
\mu_1 &= 4\int_0^1 \sqrt{1-x^2}dx = \pi,\\
\mu_2 &= \int_0^1 \sqrt{-\log x}dx = \frac{1}{2}\sqrt{\pi},\\
\mu_3 &= \mathbf{P}(Z>4), \text{ where } Z \sim \mathbf{N}(0,1)\\
\mu_4 &= \mathbf{E}(e^Z-5)^+, \text{ where } Z \sim \mathbf{N}(0,1)
\end{aligned}$$

In all exercises use $n = 10^6$ samples, and compute the corresponding errors with 0,95 % confidence:

$$\epsilon = \frac{1{,}96s}{\sqrt{n}}.$$

where $s = \widehat{\sigma}$.

Exercise 1. Compute μ_1 with the following methods:

(a) Acceptance rejection on the square $[0,1]^2$.

(b) Sample mean method.

(c) Use antithetic variables.

(d) Use control variates, with $g(x) = 1 - x$. First find the optimal c^* with a small sample (10^3) and then run your algorithm with this estimation.

(e) Use the stratified method suggested in Lecture 3, i.e. using, for $\mu = \int_0^1 h(x)dx$ the formula

$$\widehat{\mu} = \frac{1}{n}\sum_{j=1}^{n} h\left(\frac{U_j + j - 1}{n}\right).$$

You can combine with antithetic variates (In this case it is no direct to obtain an error estimate).

(f) Do you know a deterministic method to compute this integral? .For instance, compute a Riemann sum, or use the trapezoidal rule:

$$\widehat{\mu} = \frac{1}{2n}(f(0) + f(1)) + \frac{1}{n}\sum_{j=1}^{n-1} f(j/n).$$

Exercise 2. Compute μ_2 with the following methods:

(a) Can you use acceptance rejection method with uniform variables?

(b) Use the sample mean method.

(c) Use antithetic variables.

(d) Use control variates, with $g(x) = -\log x$, taking into account that $\int_0^1 -\log(x)dx = 1$. First find the optimal c^* with a small sample (10^3) and then run your algorithm with this estimation.

Exercise 3. Compute μ_3 with the following methods:

(a) First use crude MC. As the probability is very small, a large n is necessary.

(b) Check if the antithetic variates method improves the situation.

(c) Use the importance sampling method, based in the following idenantity.

$$\begin{aligned}\mathbf{P}(Z > 4) &= \int_4^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \int_4^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{e^{-(x-4)^2/2}}{e^{-(x-4)^2/2}} dx \\ &= \int_4^\infty e^{-4x+8} \frac{e^{-(x-4)^2/2}}{\sqrt{2\pi}} dx = \mathbf{E} e^{-4X+8} \mathbf{1}_{\{X>4\}}.\end{aligned}$$

where $X \sim \mathbf{N}(4,1)$.

Exercise 4. Compute μ_4 with the following methods:

(a) First use crude MC.

(b) Use antithetic variates.

(c) Now we use control variates in the following way. Check the following identity:

$$\mathbf{E}(e^Z - K)^+ = \mathbf{E}(e^Z) - K + \mathbf{E}(K - e^Z)^+.$$

(here $x^+ = \text{máx}(0, x)$, and you can use that $x = x^+ - (-x)^+$). Then, computing $\mathbf{E}e^Z = e^{1/2}$, we find the price of the *put option*, given by

$$P(K) = \mathbf{E}(K - e^Z)^+.$$

You can use also antithetic variates in this situation.

(d) Importance sampling can be implemented in the following way. Check the following identity:

$$\mathbf{E}(e^Z - K)^+ = \int_{\mathbf{R}} (e^x - K)^+ \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = e^{1/2} \int_{\mathbf{R}} (1 - Ke^{-x})^+ \frac{1}{\sqrt{2\pi}} e^{-(x-1)^2/2} dx.$$

that is known as *put-call duality*.

$$P(K) = e^{1/2}\mathbf{E}(1 - Ke^{-X})^+.$$

where $X \sim \mathbf{N}(1, 1)$. You can use also antithetic variates in this situation.

Exercise 5. We want to compute the integral by simulation:

$$\mu = \int_0^1 (1 - x)e^{-x^2} dx$$

In all cases provide the 95 % error of estimation.

(a) Estimate μ using the acceptance-rejection method on the square $[0, 1]^2$.

(b) Estimate μ using uniform random variables, by the sample mean method.

(c) Estimate μ using random variables with density $f(x)$ of exercise 1.

(d) Use antithetic variables.

(e) Use control variates, with $g(x) = 1 - x$. First find the optimal c^* with a small sample (10^3) and then run your algorithm with this estimation.

Exercise 6. We want to estimate by simulation:

$$\mu = \int_0^1 \frac{1}{2\sqrt{x + x^2}} dx = \int_0^1 \frac{1}{2\sqrt{x}} \frac{1}{\sqrt{1 + x}} dx.$$

In all cases provide the 95 % error of estimation.

(a) Estimate μ using uniform random variables, by the sample mean method.

(b) Estimate μ using random variables with density

$$f(x) = \begin{cases} 1/(2\sqrt{x}), & \text{when } 0 \le x \le 1, \\ 0, & \text{in other case.} \end{cases}$$

(c) Use antithetic variables.

(d) Use control variates, with $g(x) = 1 - x$. First find the optimal c^* with a small sample (10^3) and then run your algorithm with this estimation.

Exercise 7. Volume of a sphere in $\mathbf{R}^6$.

1. Evaluate by the method of Monte Carlo the volume of a unit sphere in dimension 6. (Compare with the true value $\pi^3/6$).

2. We now use the method of Monte-Carlo with a different law Let Q be the law in $\mathbf{R}^6$ with density

$$q(x) = \frac{1}{Z(\alpha)} \exp(-\alpha \sum_{i=1}^{6} |x_i|)$$

   and $Z(\alpha)$ is the normalizing constant.

3. Compute $Z(\alpha)$.

4. Plot a code to simulate a point in $\mathbf{R}^6$ with law Q.

5. Devise a method to compute the volume of the sphere using the distribution Q using importance sampling. Compute the error of the method as a function of α

6. Determine a reasonable interval for α and propose a value to minimize the error.