Stochastic Processes and Financial Mathematics (37363)

Chapter 3

Methods of stochastic simulation

Alex Novikov and Scott Alexander

School of Mathematical and Physical Sciences, UTS

Autumn 2025

Chapter outline

Topics:

- Convergence of RVs
- Limit theorems
- Stochastic simulation
  - Monte Carlo approximation of integrals
  - MC variance reduction
- Simulation of RVs

Convergence of RVs

We are interested in two of the most important limit theorems in probability, the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT), both of which are very important in applications.

The LLN states that, given certain conditions, the **arithmetic mean** of a sufficiently large number of independent and identically distributed (iid) RVs, each with a well-defined (finite) expected value will be **approximately equal to a constant**.

The CLT states that, given certain conditions, the **normalised arithmetic mean** of a sample of iid RVs will be **approximately normally distributed, regardless of the underlying distribution**.

Both LLN and CLT are used to justify approximations obtained by simulation of RVs using the Monte Carlo and other approaches.

Convergence of RVs

Approximations of RVs are often of the form

$$\xi_n \approx \xi$$

where n, the sample size of simulations, is a parameter of interest.

Example.
As discussed in Chapter 1, we have the convergence of frequency to probability

$$\frac{n(A)}{n} \to P(A) \quad \text{as } n \to \infty$$

and so for sufficiently large n

$$\frac{n(A)}{n} \approx P(A).$$

In applications other parameters could be of interest, e.g. sample mean etc.

Convergence of RVs

To prove many properties and theorems, the convergence of sequences of RVs must be considered.

The proximity of RVs ξ_n and ξ may be understood in a number of ways.

Typical measures of proximity include probability

$$P(|\xi_n - \xi| > \varepsilon)$$

and mean square

$$E[|\xi_n - \xi|^2]$$

which give rise to weaker forms of convergence than pointwise convergence encountered in real analysis.

We are going to define three types of convergence.

Convergence of RVs

Definition 1 (convergence in probability)

A sequence of RVs ξ_n converges to ξ in probability if for any $\varepsilon > 0$

$$\lim_{n\to\infty} P(|\xi_n - \xi| > \varepsilon) = 0.$$

To indicate this form of convergence we write

$$\xi_n \xrightarrow{P} \xi \quad \text{or} \quad \lim_{n\to\infty} \xi_n \stackrel{P}{=} \xi.$$

Other properties follow from this definition.

Proposition 1

If $\xi_n \xrightarrow{P} \xi$ and $\eta_n \xrightarrow{P} \eta$ then

$$(\xi_n + \eta_n) \xrightarrow{P} \xi + \eta \quad \text{and} \quad (\xi_n \eta_n) \xrightarrow{P} \xi\eta.$$

Convergence of RVs

A stronger form of convergence is "mean square" or "L^2".

Definition 2 (convergence in mean square)

A sequence of RVs ξ_n converges to ξ in mean square if

$$\lim_{n\to\infty} E[|\xi_n - \xi|^2] = 0$$

where $E[|\xi_n|^2], E[|\xi|^2] < \infty$. To indicate this form of convergence we write

$$\xi_n \xrightarrow{L^2} \xi \quad \text{or} \quad \lim_{n\to\infty} \xi_n \stackrel{L^2}{=} \xi.$$

The notation "L^2" refers to the L^2 function space. The interested reader is referred to [Rudin, 1976].

Note that $\xi_n \xrightarrow{L^2} \xi \Rightarrow \xi_n \xrightarrow{P} \xi$.

Convergence of RVs

The strongest form of convergence in probability is "almost sure".

Definition 3 (almost sure convergence)

A sequence of RVs ξ_n converges to ξ almost surely (or with probability one) if

$$P(\lim_{n\to\infty} \xi_n = \xi) = 1.$$

To indicate this form of convergence we write

$$\xi_n \overset{a.s.}{\to} \xi \quad \text{or} \quad \lim_{n\to\infty} \xi_n \overset{a.s.}{=} \xi.$$

This is weaker than pointwise convergence (sure convergence) encountered in real analysis. The details are technical and involve measure theory – the interested reader is referred to [Chung, 2001].

Note that $\xi_n \overset{a.s.}{\to} \xi \Rightarrow \xi_n \overset{P}{\to} \xi$.

Convergence of RVs

A particularly useful result is the Chebyshev-Markov inequality, which relates ideas from convergence in both probability and mean square.

Proposition 2 (Chebyshev-Markov inequality)

For any RVs ξ_n, ξ and $\varepsilon > 0$

$$P(|\xi_n - \xi| > \varepsilon) \leq \frac{E[|\xi_n - \xi|^2]}{\varepsilon^2}.$$

Proof.
We see that

$$\begin{aligned} E[|\xi_n - \xi|^2] &\geq E[|\xi_n - \xi|^2 I(|\xi_n - \xi| > \varepsilon)] \\ &\geq E[\varepsilon^2 I(|\xi_n - \xi| > \varepsilon)] = \varepsilon^2 E[I(|\xi_n - \xi| > \varepsilon)] \\ &= \varepsilon^2 P(|\xi_n - \xi| > \varepsilon). \end{aligned}$$

Convergence of RVs

As an aside we will show why the last step of the previous proof holds.

Recall the indictor function $I(X > a)$, with X a RV and $a \in \mathbb{R}$, is defined as

$$I(X > a) = \begin{cases} 1 & \text{if } X > a, \\ 0 & \text{if } X \leq a. \end{cases}$$

Now suppose that X is continuous with PDF f_X and observe

$$\begin{aligned} E[I(X > a)] &= \int_{-\infty}^{\infty} I(x > a) f_X(x) dx = \int_a^{\infty} f_X(x) dx \\ &= P(X > a). \end{aligned}$$

The result also holds if X has no PDF or if X is discrete.

Limit theorems

With these results we can derive the first of two very important limit theorems.

Theorem 1 (weak law of large numbers (WLLN) – Markov)

Let $X_1, X_2, \ldots, X_n$ be a sequence of uncorrelated RVs with

$$E[X_i] = m \quad \text{and} \quad \text{var}(X_i) \leq c < \infty$$

for all $i \in \{1, 2, \ldots, n\}$ where $m \in \mathbb{R}$ and $c > 0$. Then as $n \to \infty$

$$\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{P} m.$$

That is, the sample mean converges to the population mean as the sample size grows ever larger.

Proof.
Taking $\xi_n = \frac{1}{n}\sum_{i=1}^n X_i$ and

$$\xi = E[\xi_n] = E\Big[\frac{1}{n}\sum_{i=1}^n X_i\Big] = \frac{1}{n}\sum_{i=1}^n E[X_i] = \frac{1}{n}\sum_{i=1}^n m = m$$

in the Chebyshev-Markov inequality (Proposition 2) gives

$$P\Big(\Big|\frac{1}{n}\sum_{i=1}^n X_i - m\Big| > \varepsilon\Big) \leq \frac{E\Big[\Big|\frac{1}{n}\sum_{i=1}^n X_i - m\Big|^2\Big]}{\varepsilon^2}$$

for any $\varepsilon > 0$ where

$$E\Big[\Big|\frac{1}{n}\sum_{i=1}^n X_i - m\Big|^2\Big] = \text{var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n (\text{var}(X_i)) \leq \frac{cn}{n^2} = \frac{c}{n}$$

with second equality following from the zero covariance assumption.

Limit theorems

That is

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - m\right| > \varepsilon\right) \leq \frac{c}{n\epsilon^2}$$

so

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - m\right| > \epsilon\right) = 0$$

and we have by Definition 1 that $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} m$.

There is a version of this theorem by Kolmogorov called the strong law of large numbers (SLLN) where the convergence in probability is strengthened to almost sure convergence.

Limit theorems

Finally, the other very important limit theorem.

Theorem 2 (central limit theorem (CLT))

Let $X_1, X_2, \ldots, X_n$ be a sequence of iid RVs with

$$E[X_i] = m \quad \text{and} \quad \text{var}(X_i) = \sigma^2 < \infty$$

for all $i \in \{1, 2, \ldots, n\}$ and define

$$\xi_n = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - m}{\sigma/\sqrt{n}}$$

and the standard normal, or $N(0,1)$, distribution function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du.$$

Then for all $x \in \mathbb{R}$

$$\lim_{n\to\infty} P(\xi_n \leq x) \to \Phi(x).$$

Stochastic simulation

Monte-Carlo is a method of approximating the solution of complex physical or mathematical systems based on use of random variables.

The method was adopted and improved by John von Neumann and Stanislaw Ulam during the "Manhattan Project" of World War II, which produced the first atomic bomb.

Stanislaw Ulam (1909-1985). Polish-born US mathematician. Migrated to USA in 1936 and joined the Institute of Advanced Study at Princeton.

John von Neumann (1903-1957). Hungarian born US scientist and mathematician, a pioneer of computer design, was a professor at Princeton University from 1931.

Stochastic simulation

Books on Monte Carlo methods

- *Monte Carlo: Concepts, Algorithms, and Applications* [Fishman, 1995]
- *Monte Carlo Methods in Financial Engineering* [Glasserman, 2004]
- *Monte Carlo Methods in Finance* [Jäckel, 2002]

Stochastic simulation – Monte Carlo approximation

Let $\boldsymbol{X} = (X_1, \ldots, X_m)^T$ be a random vector on (Ω, F, P) and $g : \mathbb{R}^m \to \mathbb{R}$ be a given function with properly defined expectation

$$G := E[g(\boldsymbol{X})].$$

If we can generate a sequence of independent RVs $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, all from the same distribution as $\boldsymbol{X}$ (i.e. iid), then according to the WLLN (Theorem 1)

$$G_n := \frac{1}{n} \sum_{i=1}^{n} g(\boldsymbol{X}_i) \xrightarrow{P} E[g(\boldsymbol{X})] = G$$

which can be improved to almost sure convergence under the SLLN.

Stochastic simulation – Monte Carlo approximation

Note that G_n is an unbiased estimator of G as

$$E[G_n] = E\Big[\frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{X}_i)\Big] = \frac{1}{n}\sum_{i=1}^{n} E[g(\boldsymbol{X}_i)] = \frac{1}{n}\sum_{i=1}^{n} E[g(\boldsymbol{X})]$$
$$= E[g(\boldsymbol{X})] = G.$$

In particular, if $\boldsymbol{X}$ has PDF (density) $f_{\boldsymbol{X}}$ then

$$G = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_m) f_{\boldsymbol{X}}(x_1, \dots, x_m) dx_1 \cdots dx_m$$

and so we have an algorithm for approximating multidimensional integrals!

Stochastic simulation – Monte Carlo approximation

To estimate accuracy of the approximation, we assume

$$\sigma^2(g) := \mathrm{var}(g(\boldsymbol{X})) < \infty$$

and note that

$$\begin{aligned}\mathrm{var}(G_n) &= \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{X}_i)\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}\left(g(\boldsymbol{X}_i)\right) \quad \text{(independence)} \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}\left(g(\boldsymbol{X})\right) = \frac{\sigma^2(g)}{n}.\end{aligned}$$

Then by the CLT (Theorem 2)

$$\lim_{n\to\infty} \frac{\frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{X}_i) - E[g(\boldsymbol{X})]}{\sqrt{\mathrm{var}(g(\boldsymbol{X}))/n}} = \lim_{n\to\infty} \frac{G_n - G}{\sigma(g)/\sqrt{n}} \sim N(0,1).$$

Stochastic simulation – Monte Carlo approximation

This can be rearranged to give the $(1-\alpha)\%$ confidence interval, or error bound,

$$|G - G_n| \leq z_{1-\alpha/2}\frac{\sigma(g)}{\sqrt{n}}$$

where $z_{1-\alpha/2}$ is the $1-\frac{\alpha}{2}$ quantile from the $N(0,1)$ distribution.

In practice though, $\sigma^2(g)$ will rarely be known, but it can be estimated as

$$\widehat{\sigma}_n^2(g) := \frac{1}{n}\sum_{i=1}^{n} g^2(\boldsymbol{X}_i) - (G_n)^2 \quad \xrightarrow{P} \quad E[g^2(\boldsymbol{X})] - G^2 = \sigma^2(g).$$

Definition 4 (big O notation)

Let the functions f, r be defined on some unbounded subset of $\mathbb{R}_{>0}$ where r is positive and monotone. Then we write

$$f(x) = O(r(x)) \text{ as } x \to \infty$$

if there exists a $c > 0$ and sufficiently large $x_0 \in \mathbb{R}_{>0}$ such that

$$|f(x)| \leq cr(x) \text{ for all } x \geq x_0.$$

Using this definition we can write

$$|G_n - G| = O(n^{-\frac{1}{2}})$$

with probability close to one.

Comparison to quadrature.
The most commonly encountered multivariate quadrature (numerical integration) techniques is Simpson's rule.

Using Simpson's rule to estimate G results in an approximation error which is $O(n^{-\frac{4}{m}})$, although other conditions must also be satisfied to use this technique.

We see for $n > 1$ that $n^{-\frac{1}{2}} < n^{-\frac{4}{m}}$ when $m > 8$.

That is, for integration problems involving $m > 8$ dimensions, the Monte-Carlo approach is asymptotically faster than Simpson's rule!

Stochastic simulation – Monte Carlo approximation

Example.
Take

$$G = \int_0^1 xe^{-x}dx \approx 0.26424.$$

Solution 1. If we consider the RV $X \sim U(0,1)$ with PDF $f_X(x) = I(0 \leq x \leq 1)$ and the function $g(x) = xe^{-x}$, then this integral can be re-written as

$$\begin{aligned} G = E[g(X)] &= \int_{-\infty}^{\infty} xe^{-x} I(0 \leq x \leq 1)dx \\ &= \int_0^1 xe^{-x}dx \approx 0.264241 \end{aligned}$$

with

$$\begin{aligned} \sigma^2(g) &= \text{var}(g(X)) = E[g^2(X)] - E[g(X)]^2 \\ &= \int_{-\infty}^{\infty} x^2 e^{-2x} I(0 \leq x \leq 1)dx - G^2 = \int_0^1 x^2 e^{-2x}dx - G^2 \\ &\approx 0.0110075. \end{aligned}$$

Stochastic simulation – Monte Carlo approximation

Solution 2. If we consider the RV $X \sim \text{Exp}(1)$ with PDF $f_X(x) = e^{-x} I(x \geq 0)$ and the function $g(x) = xI(x \leq 1)$, then this integral can be re-written as

$$\begin{aligned} G = E[g(X)] &= \int_{-\infty}^{\infty} xI(x \leq 1) e^{-x} I(x \geq 0) dx \\ &= \int_0^1 x e^{-x} dx \approx 0.264241 \end{aligned}$$

with

$$\begin{aligned} \sigma^2(g) = \text{var}(g(X)) &= E[g^2(X)] - E[g(X)]^2 \\ &= \int_{-\infty}^{\infty} x^2 I^2(x \leq 1) e^{-x} I(x \geq 0) dx - G^2 \\ &= \int_0^1 x^2 e^{-x} dx - G^2 \approx 0.0907794 \end{aligned}$$

which is not as accurate as Solution 1.

Stochastic simulation – MC variance reduction

In the section above we introduced the "cude Monte Carlo" estimator for $G = E[g(\boldsymbol{X})]$, namely

$$G_n = \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{X}_i).$$

Besides simplicity, this has several nice properties:

1. it is consistent, i.e. $G_n \xrightarrow{P} E[g(\boldsymbol{X})]$
2. it is unbiased, i.e. $E[G_n] = E[g(\boldsymbol{X})]$
3. and it is asymptotically normal, i.e.

$$P\left(|G_n - G| < z\frac{\sigma(g)}{\sqrt{n}}\right) \to 2\Phi(z) - 1$$

where $\sigma^2(g) = \mathrm{var}(g(\boldsymbol{X}))$ and Φ is the $N(0,1)$ distribution function.

There are many methods for reduction of variance $\sigma^2(g)$. We shall discuss only two of the most popular: control variates and antithetic variates.

CONTROL VARIATES

A control variate is a RV $q(\boldsymbol{X})$ such that the constant

$$Q = E[q(\boldsymbol{X})]$$

is known.

Consider now the estimator

$$Q_n = \frac{1}{n}\sum_{i=1}^{n} q(\boldsymbol{X}_i)$$

which allows the construction of another estimator for G

$$\widetilde{G}_n = G_n - a(Q_n - Q)$$

where $a \in \mathbb{R}$ is a parameter chosen to minimise the variance of $\tilde{G}_n$.

Stochastic simulation – MC variance reduction

Since $G_n \xrightarrow{P} G$ and $Q_n \xrightarrow{P} Q$ we have

$$\widetilde{G}_n \xrightarrow{P} G.$$

Also, $\widetilde{G}_n$ is an unbiased estimator of G as

$$\begin{aligned}
E[\widetilde{G}_n] &= E[G_n - a(Q_n - Q)] \\
&= E[G_n] - aE[Q_n] + aQ \quad \text{(linearity, non-random terms)} \\
&= G - aE\Big[\frac{1}{n}\sum_{i=1}^{n} q(\boldsymbol{X}_i)\Big] + aQ \\
&= G - a\frac{1}{n}\sum_{i=1}^{n} E[q(\boldsymbol{X}_i)] + aQ \\
&= G - a\frac{1}{n}\sum_{i=1}^{n} E[q(\boldsymbol{X})] + aQ \\
&= G - aQ + aQ = G.
\end{aligned}$$

Stochastic simulation – MC variance reduction

Now assume that

$$\sigma^2(q) := \mathrm{var}(q(\boldsymbol{X})) < \infty$$

and define

$$\begin{aligned}\sigma^2(g - aq) &:= \mathrm{var}(g(\boldsymbol{X}) - aq(\boldsymbol{X})) \\ &= \sigma^2(g) + a^2\sigma^2(q) - 2a\,\mathrm{cov}(g(\boldsymbol{X}), q(\boldsymbol{X}))\end{aligned}$$

so that

$$\begin{aligned}\mathrm{var}(\widetilde{G}_n) &= \mathrm{var}(G_n - a(Q_n - Q)) \\ &= \mathrm{var}(G_n - aQ_n) \quad \text{(remove non-random terms)} \\ &= \mathrm{var}(G_n) + a^2\,\mathrm{var}(Q_n) - 2a\,\mathrm{cov}(G_n, Q_n) \\ &= \frac{\sigma^2(g - aq)}{n}.\end{aligned}$$

Then by the CLT (Theorem 2)

$$\lim_{n\to\infty} \frac{\widetilde{G}_n - G}{\sigma(g - aq)/\sqrt{n}} \sim N(0, 1).$$

Stochastic simulation – MC variance reduction

We see that like the crude MC estimator, the control variate estimator:

1. is consistent, i.e. $\widetilde{G}_n \xrightarrow{P} E[g(\boldsymbol{X})]$
2. is unbiased, i.e. $E[\widetilde{G}_n] = E[g(\boldsymbol{X})]$
3. and is asymptotically normal, i.e.

$$P\left(|\widetilde{G}_n - G| < z\frac{\sigma(g - aq)}{\sqrt{n}}\right) \to 2\Phi(z) - 1$$

where Φ is the $N(0, 1)$ distribution function.

Stochastic simulation – MC variance reduction

It remains to find the optimum value of a, a^*, where the minimum of $\sigma^2(g - aq)$ is attained.

This function is quadratic in a, so a^* satisfies

$$\frac{d}{da}\sigma^2(g - aq)|_{a=a^*} = 0$$

or

$$a^*\sigma^2(q) - \text{cov}(g(\boldsymbol{X}), q(\boldsymbol{X})) = 0$$

so

$$a^* = \frac{\text{cov}(g(\boldsymbol{X}), q(\boldsymbol{X}))}{\sigma^2(q)}.$$

Stochastic simulation – MC variance reduction

The minimum variance estimator $\widetilde{G}_n$ has

$$\begin{aligned}
\mathrm{var}(\widetilde{G}_n) &= \frac{\sigma^2(g - a^* q)}{n} \\
&= \frac{1}{n}\Big(\sigma^2(g) + \Big(\frac{\mathrm{cov}(g(\boldsymbol{X}), q(\boldsymbol{X}))}{\sigma^2(q)}\Big)^2 \sigma^2(q) \\
&\qquad - 2\frac{\mathrm{cov}(g(\boldsymbol{X}), q(\boldsymbol{X}))}{\sigma^2(q)} \mathrm{cov}(g(\boldsymbol{X}), q(\boldsymbol{X}))\Big) \\
&= \frac{1}{n}\Big(\sigma^2(g) - \frac{\mathrm{cov}(g(\boldsymbol{X}), q(\boldsymbol{X}))^2}{\sigma^2(q)}\Big) \\
&= \frac{\sigma^2(g)}{n}(1 - \rho^2_{g,q})
\end{aligned}$$

where

$$\rho_{g,q} := \frac{\mathrm{cov}(g(\boldsymbol{X}), q(\boldsymbol{X}))}{\sigma(g)\sigma(q)}.$$

It is now clear we should choose a control variate – we choose a $q(\boldsymbol{X})$ strongly correlated (positively or negatively) with $g(\boldsymbol{X})$.

Stochastic simulation – MC variance reduction

For example, if $|\rho_{g,q}| = \frac{9}{10}$ then

$$\text{var}(\widetilde{G}_n) = \frac{\sigma^2(g - a^* q)}{n} = \frac{19}{100}\frac{\sigma^2(g)}{n} = \frac{19}{100}\,\text{var}(G_n).$$

Instead of reducing MC variance the estimator can be used to reduce the number of simulations.

Let m and n be the number of simulations used in the crude and control variate MC estimators and note that

$$\begin{aligned} \text{var}(G_m) = \text{var}(\widetilde{G}_n) \quad &\Rightarrow \quad \frac{\sigma^2(g)}{m} = \frac{\sigma^2(g)(1-\rho_{g,q}^2)}{n} \\ &\Rightarrow \quad n = m(1-\rho_{g,q}^2) = \frac{19}{100}m. \end{aligned}$$

Even larger reductions in variance (or number of simulations) can be achieved with the use of two or more control variates.

Example.
Recall our earlier example using the RV $X \sim U(0,1)$ with PDF $f_X(x) = I(0 \leq x \leq 1)$ and the function $g(x) = xe^{-x}$, where we investigated crude MC estimation of

$$G = E[g(X)] = \int_{-\infty}^{\infty} xe^{-x} I(0 \leq x \leq 1) dx$$
$$= \int_0^1 xe^{-x} dx \approx 0.264241$$

with

$$\sigma^2(g) = \text{var}(g(X)) = E[g^2(X)] - E[g(X)]^2$$
$$= \int_{-\infty}^{\infty} x^2 e^{-2x} I(0 \leq x \leq 1) dx - G^2$$
$$= \int_0^1 x^2 e^{-2x} dx - G^2 \approx 0.0110075.$$

Stochastic simulation – MC variance reduction

Now we calculate the reduction in variance that can be obtained using the control variate $q(x) = x$ such that

$$\begin{aligned} Q = E[q(X)] &= \int_{-\infty}^{\infty} xI(0 \leq x \leq 1)dx \\ &= \int_0^1 xdx = 0.5 \end{aligned}$$

with

$$\begin{aligned} \sigma^2(q) = \text{var}(q(X)) &= E[q^2(X)] - E[q(X)]^2 \\ &= \int_{-\infty}^{\infty} x^2 I(0 \leq x \leq 1)dx - Q^2 \\ &= \int_0^1 x^2 dx - Q^2 \approx 0.0833333. \end{aligned}$$

Stochastic simulation – MC variance reduction

So

$$\begin{aligned}\text{cov}(g(X), q(X)) &= E[g(X)q(X)] - E[g(X)]E[q(X)] \\ &= \int_{-\infty}^{\infty} xe^{-x} xI(0 \leq x \leq 1)dx - GQ \\ &= \int_{0}^{1} x^2 e^{-x} dx - GQ \approx 0.0284822\end{aligned}$$

and

$$a^* = \frac{\text{cov}(g(X), q(X))}{\text{var}(q(X))} \approx 0.341787$$

giving

$$\rho_{g,q} = \frac{\text{cov}(g(X), q(X))}{\sqrt{\text{var}(g(X))\,\text{var}(q(X))}} \approx 0.940416.$$

We see that the ratio of variance for the two MC estimators

$$\frac{\text{var}(G_n)}{\text{var}(\widetilde{G}_n)} = \frac{\sigma^2(g)}{\sigma^2(g - a^* q)} = \frac{1}{1 - \rho_{g,q}^2} \approx 8.64913.$$

ANTITHETIC VARIATES

Antithetic variates are two RVs $g_1(\boldsymbol{X})$ and $g_2(\boldsymbol{X})$ with the same distribution so that

$$G = E[g_1(\boldsymbol{X})] = E[g_2(\boldsymbol{X})] = \frac{E[g_1(\boldsymbol{X}) + g_2(\boldsymbol{X})]}{2}$$

but with

$$\mathrm{cov}(g_1(\boldsymbol{X}), g_2(\boldsymbol{X})) < 0.$$

These allow the construction of another estimator for G

$$\widehat{G}_n = \frac{1}{2n} \sum_{i=1}^{n} \big(g_1(\boldsymbol{X}_i) + g_2(\boldsymbol{X}_i)\big).$$

Stochastic simulation – MC variance reduction

In a similar way as we showed $G_n \xrightarrow{P} G$ we have

$$\widehat{G}_n \xrightarrow{P} G.$$

Also, $\widehat{G}_n$ is an unbiased estimator of G as

$$\begin{aligned} E[\widehat{G}_n] &= E\Big[\frac{1}{2n}\sum_{i=1}^{n}\big(g_1(\boldsymbol{X}_i) + g_2(\boldsymbol{X}_i)\big)\Big] \\ &= \frac{1}{2n}\sum_{i=1}^{n}\big(E[g_1(\boldsymbol{X}_i)] + E[g_2(\boldsymbol{X}_i)]\big) \\ &= \frac{1}{2n}\sum_{i=1}^{n}\big(E[g_1(\boldsymbol{X})] + E[g_2(\boldsymbol{X})]\big) \\ &= \frac{1}{2n}\sum_{i=1}^{n} 2G = G. \end{aligned}$$

Stochastic simulation – MC variance reduction

Now assume that

$$\sigma^2(g_1) := \mathrm{var}(g_1(\boldsymbol{X})) < \infty$$

and

$$\sigma^2(g_2) := \mathrm{var}(g_2(\boldsymbol{X})) < \infty$$

and define

$$\begin{aligned}\sigma^2(g_1 + g_2) &:= \mathrm{var}(g_1(\boldsymbol{X}) + g_2(\boldsymbol{X})) \\ &= \sigma^2(g_1) + \sigma^2(g_2) + 2\,\mathrm{cov}(g_1(\boldsymbol{X}), g_2(\boldsymbol{X}))\end{aligned}$$

so that

$$\mathrm{var}(\widehat{G}_n) = \frac{\sigma^2(g_1 + g_2)}{4n}.$$

Then by the CLT (Theorem 2)

$$\lim_{n\to\infty} \frac{\widehat{G}_n - G}{\sigma(g_1 + g_2)/\sqrt{4n}} \sim N(0, 1).$$

Stochastic simulation – MC variance reduction

We see that like the crude MC estimator, the antithetic variate estimator:

1. is consistent, i.e. $\widehat{G}_n \xrightarrow{P} G$
2. is unbiased, i.e. $E[\widehat{G}_n] = G$
3. and is asymptotically normal, i.e.

$$P\left(|\widehat{G}_n - G| < z\frac{\sigma(g_1 + g_2)}{\sqrt{4n}}\right) \to 2\Phi(z) - 1$$

where Φ is the $N(0, 1)$ distribution function.

Stochastic simulation – MC variance reduction

If $\mathrm{cov}(g_1(\boldsymbol{X}), g_2(\boldsymbol{X})) < 0$ then

$$\begin{aligned}\mathrm{var}(\widehat{G}_n) &= \frac{\sigma^2(g_1 + g_2)}{4n} = \frac{\sigma^2(g_1) + \sigma^2(g_2) + 2\,\mathrm{cov}(g_1(\boldsymbol{X}), g_2(\boldsymbol{X}))}{4n} \\ &= \frac{2\sigma^2(g_1) + 2\,\mathrm{cov}(g_1(\boldsymbol{X}), g_2(\boldsymbol{X}))}{4n} \\ &< \frac{\sigma^2(g_1)}{2n} = \mathrm{var}(G_{2n}^{(1)})\end{aligned}$$

where $\mathrm{var}(G_{2n}^{(1)})$ is the crude MC estimator of $E[g_1(\boldsymbol{X})]$ using $2n$ simulations, which is comparable to the estimator $\frac{E[g_1(\boldsymbol{X}) + g_2(\boldsymbol{X})]}{2}$ using n simulations.

As stated earlier, there are many approaches to variance reduction (importance sampling, stratified sampling, measure transform etc.) – the interested reader is referred to [Borokov, 2014].

Simulation of RVs

INVERSE TRANSFORMATION METHOD – CONTINUOUS RV.
For any continuous strictly monotonic distribution function

$$F_X(x) = P(X \leq x)$$

denote by $F_X^{-1}(u)$, for any u such that $0 < u < 1$, its inverse function so that

$$F_X(F_X^{-1}(u)) = u.$$

If $U \sim U(0,1)$ then the RV

$$X = F_X^{-1}(U)$$

has the distribution function F_X.

Example (simulating an exponential RV).
An exponential RV X has PDF $f_X(x) = \lambda e^{-\lambda x}$, $x \geq 0$ and $\lambda > 0$, with distribution function

$$F_X(x) = \int_0^x f_X(z)dz = \int_0^x \lambda e^{-\lambda z} dz = 1 - e^{-\lambda x}.$$

Simulation of RVs

If $U \sim \text{Uniform}(0,1)$ then $1-U$ is also Uniform$(0,1)$. To see this note the CF of U is

$$E[e^{izU}] = \int_0^1 e^{izu} du = \frac{1}{iz} e^{izu}|_0^1 = \frac{1}{iz}(e^{iz} - 1)$$

and the CF of $1-U$ is

$$E[e^{iz(1-U)}] = e^{iz} \int_0^1 e^{-izu} du = -\frac{e^{iz}}{iz} e^{-izu}|_0^1 = \frac{1}{iz}(e^{iz} - 1).$$

Now solve $F_X(x)$ for x to obtain the inverse function

$$F_X^{-1}(F_X(x)) = x = -\frac{\ln(1 - F_X(x))}{\lambda}.$$

But $X = F_X^{-1}(U) \Rightarrow F_X(X) = U$ so

$$F_X^{-1}(U) = X = -\frac{\log(1-U)}{\lambda} \sim -\frac{\log(U)}{\lambda} \sim Exp(\lambda).$$

Simulation of RVs

There are many other methods of generating random numbers with specific distributions.

INVERSE TRANSFORMATION METHOD – DISCRETE RV.
If we need to simulate a r.v. X having probability mass function

$$P(X = x_j) = p_j, \quad j = 0, 1, \ldots,$$

we can use the following discrete analog of the inverse transform technique.

Let $U \sim U(0, 1)$ and set

$$X = \begin{cases} x_1, & U < p_1 \\ x_2, & p_1 < U < p_1 + p_2 \\ \vdots & \vdots \\ x_j, & \sum_i^{j-1} p_i < U < \sum_i^j p_i \\ \vdots & \vdots \end{cases}.$$

Simulation of RVs

The RV X has the required distribution because

$$P(X = x_j) = P\Big(\sum_{i=1}^{j-1} p_i < U < \sum_{i=1}^{j} p_i\Big) = \sum_{i=1}^{j} p_i - \sum_{i=1}^{j-1} p_i = p_j.$$

BOX-MULLER METHOD FOR NORMAL RVS

If $U_1, U_2 \sim \text{Uniform}(0,1)$ and independent then

$$\xi_1 = \sqrt{-2\ln(U_1)}\cos(2\pi U_2) \sim N(0,1)$$
$$\xi_2 = \sqrt{-2\ln(U_1)}\sin(2\pi U_2) \sim N(0,1)$$

and independent.

The interested reader is referred to [Glasserman, 2004].

Simulation of RVs

Having a method for generating the vector

$$\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)^T \sim N(\mathbf{0}, \boldsymbol{I_n})$$

of independent RVs $\xi_i \sim N(0,1)$ we may generate the normal random vector

$$\boldsymbol{X} = (X_1, \ldots, X_n)^T \sim N(\boldsymbol{m}, \boldsymbol{Q})$$

where $\boldsymbol{m} \in \mathbb{R}^n$ and $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$ where $\boldsymbol{u} \cdot \boldsymbol{Qu} > 0$ for all $\boldsymbol{u} \neq \mathbf{0}$.

From Chapter 2 we know we can do this via

$$\boldsymbol{X} = \boldsymbol{m} + \boldsymbol{H}^T \boldsymbol{\xi}$$

using the Cholesky decomposition

$$\boldsymbol{Q} = \boldsymbol{H}^T \boldsymbol{H}$$

where $\boldsymbol{H} \in \mathbb{R}^{n \times n}$ is an upper triangle matrix and $\boldsymbol{H}^T$ is a lower triangle matrix.

References I


Borokov, K. (2014).
Elements of Stochastic Modelling.
World Scientific, ?, 2nd edition.


Chung, K. (2001).
A Course in Probability Theory.
Academic Press, San Diego, CA, 3rd edition.


Fishman, G. (1995).
Monte Carlo: Concepts, Algorithms, and Applications.
Springer-Verlag, New York City, NY.


Glasserman, P. (2004).
Monte Carlo Methods in Financial Engineering.
Spriner, New York City, NY.


Jäckel, P. (2002).
Monte Carlo Methods in Finance.
Wiley, Sussex, England.

References II


Rudin, W. (1976).
Principles of Mathematical Analysis.
McGraw-Hill, New York City, NY, 3rd edition.