Introduction to variance reduction methods

Bernard Lapeyre
Halmstad, January 2007

All the results of the preceding lecture show that the ratio $\sigma/\sqrt{N}$ governs the accuracy of a Monte-Carlo method with N simulations. An obvious consequence of this fact is that one always has interest to rewrite the quantity to compute as the expectation of a random variable which has a smaller variance : this is the basic idea of variance reduction techniques. For complements, we refer the reader to [**?**],[**?**],[**?**] or [**?**].

Suppose that we want to evaluate $\mathbf{E}(X)$. We try to find an alternative representation for this expectation as

$$\mathbf{E}(X) = \mathbf{E}(Y) + C,$$

using a random variable Y with lower variance and C a known constant. A lot of techniques are known in order to implement this idea. This paragraph gives an introduction to some standard methods.

1 Control variates

The basic idea of control variate is to write $\mathbf{E}(f(X))$ as

$$\mathbf{E}(f(X)) = \mathbf{E}(f(X) - h(X)) + \mathbf{E}(h(X)),$$

where $\mathbf{E}(h(X))$ can be explicitly computed and $\text{Var}(f(X) - h(X))$ is smaller than $\text{Var}(f(X))$. In these circumstances, we use a Monte-Carlo method to estimate $\mathbf{E}(f(X) - h(X))$, and we add the value of $\mathbf{E}(h(X))$. Let us illustrate this principle by several financial examples.

Using call-put arbitrage formula for variance reduction Let S_t be the price at time t of a given asset and denote by C the price of the European call option

$$C = \mathbf{E}\left(e^{-rT}(S_T - K)_+\right),$$

and by P the price of the European put option

$$P = \mathbf{E}\left(e^{-rT}(K - S_T)_+\right).$$

There exists a relation between the price of the put and the call which does not depend on the models for the price of the asset,namely, the “call-put arbitrage formula” :

$$C - P = \mathbf{E}\left(e^{-rT}(S_T - K)\right) = S_0 - Ke^{-rT}.$$

This formula (easily proved using linearity of the expectation) can be used to reduce the variance of a call option since

$$C = \mathbf{E}\left(e^{-rT}(K - S_T)_+\right) + S_0 - Ke^{-rT}.$$

The Monte-Carlo computation of the call is then reduced to the computation of the put option.

Remark 1.1 For the Black-Scholes model explicit formulas for the variance of the put and the call options can be obtained. In most cases, the variance of the put option is smaller than the variance of the call since the payoff of the put is bounded whereas the payoff of the call is not. Thus, one should compute put option prices even when one needs a call prices.

Remark 1.2 Observe that call-put relations can also be obtained for Asian options or basket options.

For example, for Asian options, set $\bar{S}_T = \frac{1}{T}\int_0 S_s ds$. We have :

$$\mathbf{E}\left(\left(\bar{S}_T - K\right)_+\right) - \mathbf{E}\left(\left(K - \bar{S}_T\right)_+\right) = \mathbf{E}\left(\bar{S}_T\right) - K,$$

and, in the Black-Scholes model,

$$\mathbf{E}\left(\bar{S}_T\right) = \frac{1}{T}\int_0^T \mathbf{E}(S_s)ds = \frac{1}{T}\int_0^T S_0 e^{rs} ds = S_0 \frac{e^{rT} - 1}{rT}.$$

The Kemma and Vorst method for Asian options The price of an average (or Asian) put option with fixed strike is

$$M = \mathbf{E}\left(e^{-rT}\left(K - \frac{1}{T}\int_0^T S_s ds\right)_+\right).$$

Here $(S_t, t \geq 0)$ is the Black Scholes process

$$S_t = x\exp\left(\left(r - \frac{\sigma^2}{2}\right)t + \sigma W_t\right).$$

If σ and r are small enough, we can hope that

$$\frac{1}{T}\int_0^T S_s ds \text{ “is not too far from” } \exp\left(\frac{1}{T}\int_0^T \log(S_s)ds\right).$$

This heuristic argument suggests to use Y

$$Y = e^{-rT}\left(K - \exp(Z)\right)_+,$$

with $Z = \frac{1}{T}\int_0^T \log(S_s)ds$ as a control variate. As the random variable Z is Gaussian, we can explicitly compute

$$\mathbf{E}\left(e^{-rT}\left(K - e^Z\right)_+\right).$$

This is done by using the formula

$$\mathbf{E}\left(\left(K - e^Z\right)_+\right) = KN(-d) - e^{\mathbf{E}(Z) + \frac{1}{2}\text{Var}(Z)} N(-d - \sqrt{\text{Var}(Z)}),$$

where $d = \frac{\mathbf{E}(Z) - \log(K)}{\sqrt{\text{Var}(Z)}}$. For a proof of this formula, see exercise **??**, and use the call put parity relation.

This method is proposed in [**?**] and is very efficient when $\sigma \approx 0.3$ by year, $r \approx 0.1$ by year and $T \approx 1$ year. Of course, if the value of σ and r are larger, the gain obtained with this control variate is less significant but this method may remain useful.

Basket options. A very similar idea can be used for pricing basket options. Assume that, for $i = 1, \ldots, d$

$$S_T^i = x_i e^{\left(r - \frac{1}{2}\sum_{j=1}^p \sigma_{ij}^2\right) T + \sum_{j=1}^p \sigma_{ij} W_T^j},$$

where $W^1, \ldots, W^p$ are independent Brownian motions. Let a_i, $1 \le i \le p$, be positive real numbers such that $a_1 + \cdots + a_d = 1$. We want to compute a put option on a basket

$$\mathbf{E}\left((K - X)_+\right),$$

where $X = a_1 S_T^1 + \cdots + a_d S_T^d$. The idea is to approximate

$$\frac{X}{m} = \frac{a_1 x_1}{m} e^{rT + \sum_{j=1}^p \sigma_{1j} W_T^j} + \cdots + \frac{a_d x_d}{m} e^{rT + \sum_{j=1}^p \sigma_{dj} W_T^j}$$

where $m = a_1 x_1 + \cdots + a_d x_d$, by $\frac{Y}{m}$ where Y is the log-normal random variable

$$Y = m e^{\sum_{i=1}^d \frac{a_i x_i}{m}\left(rT + \sum_{j=1}^p \sigma_{ij} W_T^j\right)}.$$

As we can compute an explicit formula for

$$\mathbf{E}\left[(K - Y)_+\right],$$

one can to use the control variate $Z = (K - Y)_+$ and sample $(K - X)_+ - (K - Y)_+$.

A random volatility model. Consider the pricing of an option in a Black and Scholes model with stochastic volatility. The price $(S_t, t \ge 0)$ is the solution of the stochastic differential equation

$$dS_t = S_t\left(rdt + \sigma(Y_t)dW_t\right), S(0) = x,$$

where σ is a bounded function and Y_t is solution of another stochastic differential equation

$$dY_t = b(Y_t)dt + c(Y_t)dW'_t, Y_0 = y,$$

where $(W_t, t \ge 0)$ et $(W'_t, t \ge 0)$ are two independent Brownian motions. We want to compute

$$\mathbf{E}\left(e^{-rT} f(S_T)\right).$$

If the volatility of the volatility (i.e. $c(Y_t)$) is not too large, σ_t remains near its initial value σ_0. This suggests to use the control variate $e^{-rT} f(\bar{S}_T)$ where $\bar{S}_T$ is the solution of

$$d\bar{S}_t = \bar{S}_t\left(rdt + \sigma_0 dW_t\right), S(0) = x,$$

since $\mathbf{E}\left(e^{-rT} f(\bar{S}_T)\right)$ can be obtained by an explicit Black and Scholes formula, and to sample

$$e^{-rT} f(S_T) - e^{-rT} f(\bar{S}_T).$$

It is easy to check by simulation, using the standard estimate for the variance, that this procedure actually reduce the variance.

Using the hedge as a control variate. Another idea is to use an approximate hedge of the option as a control variate. Let $(S_t, t \geq 0)$ be the price of the asset. Assume that the price of the option at time t can be expressed as $C(t, S_t)$ (this fact is satisfied for Markovian models). Assume that, as in the previous example, an explicit approximation $\bar{C}(t,x)$ of $C(t,x)$ is known. Then one can use the control variate

$$Y = \sum_{k=1}^{N} \frac{\partial \bar{C}}{\partial x}(t_k, S_{t_k}) \left(\left(S_{t_{k+1}} - S_{t_k} \right) - \mathbf{E}\left(S_{t_{k+1}} - S_{t_k} \right) \right).$$

If $\bar{C}$ is closed to C and if N is large enough, a very large reduction of the variance can be obtained.

2 Importance sampling

Importance sampling is another variance reduction procedure. It is obtained by changing the sampling law.

We start by introducing this method in a very simple context. Suppose we want to compute

$$\mathbf{E}(g(X)),$$

X being a random variable following the density $f(x)$ on $\mathbf{R}$, then

$$\mathbf{E}(g(X)) = \int_{\mathbf{R}} g(x) f(x) dx.$$

Let $\tilde{f}$ be another density such that $\tilde{f}(x) > 0$ and $\int_{\mathbf{R}} \tilde{f}(x)dx = 1$. Clearly one can write $\mathbf{E}(g(X))$ as

$$\mathbf{E}(g(X)) = \int_{\mathbf{R}} \frac{g(x)f(x)}{\tilde{f}(x)} \tilde{f}(x) dx = \mathbf{E}\left(\frac{g(Y)f(Y)}{\tilde{f}(Y)} \right),$$

where Y has density $\tilde{f}(x)$ under $\mathbf{P}$. We thus can approximate $\mathbf{E}(g(X))$ by

$$\frac{1}{n}\left(\frac{g(Y_1)f(Y_1)}{\tilde{f}(Y_1)} + \cdots + \frac{g(Y_n)f(Y_n)}{\tilde{f}(Y_n)} \right),$$

where $(Y_1, \ldots, Y_n)$ are independant copies of Y. Set $Z = g(Y)f(Y)/\tilde{f}(Y)$. We gave decreased the variance of the simulation if $\text{Var}(Z) < \text{Var}(g(X))$. It is easy to compute the variance of Z as

$$\text{Var}(Z) = \int_{\mathbf{R}} \frac{g^2(x)f^2(x)}{\tilde{f}(x)} dx - \mathbf{E}(g(X))^2.$$

From this and an easy computation it follows that if $g(x) > 0$ and $\tilde{f}(x) = g(x)f(x)/\mathbf{E}(g(X))$ then $\text{Var}(Z) = 0$! Of course this result cannot be used in practice as it relies on the exact knowledge of $\mathbf{E}(g(X))$, which is the exactly what we want to compute. Nevertheless, it leads to a heuristic approach : choose $\tilde{f}(x)$ as a good approximation of $|g(x)f(x)|$ such that $\tilde{f}(x)/\int_{\mathbf{R}} \tilde{f}(x)dx$ can be sampled easily.

An elementary financial example Suppose that G is a Gaussian random variable with mean zero and unit variance, and that we want to compute

$$\mathbf{E}\left(\phi(G)\right),$$

for some function ϕ. We choose to sample the law of $\tilde{G} = G + m$, m being a real constant to be determined carefully. We have :

$$\mathbf{E}\left(\phi(G)\right) = \mathbf{E}\left(\phi(\tilde{G})\frac{f(\tilde{G})}{\tilde{f}(\tilde{G})}\right) = \mathbf{E}\left(\phi(\tilde{G})e^{-m\tilde{G}+\frac{m^2}{2}}\right).$$

This equality can be rewritten as

$$\mathbf{E}\left(\phi(G)\right) = \mathbf{E}\left(\phi(G+m)e^{-mG-\frac{m^2}{2}}\right).$$

Suppose we want to compute a European call option in the Black and Scholes model, we have

$$\phi(G) = \left(\lambda e^{\sigma G} - K\right)_+,$$

and assume that $\lambda << K$. In this case, $\mathbf{P}(\lambda e^{\sigma G} > K)$ is very small and unlikely the option will be exercised. This fact can lead to a very large error in a standard Monte-Carlo method. In order to increase to exercise probability, we can use the previous equality

$$\mathbf{E}\left(\left(\lambda e^{\sigma G} - K\right)_+\right) = \mathbf{E}\left(\left(\lambda e^{\sigma(G+m)} - K\right)_+ e^{-mG-\frac{m^2}{2}}\right),$$

and choose $m = m_0$ with $\lambda e^{\sigma m_0} = K$, since

$$\mathbf{P}\left(\lambda e^{\sigma(G+m_0)} > K\right) = \frac{1}{2}.$$

This choice of m is certainly not optimal; however it drastically improves the efficiency of the Monte-Carlo method when $\lambda << K$ (see exercise **??** for a mathematical hint of this fact).

The multidimensional case Monte-Carlo simulations are really useful for problems with large dimension, and thus we have to extend the previous method to multidimensional setting. The ideas of this section come from [**?**].

Let us start by considering the pricing of index options. Let σ be a $n \times d$ matrix and $(W_t, t \geq 0)$ a d-dimensional Brownian motion. Denote by $(S_t, t \geq 0)$ the solution of

$$\left\{\begin{array}{ccc} dS_t^1 & = & S_t^1\left(rdt + [\sigma dW_t]_1\right) \\ & \dots & \\ dS_t^n & = & S_t^n\left(rdt + [\sigma dW_t]_n\right) \end{array}\right.$$

where $[\sigma dW_t]_i = \sum_{j=1}^{d} \sigma_{ij} dW_t^j$.

Moreover, denote by I_t the value of the index

$$I_t = \sum_{i=1}^{n} a_i S_t^i,$$

where $a_1, \ldots, a_n$ is a given set of positive numbers such that $\sum_{i=1}^n a_i = 1$. Suppose that we want to compute the price of a European call option with payoff at time T given by

$$h = (I_T - K)_+ .$$

As

$$S_T^i = S_0^i \exp\left(\left(r - \frac{1}{2}\sum_{j=1}^d \sigma_{ij}^2\right) T + \sum_{j=1}^d \sigma_{ij} W_T^j\right),$$

there exists a function ϕ such that

$$h = \phi\left(G_1, \ldots, G_d\right),$$

where $G_j = W_T^j / \sqrt{T}$. The price of this option can be rewritten as

$$\mathbf{E}\left(\phi(G)\right)$$

where $G = (G_1, \ldots, G_d)$ is a d-dimensional Gaussian vector with unit covariance matrix.

As in the one dimensional case, it is easy (by a change of variable) to prove that, if $m = (m_1, \ldots, m_d)$,

$$\mathbf{E}\left(\phi(G)\right) = \mathbf{E}\left(\phi(G+m) e^{-m.G - \frac{|m|^2}{2}}\right), \tag{1}$$

where $m.G = \sum_{i=1}^d m_i G_i$ and $|m|^2 = \sum_{i=1}^d m_i^2$. In view of **??**, the variance $V(m)$ of the random variable

$$X_m = \phi(G+m) e^{-m.G - \frac{|m|^2}{2}}$$

is

$$\begin{aligned} V(m) &= \mathbf{E}\left(\phi^2(G+m) e^{-2m.G - |m|^2}\right) - \mathbf{E}\left(\phi(G)\right)^2, \\ &= \mathbf{E}\left(\phi^2(G+m) e^{-m.(G+m) + \frac{|m|^2}{2}} e^{-m.G - \frac{|m|^2}{2}}\right) - \mathbf{E}\left(\phi(G)\right)^2, \\ &= \mathbf{E}\left(\phi^2(G) e^{-m.G + \frac{|m|^2}{2}}\right) - \mathbf{E}\left(\phi(G)\right)^2. \end{aligned}$$

The reader is refered to [**?**] for an almost optimal way to choose the parameter m based on this representation.

We now extend this sort of techniques to the case of path dependent options. We use the Girsanov theorem.

The Girsanov theorem and path dependent options Let $(S_t, t \geq 0)$ be the solution of

$$dS_t = S_t\left(r dt + \sigma dW_t\right), S_0 = x,$$

where $(W_t, t \geq 0)$ is a Brownian motion under a probability $\mathbf{P}$. We want to compute the price of a path dependent option which payoff is given by

$$\phi(S_t, t \leq T) = \psi(W_t, t \leq T).$$

Common examples of such a situation are

- Asian options whose payoff is given by $f(S_T, \int_0^T S_s ds)$,
- Maximum options whose payoff is given by $f(S_T, \max_{s \le T} S_s)$.

We start by considering an elementary importance sampling technique. It is a straightforward extension of the technique used in the preceding example. For every real number λ define the process $(W_t^\lambda, t \le T)$ as

$$W_t^\lambda := W_t + \lambda t.$$

According to Girsanov theorem $(W_t^\lambda, t \le T)$ is a Brownian motion under the probability law $\mathbf{P}^\lambda$ defined by

$$\mathbf{P}^\lambda(A) = \mathbf{E}(L_T^\lambda \mathbf{1}_A), A \in \mathcal{F}_T,$$

where $L_T^\lambda = e^{-\lambda W_T - \frac{\lambda^2 T}{2}}$. Denote $\mathbf{E}^\lambda$ the expectation under this new probability $\mathbf{P}^\lambda$. For every bounded function ψ we have

$$\mathbf{E}\left(\psi(W_t, t \le T)\right) = \mathbf{E}^\lambda\left(\psi(W_t^\lambda, t \le T)\right) = \mathbf{E}\left(L_T^\lambda \psi(W_t^\lambda, t \le T)\right),$$

and thus

$$\mathbf{E}\left(\psi(W_t, t \le T)\right) = \mathbf{E}\left(e^{-\lambda W_T - \frac{\lambda^2 T}{2}} \psi(W_t + \lambda t, t \le T)\right).$$

For example, if we want to compute the price of fixed strike Asian option given by

$$P = \mathbf{E}\left(e^{-rt}\left(\frac{1}{T}\int_0^T x e^{\left(r - \frac{\sigma^2}{2}\right)s + \sigma W_s} ds - K\right)_+\right),$$

we can use the previous equality to obtain

$$P = \mathbf{E}\left(e^{-rt - \lambda W_T - \frac{\lambda^2 T}{2}}\left(\frac{1}{T}\int_0^T x e^{\left(r - \frac{\sigma^2}{2}\right)s + \sigma(W_s + \lambda s)} ds - K\right)_+\right).$$

This representation can be used in case of a deep out of the money option (that is to say, $x << K$). Then λ is chosen such that

$$\frac{x}{T}\int_0^T e^{\left(r - \frac{\sigma^2}{2}\right)s + \sigma \lambda s} ds = K.$$

3 Antithetic variables

The use of antithetic variables is widespread in Monte-Carlo simulation. This technique is often efficient but its gains are less dramatic than other variance reduction techniques.

We begin by considering a simple and instructive example. Let

$$I = \int_0^1 g(x) dx.$$

If U follows a uniform law on the interval $[0, 1]$, then $1 - U$ has the same law as U, and thus

$$I = \frac{1}{2}\int_0^1 (g(x) + g(1 - x)) dx = \mathbf{E}\left(\frac{1}{2}(g(U) + g(1 - U))\right).$$

Therefore one can draw n independent random variables $U_1, \ldots, U_n$ following a uniform law on $[0,1]$, and approximate I by

$$\begin{array}{rcl} I_{2n} & = & \frac{1}{n}\left(\frac{1}{2}(g(U_1)+g(1-U_1))+\cdots+\frac{1}{2}(g(U_n)+g(1-U_n))\right) \\ & = & \frac{1}{2n}\left(g(U_1)+g(1-U_1)+\cdots+g(U_n)+g(1-U_n)\right). \end{array}$$

We need to compare the efficiency of this Monte-Carlo method with the standard one with $2n$ drawings

$$\begin{array}{rcl} I_{2n}^0 & = & \frac{1}{2n}\left(g(U_1)+g(U_2)+\cdots+g(U_{2n-1})+g(U_{2n})\right) \\ & = & \frac{1}{n}\left(\frac{1}{2}(g(U_1)+g(U_2))+\cdots+\frac{1}{2}(g(U_{2n-1})+g(U_{2n}))\right). \end{array}$$

We will now compare the variances of I_{2n} and I_{2n}^0. Observe that in doing this we assume that most of numerical work relies in the evaluation of f and the time devoted to the simulation of the random variables is negligible. This is often a realistic assumption.

An easy computation shows that the variance of the standard estimator is

$$\mathrm{Var}(I_{2n}^0) = \frac{1}{2n}\mathrm{Var}\left(g(U_1)\right),$$

whereas

$$\begin{aligned} \mathrm{Var}(I_{2n}) &= \frac{1}{n}\mathrm{Var}\left(\frac{1}{2}(g(U_1)+g(1-U_1))\right) \\ &= \frac{1}{4n}\left(\mathrm{Var}(g(U_1))+\mathrm{Var}(g(1-U_1))+2\mathrm{Cov}(g(U_1),g(1-U_1))\right) \\ &= \frac{1}{2n}\left(\mathrm{Var}(g(U_1)+\mathrm{Cov}(g(U_1),g(1-U_1))\right). \end{aligned}$$

Obviously, $\mathrm{Var}(I_{2n}) \leq \mathrm{Var}(I_{2n}^0)$ if and only if $\mathrm{Cov}(g(U_1), g(1-U_1)) \leq 0$. One can prove that if f is a monotonic function this is always true (see **??** for a proof) and thus the Monte-Carlo method using antithetic variables is better than the standard one.

This ideas can be generalized in dimension greater than 1, in which case we use the transformation

$$(U_1, \ldots, U_d) \to (1-U_1, \ldots, 1-U_d).$$

More generaly, if X is a random variable taking its values in $\mathbf{R}^d$ and T is a transformation of $\mathbf{R}^d$ such that the law of $T(X)$ is the same as the law of X, we can construct an antithetic method using the equality

$$\mathbf{E}(g(X)) = \frac{1}{2}\mathbf{E}\left(g(X)+g(T(X))\right).$$

Namely, if $(X_1, \ldots, X_n)$ are independent and sampled along the law of X, we can consider the estimator

$$I_{2n} = \frac{1}{2n}\left(g(X_1)+g(T(X_1))+\cdots+g(X_n)+g(T(X_n))\right)$$

and compare it to

$$I_{2n}^0 = \frac{1}{2n}\left(g(X_1)+g(X_2))+\cdots+g(X_{2n-1})+g(X_{2n})\right).$$

The same computations as before prove that the estimator I_{2n} is better than the crude one if and only if $\mathrm{Cov}(g(X), g(T(X))) \leq 0$. We now show a few elementary examples in finance.

A toy financial example. Let G be a standard Gaussian random variable and consider the call option

$$\mathbf{E}\left(\left(\lambda e^{\sigma G}-K\right)_+\right).$$

Clearly the law of $-G$ is the same as the law of G, and thus the function T to be considered is $T(x)=-x$. As the payoff is increasing as a function of G, the following antithetic estimator certainly reduces the variance :

$$I_{2n}=\frac{1}{2n}\left(g(G_1)+g(-G_1)+\cdots+g(G_n)+g(-G_n)\right),$$

where $g(x)=\left(\lambda e^{\sigma x}-K\right)_+$.

Antithetic variables for path-dependent options. Consider the path dependent option with payoff at time T

$$\psi\left(S_s, s\leq T\right),$$

where $(S_t, t\geq 0)$ is the lognormal diffusion

$$S_t = x\exp\left((r-\frac{1}{2}\sigma^2)t+\sigma W_t\right).$$

As the law of $(-W_t, t\geq 0)$ is the same as the law of $(W_t, t\geq 0)$ one has

$$\mathbf{E}\left(\psi\left(x\exp\left((r-\frac{1}{2}\sigma^2)s+\sigma W_s\right), s\leq T\right)\right) = \mathbf{E}\left(\psi\left(x\exp\left((r-\frac{1}{2}\sigma^2)s-\sigma W_s\right), s\leq T\right)\right),$$

and, for appropriate functionals ψ, the antithetic variable method may be efficient.

4 Stratification methods

These methods are widely used in statistics (see [**?**]). Assume that we want to compute the expectation

$$I=\mathbf{E}(g(X))=\int_{\mathbf{R}^d} g(x)f(x)dx,$$

where X is a $\mathbf{R}^d$ valued random variable with density $f(x)$.

Let $(D_i, 1\leq i\leq m)$ be a partition of $\mathbf{R}^d$. I can be expressed as

$$I=\sum_{i=1}^m \mathbf{E}(\mathbf{1}_{\{X\in D_i\}}g(X))=\sum_{i=1}^m \mathbf{E}(g(X)|X\in D_i)\mathbf{P}(X\in D_i),$$

where

$$\mathbf{E}(g(X)|X\in D_i)=\frac{\mathbf{E}(\mathbf{1}_{\{X\in D_i\}}g(X))}{\mathbf{P}(X\in D_i)}.$$

Note that $\mathbf{E}(g(X)|X \in D_i)$ can be interpreted as $\mathbf{E}(g(X^i))$ where X^i is a random variable whose law is the law of X conditioned by X belongs to D_i, whose density is

$$\frac{1}{\int_{D_i} f(y)dy} \mathbf{1}_{\{x \in D_i\}} f(x)dx.$$

Remark 4.1 The random variable X^i is easily simulated using an acceptance rejection procedure. But this method is clearly unefficient when $\mathbf{P}(X \in D_i)$ is small.

When the numbers $p_i = \mathbf{P}(X \in D_i)$ can be explicitly computed, one can use a Monte-Carlo method to approximate each conditional expectation $I_i = \mathbf{E}(g(X)|X \in D_i)$ by

$$\tilde{I}_i = \frac{1}{n_i}\left(g(X_1^i) + \cdots + g(X_{n_i}^i)\right),$$

where $(X_1^i, \ldots, X_{n_i}^i)$ are independent copies of X^i. An estimator $\tilde{I}$ of I is then

$$\tilde{I} = \sum_{i=1}^m p_i \tilde{I}_i.$$

Of course the samples used to compute $\tilde{I}_i$ are supposed to be independent and so the variance of $\tilde{I}$ is

$$\sum_{i=1}^m p_i^2 \frac{\sigma_i^2}{n_i},$$

where σ_i^2 be the variance of $g(X^i)$.

Fix the total number of simulations $\sum_{i=1}^m n_i = n$. This minimization the variance above, one must choose

$$n_i = n \frac{p_i \sigma_i}{\sum_{i=1}^m p_i \sigma_i}.$$

For this values of n_i, the variance of $\tilde{I}$ is given in this case by

$$\frac{1}{n}\left(\sum_{i=1}^m p_i \sigma_i\right)^2.$$

Note that this variance is smaller than the one obtained without stratification. Indeed,

$$\begin{aligned}
\text{Var}\,(g(X)) &= \mathbf{E}\left(g(X)^2\right) - \mathbf{E}\left(g(X)\right)^2 \\
&= \sum_{i=1}^m p_i \mathbf{E}\left(g^2(X)|X \in D_i\right) - \left(\sum_{i=1}^m p_i \mathbf{E}\left(g(X)|X \in D_i\right)\right)^2 \\
&= \sum_{i=1}^m p_i \text{Var}\left(g(X)|X \in D_i\right) + \sum_{i=1}^m p_i \mathbf{E}\left(g(X)|X \in D_i\right)^2 \\
&\qquad - \left(\sum_{i=1}^m p_i \mathbf{E}\left(g(X)|X \in D_i\right)\right)^2.
\end{aligned}$$

Using the convexity inequality for x^2 we obtain $\left(\sum_{i=1}^m p_i a_i\right)^2 \leq \sum_{i=1}^m p_i a_i^2$ if $\sum_{i=1}^m p_i = 1$, and the inequality

$$\text{Var}\,(g(X)) \geq \sum_{i=1}^m p_i \text{Var}\,(g(X)|X \in D_i) \geq \left(\sum_{i=1}^m p_i \sigma_i\right)^2,$$

follows.

Remark 4.2 The optimal stratification involves the σ_i's which are seldom explicitly known. So one needs to estimate these σ_i's by Monte-Carlo simulations.

Moreover note that arbitrary choices of n_i may *increase* the variance. Common way to circumvent this difficulty is to choose

$$n_i = np_i.$$

The corresponding variance

$$\frac{1}{n}\sum_{i=1}^m p_i \sigma_i^2,$$

is always smaller than the original one as $\sum_{i=1}^m p_i \sigma_i^2 \leq \text{Var}\,(g(X))$. This choice is often made when the probabilities p_i can be computed. For more considerations on the choice of the n_i and also, for hints on suitable choices of the sets D_i, see [**?**].

A toy example in finance In the standard Black and Scholes model the price of a call option is

$$\mathbf{E}\left(\left(\lambda e^{\sigma G} - K\right)_+\right).$$

It is natural to use the following strata for G : either $G \leq d = \frac{\log(K/\lambda)}{\sigma}$ or $G > d$. Of course the variance of the stratum $G \leq d$ is equal to zero, so if you follow the optimal choice of number, you do not have to simulate points in this stratum : all points have to be sampled in the stratum $G \geq d$! This can be easily done by using the (numerical) inverse of the distribution function of a Gaussian random variable.

Of course, one does not need Monte-Carlo methods to compute call options for the Black and Scholes models; we now consider a more convincing example.

Basket options Most of what follows comes from [**?**]. The computation of an European basket option in a multidimensional Black-Scholes model can be expressed as

$$\mathbf{E}(h(G)),$$

for some function h and for $G = (G_1, \ldots, G_n)$ a vector of independent standard Gaussian random variables. Choose a vector $u \in \mathbf{R}^n$ such that $|u| = 1$ (note that $< u, G >= u_1 G_1 + \cdots + u_n G_n$ is also a standard Gaussian random variable.). Then choose a partition $(B_i, 1 \leq i \leq n)$ of $\mathbf{R}$ such that

$$\mathbf{P}(< u, G >\in B_i) = \mathbf{P}(G_1 \in B_i) = 1/n.$$

This can be done by setting

$$B_i =]N^{-1}((i-1)/n), N^{-1}(i/n)],$$

where N is the distribution function of a standard Gaussian random variable and N^{-1} is its inverse. We then define the strata by setting

$$D_i = \{< u, x > \in B_i\}.$$

In order to implement our stratification method we need to solve two simulation problems

- sample a Gaussian random variable $< u, G >$ given that $< u, G >$ belongs to B_i,
- sample a new vector G knowing the value $< u, G >$.

The first problem is easily solved since the law of

$$N^{-1}\left(\frac{i-1}{N} + \frac{U}{N}\right), \tag{2}$$

is precisely the law a standard Gaussian random variable conditioned to be in B_i.

To solve the second point, observe that

$$G- < u, G > u$$

is a Gaussian vector independent of $< u, G >$ with covariance matrix $I - u \otimes u'$ (where $u \otimes u'$ denotes the matrix defined by $(u \otimes u')_{ij} = u_i u_j$). Let Y be a copy of the vector G. Obviously $Y- < u, Y > u$ is independent of G and has the same law as $G- < u, G > u$. So

$$G =< u, G > u + G- < u, G > u \text{ and } < u, G > u + Y- < u, Y > u,$$

have the same probability law. This leads to the following simulation method of G given $< u, G >= \lambda$:

- sample n independent standard Gaussian random variables Y^i,
- set $G = \lambda u + Y- < u, Y > u$.

To make this method efficient, the choice of the vector u is crucial : an almost optimal way to choose the vector u can be found in [**?**].

5 Mean value or conditioning

This method uses the well known fact that conditioning reduces the variance. Indeed, for any square integrable random variable Z, we have

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

where Y is any random variable defined on the same probability space as Z. It is well known that $\mathbf{E}(Z|Y)$ can be written as

$$\mathbf{E}(Z|Y) = \phi(Y),$$

for some measurable function ϕ. Suppose in addition that Z is square integrable. As the conditional expectation is a L^2 projection

$$\mathbf{E}\left(\phi(Y)^2\right) \leq \mathbf{E}(Z^2),$$

and thus $\mathrm{Var}(\phi(Y)) \leq \mathrm{Var}(Z)$.

Of course the practical efficiency of simulating $\phi(Y)$ instead of Z heavily relies on an explicit formula for the function ϕ. This can be achieved when $Z = f(X,Y)$, where X and Y are independent random variables. In this case, we have

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

where $\phi(y) = \mathbf{E}(f(X,y))$.

A basic example. Suppose that we want to compute $\mathbf{P}\left(X \leq Y\right)$ where X and Y are independent random variables. This situation occurs in finance, when one computes the hedge of an exchange option (or the price of a digital exchange option).

Using the preceding, we have

$$\mathbf{P}\left(X \leq Y\right) = \mathbf{E}\left(F(Y)\right),$$

where F is the distribution function of X. The variance reduction can be significant, especially when the probability $\mathbf{P}\left(X \leq Y\right)$ is small.

A financial example : a stochastic volatility model. Let $(W_t, t \geq 0)$ be a Brownian motion. Assume that $(S_t, t \geq 0)$ follows a log-normal model with random volatility

$$dS_t = S_t\left(rdt + \sigma_t dW_t^1\right), S_0 = x,$$

where $(\sigma_t, t \geq 0)$ is a given continuous stochastic process independent of the Brownian motion $(W_t, t \geq 0)$. We want to compute the option price

$$\mathbf{E}\left(e^{-rT} f(S_T)\right),$$

where f is a given function. Clearly S_T can be expressed as

$$S_T = x\exp\left(rT - \int_0^T \sigma_t^2/2dt + \int_0^T \sigma_t dW_t^1\right).$$

As the processes $(\sigma_t, t \geq 0)$ and $(W_t, t \geq 0)$ are independent, we have

$$\int_0^T \sigma_t dW_t \text{ is equal in law to } \sqrt{\frac{1}{T}\int_0^T \sigma_t^2 dt} \times W_T.$$

Conditioning with respect to the process σ, we obtain

$$\mathbf{E}\left(e^{-rT} f(S_T)\right) = \mathbf{E}\left(\psi(\sigma_t, 0 \leq t \leq T)\right),$$

where, for a fixed volatility path $(v_t, 0 \le t \le T)$,

$$\psi(v_t, 0 \le t \le T) = \mathbf{E}\left(e^{-rT} f\left(x e^{rT - \int_0^T \frac{v_t^2}{2} dt + \sqrt{\frac{1}{T}\int_0^T v_t^2 dt} \times W_T}\right)\right)$$
$$= \phi\left(\sqrt{\frac{1}{T}\int_0^T v_t^2 dt}\right),$$

where $\phi(\sigma)$ is the price of the option in the standard Black and Scholes model with volatility σ, that is

$$\phi(\sigma) = \mathbf{E}\left(e^{-rT} f\left(x \exp\left(\left(r - \frac{\sigma^2}{2}\right) T + \sigma W_T\right)\right)\right).$$

6 Exercises and problems

Exercise 6.1 Let Z be a Gaussian random variable and K a positive real number.

1. Let $d = \frac{\mathbf{E}(Z) - \log(K)}{\sqrt{\text{Var}(Z)}}$, prove that

$$\mathbf{E}\left(\mathbf{1}_{Z \ge \log(K)}(e^Z)\right) = e^{\mathbf{E}(Z) + \frac{1}{2}\text{Var}(Z)} N\left(d + \sqrt{\text{Var}(Z)}\right).$$

2. Prove the formula (Black and Scholes formula)

$$\mathbf{E}\left(\left(e^Z - K\right)_+\right) = e^{\mathbf{E}(Z) + \frac{1}{2}\text{Var}(Z)} N\left(d + \sqrt{\text{Var}(Z)}\right) - K N(d),$$

Exercise 6.2 Consider the case of a European call in the Black and Scholes model with a stochastic interest rate. Suppose that the price of the stock is 1, and the option price at time 0 is given $\mathbf{E}(Z)$ with Z defined by

$$Z e^{-\int_0^T r_\theta d\theta}\left[e^{\int_0^T r_\theta d\theta - \frac{\sigma^2}{2} T + \sigma W_T} - K\right]_+.$$

1. Prove that the variance of Z is bounded by $\mathbf{E} e^{-\sigma^2 T + 2\sigma W_T}$.

2. Prove that $\mathbf{E} e^{-\frac{1}{2}\gamma^2 T + \gamma W_T} = 1$, and deduce an estimate for the variance of Z

Exercise 6.3 Let λ and K be two real positive numbers such that $\lambda < K$ and X_m be the random variable

$$X_m = \left(\lambda e^{\sigma(G+m)} - K\right)_+ e^{-mG - \frac{m^2}{2}}.$$

We denote its variance by σ_m^2. Give an expression for the derivative of σ_m^2 with respect to m as an expectation, then deduce that σ_m^2 is a decreasing function of m when $m \le m_0 = \log(K/\lambda)/\sigma$.

Problem 6.4 The aim of this problem is to prove that the antithetic variable method decreases the variance for a function which is monotonous with respect to each of its arguments.

1. Let f and g be two increasing functions from $\mathbf{R}$ to $\mathbf{R}$. Prove that, if X and Y are two real random variables then we have

$$\mathbf{E}\left(f(X)g(X)\right) + \mathbf{E}\left(f(Y)g(Y)\right) \geq \mathbf{E}\left(f(X)g(Y)\right) + \mathbf{E}\left(f(Y)g(X)\right).$$

2. Deduce that, if X is a real random variable, then

$$\mathbf{E}\left(f(X)g(X)\right) \geq \mathbf{E}\left(f(X)\right)\mathbf{E}\left(g(X)\right).$$

3. Prove that if $X_1, \ldots, X_n$ are n independent random variables then

$$\mathbf{E}\left(f(X_1, \ldots, X_n)g(X_1, \ldots, X_n)|X_n\right) = \phi(X_n),$$

where ϕ is a function which can be computed as an expectation.

4. Deduce from this property that if f and g are two increasing (with respect each of its argument) functions then

$$\mathbf{E}\left(f(X_1, \ldots, X_n)g(X_1, \ldots, X_n)\right) \\ \geq \mathbf{E}\left(f(X_1, \ldots, X_n)\right)\mathbf{E}\left(g(X_1, \ldots, X_n)\right).$$

5. Let h be a function from $[0,1]^n$ in $\mathbf{R}$ which is monotonous with respect to each of its arguments. Let $U_1, \ldots, U_n$ be n independent random variables following the uniform law on $[0,1]$. Prove that

$$\text{Cov}\left(h(U_1, \ldots, U_n)h(1-U_1, \ldots, 1-U_n)\right) \leq 0,$$

and deduce that in this case the antithetic variable method decreases the variance.

Problem 6.5 Let $W_t = (W_t^1, W_t^2)$ is a pair of independent Brownian motions, α_1,α_2,B_1,B_2 and A be regular functions from $\mathbf{R}^+$ in $\mathbf{R}$ and let σ be the 2×2-matrix given by

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}.$$

Assume that the price of a financial asset S_t can be writen as

$$S_t = \exp\left(A_t + B_1(t)X_t^1 + B_2(t)X_t^2\right),$$

where $X_t = (X_t^1, X_t^2)$ is a solution of

$$dX_t = \begin{pmatrix} \alpha_1(t) \\ \alpha_2(t) \end{pmatrix} dt + \sigma dW_t, X_0 = x_0.$$

The aim of this problem is to compute

$$V_0 = \mathbf{E}\left(e^{-\int_0^T X_s^1 ds} f(S_T)\right).$$

1. Prove that V_0 can be expressed as

$$V_0 = \mathbf{E}\left(e^{-\lambda_1 \int_0^T W_s^1 ds - \lambda_2 \int_0^T W_s^2 ds} \phi(T, W_T^1, W_T^2)\right).$$

   Give an expression for λ_1,λ_2 and ϕ.

2. Prove that V_0 can be expressed as $V_0 = u(0,0,0)$, where $u(t, x_1, x_2)$ is a regular solution of a parabolic equation.

3. Compute the law of the pair $\left(\int_0^T W_s^1 ds, W_T^1\right)$ and deduce a simulation method for it.

4. Propose a Monte-Carlo method for the computation of V_0.

5. Propose a quasi-Monte-Carlo method for the same problem using N^{-1}, the inverse of the repartition function of a standard Gaussian random variable.

6. Let f and g be two continuous functions. Compute the law of

$$\left(\int_0^T f(s) dW_s^1, \int_0^T g(s) dW_s^1\right)$$

   and propose a simulation method for this pair.

7. We assume now that σ is a deterministic matrix which depends on t. Construct a Monte-Carlo method for the computation of V_0, avoiding the simulation of the trajectory of the process $(X_s, s \geq 0)$.

Problem 6.6 Let X and Y be independent real random variables. Let F and G be the distribution functions of X and G respectively We want to compute by a Monte-Carlo method the probability

$$\theta = \mathbf{P}\left(X + Y \leq t\right).$$

1. Propose a variance reduction procedure using a conditioning method.

2. We assume that F and G are (at least numerically) easily invertible. Explain how to implement the antithetic variates methods. Why does this method decrease the variance in this case ?

3. Assume that h is a function such that $\int_0^1 |h(s)|^2 ds < +\infty$. Let $(U_i, i \geq 1)$ be a sequence of independent random variates with a uniform distribution on $[0,1]$. Prove that $\frac{1}{N}\sum_{i=1}^N h\left((i-1+U_i)/n\right)$ has a lower variance than $\frac{1}{N}\sum_{i=1}^N h\left(U_i\right)$.

Problem 6.7 This problem presents methods to compute the price of a two-asset option, when the correlation depends on time.

We denote by S_t^1 and S_t^2 the prices of two assets being the solutions of

$$\begin{cases} dS_t^1 = S_t^1\left(rdt + \sigma_1 d\bar{W}_t^1\right), S_0^1 = x_1, \\ dS_t^2 = S_t^2\left(rdt + \sigma_2 d\bar{W}_t^2\right), S_0^2 = x_2. \end{cases}$$

where r, σ_1, σ_2, x_1, x_2 are real numbers. Moreover, we assume that $(\bar{W}^1_t, t \geq 0)$ are $(\bar{W}^2_t, t \geq 0)$ two Brownian motions with respect to a given filtration $(\mathcal{F}_t, t \geq 0)$, such that

$$d < \bar{W}^1, \bar{W}^2 >_t = \rho(t)dt.$$

Here $\rho(t)$ is a deterministic function such that

$$|\rho(t)| \leq \rho_0 < 1.$$

We want to compute the price at time t of the option with payoff $h(S^1_s, S^2_s, s \leq T)$, given by

$$V_t = \mathbf{E}\left(e^{-r(T-t)} h(S^1_s, S^2_s, s \leq T) | \mathcal{F}_t\right).$$

1. Let $(W^1_t, t \geq 0)$ and $(W^2_t, t \geq 0)$ be two independent Brownian motions with respect to $(\mathcal{F}_t, t \geq 0)$. Explain how to construct two Brownian motions $\bar{W}^1$ and $\bar{W}^2$ satisfying the previous assumptions.

2. Compute the law of $(\int_0^T \rho(s) dW^1_s, W^1_T)$ and deduce an efficient simulation method for this vector. What happens when $\rho(t)$ does not depend on t ?

3. Explain how to efficiently simulate the vector (S^1_T, S^2_T). Propose a Monte-Carlo method to compute the price of an option with payoff at time T, $f(S^1_T, S^2_T)$.

Problem 6.8 This exercise prove (part of) Girsanov theorem. Let $(W_t, 0 \leq t \leq T)$ be a Brownian motion. Define L_T by

$$L_T = e^{-\lambda W_T - \frac{\lambda^2}{2} T},$$

1. Prove that
$$\mathbf{E}\left(L_T f(W_T + \lambda T)\right) = \mathbf{E}\left(f(W_T)\right)$$

2. Define $\tilde{P}$ on all set A in $\mathcal{F}_T = \sigma\left(W_t, t \leq T\right)$ by
$$\tilde{P}\left(A\right) = \mathbf{E}\left(L_T \mathbf{1} A\right).$$
Prove that $\tilde{P}$ define a probability and that
$$\tilde{\mathbf{E}}\left(X\right) = \mathbf{E}\left(L_T X\right)$$
for every $\mathcal{F}_T$-measurable bounded random variable X.

3. Prove that the law of $\tilde{W}_T = W_T + \lambda T$ under the probability $\tilde{\mathbf{P}}$ is identical to the one of W_T under $\mathbf{P}$.

4. What is the law of $\tilde{W}_t$ under $\tilde{\mathbf{P}}$?

5. Compute the law of $(\tilde{W}_t, \tilde{W}_T - \tilde{W}_t)$, then the law of $(\tilde{W}_t, \tilde{W}_T)$ under $\tilde{\mathbf{P}}$.

6. More generally show that the law of $(\tilde{W}_{t_1}, \ldots, \tilde{W}_{t_n})$ under $\tilde{\mathbf{P}}$ is the same as the law of $(W_{t_1}, \ldots, W_{t_n})$ under $\mathbf{P}$.

Problem 6.9 Let Z be a random variable given by

$$Z = \lambda_1 e^{\beta_1 X_1} + \lambda_2 e^{\beta_2 X_2},$$

where (X_1, X_2) is a couple of real random variables and λ_1, λ_2, β_1 and β_2 are real positive numbers. This problem studies various methods to compute the price of an index option given by $p = \mathbf{P}\left(Z > t\right)$.

1. In this question, we assume that (X_1, X_2) is a Gaussian vector with mean 0 such that Var$(X_1) =$ Var$(X_2) = 1$ and Cov$(X_1, X_2) = \rho$, with $|\rho| \leq 1$. Explain how to simulate random samples along the law of Z. Describe a Monte-Carlo method allowing to estimate p and explain how to estimate the error of the method.

2. Explain how to use low discrepancy sequences to compute p.

3. We assume that X_1 and X_2 are two independent Gaussian random variables with mean 0 and variance 1. Let m be a real number. Prove that p can be written as

   $$p = \mathbf{E}\left[\phi(X_1, X_2)\mathbf{1}_{\left\{\lambda_1 e^{\beta_1(X_1+m)} + \lambda_2 e^{\beta_2(X_2+m)} \geq t\right\}}\right],$$

   for some function ϕ. How can we choose m such that

   $$\mathbf{P}(\lambda_1 e^{\beta_1(X_1+m)} + \lambda_2 e^{\beta_2(X_2+m)} \geq t) \geq \frac{1}{4}?$$

   Propose a new Monte-Carlo method which allows to compute p. Explain how to check on the drawings that the method does reduce the variance.

4. Assuming now that X_1 and X_2 are two independent random variables with distribution functions $F_1(x)$ and $F_2(x)$ respectively. Prove that

   $$p = \mathbf{E}\left[1 - G_2\left(t - \lambda_1 e^{\beta_1 X_1}\right)\right],$$

   where $G_2(x)$ is a function such that the variance of

   $$1 - G_2\left(t - \lambda_1 e^{\lambda_1 X_1}\right),$$

   is always less than the variance of $\mathbf{1}_{\left\{\lambda_1 e^{\beta_1 X_1} + \lambda_2 e^{\lambda_2 X_2} > t\right\}}$. Propose a new Monte-Carlo method to compute p.

5. We assume again that (X_1, X_2) is a Gaussian vector with mean 0 and such that Var$(X_1) =$ Var$(X_2) = 1$ and Cov$(X_1, X_2) = \rho$, with $|\rho| \leq 1$. Prove that $p = \mathbf{E}\left[1 - F_2\left(\phi(X_1)\right)\right]$ where F_2 is the repartition function of X_2 and ϕ a function to be computed.

   Deduce a variance reduction method computing p.

References

[Coc77] Cochran. *Sampling Techniques*. Wiley Series in Probabilities and Mathematical Statistics, 1977.

[GHS99] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Asymptotically optimal importance sampling and stratification for pricing path dependent opions. *Mathematical Finance*, 9(2):117–152, April 1999.

[HH79] J. Hammersley and D. Handscomb. *Monte Carlo Methods*. Chapman and Hall, London, 1979.

[KS91] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York, second edition, 1991.

[KV90] A.G.Z Kemna and A.C.F. Vorst. A pricing method for options based on average asset values. *J. Banking Finan.*, pages 113–129, March 1990.

[KW86] M.H. Kalos and P.A. Whitlock. *Monte Carlo Methods*. John Wiley & Sons, 1986.

[New94] N.J. Newton. Variance reduction for simulated diffusions. *SIAM J. Appl. Math.*, 54(6):1780–1805, 1994.

[Ra97] L. C. G. (ed.) Rogers and al., editeurs. *Monte Carlo methods for stochastic volatility models.* Cambridge University Press., 1997.

[Rip87] B.D. Ripley. *Stochastic Simulation*. Wiley, 1987.

[Rub81] R.Y. Rubinstein. *Simulation and the Monte Carlo Method*. Wiley Series in Probabilities and Mathematical Statistics, 1981.

[RY91] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Springer-Verlag, Berlin, 1991.