1 Statistical Analysis of Simulated Data

We want to estimate θ. A Monte Carlo algorithm is based on a random variable X such that $E[X] = \theta$, and independent simulations of the algorithm gives independent realizations of the random variable $X : X_1, ..., X_n$. The average of these values

$$\bar{X} = \frac{X_1 + ... + X_n}{n}$$

is the *sample mean estimator* for θ. We have:

1. $E[\bar{X}] = E[X] = \theta$
2. $Var(\bar{X}) = \sigma^2_{\bar{X}} = Var(X)/n = \sigma^2_X/n$

Question How do we decide when to stop simulation? What is a good value for n?

To answer this question, we need to know how good $\bar{X}$ is as an estimator for θ. Consider the *mean square error* of $\bar{X}$:

$$E[(\bar{X} - \theta)]^2 = Var\bar{X} = \sigma^2_X/n.$$

If σ^2_X/n is smaller, $\bar{X}$ will be "closer" to θ. This statement can be made precise using the Central Limit Theorem (CLT). If n is large, $\bar{X}$ will be approximately normal with

$$\frac{\bar{X} - \theta}{\sigma_X/\sqrt{n}} \approx \mathbb{N}(0, 1)$$

and thus, the probability that $\bar{X}$ is farther from θ by c units (normalized by $\sigma_X/\sqrt{n}$) is:

$$P\left\{\left|\frac{\bar{X} - \theta}{\sigma_X/\sqrt{n}}\right| > c\right\} = P\{|Z| > c\} = P\{Z > c\} + P\{Z < -c\} = 2(1 - \phi(c))$$

For example, if we choose $c = 1.96$, then $\phi(c) = 0.975$ and $2(1 - \phi(c)) = 0.05$, which means with 95% probability, the sample mean and θ will be close to each other within $1.96\sigma_X/\sqrt{n}$ units.

Therefore, we want to have a sufficiently large n that will make $1.96\sigma_X/\sqrt{n}$ as small as we want. However, in practice, we do not know σ; it has to be estimated.

Fact $S^2_X = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$ is an unbiased estimator for σ^2_X.

When to stop We have $E[X] = \theta$, and use simulation to obtain data: $x_1, x_2, ..., x_n, ...$. For each n, we compute $\bar{X}$ and S_X. Stop when $S_X/\sqrt{n}$ is less than an acceptable value. (Note: since S_X is an estimate only, take $n \geq 30$)

Example You have a Monte Carlo algorithm for the price of an American put option. You want to be at least 95% certain that your estimate for the option price will not differ from the true price by more than one cent. How many samples should you generate?

Solution We have

$$P\left\{\left|\frac{\bar{X}-\theta}{\sigma_X/\sqrt{n}}\right| < 1.96\right\} = P\left\{\left|\bar{X}-\theta\right| < 1.96\sigma_X/\sqrt{n}\right\} = 95\%$$

Then, generate n samples, $n \geq 30$, $X_1, ..., X_n$, such that $1.96 S_X/\sqrt{n} < 0.01\$$, where S_X is the sample standard deviation of the data measured in dollars.

A recursive computation of S_X would be very helpful in our stopping procedure, so that we do not have to compute sample variance from scratch every time a new data is generated. Consider the time when we have n data values, and compute the sample mean and variance (we index them by n, the current "time", and drop the index X used earlier):

$$\begin{aligned}\bar{X}_n &= \frac{X_1 + ... + X_n}{n} \\ S_n^2 &= \sum_{i=1}^{n} \frac{(X_i - \bar{X}_n)^2}{n-1}\end{aligned}$$

Now generate a new data value, X_{n+1}. The new sample mean and variance can be computed recursively as follows:

$$\begin{aligned}\bar{X}_{n+1} &= \bar{X}_n + \frac{X_{n+1} - \bar{X}_n}{n+1} \\ S_{n+1}^2 &= \left(1 - \frac{1}{n}\right) S_n^2 + (n+1)(\bar{X}_{n+1} - \bar{X}_n)^2\end{aligned} \quad (1)$$

1.1 Confidence intervals

Given a sample $X_1, ..., X_n$, we can compute $\bar{X}$ to obtain a "point" estimate for θ. Another approach with more information is to compute an interval that contains θ with a known probability. Let z_α be such that $P\{Z > z_\alpha\} = \alpha$ where Z is the standard normal. From symmetry, we have $P\{Z < -z_\alpha\} = \alpha$. Therefore, $P\{-z_\alpha < Z < z_\alpha\} = 1 - 2\alpha$. Now, replacing α by $\alpha/2$, and noting that $\frac{\bar{X}-\theta}{S_X/\sqrt{n}}$ is approximately a standard normal, we obtain

$$P\{-z_{\alpha/2} < \frac{\bar{X}-\theta}{S_X/\sqrt{n}} < z_{\alpha/2}\} \approx 1 - \alpha,$$

which, after some algebra, can be written as

$$P\left\{\bar{X} - z_{\alpha/2}\frac{S_X}{\sqrt{n}} < \theta < \bar{X} + z_{\alpha/2}\frac{S_X}{\sqrt{n}}\right\} \approx 1 - \alpha.$$

If the observed values of $\bar{X}$ and S_X are $\bar{x}$ and s, we call the interval $(\bar{x} - z_{\alpha/2}s/\sqrt{n}, \bar{x} + z_{\alpha/2}s/\sqrt{n})$ an (approximate) $100(1-\alpha)$ percent confidence interval estimate of θ.

For example, if $\alpha = 0.05$, then $100(1-\alpha) = 95\%$ and $z_{\alpha/2} = 1.96$, giving an interval $(\bar{x} - 1.96s/\sqrt{n}, \bar{x} + 1.96s/\sqrt{n})$, which either contains θ or not. If 100 such intervals are obtained independently, then approximately 95 of them will contain θ.

2 Importance Sampling

Consider the problem of estimating the definite integral (I suppress the domain of integration, $\mathcal{R}^n$, for simplicity, below)

$$E[h(X)] = \int h(x)f(x)dx = I$$

where X is a random vector with density f and $h : \mathcal{R}^n \to \mathcal{R}$. Let g be another density on $\mathcal{R}^n$ such that

$$f(x) > 0 \Rightarrow g(x) > 0 \text{ for all } x \in \mathcal{R}^n.$$

Observe that

$$\int h(x)\frac{f(x)}{g(x)}g(x)dx = I$$

and this integral can be viewed as $E\left[h(X)\frac{f(X)}{g(X)}\right]$ where X has density g. The importance sampling estimator (with respect to g) of I is defined as

$$\theta = \frac{1}{N}\sum_{i=1}^{N} h(X_i)\frac{f(X_i)}{g(X_i)}$$

where $X_1, ..., X_N$ are independent draws from density g. Let ν be the "crude" Monte Carlo estimator of I, i.e.,

$$\nu = \frac{1}{N}\sum_{i=1}^{N} h(X_i)$$

where $X_1, ..., X_N$ are independent draws from density f.

Theorem 1 *θ is an unbiased estimator of I and*

$$Var\theta = \frac{1}{N}\left(\int \frac{h^2(x)f^2(x)}{g(x)}dx - I^2\right) \quad (2)$$

Proof. *The fact that θ is an unbiased estimator of I is obvious. For the* *variance, we have*

$$\begin{aligned}
Var\theta &= \frac{1}{N^2}\sum_{i=1}^{N} Var\left(h(X_i)\frac{f(X_i)}{g(X_i)}\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\left(E\left[h^2(X_i)\frac{f^2(X_i)}{g^2(X_i)}\right] - E\left[h(X_i)\frac{f(X_i)}{g(X_i)}\right]^2\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\left(\int h^2(x)\frac{f^2(x)}{g^2(x)}g(x)dx - I^2\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\left(\int h^2(x)\frac{f^2(x)}{g(x)}dx - I^2\right) \\
&= \frac{1}{N}\left(\int \frac{h^2(x)f^2(x)}{g(x)}dx - I^2\right).
\end{aligned}$$

■

The main question in importance sampling is this: how can we choose g that will minimize $Var\theta$?

Theorem 2 *The minimum of $Var\theta$ is obtained when*

$$g(x) = \frac{|h(x)f(x)|}{\int |h(x)f(x)|dx} \tag{3}$$

and this minimum value is

$$Var\theta = \frac{1}{N}\left(\int \frac{h^2(x)f^2(x)}{g(x)}dx - I^2\right) = \frac{1}{N}\left[\left(\int |h(x)f(x)|dx\right)^2 - I^2\right] \tag{4}$$

Proof. *If we substitute 3 in the expression for $Var\theta$ in Theorem 1, we get*

$$\begin{aligned}
Var\theta &= \frac{1}{N}\left(\int \frac{h^2(x)f^2(x)}{g(x)}dx - I^2\right) \\
&= \frac{1}{N}\left[\left(\int |h(x)f(x)|dx\right)\left(\int \frac{h^2(x)f^2(x)}{|h(x)f(x)|}dx\right) - I^2\right] \\
&= \frac{1}{N}\left[\left(\int |h(x)f(x)|dx\right)^2 - I^2\right].
\end{aligned}$$

To show that this is the minimum variance, it suffices to prove (compare equations (2) and (4))

$$\left(\int |h(x)f(x)|dx\right)^2 \leq \int \frac{h^2(x)f^2(x)}{g(x)}dx.$$

Using Cauchy-Schwarz inequality, we have

$$\begin{aligned}\left(\int |h(x)|f(x)dx\right)^2 &= \left(\int \frac{|h(x)|}{g(x)^{1/2}} f(x) g(x)^{1/2} dx\right)^2 \\ &\leq \left(\int \frac{h^2(x)}{g(x)} f^2(x)dx\right)\left(\int g(x)dx\right).\end{aligned}$$

■

Corollary 3 *If $h(x) > 0$, then the optimal density is*

$$g(x) = \frac{|h(x)f(x)|}{\int |h(x)f(x)|dx} = \frac{h(x)f(x)}{\int h(x)f(x)dx} = \frac{h(x)f(x)}{I}$$

Remark 1 *This shows that the optimal density requires the knowledge of I, which was what we were trying to estimate in the first place! Is this then a useless results? No, because it tells us that a good strategy is to try to sample points in proportion to $h(x)f(x)$. I is the optimal constant of proportion.*

2.1 A Theoretical Model for Importance Sampling in more Abstract Spaces

In many applications, including finance, one needs to compute $E[X]$ where X is defined on a probability space Ω, and Ω consists of Markov chains. In such problems, we simulate paths

$$w : S(t_0), S(t_1), ..., S(t_m) \tag{5}$$

and we are only given the conditional distributions (transition densities for the Markov chain). The probability of a path like above is

$$f_1(S(t_0), S(t_1)) \times f_2(S(t_1), S(t_2)) \times ... \times f_m(S(t_{m-1}), S(t_m))$$

In an application of importance sampling to such problems, one changes the underlying probability measure on Ω, indirectly, by replacing the transition densities f_i by g_i, so that the chain (5) is now simulated using g_i. Consider a function h defined on Markov chains of $m-$steps, i.e.,

$$h := h(S(t_1), ..., S(t_m))$$

(assuming $S(t_0)$ is fixed). Under the original measure (for simplicity, assume the densities are discrete)

$$\begin{aligned} E[h(S(t_1), ..., S(t_m))] &= \sum_{w=S(t_0),...,S(t_m)} h(w) \Pr(w) \\ &= \sum_w h(w) f_1(S(t_0), S(t_1)) \times ... \times f_m(S(t_{m-1}), S(t_m)) \end{aligned}$$

To make sure this expectation remains the same when we change the densities to g_i, we multiply by "weights" or "the likelihood ratios"

$$\frac{f_1(S(t_0), S(t_1))}{g_1(S(t_0), S(t_1))} \times ... \times \frac{f_m(S(t_{m-1}), S(t_m))}{g_m(S(t_{m-1}), S(t_m))} = \prod_{i=1}^{m} \frac{f_i(S(t_{i-1}), S(t_i))}{g_i(S(t_{i-1}), S(t_i))} \tag{6}$$

so that

$$\begin{aligned} & \tilde{E}\left[h(S(t_1), ..., S(t_m)) \prod_{i=1}^{m} \frac{f_i(S(t_{i-1}), S(t_i))}{g_i(S(t_{i-1}), S(t_i))}\right] \qquad (7) \\ = & \sum_{w=S(t_0),...,S(t_m)} \left(h(w) \prod_{i=1}^{m} \frac{f_i(S(t_{i-1}), S(t_i))}{g_i(S(t_{i-1}), S(t_i))}\right) g_1(S(t_0), S(t_1)) \times ... \times g_m(S(t_{m-1}), S(t_m)) \\ = & E[h(S(t_1), ..., S(t_m))] \end{aligned}$$

assuming the latter expectation exists and finite.

More specifically, in many problems from finance, we often simulate a (price) path $S(t_0), ..., S(t_m)$ using a recursion

$$S(t_{i+1}) = G(S(t_i), X_{i+1})$$

where the random vectors $X_1, X_2, ...$ are independent and from a common distribution (normal, in Black-Scholes-Merton model). The probability of the path

$$S(t_0), S(t_1), ..., S(t_m) = S(t_0), \underbrace{G(S(t_0), X_1)}_{S(t_1)}, \underbrace{G(G(S(t_0), X_1), X_2)}_{S(t_2)}, ...$$

is

$$f(X_1) \times f(X_2) \times ... \times f(X_m)$$

where f is the common density of X_i. The likelihood ratio (6) simplifies to

$$\prod_{i=1}^{m} \frac{f(X_i)}{g(X_i)}$$

where g is the new common density obtained after the original measure is changed. We assume that the independence of X_i is preserved while the measure is changed. Equation (7) simplifies to

$$\tilde{E}\left[h(S(t_1), ..., S(t_m)) \prod_{i=1}^{m} \frac{f(X_i)}{g(X_i)}\right] = E[h(S(t_1), ..., S(t_m))] \tag{8}$$

Remark 2 *In some problems m is not fixed but a random variable. For example, barrier options. The equations (7) & (8) are still valid if m is replaced by a stopping time. See Glynn and Iglehart [1] for details.*

2.1.1 A strange behavior of the likelihood ratios

Assume $\tilde{E}\left[\left|\log \frac{f(X_1)}{g(X_1)}\right|\right] < \infty$. From Strong Law of Large Numbers,

$$\frac{1}{m}\sum_{i=1}^{m} \log \frac{f(X_i)}{g(X_i)} \to E\left[\log \frac{f(X_1)}{g(X_1)}\right] = c$$

with $\tilde{P}-$probability 1. By Jensen's inequality

$$c = E\left[\log \frac{f(X_1)}{g(X_1)}\right] \leq \log E\left[\frac{f(X_1)}{g(X_1)}\right] = \log \int \frac{f(x)}{g(x)} g(x)dx = 0.$$

Note that $c = 0$ iff $f(X_1) = g(X_1)$ with $\tilde{P}-$probability 1, since log is strictly concave. If $c < 0$, then from

$$\frac{1}{m}\sum_{i=1}^{m} \log \frac{f(X_i)}{g(X_i)} \to c < 0$$

we deduce

$$\sum_{i=1}^{m} \log \frac{f(X_i)}{g(X_i)} \to -\infty$$

which implies

$$\prod_{i=1}^{m} \frac{f(X_i)}{g(X_i)} \to 0$$

as $m \to \infty$, with $\tilde{P}-$probability 1.

So, although $E\left[\prod_{i=1}^{m} \frac{f(X_i)}{g(X_i)}\right] = 1$ for any $m < \infty$, $\prod_{i=1}^{\infty} \frac{f(X_i)}{g(X_i)} = 0$ with $\tilde{P}-$probability 1. Heuristically, this means the likelihood ratio takes increasingly large values with small but non-negligible probabilities. An interesting application would be to investigate the distribution of the likelihood ratio in a model problem.

2.2 Exponential change of measure

Here is the question we want to discuss in this section: How can we select an appropriate density g in importance sampling? A commonly used technique involves the so-called tilted densities (also called exponential tilting, exponential twisting, exponential change of measure).

Definition 1 *A density function*

$$f_t(x) = \frac{e^{tx} f(x)}{M(t)}$$

where $M(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x)dx$ *is the moment generating function, is called a tilted density of* f*, where* $-\infty < t < \infty$.

Let the random variable X have density f, and X_t have density f_t. If $t > 0, X_t$ tends to be larger than X, and if $t < 0, X_t$ tends to be smaller than X.

Example 1 (Exponential distribution) *The exponential density is* $f(x) = \lambda e^{-\lambda x}, x0$. *Therefore*

$$f_t(x) = \frac{e^{tx}\lambda e^{-\lambda x}}{M(t)} = Ce^{-(\lambda - t)x}$$

where, the constant C, *must be equal to* $\lambda - t$, *and* f_t *is an exponential density with rate* $\lambda - t$ *if* $\lambda - t > 0$. *(Also recall that* $M(t) = \lambda/(\lambda - t), t < \lambda$.*)*

Example 2 (Normal distribution) *The normal density is* $f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-(x-\mu)^2/2\sigma^2)$. *Then*

$$f_t(x) = \frac{1}{M(t)}\exp(tx)\exp(-(x-\mu)^2/2\sigma^2)\frac{1}{\sqrt{2\pi}\sigma}$$

where $M(t) = \exp(\mu t + \sigma^2 t^2/2)$. *Simplifying*

$$f_t(x) = \exp\left(tx - \frac{(x-\mu)^2}{2\sigma^2} - \mu t - \frac{\sigma^2 t^2}{2}\right)\frac{1}{\sqrt{2\pi}\sigma}$$

Note that

$$\begin{aligned} & 2\sigma^2 tx - x^2 + 2x\mu - \mu^2 - 2\sigma^2\mu t - \sigma^4 t^2 \\ = \quad & -\left(x^2 - 2x(\mu + \sigma^2 t) + \sigma^4 t^2 + 2\sigma^2\mu t + \mu^2\right) \\ = \quad & -\left(x - (\mu + \sigma^2 t)\right)^2 \end{aligned}$$

and thus

$$f_t(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-\left(x - (\mu + \sigma^2 t)\right)^2}{2\sigma^2}\right) \sim \mathbf{N}(\mu + \sigma^2 t, \sigma^2).$$

Let μ_t be the mean of the random variable with density f_t. Then

$$\mu_t = \int x f_t(x)dx = \int x\frac{e^{tx}f(x)}{M(t)}dx.$$

Observe that since $M(t) = \int_{-\infty}^{\infty} e^{tx}f(x)dx$, we have $M'(t) = \int_{-\infty}^{\infty} xe^{tx}f(x)dx$. Therefore

$$\mu_t = \frac{M'(t)}{M(t)} = (\log M(t))' = \psi'(t)$$

where $\psi(t)$ is defined next.

Definition 2 *The function* $\psi(t) = \log M(t)$ *is called the cumulant generating function of the distribution.*

Lemma 1 *1.* $\psi'(t) = \mu_t$

2. $\psi"(t) = \sigma_t^2 > 0$ *(so* $\psi(t)$ *is convex)*

3. $\psi(0) = 0$

Example 3 *Let* $X_1, X_2, ...,$ *i.i.d. sequence of* $\mathbf{N}(\mu, 1)$ *variables, with* $\mu < 0$. *Let* $S_n = \sum_{i=1}^{n} X_i$ *and put*

$$N = Min\{n : \text{ either } S_n < -A \text{ or } S_n > B\}.$$

N *is the first time* S_n *crosses one of the lines* $S = B$, *or* $S = -A$. *We want to estimate*

$$\theta = P\{S_N > B\},$$

the probability that the partial sum exceeds B, *before going below* $-A$.

Our first idea might be to simulate $X_1, X_2, ...$ *until the sum exceeds* B *or goes below* $-A$, *and compute*

$$I = \begin{cases} 1 \text{ if } S_N > B \\ 0 \text{ otherwise} \end{cases}$$

However, this is not a very efficient approach: note that $E[S_N] = N\mu < 0$ *and thus* $P\{S_N > B\}$ *will be "small".*

Here is how we can use importance sampling in this example. Simulate $X_1, X_2, ..., X_N$ *from* $\mathbf{N}(-\mu, 1)$ *instead of* $\mathbf{N}(\mu, 1)$. *Observe that if we twist the normal density of* $\mathbf{N}(\mu, 1)$ *by* $t = -2\mu$, *we obtain the density of* $\mathbf{N}(-\mu, 1)$. *Let* $f_{-2\mu}(x)$ *be the twisted density for* $\mathbf{N}(-\mu, 1)$, *and* $f(x)$ *be the density for* $\mathbf{N}(\mu, 1)$. *Then the likelihood ratio is:*

$$\frac{f(x)}{f_{-2\mu}(x)} = \frac{\exp\left(-(x-\mu)^2/2\right)}{\exp\left(-(x+\mu)^2/2\right)} = \exp(2x\mu)$$

and thus

$$\prod_{i=1}^{N} \frac{f(x_i)}{f_{-2\mu}(x_i)} = \prod_{i=1}^{N} e^{2\mu x_i} = e^{2\mu \sum_{i=1}^{N} x_i} = e^{2\mu S_N}.$$

Therefore, the importance sampling estimator for θ *is*

$$Ie^{2\mu S_N}$$

where $X_1, X_2, ..., X_N$ *are generated from* $f_{-2\mu}(x)$.*Observe that:*

$$\tilde{E}[Ie^{2\mu S_N}] = \theta \text{ and } E[I] = \theta$$

however, $Var(Ie^{2\mu S_N})$ *is significantly lower than* $VarI$. *Indeed, we have*

$$Ie^{2\mu S_N} = \begin{cases} 0 \text{ if } S_N < -A \\ e^{2\mu S_N} < e^{2\mu B} \text{ if } S_N > B \end{cases}$$

whereas

$$I = \begin{cases} 0 \text{ if } S_N < -A \\ 1 \text{ otherwise} \end{cases}.$$

Remark 3 *In the previous example, we have*

$$\theta = P\{S_N > B\} = \tilde{E}[Ie^{2\mu S_N}] < \tilde{E}[Ie^{2\mu B}] < e^{2\mu B}$$

Therefore, $P\{$ *partial sum cross* B *before* $-A\} \leq e^{2\mu B}$. *Since this is true for any* $A > 0$, *we conclude* $P\{$ *partial sum ever cross* $B\} \leq e^{2\mu B}$.

2.2.1 Down-and-in barrier options

The payoff function for a down-and-in barrier option is

$$1\{S(T) > K\} \cdot 1\{\min_{1\leq k\leq m} S(t_k) < H\} \tag{9}$$

where K is the strike price, and H is the barrier. The option is observed at $t_1, t_2, ..., t_m = T$. We assume the stock price follows the lognormal model.

The difficulty with a crude Monte Carlo estimation of the barrier option price is the following: If $S(0) >> H$, then only a few stock price paths will ever cross H. Thus, most of the paths will return the value 0, and only a few will return a non-zero option estimate. This makes simulation inefficient. The importance sampling idea will be to change the underlying measure so that more of the paths cross the barrier, and hence we obtain more nonzero estimates. Let's start by recalling the lognormal model:

$$S(t_n) = S(0)\exp\left(\underbrace{(r - \sigma^2/2)t_n + \sigma\sqrt{t_n}Z}_{L_n}\right)$$

Let $t_n = nh$, so we consider equally spaced time points. Then we get

$$L_n = (r - \sigma^2/2)t_n + \sigma\sqrt{t_n}Z =^d \mathbf{N}((r - \sigma^2/2)nh, \sigma^2 nh) =^d \sum_{i=1}^{n} X_i$$

where X_i is $\mathbf{N}((r - \sigma^2/2)h, \sigma^2 h)$.

So a general description of the stock price model is:

$$S(t_n) = S(0)\exp(L_n), \text{ where } L_n = \sum_{i=1}^{n} X_i$$

and $X_1, ..., X_n$ are i.i.d. random variables, and $L_0 = 0$. Observe that

$$\begin{aligned} S(T) &> K \Leftrightarrow S(0)\exp(L_m) > K \\ &\Leftrightarrow L_m > \log(K/S(0)) = c \end{aligned}$$

and

$$\begin{aligned} \min_{1\leq k\leq m} S(t_k) &< H \Leftrightarrow \min_{1\leq k\leq m} S(0)\exp(L_k) < H \\ &\Leftrightarrow \min_{1\leq k\leq m} \exp(L_k) < H/S(0) \\ &\Leftrightarrow \min_{1\leq k\leq m} L_k < \log(H/S(0)) = -b \end{aligned}$$

where b is some positive constant.

Let τ be the first time L_n drops below $-b$, and set $\tau = \infty$ if it never does. The payoff function (9) then can be written as

$$1\{L_m > c, \tau < m\}.$$

The importance sampling idea is this: until the barrier is crossed, exponentially twist the distribution as

$$f_{t^-}(x) = \frac{e^{t^- x} f(x)}{M(t^-)}$$

where t^- is chosen such that the drift is negative. This will make the price paths go down the barrier faster. And after the barrier is crossed, exponentially twist the distribution as

$$f_{t^+}(x) = \frac{e^{t^+ x} f(x)}{M(t^+)}$$

where t^+ is chosen such that the drift is positive. This will make the price paths go up to the exercise price faster, so that the payoff function will have nonzero value. Since we are using the GBM model, $f(x)$ in the above equations is the normal density $\mathbf{N}((r-\sigma^2/2)h, \sigma^2 h)$. Therefore f_{t^-} and f_{t^+} are also normal densities. Therefore, we will simulate $X_1, ..., X_\tau$ from $f_{t^-}(x)$ and $X_{\tau+1}, ..., X_m$ from $f_{t^+}(x)$.

The likelihood ratios are:

$$\begin{aligned} & \frac{f(x_1)f(x_2)...f(x_\tau)}{f_{t^-}(x_1)f_{t^-}(x_2)...f_{t^-}(x_\tau)} \frac{f(x_{\tau+1})f(x_{\tau+2})...f(x_m)}{f_{t^+}(x_{\tau+1})f_{t^+}(x_{\tau+2})...f_{t^+}(x_m)} \qquad (10) \\ = & \frac{M(t^-)^\tau}{\exp(t^-(X_1 + ... + X_\tau))} \frac{M(t^+)^{m-\tau}}{\exp(t^+(X_{\tau+1} + ... + X_m))} \\ = & M(t^-)^\tau M(t^+)^{m-\tau} \exp(-t^- L_\tau) \exp(-t^+(L_m - L_\tau)) \\ = & \left[\frac{M(t^-)}{M(t^+)}\right]^\tau M(t^+)^m \exp\left((t^+ - t^-)L_\tau - t^+ L_m\right) \end{aligned}$$

We now discuss how to choose t^-, t^+ from some heuristics. We claim $L_\tau \approx -b$ and $L_m \approx c$, for large b and c. In other words, undershoot below $-b$ and overshoot above c will be small. This means that most of the variability in the above likelihood ratio comes from $(M(t^-)/M(t^+))^\tau$. Thus, we choose t^- and t^+ so that $M(t^-) = M(t^+)$. Then the likelihood ratio (10) simplifies to

$$M(t^+)^m \exp\left((t^+ - t^-)L_\tau - t^+ L_m\right).$$

Recall that $\psi(t) = \log M(t)$, and $\psi'(t^-), \psi'(t^+)$ are the drifts (means) of the densities f_{t^-}, f_{t^+}. Now let L_n go to the barrier at a fixed rate - given by the twisted mean of each X_i, and impose

$$\tau\psi'(t^-) = -b.$$

And similarly, let L_n go to the exercise level at a fixed rate, imposing

$$(m - \tau)\psi'(t^+) = c + b.$$

These two conditions can be written as

$$\frac{-b}{\psi'(t^-)} + \frac{c+b}{\psi'(t^+)} = m.$$

Then solve the equations

$$\begin{aligned} M(t^-) &= M(t^+) \text{ (or, } \psi(t^-) = \psi(t^+)) \\ \frac{-b}{\psi'(t^-)} + \frac{c+b}{\psi'(t^+)} &= m \end{aligned} \tag{11}$$

to obtain t^- and t^+.

In our GBM model, we have

$$\begin{aligned} X_n &\sim \mathbf{N}((r-\sigma^2/2)h, \sigma^2 h) \\ M(t) &= \exp\left((r-\sigma^2/2)ht + \sigma^2 h t^2/2\right) \\ \psi(t) &= (r-\sigma^2/2)ht + \sigma^2 h t^2/2 \end{aligned}$$

Note the form of the functions: $\psi(t) = At + Bt^2, \psi'(t) = A + 2Bt$, and the minimum of $\psi(t)$ is attained when $t = -A/2B$. There is symmetry about the line $t = -A/2B$. Hence, $\psi(t^-) = \psi(t^+)$ implies $t^+ + A/2B = -A/2B - t^-$, or, $t^+ + t^- = -A/B$. Using the symmetry, it is easy to show that $\psi'(t^-) = -\psi'(t^+)$. Then, from the second equation of (11), we get $\psi'(t^+) = (c+2b)/m$. Then, from $\psi'(t^+) = A + 2Bt^+ = (c+2b)/m$ we solve

$$t^+ = \left(\frac{1}{2} - \frac{r}{\sigma^2}\right) + \left(\frac{2b+c}{m\sigma^2 h}\right). \tag{12}$$

Similarly, we obtain

$$t^- = \left(\frac{1}{2} - \frac{r}{\sigma^2}\right) - \left(\frac{2b+c}{m\sigma^2 h}\right). \tag{13}$$

In summary, the importance sampling estimator for the option price is

$$e^{-rT} M(t^+)^m \exp\left((t^+ - t^-)L_\tau - t^+ L_m\right) 1\{L_m > c, \tau < m\}$$

where t^+, t^- are given by equations (12) and (13). $X_1, ..., X_\tau$ is generated from $f_{t^-}(x)$ and $X_{\tau+1}, ..., X_m$ from $f_{t^+}(x)$, where $f_{t^-}(x)$ is the density function for $\mathbf{N}((r-\sigma^2/2)h + \sigma^2 h t^-, \sigma^2 h)$, and $f_{t^+}(x)$ is the density function for $\mathbf{N}((r-\sigma^2/2)h + \sigma^2 h t^+, \sigma^2 h)$.

In the following example, the option have the following parameters: $\sigma = 0.3$, $r = 0.1$, $S(0) = 100$, $K = 100$, $H = 91, T = 0.2$, and $m = 50$.

N	Crude MC	ImpSamp-h	CondExp	Combined-h
5K	1.66×10^{-3}	3.56×10^{-4}	4.66×10^{-4}	9.78×10^{-5}
10K	9.73×10^{-4}	1.74×10^{-4}	4.17×10^{-4}	5.11×10^{-5}
50K	2.0×10^{-4}	4.90×10^{-5}	6.98×10^{-5}	1.66×10^{-5}
cross barrier	62%	86%	40%	86%
cross exercise	29%	42%	n/a	n/a

$H = 91, price = 0.3670447223$

More details can be found in [2].

3 Control Variates

Let Y be an estimator for μ, i.e., $E[Y] = \mu$. The random variable C is called a control variate for Y if C has known mean, μ_C, and C is correlated with Y.

Now consider

$$Y(\beta) = Y - \beta(C - \mu_C).$$

Note that $E[Y(\beta)] = E[Y] = \mu$. For its variance, we have

$$\begin{aligned} VarY(\beta) &= VarY + \beta^2 VarC - 2Cov(Y, \beta(C - \mu_C)) \\ &= VarY + \beta^2 VarC - 2Cov(Y, \beta C) + 2Cov(Y, \beta\mu_C) \\ &= VarY - 2\beta Cov(Y, C) + \beta^2 VarC. \end{aligned}$$

Therefore, if $2\beta Cov(Y, C) > \beta^2 VarC$, then variance reduction is achieved. Using Calculus, you can verify that the value of β that minimizes $VarY - 2\beta Cov(Y, C) + \beta^2 VarC$ is

$$\beta^* = \frac{Cov(Y, C)}{VarC}$$

and the minimum variance that corresponds to β^* is

$$\begin{aligned} VarY(\beta^*) &= VarY - 2\frac{Cov(Y,C)}{VarC}Cov(Y, C) + \frac{(Cov(Y,C))^2}{(VarC)^2}VarC \\ &= \left(1 - 2\frac{(Cov(Y,C))^2}{(VarY)(VarC)} + \frac{(Cov(Y,C))^2}{(VarY)(VarC)}\right)VarY \\ &= (1 - \rho_{YC}^2)VarY \end{aligned}$$

where ρ_{YC} is the correlation coefficient between Y and C.

In general, β^* will not be known. In an application, we can estimate β^* from a sample:

$$\hat{\beta}_n^* = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(C_i - \bar{C})}{\sum_{i=1}^n (C_i - \bar{C})^2}$$

where $\bar{Y}, \bar{C}$, are the sample means. Note that, if we draw a least-squares regression line to the data $(C_i, Y_i), i = 1, ..., n$, then the slope of the regression line is $\hat{\beta}_n^*$.

Another type of control variate is one for which the mean $E[C]$ is unknown but equal to μ, i.e., $E[Y] = E[C] = \mu$. Then, any linear combination

$$Y(\beta) = \beta Y + (1 - \beta)C$$

is again an unbiased estimator of μ and if $Y \& C$ are correlated, variance reduction will be achieved.

Example 4 *Consider pricing a European call option, where we estimate the expected value of the discounted payoff function*

$$Y = e^{-rT}(S(T) - K)^+$$

by Monte Carlo. We will use the underlying asset as control. Let $C = S(T)$, *and then* $E[C] = e^{rT}S(0)$. *The control variate estimator is:*

$$\begin{aligned} Y(\beta) &= Y - \beta(C - \mu_C) \\ &= e^{-rT}(S(T)-K)^+ - \beta(S(T) - e^{rT}S(0)) \end{aligned}$$

We estimate $E[Y(\beta)]$ *using*

$$\frac{1}{N}\sum_{i=1}^{N} \underbrace{e^{-rT}(S_i(T)-K)^+}_{Y_i} - \hat{\beta}_N^*(\underbrace{S_i(T)}_{C_i} - \underbrace{e^{rT}S(0)}_{\mu_C})$$

where

$$\hat{\beta}_N^* = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(C_i - \bar{C})}{\sum_{i=1}^{n}(C_i - \bar{C})^2}.$$

Here are some numerical results:

K	40	40	60
N	10	1000	10
$\hat{\beta}_N^*$	0.882	0.878	0.154
$\hat{\rho}$	1	0.9997	0.795
$\hat{\rho}^2$	1	0.99	0.63
Crude MC	12.5624	14.7728	0.573679
Control MC	14.7001	14.7158	1.24738
BSM value	14.7218	14.7218	1.82405

Recall that

$$\frac{VarY(\beta^*)}{VarY} = 1 - \rho_{YC}^2.$$

Then, we have

$$\begin{aligned} \rho_{YC}^2 &= 0 \Rightarrow VarY(\beta^*) = VarY \\ \rho_{YC}^2 &= 1 \Rightarrow VarY(\beta^*) = 0 \\ \rho_{YC}^2 &= 0.63 \Rightarrow \text{ 63\% of variance} \\ &\quad \text{is eliminated by using the control} \end{aligned}$$

Also observe that if K *is small, we get higher education. Why is that?*

3.1 Asian options

Arithmetic Asian options with discrete monitoring have payoff functions

$$\begin{aligned} \text{call payoff} &: \quad (\bar{S}_A - K)^+ \\ \text{put payoff} &: \text{-} \quad (K - \bar{S}_A)^+ \end{aligned} \tag{14}$$

where

$$\bar{S}_A = \frac{1}{n}\sum_{i=1}^{n} S(t_i) \tag{15}$$

and K is the exercise price.

There are other payoff functions as well, for example, $(S(T)-\bar{S})^+$ for call, and $(\bar{S}-S(T))^+$ for put. However, here we will only consider the payoffs (14).

Note that the asset prices are only observed at a fixed set of dates $t_1,...,t_n$. There is no exact formula for these options. If the asset is observed continuously, then the discrete average (15) is replaced by the continuous average $\bar{S}_A = \left(\int_u^t S(\tau)d\tau\right)/(t-u)$ over an interval (u,t). Some analytical results exist for this case.

Geometric Asian options have the same payoff functions (14) but the arithmetic average $\bar{S}_A$ is replaced by the geometric average

$$\bar{S}_G = \left(\prod_{i=1}^n S(t_i)\right)^{1/n} \tag{16}$$

in the discrete case, and $\bar{S}_G = \exp\left(\int_u^t \log S(\tau)d\tau\right)$ in the continuous case. There are analytical formulas for geometric Asian options.

We will develop a Monte Carlo algorithm for the discrete Asian arithmetic (call) option, using the discrete Asian geometric (call) option, whose price can be computed exactly, as a control variate. First, we derive a formula for the Asian geometric option.

The stock price model is the GBM model:

$$S(t_i) = S(0)\exp\left((r-\sigma^2/2)t_i + \sigma W(t_i)\right) \tag{17}$$

where $W(t)$ is the standard Brownian motion on $[0,T]$. Multiplying (17) as $i=1,...,n$, we obtain

$$\bar{S}_G = \left(\prod_{i=1}^n S(t_i)\right)^{1/n} = S(0)\exp\left(\left(r-\frac{\sigma^2}{2}\right)\left(\frac{1}{n}\sum_{i=1}^n t_i\right) + \frac{\sigma}{n}\sum_{i=1}^n W(t_i)\right) \tag{18}$$

We need to find the distribution of $\sum_{i=1}^n W(t_i)$.

Recall that $(W(t_1),...,W(t_n))$ can be written as a linear combination of i.i.d. normal variables and thus it has multivariate normal distribution with mean vector 0 and covariance matrix C where $c_{ij} = \min(t_i,t_j)$.

Lemma 2 $\sum_{i=1}^n W(t_i) \sim \mathbf{N}\left(0, \sum_{i=1}^n (2n-(2i-1))t_i\right)$
Proof. *If $X \sim \mathbf{N}(\mu,\Sigma)$, then $AX \sim \mathbf{N}(A\mu, A\Sigma A^T)$ where μ is an $n-$vector, Σ is the $n\times n$ covariance matrix, and A is any $k\times n$ matrix. Take $A = (1,1,...,1)_{1\times n}, X = (W(t_1),...,W(t_n))_{n\times 1}$. Then $AX = \sum_{i=1}^n W(t_i)$ and*

$$ACA^T = \sum_{i,j}\min(t_i,t_j) = \sum_{i=1}^n (2n-(2i-1))t_i.$$

■

To price the Asian geometric option, we need to compute $E[e^{-rT}(\bar{S}_G - K)^+]$.The following lemma will help.

Lemma 3 *Let X have lognormal distribution such that $Var(\log X) = s^2$. Then*

$$E[(X-K)^+] = E[X]\Phi(d_1) - K\Phi(d_2)$$

where

$$\begin{aligned} d_1 &= \frac{\log(E[X]/K) + s^2/2}{s} \\ d_2 &= \frac{\log(E[X]/K) - s^2/2}{s} \end{aligned}$$

and Φ is the standard normal c.d.f.

Recall that if Y is a lognormal random variable with parameters $\bar{\mu}, \bar{\sigma}^2$ (i.e., $\log Y \sim \mathbf{N}(\bar{\mu}, \bar{\sigma}^2)$) then $E[Y] = \exp(\bar{\mu} + \bar{\sigma}^2/2)$. In our problem, $\bar{S}_G$ is lognormal with

$$\log \bar{S}_G = \mathbf{N}\left(\log S(0) + \left(r - \frac{\sigma^2}{2}\right)\left(\frac{1}{n}\sum_{i=1}^{n} t_i\right), \frac{\sigma^2}{n^2}\sum_{i=1}^{n}(2n-(2i-1))t_i\right),$$

which follows from (18) and Lemma 2. Therefore,

$$\begin{aligned} E[\bar{S}_G] &= \exp\left(\log S(0) + \left(r - \frac{\sigma^2}{2}\right)\left(\frac{1}{n}\sum_{i=1}^{n} t_i\right) + \frac{\sigma^2}{2n^2}\sum_{i=1}^{n}(2n-(2i-1))t_i\right) \quad (19) \\ &= S(0)\exp\left(\left(r - \frac{\sigma^2}{2}\right)\left(\frac{1}{n}\sum_{i=1}^{n} t_i\right) + \frac{\sigma^2}{2n^2}\sum_{i=1}^{n}(2n-(2i-1))t_i\right) \end{aligned}$$

and

$$s^2 = Var(\log \bar{S}_G) = \frac{\sigma^2}{n^2}\sum_{i=1}^{n}(2n-(2i-1))t_i. \quad (20)$$

In conclusion, the price of the discrete Asian geometric call option observed at $t_1, ..., t_n$, with payoff $(\bar{S}_G - K)^+$, where $\bar{S}_G = \left(\prod_{i=1}^{n} S(t_i)\right)^{1/n}$, is

$$E[\bar{S}_G]\Phi(d_1) - K\Phi(d_2).$$

Here $E[\bar{S}_G]$ is given by (19) and

$$\begin{aligned} d_1 &= \frac{\log(E[\bar{S}_G]/K) + s^2/2}{s} \\ d_2 &= \frac{\log(E[\bar{S}_G]/K) - s^2/2}{s} \end{aligned}$$

and s^2 is given by (20), Φ is the standard normal c.d.f.

3.1.1 Using Asian geometric option as a control

The crude Monte Carlo algorithm for the arithmetic Asian call option is

$$Y = e^{-rT}(\bar{S}_A - K)^+,$$

and the control is

$$C = e^{-rT}(\bar{S}_G - K)^+.$$

$E[C]$ is computed from the formula we derived above. The control variate estimator is

$$\begin{aligned} Y(\beta) &= Y - \beta(C - \mu_C) \\ &= e^{-rT}(\bar{S}_A - K)^+ - \beta(e^{-rT}(\bar{S}_G - K)^+ - E[C]). \end{aligned}$$

We estimate $E[Y(\beta)]$ using the sample mean

$$\frac{1}{N}\sum Y_i - \hat{\beta}^*_N(C_i - E[C])$$

where

$$\hat{\beta}^*_N = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(C_i - \bar{C})}{\sum_{i=1}^n (C_i - \bar{C})^2}.$$

4 Conditional Monte Carlo

We will first review some facts from probability theory. We start with the definition of conditional expectation. Let X, Y be random variables. The conditional expectation of X, given that $Y = y$, is defined as

$$E[X \mid Y = y] = \begin{cases} \sum_x xP\{X = x \mid Y = y\} = \sum_x x\frac{P\{X=x,Y=y\}}{P\{Y=y\}}, & \text{discrete case} \\ \frac{\int xf(x,y)dx}{\int f(x,y)dx}, & \text{continuous case} \end{cases}$$

We now define a random variable, as a function of Y, as follows:

$$E[X \mid Y] = h(Y)$$

The value of this random variable at $Y = y$ is $E[X \mid Y = y]$.

Theorem 4

$$E[E[X \mid Y]] = E[X]$$

Proof. *We will prove this for the discrete case. We want to show that*

$$\sum_y h(y)P\{Y = y\} = \sum_y E[X \mid Y = y]P\{Y = y\} = E[X].$$

The left-hand side simplifies as:

$$\begin{aligned}
\sum_y \left(\sum_x x P\{X = x \mid Y = y\} \right) P\{Y &= y\} = \sum_y \sum_x x P\{X = x, Y = y\} \\
&= \sum_x x \underbrace{\sum_y P\{X = x, Y = y\}}_{P\{X=x\}} \\
&= E[X]
\end{aligned}$$

where the change of summation is valid if the expectation exist. ■

Definition 3 *The conditional variance of X given Y is defined as*

$$Var(X \mid Y) = E\left[(X - E[X|Y])^2 \mid Y\right].$$

$Var(X \mid Y)$ is a function of Y - it is a random variable. When $Y = y$, it takes the value

$$Var(X \mid Y = y) = E\left[(X - E[X|Y = y])^2 \mid Y = y\right].$$

Lemma 4

$$Var(X|Y) = E[X^2|Y] - E[X|Y]^2$$

Taking expectations of both sides, we get

$$\begin{aligned}
E[Var(X|Y)] &= E[E[X^2|Y]] - E[E[X|Y]^2] \\
&= E[X^2] - E[E[X|Y]^2]
\end{aligned} \tag{21}$$

We also have the following relationship

$$\begin{aligned}
Var(E[X|Y]) &= E[E[X|Y]^2] - (E[E[X|Y]])^2 \\
&= E[E[X|Y]^2] - (E[X])^2
\end{aligned} \tag{22}$$

If we add equations (21) and (22), we obtain

$$E[Var(X|Y)] + Var(E[X|Y]) = E[X^2] - (E[X])^2,$$

which gives us the "conditional variance formula":

$$VarX = E[Var(X|Y)] + Var(E[X|Y]).$$

The conditional variance formula implies that

$$Var(E[X|Y]) \leq VarX$$

Observe that

1. X and $E[X|Y]$ are unbiased estimators of θ, i.e., $E[X] = E[E[X|Y]] = \theta$;

2. Variance of $E[X|Y]$ is smaller or equal to the variance of X.

The conditional Monte Carlo method estimates θ using $E[X|Y]$. The crude Monte Carlo method generates values of X, i.e., $X_1, ..., X_n$, and compute the sample mean of the X_i. The conditional Monte Carlo generates values of Y (think of Y as the output produced in the first stage of simulation, prior to generating X), and estimates θ by

$$\frac{1}{N}\sum_{i=1}^{N} E[X \mid Y = y_i].$$

4.1 Down-and-in barrier options

Consider a down-and-in European call option written on a stock $S(t)$, whose prices are observed at discrete time steps $0 = t_0 < t_1 < \cdots < t_m = T$. The barrier price will be denoted by H, the exercise price by K, and the expiry by T. The option becomes a vanilla European call option when the stock "crosses" the barrier, i.e., if $S(t_i) < H$ for some $t_i < T$. We will assume that $S(0) > H$. The time the stock crosses the barrier is t_τ, where τ is a nonnegative random variable.

The payoff of the down-and-in European call option at time T is the product

$$\mathbf{1}(\tau < m) \cdot (S(T) - K, 0)^+, \tag{23}$$

where $\mathbf{1}(\,\cdot\,)$ denotes the indicator function and $(x, y)^+$ denotes the maximum of x and y. The price of the option at time 0 is the discounted expected payoff

$$I = \mathbf{E}\left[e^{-rT}\mathbf{1}(\tau < m)(S(T) - K, 0)^+\right] \tag{24}$$

where r is the risk-free interest rate and the expectation is conditioned on the initial stock price $S(0)$.

The crude Monte Carlo estimates I by simulating N stock price paths, averaging over the values

$$\begin{cases} (Y - K, 0)^+ \text{ if } \tau < m \text{ and } S(T) = Y \\ \qquad\qquad 0 \text{ otherwise} \end{cases}$$

and then discounting this average by multiplying by e^{-rT}.

The conditional expectation Monte Carlo is based on the following observation: If it is possible to price the down-and-in European call option analytically when it crosses the barrier, then one should use this analytical formula instead of simulating the price path until expiry. An example is the lognormal model for $S(t)$, for which the Black-Scholes-Merton formula can be used to price the European call option, when the stock crosses the barrier. To proceed, we condition the expectation (24) on τ and $S(t_\tau)$, and consider the case when $\tau < m$:

$$\mathbf{E}\left[(S(T) - K, 0)^+\right] = \mathbf{E}[\mathbf{E}[(S(T) - K, 0)^+ \mid \tau, S(t_\tau)]]. \tag{25}$$

To compute the inner expectation, we assume the lognormal model for $S(t)$, and obtain

$$\begin{aligned}\mathbf{E}[(S(T)-K,0)^+ \mid \tau=k, S(t_k)] &= \mathbf{E}[(S(T)-K,0)^+ \mid S(t_k)] \\ &= e^{r(T-t_k)} BSM(S(t_k), t_k, T)\end{aligned}$$

where $BSM(S(t_k), t_k, T)$ is the Black-Scholes-Merton formula for the price of a European call option at time t_k, with expiry T, and initial stock price $S(t_k)$:

$$\begin{aligned}BSM(S(t_k), t_k, T) &= \mathbf{E}[e^{-r(T-t_k)}(S(T)-K)^+ \mid S(t_k)=\mathcal{S}] \\ &= \mathcal{S}\Phi(d_1) - Ke^{-r(T-t_k)}\Phi(d_2)\end{aligned}$$

with

$$d_1 = d_2 + \sigma\sqrt{T-t_k} = \frac{\log(\mathcal{S}/K) + (r+\sigma^2/2)(T-t_k)}{\sigma\sqrt{T-t_k}}.$$

Therefore, the inner expectation in (25) is

$$\mathbf{E}[(S(T)-K,0)^+ \mid \tau, S(t_\tau)] = e^{r(T-t_\tau)} BSM(S(t_\tau), t_\tau, T)$$

and (25) simplifies to

$$\mathbf{E}\left[(S(T)-K,0)^+\right] = \mathbf{E}[e^{r(T-t_\tau)} BSM(S(t_\tau), t_\tau, T)].$$

So, the conditional Monte Carlo estimator simulates N paths and averages over the values

$$\begin{cases} e^{r(T-t_k)} BSM(S(t_k), t_k, T) \text{ if } \tau = k < m \\ \qquad\qquad 0 \text{ otherwise} \end{cases}.$$

Discounting this average by e^{-rT} gives an unbiased estimate for the option value I. In other words, we estimate the option value by

$$e^{-rT}\left(\frac{1}{N}\sum_{i=1}^{N} e^{r(T-t_\tau)} BSM(S^{(i)}(t_\tau), t_\tau, T) 1(\tau < m)\right) = \tag{26}$$
$$\frac{1}{N}\sum_{i=1}^{N} e^{-rt_\tau} BSM(S^{(i)}(t_\tau), t_\tau, T) 1(\tau < m)$$

where $S^{(i)}$ is the ith stock price path, and $S^{(i)}(t_\tau)$ is the price observed at the time when the path crosses the barrier. If the barrier is not crossed by the expiry, the estimator value is zero[1].

The following table was discussed earlier when we covered importance sampling. The option has the following parameters: $\sigma = 0.3$, $r = 0.1$, $S(0) = 100$,

[1] This formula was erroneously written with the discount factor e^{-rT} in Boyle et al. [4] (page 1289), and with no discounting factor in Ross & Shanthikumar [3] (equation 2, page 321).

$K = 100$, $H = 91, T = 0.2$, and $m = 50$. The table displays the relative mean square error of fifty independent estimates.

N	Crude MC	ImpSamp-h	CondExp	Combined-h
5K	1.66×10^{-3}	3.56×10^{-4}	4.66×10^{-4}	9.78×10^{-5}
10K	9.73×10^{-4}	1.74×10^{-4}	4.17×10^{-4}	5.11×10^{-5}
50K	2.0×10^{-4}	4.90×10^{-5}	6.98×10^{-5}	1.66×10^{-5}
cross barrier	62%	86%	40%	86%
cross exercise	29%	42%	n/a	n/a

$H = 91, price = 0.3670447223$

Remark 4 *For the down-and-in barrier option, one can combine importance sampling and conditional expectation estimators. In this approach, we would twist the distribution so that the random walk crosses the barrier with a larger probability, and use the BSM formula to price the option when it crosses the barrier. The "Combined-h" in the table above refers to this combined estimator. For details see Ökten et. al. [2].*

The Monte Carlo estimates in the above table can be improved if we use randomized quasi-Monte Carlo methods. In the following tables, we use scrambled Faure sequences, for a small dimensional problem with $m = 5$.

N	CondExp	ImpSamp-h	Combined-h
5K	2.75×10^{-5}	9.16×10^{-5}	1.19×10^{-5}
10K	9.16×10^{-6}	4.97×10^{-5}	5.87×10^{-6}
50K	1.09×10^{-6}	6.69×10^{-6}	6.78×10^{-7}

RQMC results. $m = 5, H = 93, price = 0.3443581039$

N	CondExp	ImpSamp-h	Combined-h
5K	6.02×10^{-4}	1.03×10^{-3}	2.37×10^{-4}
10K	1.93×10^{-4}	4.53×10^{-4}	1.04×10^{-4}
50K	7.89×10^{-5}	8.26×10^{-5}	3.59×10^{-5}

MC results. $m = 5, H = 93, price = 0.3443581039$

N	CondExp	ImpSamp-h	Combined-h
5K	22	11	20
10K	21	9	18
50K	72	12	53

MC/RQMC ratios. $m = 5$

5 Antithetic Variates

This is probably the most commonly used, and the simplest to implement, of all variance reduction techniques. We seek two unbiased estimators X and Y for some unknown parameter I, having strong correlation. Then,

$$\frac{1}{2}(X + Y)$$

is an unbiased estimator for I, and

$$Var\left(\frac{X+Y}{2}\right) = \frac{1}{4}VarX + \frac{1}{4}VarY + \frac{1}{2}Cov(X,Y).$$

Therefore, if $Cov(X,Y)$ is strongly negative, then this method can reduce variance.

As an example, let $I = \int_0^1 g(x)dx$. Set $X = g(U)$ and $Y = g(1-U)$, where U is a uniform random variable on $(0,1)$. Then

$$I = \frac{1}{2}\int_0^1 \left(g(u) + g(1-u)\right) du$$

and we estimate I by

$$\eta = \frac{1}{2N}\sum_{i=1}^{N}(g(u_i) + g(1-u_i))$$

where $u_1, u_2, ...$ are pseudorandom numbers.

Observe that the time required for one computation of η takes twice as time as required by the crude Monte Carlo

$$\gamma = \frac{1}{N}\sum_{i=1}^{N} g(u_i).$$

If we define the efficiency of a Monte Carlo simulation as the computing time times the variance, then the following theorem shows that estimating η is more efficient than the crude Monte Carlo.

Theorem 5 *Let $g(x)$ be a continuous and monotonic function with continuous first derivatives. Then*

$$Var(\eta) \leq \frac{1}{2}Var(\gamma).$$

Proof. *See Rubinstein [5] for a proof.* ■

In general, let

$$I = \int_{-\infty}^{\infty} g(x)f(x)dx$$

where f is a probability density function on $(-\infty, \infty)$, and F the corresponding distribution function. Then we estimate I by

$$\eta = \frac{1}{2N}\sum_{i=1}^{N}(g(x_i) + g(y_i))$$

where

$$x_i = F^{-1}(u_i), y_i = F^{-1}(1-u_i).$$

The fact that the variables x_i, y_i are negatively correlated is a consequence of the corollary below.

Theorem 6 *If $X_1, ..., X_n$ are independent random variables, then for any increasing functions f and g of n variables*

$$E[f(X)g(X)] \geq E[f(X)]E[g(X)]$$

where $X = (X_1, ..., X_n)$.
Proof. *See Ross [6] for a proof.* ■

Corollary 7 *If $h(x_1, ..., x_n)$ is a monotone function of each of its variables, then for a set $U_1, ..., U_n$ of independent random variables*

$$Cov(h(U_1, ..., U_n), h(1 - U_1, ..., 1 - U_n)) \leq 0$$

Proof. *Assume that h is increasing in its first r arguments and decreasing in its final $n - r$. Then*

$$\begin{aligned} f(x_1, ..., x_n) &= h(x_1, ..., x_r, 1 - x_{r+1}, ..., 1 - x_n) \\ g(x_1, ..., x_n) &= -h(1 - x_1, ..., 1 - x_r, x_{r+1}, ..., x_n) \end{aligned}$$

are both increasing, and

$$Cov(f(U_1, ..., U_n), g(1 - U_1, ..., 1 - U_n)) \geq 0$$

from the above theorem, or

$$Cov(h(U_1, ..., U_r, 1 - U_{r+1}, ..., 1 - U_n), h(1 - U_1, ..., 1 - U_r, U_{r+1}, ..., U_n)) \leq 0.$$

Finally, observe that the joint distribution of the random vector $h(U_1, ..., U_n), h(1-U_1, ..., 1-U_n)$ is the same as the joint distribution of $h(U_1, ..., U_r, 1-U_{r+1}, ..., 1-U_n), h(1 - U_1, ..., 1 - U_r, U_{r+1}, ..., U_n)$. ■

References

[1] Glynn and Iglehart, "Importance Sampling for Stochastic Simulations", Management Science 35, 1367-1392.

[2] G. Ökten, E. Salta, A. Göncü, "On Pricing Discrete Barrier Options Using Conditional Expectation and Importance Sampling Monte Carlo", Mathematical and Computer Modelling, Vol. 47, pp. 484-494, 2008.

[3] S. M. Ross and J. G. Shanthikumar, Pricing exotic options: monotonicity in volatility and efficient simulation, *Probability in the Engineering and Informational Sciences* **14**, 317-326, (2000).

[4] P. Boyle, M. Broadie, and P. Glasserman, Monte Carlo methods for security pricing, *Journal of Economic Dynamics and Control* **21**, 1267-1321, (1997).

[5] R. Y. Rubinstein, Simulation and the Monte Carlo Method, John Wiley & Sons, 1981.

[6] S. Ross, Simulation, Academic Press, 4th edition, 2006.

6 Problems

1. Consider the Bernoulli probability mass function

$$f(x) = p^x(1-p)^{1-x}; x = 0, 1$$

Find the corresponding tilted mass function $f_t(x)$. Is the tilted mass function a Bernoulli mass function?

2. If $0 \leq X \leq a$, show that

   (a) $E[X^2] \leq aE[X]$

   (b) $Var(X) \leq E[X](a - E[X])$

   (c) $Var(X) \leq a^2/4$
   This result shows that to minimize variance, we can try minimizing the upper bound on X. This result can be useful in finding optimal parameters for tilted densities.

3. In proving Lemma 2, we used the identity

$$\sum_{i,j} \min(t_i, t_j) = \sum_{i=1}^{n} (2n - (2i-1))t_i$$

without proof. Prove this identity.

4. We briefly discussed estimators of the type

$$\beta Y + (1-\beta)C \tag{27}$$

where $E[Y] = E[C] = \theta$. First, show that the minimum variance is attained if the constant β is chosen as

$$\beta^* = \frac{Var(C) - Cov(Y,C)}{Var(Y) + Var(C) - 2Cov(Y,C)}.$$

Now let C be the conditional expectation estimator, i.e., $C = E[Y|Z]$, and (27) becomes

$$\beta Y + (1-\beta)E[Y|Z].$$

Show that $\beta^* = 0$, which means that no further improvement is possible by combining Y and $E[Y|Z]$.

5. In the following, we will use Monte Carlo simulation to estimate $\theta = E[e^U] = \int_0^1 e^x dx$.

   (a) What is the sample mean estimator and antithetic variates estimator for θ? Find the variance of these two estimators, and find how much variance reduction is achieved by the antithetic variates estimator over the sample mean estimator.

(b) What is the control variates estimator for θ, if we take $f(U) = U$ as the control? Find the variance of this estimator, and compare it to the ones found in (a).

(c) Write a *computer* program for the three estimators mentioned in parts (1) and (2). Compare the estimators numerically, by comparing the absolute value of the error produced by each estimator for $N = 10000, 20000, ..., 50000$. You can use your favorite random number generator.