IEOR E4703: Monte-Carlo Simulation

Simulation Efficiency and an Introduction to Variance Reduction Methods

Martin Haugh

Department of Industrial Engineering and Operations Research
Columbia University
Email: martin.b.haugh@gmail.com

Outline

Simulation Efficiency

As usual we wish to estimate $\theta := \mathbb{E}[h(\mathbf{X})]$. Standard simulation algorithm is:

1. Generate $\mathbf{X}_1, \ldots, \mathbf{X}_n$
2. Estimate θ with $\hat{\theta}_n = \sum_{j=1}^n Y_j/n$ where $Y_j := h(\mathbf{X}_j)$.
3. Approximate $100(1-\alpha)\%$ confidence intervals are then given by

$$\left[\hat{\theta}_n - z_{1-\alpha/2}\frac{\hat{\sigma}_n}{\sqrt{n}}\ ,\ \hat{\theta}_n + z_{1-\alpha/2}\frac{\hat{\sigma}_n}{\sqrt{n}}\right].$$

Can measure quality of $\hat{\theta}_n$ by the half-width HW of the CI

$$HW = z_{1-\alpha/2}\sqrt{\frac{\mathrm{Var}(Y)}{n}}.$$

Would like HW to be small but sometimes this is difficult to achieve.

So often imperative to address the issue of simulation efficiency. There are a number of things we can do:

Simulation Efficiency

1. Develop a good simulation algorithm.
2. Program carefully to minimize storage requirements.

   e.g. Do not need to store all Y_j's. Instead just store $\sum Y_j$ and $\sum Y_j^2$ to compute $\hat{\theta}_n$ and approximate CI's.
3. Program carefully to minimize execution time.
4. Decrease variability of simulation output that we use to estimate θ.

   Techniques used to do this are called variance reduction techniques.

Will now study some of the simplest variance reduction techniques, and assume we are doing items (1) to (3) as well as possible.

But first we should first discuss a measure of simulation efficiency.

Measuring Simulation Efficiency

Suppose there are two r.vars, W and Y, such that $\mathbb{E}[W] = \mathbb{E}[Y] = \theta$.

Let M_w and M_y denote methods of estimating θ by simulating the W_i's and Y_i's, respectively.

Question: Which method is more efficient, M_w or M_y?

To answer this, let n_w and n_y be the number of samples of W and Y, respectively, that are needed to achieve a half-width, HW. Then

$$
\begin{aligned}
n_w &= \left(\frac{z_{1-\alpha/2}}{HW}\right)^2 \mathrm{Var}(W) \\
n_y &= \left(\frac{z_{1-\alpha/2}}{HW}\right)^2 \mathrm{Var}(Y).
\end{aligned}
$$

Let E_w and E_y denote the amount of **computational effort** required to produce one sample of W and Y, respectively.

Measuring Simulation Efficiency

Then the total effort expended by M_w and M_y, respectively, to achieve HW are

$$
\begin{aligned}
TE_w &= \left(\frac{z_{1-\alpha/2}}{HW}\right)^2 \mathsf{Var}(W)\ E_w \\
TE_y &= \left(\frac{z_{1-\alpha/2}}{HW}\right)^2 \mathsf{Var}(Y)\ E_y.
\end{aligned}
$$

Say M_w is **more efficient** than M_y if $TE_w < TE_y$. This occurs if and only if

$$\mathsf{Var}(W)E_w < \mathsf{Var}(Y)E_y. \tag{1}$$

Will therefore use $\mathsf{Var}(W)E_w$ as a measure of M_w's efficiency of the.

Note that (1) implies we cannot conclude that M_w is better than M_y simply because $\mathsf{Var}(W) < \mathsf{Var}(Y)$.

But often the case that $E_w \approx E_y$ and $\mathsf{Var}(W) << \mathsf{Var}(Y)$.

In such cases it is clear that using M_w provides a substantial improvement over using M_y.

Control Variates

We wish to estimate $\theta := \mathbb{E}[Y]$ where $Y = h(\mathbf{X})$ is the output of a simulation experiment.

Suppose Z is also an output (or that we can easily output it if we wish).

Also assume **we know** $\mathbb{E}[Z]$.

Then can construct many unbiased estimators of θ:

1. $\hat{\theta} = Y$, our usual estimator
2. $\hat{\theta}_c := Y + c(Z - \mathbb{E}[Z])$.

Variance of $\hat{\theta}_c$ satisfies

$$\mathsf{Var}(\hat{\theta}_c) = \mathsf{Var}(Y) + c^2\mathsf{Var}(Z) + 2c\,\mathsf{Cov}(Y, Z). \tag{2}$$

Can choose c to minimize this quantity and optimal value given by

$$c^* = -\frac{\mathsf{Cov}(Y, Z)}{\mathsf{Var}(Z)}.$$

Control Variates

Minimized variance satisfies

$$\begin{aligned}\text{Var}(\hat{\theta}_{c^*}) &= \text{Var}(Y) - \frac{\text{Cov}(Y,Z)^2}{\text{Var}(Z)} \\ &= \text{Var}(\hat{\theta}) - \frac{\text{Cov}(Y,Z)^2}{\text{Var}(Z)}.\end{aligned}$$

In order to achieve a variance reduction therefore only necessary that $\text{Cov}(Y,Z) \neq 0$.

New resulting Monte Carlo algorithm proceeds by generating n samples of Y and Z and then setting

$$\hat{\theta}_{c^*} = \frac{\sum_{i=1}^{n} (Y_i + c^*(Z_i - \mathbb{E}[Z]))}{n}.$$

There is a problem with this, however, as we usually do not know $\text{Cov}(Y,Z)$.

Resolve this problem by doing p pilot simulations and setting

$$\widehat{\text{Cov}}(Y,Z) = \frac{\sum_{j=1}^{p} (Y_j - \overline{Y}_p)(Z_j - \mathbb{E}[Z])}{p-1}.$$

Control Variates

If it is also the case that $\mathsf{Var}(Z)$ unknown, then can also estimate it with

$$\widehat{\mathsf{Var}}(Z) = \frac{\sum_{j=1}^{p}(Z_j - \mathbb{E}[Z])^2}{p-1}$$

and finally set

$$\hat{c}^* = -\frac{\widehat{\mathsf{Cov}}(Y, Z)}{\widehat{\mathsf{Var}}(Z)}.$$

Our control variate simulation algorithm is as follows:

Control Variate Simulation Algorithm for Estimating $\mathbb{E}[Y]$

/* Do pilot simulation first */
for $i = 1$ **to** p

    generate (Y_i, Z_i)
end for
compute $\hat{c}^*$
/* Now do main simulation */
for $i = 1$ **to** n

    generate (Y_i, Z_i)
    set $V_i = Y_i + \hat{c}^*(Z_i - \mathbb{E}[Z])$
end for
set $\hat{\theta}_{\hat{c}^*} = \overline{V}_n = \sum_{i=1}^{n} V_i/n$
set $\hat{\sigma}^2_{n,v} = \sum(V_i - \hat{\theta}_{\hat{c}^*})^2/(n-1)$
set $100(1-\alpha)\%$

$$\text{CI} = \left[\hat{\theta}_{\hat{c}^*} - z_{1-\alpha/2}\frac{\hat{\sigma}_{n,v}}{\sqrt{n}}\ ,\ \hat{\theta}_{\hat{c}^*} + z_{1-\alpha/2}\frac{\hat{\sigma}_{n,v}}{\sqrt{n}}\right]$$

e.g. Pricing an Asian Call Option

Payoff of an Asian call option given by

$$h(\mathbf{X}) := \max\left(0, \frac{\sum_{i=1}^{m} S_{iT/m}}{m} - K\right)$$

where $\mathbf{X} := \{S_{iT/m} : i = 1, \ldots, m\}$, K is the strike and T the expiration date.

Price of option then given by

$$C_a = \mathbb{E}_0^Q[e^{-rT} h(\mathbf{X})].$$

Will assume as usual that $S_t \sim GBM(r, \sigma)$ under Q.

Usual simulation method for estimating C_a is to generate n independent samples of the payoff, $Y_i := e^{-rT} h(\mathbf{X}_i)$, for $i = 1, \ldots, n$, and to set

$$\hat{C}_a = \frac{\sum_{i=1}^{n} Y_i}{n}.$$

e.g. Pricing an Asian Call Option

But could also estimate C_a using control variates and there are many possible choices:

1. $Z_1 = S_T$
2. $Z_2 = e^{-rt} \max(0, S_T - K)$
3. $Z_3 = \sum_{j=1}^{m} S_{iT/m}/m$

In each of the three cases, it is easy to compute $\mathbb{E}[Z]$.

Would also expect Z_1, Z_2 and Z_3 to have a positive covariance with Y, so that each one would be a suitable candidate for use as a control variate.

Question: Which one would lead to the greatest variance reduction?

Question: Explain why you could also use the value of the option with payoff

$$g(\mathbf{X}) := \max\left(0, \left(\prod_{i=1}^{m} S_{iT/m}\right)^{1/m} - K\right)$$

as a control variate. Do you think it would result in a substantial variance reduction?

e.g. The Barbershop

A barbershop opens for business every day at 9am and closes at 6pm.

There is only 1 barber but he's considering hiring another one.

But first he would like to estimate the **mean total time that customers spend waiting each day**.

Assume customers arrive at barbershop according to a non-homogeneous Poisson process, $N(t)$, with intensity $\lambda(t)$.

Let W_i denote waiting time of i^{th} customer.

The barber closes the shop after $T = 9$ hours but still serves any customers who have arrived before then.

Quantity that he wants to estimate is $\theta := \mathbb{E}[Y]$ where

$$Y := \sum_{j=1}^{N(T)} W_j.$$

e.g. The Barbershop

Assume the service times of customers are IID with CDF, $F(.)$, and that they are also independent of the arrival process, $N(t)$.

Usual simulation algorithm: simulate n days of operation in the barbershop, thereby obtaining n samples, $Y_1, \ldots, Y_n$, and then set

$$\hat{\theta}_n = \frac{\sum_{j=1}^{n} Y_j}{n}.$$

However, a better estimate could be obtained by using a control variate.

Let Z denote the total time customers on a given day spend **in service** so that

$$Z := \sum_{j=1}^{N(T)} S_j$$

where S_j is the service time of the j^{th} customer. Then easy to see that

$$\mathbb{E}[Z] = \mathbb{E}[S]\mathbb{E}[N(T)].$$

Intuition suggests that Z would be a good candidate to use as a control variate.

Multiple Control Variates

No reason why we should not use more than one control variate.

So suppose again that we wish to estimate $\theta := \mathbb{E}[Y]$ where Y is the output of a simulation experiment.

Also suppose that for $i = 1, \ldots, m$, Z_i is an output or that we can easily output it if we wish, and that $\mathbb{E}[Z_i]$ is known for each i.

Can then construct unbiased estimators of θ by defining

$$\hat{\theta}_c \quad := \quad Y + c_1(Z_1 - \mathbb{E}[Z_1]) + \ \ldots \ + c_m(Z_m - \mathbb{E}[Z_m]).$$

Variance of $\hat{\theta}_c$ satisfies

$$\text{Var}(\hat{\theta}_c) = \text{Var}(Y) + 2\sum_{i=1}^{m} c_i \text{Cov}(Y, Z_i) + \sum_{i=1}^{m}\sum_{i=1}^{m} c_i c_j \text{Cov}(Z_i, Z_j) \tag{3}$$

Can easily minimize $\text{Var}(\hat{\theta}_c)$ w.r.t the c_i's.

Multiple Control Variates

As before, however, optimal solutions c_i^* will involve unknown covariances (and possibly variances of the Z_i's) that will need to be estimated using a pilot simulation.

A convenient way of doing this is to observe that

$$\hat{c}_i^* = -b_i$$

where the b_i's are the least squares solution to the following linear regression:

$$Y = a + b_1 Z_1 + \ldots + b_m Z_m + \epsilon. \tag{4}$$

The simulation algorithm with multiple control variates is exactly analogous to the simulation algorithm with just a single control variate.

Antithetic Variates

As usual would like to estimate $\theta = \mathbb{E}[h(\mathbf{X})] = \mathbb{E}[Y]$.

Suppose we have generated two samples, Y_1 and Y_2.

Then an unbiased estimate of θ is given by

$$\hat{\theta} := (Y_1 + Y_2)/2$$

with

$$\text{Var}(\hat{\theta}) = \frac{\text{Var}(Y_1) + \text{Var}(Y_2) + 2\text{Cov}(Y_1, Y_2)}{4}.$$

If Y_1 and Y_2 are IID, then $\text{Var}(\hat{\theta}) = \text{Var}(Y)/2$.

However, we could reduce $\text{Var}(\hat{\theta})$ if we could arrange it so that $\text{Cov}(Y_1, Y_2) < 0$. We now describe the method of antithetic variates for doing this.

We will begin with the case where Y is a function of IID $U(0,1)$ random variables so that $\theta = \mathbb{E}[h(\mathbf{U})]$ where $\mathbf{U} = (U_1, \ldots, U_m)$ and the U_i's are IID $\sim U(0,1)$.

Usual Simulation Algorithm for Estimating θ (with $2n$ Samples)

for $i = 1$ to $2n$

  generate $\mathbf{U_i}$
  set $Y_i = h(\mathbf{U_i})$

end for

set $\hat{\theta}_{2n} = \overline{Y}_{2n} = \sum_{i=1}^{2n} Y_i/2n$

set $\hat{\sigma}^2_{2n} = \sum_{i=1}^{2n}(Y_i - \overline{Y}_{2n})^2/(2n-1)$

set approx. $100(1-\alpha)$ % CI $= \hat{\theta}_{2n} \pm z_{1-\alpha/2}\frac{\hat{\sigma}_{2n}}{\sqrt{2n}}$

Antithetic Variates

In above algorithm, however, could also have used the $1 - U_i$'s to generate sample Y values.

Can use this fact to construct another estimator of θ as follows:

1. As before, set $Y_i = h(\mathbf{U_i})$, where $\mathbf{U_i} = (U_1^{(i)}, \ldots, U_m^{(i)})$.
2. Also set $\tilde{Y}_i = h(1 - \mathbf{U_i})$, where we use $1 - \mathbf{U_i} = (1 - U_1^{(i)}, \ldots, 1 - U_m^{(i)})$.
3. Set $Z_i := (Y_i + \tilde{Y}_i)/2$.

Note that $\mathbb{E}[Z_i] = \theta$ so Z_i also unbiased estimator of θ.

If the U_i's are IID, then so too are the Z_i's and can use them as usual to compute approximate CI's for θ.

We say that U_i and $1 - U_i$ are antithetic variates.

Have the following antithetic variate simulation algorithm.

Antithetic Variate Simulation Algorithm for Estimating θ

```
for i = 1 to n
        generate U_i
        set Y_i = h(U_i) and Ỹ_i = h(1 - U_i)
        set Z_i = (Y_i + Ỹ_i)/2
end for
set θ̂_{n,a} = Z̄_n = Σ_{i=1}^n Z_i/n
set σ̂²_{n,a} = Σ_{i=1}^n (Z_i - Z̄_n)²/(n-1)
set approx. 100(1-α) % CI = θ̂_{a,n} ± z_{1-α/2} σ̂_{n,a}/√n
```

for $i = 1$ to n

 generate $\mathbf{U_i}$
 set $Y_i = h(\mathbf{U_i})$ and $\tilde{Y}_i = h(1 - \mathbf{U_i})$
 set $Z_i = (Y_i + \tilde{Y}_i)/2$

end for
set $\hat{\theta}_{n,a} = \overline{Z}_n = \sum_{i=1}^n Z_i/n$
set $\hat{\sigma}^2_{n,a} = \sum_{i=1}^n (Z_i - \overline{Z}_n)^2/(n-1)$
set approx. $100(1-\alpha)$ % CI $= \hat{\theta}_{a,n} \pm z_{1-\alpha/2} \frac{\hat{\sigma}_{n,a}}{\sqrt{n}}$

$\hat{\theta}_{a,n}$ is unbiased and SLLN implies $\hat{\theta}_{n,a} \rightarrow \theta$ w.p. 1 as $n \rightarrow \infty$.

Each of the two algorithms uses $2n$ samples so question arises as to which algorithm is better?

Both algorithms require approximately same amount of effort so comparing the two algorithms amounts to computing which estimator has a smaller variance.

Comparing Estimator Variances

Easy to see that

$$\mathsf{Var}(\hat{\theta}_{2n}) = \mathsf{Var}\left(\frac{\sum_{i=1}^{2n} Y_i}{2n}\right) = \frac{\mathsf{Var}(Y)}{2n}$$

and

$$\begin{aligned}
\mathsf{Var}(\hat{\theta}_{n,a}) &= \mathsf{Var}\left(\frac{\sum_{i=1}^{n} Z_i}{n}\right) = \frac{\mathsf{Var}(Z)}{n} \\
&= \frac{\mathsf{Var}(Y + \tilde{Y})}{4n} = \frac{\mathsf{Var}(Y)}{2n} + \frac{\mathsf{Cov}(Y, \tilde{Y})}{2n} \\
&= \mathsf{Var}(\hat{\theta}_{2n}) + \frac{\mathsf{Cov}(Y, \tilde{Y})}{2n}.
\end{aligned}$$

Therefore $\mathsf{Var}(\hat{\theta}_{n,a}) < \mathsf{Var}(\hat{\theta}_{2n})$ if and only $\mathsf{Cov}(Y, \tilde{Y}) < 0$.

Recalling that $Y = h(\mathbf{U})$ and $\tilde{Y} = h(1 - \mathbf{U})$, this means that

$$\mathsf{Var}(\hat{\theta}_{n,a}) < \mathsf{Var}(\hat{\theta}_{2n}) \Longleftrightarrow \mathsf{Cov}\left(h(\mathbf{U}), h(1 - \mathbf{U})\right) < 0.$$

When Can a Variance Reduction Be Guaranteed?

Consider first the case where $\mathbf{U}$ is a scalar uniform so $m = 1$, $\mathbf{U} = U$ and $\theta = \mathbb{E}[h(U)]$.

Suppose $h(.)$ is a **non-decreasing** function of u over $[0, 1]$.

Then if U is large, $h(U)$ will also tend to be large while $1 - U$ and $h(1 - U)$ will tend to be small. That is, $\text{Cov}(h(U), h(1 - U)) < 0$.

Can similarly conclude that if $h(.)$ is a **non-increasing** function of u then once again, $\text{Cov}(h(U), h(1 - U)) < 0$.

So for the case where $m = 1$, a sufficient condition to guarantee a variance reduction is for $h(.)$ to be a monotonic function of u on $[0, 1]$.

Consider the more general case where $m > 1$, $\mathbf{U} = (U_1, \ldots, U_m)$ and $\theta = \mathbb{E}[h(\mathbf{U})]$.

Say $h(u_1, \ldots, u_m)$ is a monotonic function of each of its m arguments if, in each of its arguments, it is non-increasing or non-decreasing.

Comparing Estimator Variances

Theorem. If $h(u_1, \ldots, u_m)$ is a monotonic function of each of its arguments on $[0,1]^m$, then for a set $\mathbf{U} := (U_1, \ldots, U_m)$ of IID $U(0,1)$ random variables

$$\text{Cov}(h(\mathbf{U}), h(1-\mathbf{U})) < 0$$

where $\text{Cov}(h(\mathbf{U}), h(1-\mathbf{U})) := \text{Cov}(h(U_1, \ldots, U_m), h(1-U_1, \ldots, 1-U_m))$.

Proof See Sheldon M. Ross's *Simulation*. □

Note that the theorem specifies **sufficient** conditions for a variance reduction, but not **necessary** conditions.

So still possible to obtain a variance reduction even if conditions of the theorem are not satisfied.

For example, if $h(.)$ is "mostly" monotonic, then a variance reduction might be still be obtained.

Non-Uniform Antithetic Variates

So far have only considered problems where $\theta = \mathbb{E}[h(\mathbf{U})]$, for $\mathbf{U}$ a vector of IID $U(0,1)$ random variables.

But often the case that $\theta = \mathbb{E}[Y]$ where $Y = h(X_1, \ldots, X_m)$, and where $(X_1, \ldots, X_m)$ is a vector of independent random variables.

Can still use antithetic variable method for such problems if we can use the **inverse transform** method to generate the X_i's.

To see this, suppose $F_i(.)$ is the CDF of X_i. If $U_i \sim U(0,1)$ then $F_i^{-1}(U_i)$ has the same distribution as X_i.

So can generate Y by generating $U_1, \ldots, U_m \sim$ IID $U(0,1)$ and setting

$$Z = h\left(F_1^{-1}(U_1), \ldots, F_m^{-1}(U_m)\right).$$

Since the CDF of any random variable is non-decreasing, it follows that $F_i^{-1}(.)$ also non-decreasing.

So if $h(x_1, \ldots, x_m)$ monotonic in each of its arguments, then $h(F_1^{-1}(U_1), \ldots, F_m^{-1}(U_m))$ also monotonic function of the U_i's.

The Barbershop Revisited

Consider again our barbershop example and suppose the barber now wants to estimate the average total waiting time, θ, of the first 100 customers.

Then $\theta = \mathbb{E}[Y]$ where $Y = \sum_{j=1}^{100} W_j$ and where W_j is the waiting time of the j^{th} customer.

For each customer, j, there is an inter-arrival time, $I_j =$ time between the $(j-1)^{th}$ and j^{th} arrivals.

There is also a service time, $S_j =$ amount of time the barber spends cutting the j^{th} customer's hair.

Therefore there is some function, $h(.)$, for which

$$Y = h(I_1, \ldots, I_{100}, S_1, \ldots, S_{100}).$$

For many queueing systems, $h(.)$ will be a monotonic function of its arguments. Why?

Antithetic variates guaranteed to give a variance reduction in these systems.

Normal Antithetic Variates

Can also generate antithetic **normal** random variates without using the inverse transform technique.

For if $X \sim \mathsf{N}(\mu, \sigma^2)$ then $\tilde{X} \sim \mathsf{N}(\mu, \sigma^2)$ also, where $\tilde{X} := 2\mu - X$.

Clearly X and $\tilde{X}$ are negatively correlated.

So if $\theta = \mathbb{E}[h(X_1, \ldots, X_m)]$ where the X_i's are IID $\mathsf{N}(\mu, \sigma^2)$ and $h(.)$ is monotonic in its arguments, then we can again achieve a variance reduction by using antithetic variates.

e.g. Normal Antithetic Variates

Suppose we want to estimate $\theta = \mathbb{E}[X^2]$ where $X \sim \mathsf{N}(2, 1)$.

Then easy to see that $\theta = 5$, but can also estimate it using antithetic variates.

Question: Is a variance reduction guaranteed? Why or why not?

Question: What would you expect if $Z \sim \mathsf{N}(10, 1)$?

Question: What would you expect if $Z \sim \mathsf{N}(0.5, 1)$?

e.g. Estimating the Price of a Knock-In Option

Wish to price a knock-in option where the payoff is given by

$$h(S_T) = \max(0, S_T - K) I_{\{S_T > B\}}$$

where B is some fixed constant and $S_t \sim GBM(r, \sigma^2)$ under Q.

Option price may be then written as

$$C_0 = \mathbb{E}_0^Q[e^{-rT} \max(0, S_T - K) I_{\{S_T > B\}}]$$

Can write $S_T = S_0 e^{(r-\sigma^2/2)T + \sigma\sqrt{T}X}$ where $X \sim \mathsf{N}(0, 1)$ so we can use antithetic variates to estimate C_0.

Question: Are we sure to get a variance reduction?

Worth emphasizing that the variance reduction that can be achieved through the use of antithetic variates is **rarely (if ever!) dramatic.**

Conditional Monte Carlo

Let $\mathbf{X}$ and $\mathbf{Z}$ be random vectors, and let $Y = h(\mathbf{X})$ be a random variable. Suppose we set

$$V = \mathbb{E}[Y|\mathbf{Z}].$$

Then V is itself a random variable that **depends** on $\mathbf{Z}$, so can write $V = g(\mathbf{Z})$ for some function, $g(\cdot)$.

Also know that

$$\mathbb{E}[V] = \mathbb{E}[\mathbb{E}[Y|\mathbf{Z}]] = \mathbb{E}[Y]$$

so if we are trying to estimate $\theta = \mathbb{E}[Y]$, could simulate V instead of Y.

To determine the better estimator we compare variances of Y and $V = \mathbb{E}[Y|\mathbf{Z}]$.

To do this, recall the **conditional variance** formula:

$$\text{Var}(Y) = \mathbb{E}[\text{Var}(Y|\mathbf{Z})] + \text{Var}(\mathbb{E}[Y|\mathbf{Z}]). \tag{5}$$

Conditional Monte Carlo

Must have (why?) $\mathbb{E}[\text{Var}(Y|\mathbf{Z})] \geq 0$ but then (5) implies

$$\text{Var}(Y) \geq \text{Var}(\mathbb{E}[Y|\mathbf{Z}]) = \text{Var}(V)$$

so we can conclude (can we?!) that V is a better estimator of θ than Y.

To see this from another perspective, suppose that to estimate θ we first have to simulate $\mathbf{Z}$ and then simulate Y given $\mathbf{Z}$.

If we can compute $\mathbb{E}[Y|\mathbf{Z}]$ exactly, then we can eliminate the additional noise that comes from simulating Y given $\mathbf{Z}$, thereby obtaining a variance reduction.

Question: Why must Y and $\mathbf{Z}$ be **dependent** for the conditional Monte Carlo method to be worthwhile?

Conditional Monte Carlo

Summarizing, we want to estimate $\theta := \mathbb{E}[h(\mathbf{X})] = \mathbb{E}[Y]$ using conditional MC.

To do so, must have another variable or vector, $\mathbf{Z}$, that satisfies:

1. $\mathbf{Z}$ can be easily simulated
2. $V := g(\mathbf{Z}) := \mathbb{E}[Y|\mathbf{Z}]$ can be computed exactly.

If these two conditions satisfied then can simulate V by first simulating $\mathbf{Z}$ and then setting $V = g(\mathbf{Z}) = \mathbb{E}[Y|\mathbf{Z}]$.

Question: It may be possible to identify the distribution of $V = g(\mathbf{Z})$. What might we do in that case?

Also possible that other variance reduction methods could be used in conjunction with conditioning.

e.g. If $g(.)$ a monotonic function of its arguments, then antithetic variates might be useful.

Conditional Monte Carlo Algorithm for Estimating θ

for $i = 1$ to n

  generate $\mathbf{Z}_i$
  compute $g(\mathbf{Z}_i) = \mathbb{E}[Y|\mathbf{Z}_i]$
  set $V_i = g(\mathbf{Z}_i)$

end for
set $\hat{\theta}_{n,cm} = \overline{V}_n = \sum_{i=1}^{n} V_i/n$
set $\hat{\sigma}^2_{n,cm} = \sum_{i=1}^{n}(V_i - \overline{V}_n)^2/(n-1)$
set approx. $100(1-\alpha)$ % CI $= \hat{\theta}_{n,cm} \pm z_{1-\alpha/2} \frac{\hat{\sigma}_{n,cm}}{\sqrt{n}}$

An Example of Conditional Monte Carlo

We wish to estimate $\theta := P(U + Z > 4)$ where $U \sim \text{Exp}(1)$ and $Z \sim \text{Exp}(1/2)$.

Let $Y := I_{\{U+Z>4\}}$ then $\theta = \mathbb{E}[Y]$ and can use conditional MC as follows.

Set $V = \mathbb{E}[Y|Z]$ so that

$$\mathbb{E}[Y|Z = z] = P(U + Z > 4|Z = z) = P(U > 4 - z) = 1 - F_u(4 - z)$$

where $F_u(.)$ is the CDF of U.

Therefore

$$1 - F_u(4 - z) = \begin{cases} e^{-(4-z)} & \text{if } 0 \leq z \leq 4, \\ 1 & \text{if } z > 4. \end{cases}$$

which implies

$$V = \mathbb{E}[Y|Z] = \begin{cases} e^{-(4-Z)} & \text{if } 0 \leq Z \leq 4, \\ 1 & \text{if } Z > 4. \end{cases}$$

An Example of Conditional Monte Carlo

Now the conditional Monte Carlo algorithm for estimating $\theta = \mathbb{E}[V]$ is:

1. Generate $Z_1, \ldots, Z_n$ all independent
2. Set $V_i = \mathbb{E}[Y|Z_i]$ for $i = 1, \ldots, n$
3. Set $\hat{\theta}_{n,cm} = \sum_{i=1}^{n} V_i/n$
4. Compute approximate CI's as usual using the V_i's.

Could also use other variance reduction methods in conjunction with conditioning.

Pricing a Barrier Option

Definition. Let $c(x, t, K, r, \sigma)$ be the Black-Scholes price of a European call option when the current stock price is x, the time to maturity is t, the strike is K, the risk-free interest rate is r and the volatility is σ.

Want to estimate price of a European option with payoff

$$h(\mathbf{X}) = \begin{cases} \max(0, S_T - K_1) & \text{if } \; S_{T/2} \leq L, \\ \max(0, S_T - K_2) & \text{otherwise.} \end{cases}$$

where $\mathbf{X} = (S_{T/2}, S_T)$.

Can write option price as

$$C_0 = \mathbb{E}_0^Q \left[e^{-rT} \left(\max(0, S_T - K_1) I_{\{S_{T/2} \leq L\}} + \max(0, S_T - K_2) I_{\{S_{T/2} > L\}} \right) \right]$$

where as usual $S_t \sim GBM(r, \sigma^2)$ under Q.

Question: How would you estimate C_0 using simulation and only **one** normal random variable per sample payoff?

Question: Could you use antithetic variates as well? Would they be guaranteed to produce a variance reduction?

An Exercise: Estimating Portfolio Credit Risk

A bank has a portfolio of $N = 100$ loans to N companies and wants to evaluate its credit risk.

Given that company n defaults, the loss for the bank is a $\mathsf{N}(\mu, \sigma^2)$ random variable X_n where $\mu = 3$, $\sigma^2 = 1$.

Defaults are dependent and described by indicators $D_1, \ldots, D_N$ and a background random variable P, such that $D_1, \ldots, D_N$ are IID Bernouilli(p) given $P = p$.

P has a Beta$(1, 19)$ distribution, i.e. P has density $(1-p)^{18}/19$, $0 < p < 1$.

How would you estimate $P(L > x)$, where $L = \sum_{n=1}^{N} D_n X_n$ is the loss, using conditional Monte Carlo, where the conditioning is on $\sum_{n=1}^{N} D_n$?