

Simulation Efficiency and an Introduction to Variance Reduction Methods

In these notes we discuss the *efficiency* of a Monte-Carlo estimator. This naturally leads to the search for more efficient estimators and towards this end we describe some simple variance reduction techniques. In particular, we describe *control variates*, *antithetic variates* and *conditional Monte-Carlo*, all of which are designed to reduce the variance of our Monte-Carlo estimators. We will defer a discussion of other variance reduction techniques such as *common random numbers*, *stratified sampling* and *importance sampling* until later.

1 Simulation Efficiency

Suppose as usual that we wish to estimate $\theta := \mathbb{E}[h(\mathbf{X})]$. Then the standard simulation algorithm is:

1. Generate $\mathbf{X}_1, \dots, \mathbf{X}_n$
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

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

where $\hat{\sigma}_n$ is the usual estimate of $\text{Var}(Y)$ based on Y_1, \dots, Y_n .

One way to measure the quality of the estimator, $\hat{\theta}_n$, is by the *half-width*, HW , of the confidence interval. For a fixed α , we have

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

We would like HW to be small, but sometimes this is difficult to achieve. This may be because $\text{Var}(Y)$ is too large, or too much computational effort is required to simulate each Y_j so that n is necessarily small, or some combination of the two. As a result, it is often imperative to address the issue of *simulation efficiency*. There are a number of things we can do:

1. Develop a good simulation algorithm.
2. Program carefully to minimize storage requirements. For example we do not need to store all the Y_j 's: we only need to keep track of $\sum Y_j$ and $\sum Y_j^2$ to compute $\hat{\theta}_n$ and approximate CI's.
3. Program carefully to minimize execution time.
4. Decrease the variability of the simulation output that we use to estimate θ . The techniques used to do this are usually called *variance reduction* techniques.

We will now study some of the simplest variance reduction techniques, and assume that we are doing items (1) to (3) as well as possible. Before proceeding to study these techniques, however, we should first describe a *measure of simulation efficiency*.

1.1 Measuring Simulation Efficiency

Suppose there are two random variables, W and Y , such that $\mathbb{E}[W] = \mathbb{E}[Y] = \theta$. Then we could choose to either simulate W_1, \dots, W_n or Y_1, \dots, Y_n in order to estimate θ . Let M_w denote the method of estimating θ by simulating the W_i 's. M_y is similarly defined. 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 we know that

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

Let E_w and E_y denote the amount of computational *effort* required to produce one sample of W and Y , respectively. Then the total effort expended by M_w and M_y , respectively, to achieve a half width HW are

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

We then say that M_w is *more efficient* than M_y if $TE_w < TE_y$. Note that $TE_w < TE_y$ if and only if

$$\text{Var}(W)E_w < \text{Var}(Y)E_y. \quad (1)$$

We will use the quantity $\text{Var}(W)E_w$ as a measure of the efficiency of the simulator, M_w . Note that (1) implies we cannot conclude that one simulation algorithm, M_w , is better than another, M_y , simply because $\text{Var}(W) < \text{Var}(Y)$; we also need to take E_w and E_y into consideration. However, it is often the case that we have two simulators available to us, M_w and M_y , where $E_w \approx E_y$ and $\text{Var}(W) \ll \text{Var}(Y)$. In such cases it is clear that using M_w provides a substantial improvement over using M_y .

2 Control Variates

Suppose you wish to determine the mean midday temperature, θ , in Grassland and that your data consists of

$$\{(T_i, R_i) : i = 1, \dots, n\}$$

where T_i and R_i are the midday temperature and daily rainfall, respectively, on some random day, D_i . Then $\theta = \mathbb{E}[T]$ is the mean midday temperature. If the D_i 's are drawn uniformly from $\{1, \dots, 365\}$, then an obvious estimator for θ is

$$\hat{\theta}_n = \frac{\sum_{i=1}^n T_i}{n}$$

and we then know that $\mathbb{E}[\hat{\theta}_n] = \theta$. Suppose, however, that we also know:

1. $\mathbb{E}[R]$, the mean daily rainfall in Grassland
2. R_i and T_i are *dependent*; in particular, it tends to rain more in the cold season

Is there any way we can exploit this information to obtain a better estimate of θ ? The answer of course, is yes.

Let $\bar{R}_n := \sum_{i=1}^n R_i/n$ and now suppose $\bar{R}_n > \mathbb{E}[R]$. Then this implies that the D_i 's over-represent the rainy season in comparison to the dry season. But since the rainy season tends to coincide with the cold season, it also means that the D_i 's over-represent the cold season in comparison to the warm season. As a result, we expect $\hat{\theta}_n < \theta$. Therefore, to improve our estimate, we should increase $\hat{\theta}_n$. Similarly, if $\bar{R}_n < \mathbb{E}[R]$, we should decrease $\hat{\theta}_n$.

In this example, rainfall is the *control variate* since it enables us to better control our estimate of θ . The principle idea behind many variance reduction techniques (including control variates) is to "use what you know"

about the system. In this example, the system is Grassland's climate, and what we know is $\mathbb{E}[R]$, the average daily rainfall. We will now study control variates more formally, and in particular, we will determine by *how much* we should increase or decrease $\hat{\theta}_n$.

2.1 The Control Variate Method

Suppose again that we wish to estimate $\theta := \mathbb{E}[Y]$ where $Y = h(\mathbf{X})$ is the output of a simulation experiment. Suppose that Z is also an output of the simulation or that we can easily output it if we wish. Finally, we assume that we know $\mathbb{E}[Z]$. Then we can construct many unbiased estimators of θ :

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

for any $c \in \mathbb{R}$. The variance of $\hat{\theta}_c$ satisfies

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

and we can choose c to minimize this quantity. Simple calculus then implies the optimal value of c is given by

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

and that the 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 it is therefore only necessary that $\text{Cov}(Y, Z) \neq 0$. The 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 $\text{Cov}(Y, Z)$. We overcome this problem by doing p *pilot* simulations and setting

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

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

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

and finally set

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

Assuming we can find a control variate, our control variate simulation algorithm is as follows. Note that the V_i 's are IID, so we can compute approximate confidence intervals as before.

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 c-hat*
/* Now do main simulation */
for i = 1 to n
    generate (Y_i, Z_i)
    set V_i = Y_i + c-hat*(Z_i - E[Z])
end for
set theta-hat_c* = V-bar_n = sum_{i=1}^n V_i/n
set sigma-hat_{n,v}^2 = sum(V_i - theta-hat_c*)^2/(n-1)
set 100(1-alpha) % CI = [theta-hat_c* - z_{1-alpha/2} * sigma-hat_{n,v}/sqrt(n), theta-hat_c* + z_{1-alpha/2} * sigma-hat_{n,v}/sqrt(n)]

```

Example 1

Suppose we wish to estimate $\theta = \mathbb{E}[e^{(U+W)^2}]$ where $U, W \sim U(0, 1)$ and IID. In our notation we then have $Y := e^{(U+W)^2}$. The usual approach is:

1. Generate U_1, \dots, U_n and W_1, \dots, W_n , all IID $\sim U(0, 1)$
2. Compute $Y_1 = e^{(U_1+W_1)^2}, \dots, Y_n = e^{(U_n+W_n)^2}$
3. Construct the estimator $\hat{\theta}_{n,y} = \sum_{j=1}^n Y_j/n$
4. Build confidence intervals $\hat{\theta}_{n,y} \pm z_{1-\alpha/2} \hat{\sigma}_{n,y}/\sqrt{n}$ where $\hat{\sigma}_{n,y}^2$ is the usual estimate of $\text{Var}(Y)$.

Now consider using the control variate technique. First we have to choose an appropriate control variate, Z . There are many possibilities including

$$\begin{aligned} Z_1 &:= U + W \\ Z_2 &:= (U + W)^2 \\ Z_3 &:= e^{U+W} \end{aligned}$$

Note that we can easily compute $\mathbb{E}[Z_i]$ for $i = 1, 2, 3$ and that it's also clear that $\text{Cov}(Y, Z_i) \neq 0$. In a simple experiment we used Z_3 , estimating \hat{c}^* on the basis of a pilot simulation with 100 samples. We reduced the variance by approximately a factor of 4. In general, a good rule of thumb is that we should not be satisfied unless we have a variance reduction on the order of a factor of 5 to 10, though often we will achieve much more. ■

Example 2 (Pricing an Asian Call Option)

Recall that the payoff of an Asian call option is given by

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

where $\mathbf{X} := \{S_{iT/m} : i = 1, \dots, m\}$, K is the strike price and T is the expiration date. The price of the option is then given by

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

where r is the risk-free interest rate and Q is the risk-neutral probability measure. We will assume as usual that $S_t \sim GBM(r, \sigma)$ under Q . The 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 $i = 1, \dots, n$, and to set

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

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

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 $\mathbb{E}[Z]$. Indeed, $\mathbb{E}[Z_2]$ is the well-studied Black-Scholes option price. We 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. Which one would lead to the greatest variance reduction? ■

Exercise 1 Referring to Example 2, 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?

Example 3 (The Barbershop)

Many application of variance reduction techniques can be found in the study of queuing, communications or inventory systems. As a simple example of a queuing system, consider the case of a barbershop where the barber opens for business every day at 9am and closes at 6pm. He is the only barber in the shop and he's considering hiring another barber to share the workload. First, however, he would like to estimate the mean total time that customers spend waiting each day.

Assume customers arrive at the barbershop according to a non-homogeneous Poisson process, $N(t)$, with intensity $\lambda(t)$, and let W_i denote the waiting time of the i^{th} customer. Then, noting that the barber closes the shop after $T = 9$ hours (but still serves any customers who have arrived before then) the quantity that he wants to estimate is $\theta := \mathbb{E}[Y]$ where

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

Assume also that the service times of customers are IID with CDF, $F(\cdot)$, and that they are also independent of the arrival process, $N(t)$. The usual simulation method for estimating θ would be to simulate n days of operation in the barbershop, thereby obtaining n samples, Y_1, \dots, Y_n , and then setting

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

However, a better estimate could be obtained by using a control variate. In particular, 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, since services times are IID and independent of the arrival process, it is easy to see that

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

which should be easily computable. Intuition suggests that Z should be positively correlated with Y and therefore it would also be a good candidate to use as a control variate. ■

2.2 Multiple Control Variates

Until now we have only considered the possibility of using one control variate but there is no reason why we should not use more than one. Towards this end, suppose again that we wish to estimate $\theta := \mathbb{E}[Y]$ where Y is the output of a simulation experiment. We also suppose that for $i = 1, \dots, m$, 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 . We can then construct unbiased estimators of θ by defining

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

where $c_i \in \mathbb{R}$. The variance of $\hat{\theta}_c$ satisfies

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

and we could easily minimize this quantity w.r.t the c_i 's. As before, however, the 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:

$$Y = a + b_1 Z_1 + \dots + b_m Z_m + \epsilon \quad (4)$$

where ϵ is an error term. Note that you must include the intercept, a , in (4). The simulation algorithm with multiple control variates is exactly analogous to the simulation algorithm with just a single control variate.

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

```

/* Do pilot simulation first */
for i = 1 to p
    generate (Y_i, Z_{1,i}, ..., Z_{m,i})
end for
compute \hat{c}_j^*, j = 1, ..., m
/* Now do main simulation */
for i = 1 to n
    generate (Y_i, Z_{1,i}, ..., Z_{m,i})
    set V_i = Y_i + \sum_{j=1}^m \hat{c}_j^* (Z_{j,i} - \mathbb{E}[Z_j])
end for
set \hat{\theta}_{c^*} = \bar{V}_n = \sum_{i=1}^n V_i / n
set \hat{\sigma}_{n,v}^2 = \sum (V_i - \hat{\theta}_{c^*})^2 / (n - 1)
set 100(1 - \alpha) % CI = [\hat{\theta}_{c^*} - z_{1-\alpha/2} \frac{\hat{\sigma}_{n,v}}{\sqrt{n}}, \hat{\theta}_{c^*} + z_{1-\alpha/2} \frac{\hat{\sigma}_{n,v}}{\sqrt{n}}]

```

3 Antithetic Variates

As usual we would like to estimate $\theta = \mathbb{E}[h(\mathbf{X})] = \mathbb{E}[Y]$, and 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 $\mathbf{U} = (U_1, \dots, U_m)$ and the U_i 's are IID $\sim U(0, 1)$. The usual Monte Carlo algorithm, assuming we use $2n$ samples, is shown below.

Usual Simulation Algorithm for Estimating θ

```

for  $i = 1$  to  $2n$ 
    generate  $\mathbf{U}_i$ 
    set  $Y_i = h(\mathbf{U}_i)$ 
end for
set  $\hat{\theta}_{2n} = \bar{Y}_{2n} = \sum_{i=1}^{2n} Y_i / 2n$ 
set  $\hat{\sigma}_{2n}^2 = \sum_{i=1}^{2n} (Y_i - \bar{Y}_{2n})^2 / (2n - 1)$ 
set approx.  $100(1 - \alpha) \% \text{ CI} = \hat{\theta}_{2n} \pm z_{1-\alpha/2} \frac{\hat{\sigma}_{2n}}{\sqrt{2n}}$ 

```

In the above algorithm, however, we could also have used the $1 - U_i$'s to generate sample Y values, since if $U_i \sim U(0, 1)$, then so too is $1 - U_i$. We can use this fact to construct another estimator of θ as follows. As before, we set $Y_i = h(\mathbf{U}_i)$, where $\mathbf{U}_i = (U_1^{(i)}, \dots, U_m^{(i)})$. We now also set $\tilde{Y}_i = h(1 - \mathbf{U}_i)$, where we use $1 - \mathbf{U}_i$ to denote $(1 - U_1^{(i)}, \dots, 1 - U_m^{(i)})$. Note that $\mathbb{E}[Y_i] = \mathbb{E}[\tilde{Y}_i] = \theta$ so that in particular, if

$$Z_i := \frac{Y_i + \tilde{Y}_i}{2},$$

then $\mathbb{E}[Z_i] = \theta$ so that Z_i is an also unbiased estimator of θ . If the U_i 's are IID, then so too are the Z_i 's and we can use them as usual to compute approximate confidence intervals for θ . We say that U_i and $1 - U_i$ are *antithetic variates* and we then have the following *antithetic variate* simulation algorithm.

Antithetic Variate Simulation Algorithm for Estimating θ

```

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

```

As usual, $\hat{\theta}_{a,n}$ is an unbiased estimator of θ and the Strong Law of Large Numbers implies $\hat{\theta}_{n,a} \rightarrow \theta$ almost surely as $n \rightarrow \infty$. Each of the two algorithms we have described above uses $2n$ samples so the question naturally arises as to which algorithm is better. Note that both algorithms require approximately the same amount of effort so that comparing the two algorithms amounts to computing which estimator has a smaller variance.

3.1 Comparing Estimator Variances

It is easy to see that

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

and

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

Therefore $\text{Var}(\hat{\theta}_{n,a}) < \text{Var}(\hat{\theta}_{2n})$ if and only $\text{Cov}(Y, \tilde{Y}) < 0$. Recalling that $Y = h(\mathbf{U})$ and $\tilde{Y} = h(1 - \mathbf{U})$, this means that

$$\text{Var}(\hat{\theta}_{n,a}) < \text{Var}(\hat{\theta}_{2n}) \iff \text{Cov}(h(\mathbf{U}), h(1 - \mathbf{U})) < 0.$$

We will soon discuss conditions that are sufficient to *guarantee* that $\text{Cov}(h(\mathbf{U}), h(1 - \mathbf{U})) < 0$, but first let's consider an example.

Example 4 (Monte Carlo Integration)

Consider the problem of estimating

$$\theta = \int_0^1 e^{x^2} dx.$$

As usual, we may then write $\theta = \mathbb{E}[e^{U^2}]$ where $U \sim U(0, 1)$. We can compare the usual raw simulation algorithm with the simulation algorithm that uses antithetic variates. Using antithetic variates in this case results in a variance reduction of approximately 75%. ■

We now discuss the circumstances under which a variance reduction can be guaranteed. Consider first the case where \mathbf{U} is a uniform random variable so that $m = 1$, $\mathbf{U} = U$ and $\theta = \mathbb{E}[h(U)]$. Suppose now that $h(\cdot)$ 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, $\text{Cov}(h(U), h(1 - U)) < 0$. We can similarly conclude that if $h(\cdot)$ 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 $h(\cdot)$ to be a *monotonic* function of u on $[0, 1]$.

Let us now consider the more general case where $m > 1$, $\mathbf{U} = (U_1, \dots, U_m)$ and $\theta = \mathbb{E}[h(\mathbf{U})]$. We say $h(u_1, \dots, u_m)$ is a monotonic function of each of its m arguments if, in each of its arguments, it is non-increasing or non-decreasing. We have the following theorem which generalizes the $m = 1$ case above.

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

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

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

Proof See Sheldon M. Ross's *Simulation*. \square

Note that the theorem specifies *sufficient* conditions for a variance reduction, but not *necessary* conditions. This means that it is still possible to obtain a variance reduction even if the conditions of the theorem are not satisfied. For example, if $h(\cdot)$ is "mostly" monotonic, then a variance reduction might be still be obtained.

3.2 Non-Uniform Antithetic Variates

So far we have only considered problems where $\theta = \mathbb{E}[h(\mathbf{U})]$, for \mathbf{U} a vector of IID $U(0, 1)$ random variables. Of course in practice, it is often the case that $\theta = \mathbb{E}[Y]$ where $Y = h(X_1, \dots, X_m)$, and where (X_1, \dots, X_m) is a vector of independent random variables. We can still use the antithetic variable method for such problems if we can use the inverse transform method to generate the X_i 's.

To see this, suppose $F_i(\cdot)$ 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 . This implies that we can generate a sample of Y by generating $U_1, \dots, U_m \sim \text{IID } U(0, 1)$ and setting

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

Since the CDF of any random variable is non-decreasing, it follows that the inverse CDFs, $F_i^{-1}(\cdot)$, are also non-decreasing. This means that if $h(x_1, \dots, x_m)$ is a monotonic function of each of its arguments, then $h(F_1^{-1}(U_1), \dots, F_m^{-1}(U_m))$ is also a monotonic function of the U_i 's. Theorem 1 therefore applies.

Antithetic variables are often very useful when studying queueing systems and we briefly describe why by revisiting Example 3.

Example 5 (The Barbershop Revisited)

Consider again our barbershop example and suppose that 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. Now for each customer, j , there is an inter-arrival time, I_j , which is the time between the $(j-1)^{\text{th}}$ and j^{th} arrivals. There is also a service time, S_j , which is the amount of time the barber spends cutting the j^{th} customer's hair. It is therefore the case that Y may be written as

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

for some function, $h(\cdot)$. Now for many queueing systems, $h(\cdot)$ will be a monotonic function of its arguments since we would typically expect Y to increase as service times increase, and decrease as inter-arrival times increase. As a result, it might be advantageous to use antithetic variates to estimate θ . (We are implicitly assuming here that the I_j 's and S_j 's can be generated using the inverse transform method.) \blacksquare

3.3 Normal Antithetic Variates

We can also generate antithetic normal random variates without using the inverse transform technique. For if $X \sim N(\mu, \sigma^2)$ then $\tilde{X} \sim N(\mu, \sigma^2)$ also, where $\tilde{X} := 2\mu - X$. Clearly X and \tilde{X} are negatively correlated. So if $\theta = \mathbb{E}[h(X_1, \dots, X_m)]$ where the X_i 's are IID $N(\mu, \sigma^2)$ and $h(\cdot)$ is monotonic in its arguments, then we can again achieve a variance reduction by using antithetic variates.

Example 6 (Normal Antithetic Variates)

Suppose we want to estimate $\theta = \mathbb{E}[X^2]$ where $X \sim N(2, 1)$. Then it is easy to see that $\theta = 5$, but we can also estimate it using antithetic variates. We have the following questions:

1. Is a variance reduction guaranteed? Why or why not?
2. What would you expect if $Z \sim N(10, 1)$?

3. What would you expect if $Z \sim N(0.5, 1)$? ■

Example 7 (Estimating the Price of a Knock-In Option)

Suppose we 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. We assume that r is the risk-free interest rate, $S_t \sim GBM(r, \sigma^2)$ under the risk-neutral measure, Q , and that S_0 is the initial stock price. Then the option price may be written as

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

so we can estimate it using simulation. Since we can write $S_T = S_0 e^{(r - \sigma^2/2)T + \sigma\sqrt{T}X}$ where $X \sim N(0, 1)$ we can use antithetic variates to estimate C_0 . Are we sure to get a variance reduction? ■

Example 8 We can also use antithetic variates to price Asian options. In this case, however, we need to be a little careful generating the stock prices if we are to achieve a variance reduction. You might want to think about this yourself. ■

It's worth emphasizing at this point that in general, the variance reduction that can be achieved through the use of antithetic variates is rarely (if ever!) dramatic. Other methods are capable of much greater variance reduction in practice.

4 Conditional Monte Carlo

We now consider the *conditional* Monte Carlo variance reduction technique. The idea here is very simple. As was the case with control variates, we use our knowledge about the system being studied to reduce the variance of our estimator. As usual, we wish to estimate $\theta = \mathbb{E}[h(\mathbf{X})]$ where $\mathbf{X} = (X_1, \dots, X_m)$. If we could compute θ analytically, then this would be equivalent to solving an m -dimensional integral. However, maybe it is possible to evaluate *part* of the integral analytically. If so, then we might be able to use simulation to estimate the other part and thereby obtain a variance reduction. In order to explore these ideas we must first review conditional expectations and variances.

Conditional Expectations and Variances

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 we may write $V = g(\mathbf{Z})$ for some function, $g(\cdot)$. We also know that

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

so if we are trying to estimate $\theta = \mathbb{E}[Y]$, one possibility would be to simulate V instead of simulating Y . In order to decide which would be a better estimator of θ , it is necessary to compare the variances of Y and $V = \mathbb{E}[Y|\mathbf{Z}]$. To do this, recall the *conditional variance* formula:

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

Now $\text{Var}(Y|\mathbf{Z})$ is also a random variable that depends on \mathbf{Z} , and since a variance is always non-negative, it must follow that $\mathbb{E}[\text{Var}(Y|\mathbf{Z})] \geq 0$. But then (5) implies

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

so we can conclude that V is a better¹ 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. Finally, we point out that in order for the conditional expectation method to be worthwhile, it must be the case that Y and \mathbf{Z} are *dependent*.

Exercise 2 Why must Y and \mathbf{Z} be dependent for the conditional Monte Carlo method to be worthwhile?

4.1 The Conditional Monte Carlo Simulation Algorithm

Summarizing the previous discussion, we want to estimate $\theta := \mathbb{E}[h(\mathbf{X})] = \mathbb{E}[Y]$ using conditional Monte Carlo. To do so, we must have another variable or vector, \mathbf{Z} , that satisfies the following requirements:

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

This means that we can simulate a value of V by first simulating a value of \mathbf{Z} and then setting² $V = g(\mathbf{Z}) = \mathbb{E}[Y|\mathbf{Z}]$. We therefore have the following algorithm for estimating θ .

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

```

It is also possible that other variance reduction methods could be used in conjunction with conditioning. For example, it might be possible to use control variates, or if $g(\cdot)$ is a monotonic function of its arguments, then antithetic variates might be useful.

Example 9 Suppose we wish to estimate $\theta := \mathbf{P}(U + Z > 4)$ where $U \sim \text{Exp}(1)$ and $Z \sim \text{Exp}(1/2)$. If we set $Y := I_{\{U+Z>4\}}$ then we see that $\theta = \mathbb{E}[Y]$. Then the usual simulation method is:

1. Generate $U_1, \dots, U_n, Z_1, \dots, Z_n$ all independent
2. Set $Y_i = I_{\{U_i+Z_i>4\}}$ for $i = 1, \dots, n$
3. Set $\hat{\theta}_n = \sum_{i=1}^n Y_i/n$
4. Compute approximate CI's as usual .

¹This assumes that the work required to generate V is not too much greater than the work required to generate Y . See the discussion in Section 1.

²It may also be possible to identify the distribution of $V = g(\mathbf{Z})$, so that we could then simulate V directly.

However, we can also use the conditioning method, which works as follows. Set $V = \mathbb{E}[Y|Z]$. Then

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

where $F_u(\cdot)$ is the CDF of U which we know is $\text{Exp}(1)$. 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

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

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

1. Generate Z_1, \dots, Z_n all independent
2. Set $V_i = \mathbb{E}[Y|Z_i]$ for $i = 1, \dots, n$
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.

Note that we might also want to use other variance reduction methods in conjunction with conditioning when we are finding $\hat{\theta}_{n,cm}$.

Before proceed with another example, we first recall the Black-Scholes formula for the price of a European call option on a non-dividend paying stock with time t price, S_t . Let r be the risk-free interest rate and suppose $S_t \sim GBM(r, \sigma^2)$ under the risk-neutral measure, Q . Then the price, C_0 , of a call option with strike K and expiration T is given by

$$\begin{aligned} C_0 &= \mathbb{E}_0^Q[e^{-rT} \max(0, S_T - K)] \\ &= S_0 \Phi(d_1) - K e^{-rT} \Phi(d_2) \end{aligned}$$

where $\Phi(\cdot)$ is the CDF of a standard normal random variable and

$$\begin{aligned} d_1 &= \frac{\log(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} \\ d_2 &= \frac{\log(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}. \end{aligned}$$

We will also need the following definition for Example 10 below.

Definition 1 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 .

Example 10 (Pricing a Barrier Option)

Suppose we want to find the price of a European option that has payoff at expiration given by

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

where $\mathbf{X} = (S_{T/2}, S_T)$. We can write the price of the option 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, we assume that $S_t \sim GBM(r, \sigma^2)$.

Question 1: How would you estimate the price of this option using simulation and only *one* normal random variable per sample payoff?

Question 2: Could you use antithetic variates as well? Would they be guaranteed to produce a variance reduction? ■

Exercise 3 (Estimating Credit Risk; from Asmussen and Glynn's Stochastic Simulation)

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 $N(\mu, \sigma^2)$ random variable X_n where $\mu = 3$, $\sigma^2 = 1$. Defaults are dependent and described by indicators D_1, \dots, D_N and a background random variable P (measuring, say, general economic conditions), such that D_1, \dots, D_N are IID Bernoulli(p) given $P = p$, and P has a Beta(1, 19) distribution, that is, density $(1 - p)^{18}/19$, $0 < p < 1$.

Estimate $P(L > x)$, where $L = \sum_{n=1}^N D_n X_n$ is the loss, using both crude Monte Carlo and conditional Monte Carlo, where the conditioning is on $\sum_{n=1}^N D_n$. For x , take

$$x = 3 \mathbb{E}[L] = 3N \mathbb{E}[P] \mathbb{E}[X] = 3 \cdot 100 \cdot 0.05 \cdot 3 = 45.$$