4

Variance Reduction Techniques

This chapter develops methods for increasing the efficiency of Monte Carlo simulation by reducing the variance of simulation estimates. These methods draw on two broad strategies for reducing variance: taking advantage of tractable features of a model to adjust or correct simulation outputs, and reducing the variability in simulation inputs. We discuss control variates, antithetic variates, stratified sampling, Latin hypercube sampling, moment matching methods, and importance sampling, and we illustrate these methods through examples. Two themes run through this chapter:

- The greatest gains in efficiency from variance reduction techniques result from exploiting specific features of a problem, rather than from generic applications of generic methods.
- Reducing simulation error is often at odds with convenient estimation of the simulation error itself; in order to supplement a reduced-variance estimator with a valid confidence interval, we sometimes need to sacrifice some of the potential variance reduction.

The second point applies, in particular, to methods that introduce dependence across replications in the course of reducing variance.

4.1 Control Variates

4.1.1 Method and Examples

The method of control variates is among the most effective and broadly applicable techniques for improving the efficiency of Monte Carlo simulation. It exploits information about the errors in estimates of known quantities to reduce the error in an estimate of an unknown quantity.

To describe the method, we let $Y_1, \dots, Y_n$ be outputs from n replications of a simulation. For example, Y_i could be the discounted payoff of a derivative security on the ith simulated path. Suppose that the Y_i are independent and identically distributed and that our objective is to estimate $\mathsf{E}[Y_i]$. The usual estimator is the sample mean $\bar{Y} = (Y_1 + \cdots + Y_n)/n$. This estimator is unbiased and converges with probability 1 as $n \to \infty$.

Suppose, now, that on each replication we calculate another output X_i along with Y_i. Suppose that the pairs (X_i, Y_i), $i = 1, \ldots, n$, are i.i.d. and that the expectation $\mathsf{E}[X]$ of the X_i is known. (We use (X, Y) to denote a generic pair of random variables with the same distribution as each (X_i, Y_i).) Then for any fixed b we can calculate

$$Y_i(b) = Y_i - b(X_i - \mathsf{E}[X])$$

from the ith replication and then compute the sample mean

$$\bar{Y}(b) = \bar{Y} - b(\bar{X} - \mathsf{E}[X]) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - b(X_i - \mathsf{E}[X])). \tag{4.1}$$

This is a control variate estimator; the observed error $\bar{X} - \mathsf{E}[X]$ serves as a control in estimating $\mathsf{E}[Y]$.

As an estimator of $\mathsf{E}[Y]$, the control variate estimator (4.1) is unbiased because

$$\mathsf{E}[\bar{Y}(b)] = \mathsf{E}\left[\bar{Y} - b(\bar{X} - \mathsf{E}[X])\right] = \mathsf{E}[\bar{Y}] = \mathsf{E}[Y]$$

and it is consistent because, with probability 1,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i(b) &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(Y_i - b(X_i - \mathsf{E}[X])) \\ &= \mathsf{E}\left[Y - b(X - \mathsf{E}[X])\right] \\ &= \mathsf{E}[Y].\end{aligned}$$

Each $Y_i(b)$ has variance

$$\begin{aligned}\mathsf{Var}[Y_i(b)] &= \mathsf{Var}\left[Y_i - b(X_i - \mathsf{E}[X])\right] \\ &= \sigma_Y^2 - 2b\sigma_X\sigma_Y\rho_{XY} + b^2\sigma_X^2 \equiv \sigma^2(b),\end{aligned} \tag{4.2}$$

where $\sigma_X^2 = \mathsf{Var}[X]$, $\sigma_Y^2 = \mathsf{Var}[Y]$, and ρ_{XY} is the correlation between X and Y. The control variate estimator $\bar{Y}(b)$ has variance $\sigma^2(b)/n$ and the ordinary sample mean $\bar{Y}$ (which corresponds to $b = 0$) has variance σ_Y^2/n. Hence, the control variate estimator has smaller variance than the standard estimator if $b^2\sigma_X < 2b\sigma_Y\rho_{XY}$.

The optimal coefficient b^* minimizes the variance (4.2) and is given by

$$b^* = \frac{\sigma_Y}{\sigma_X}\rho_{XY} = \frac{\mathsf{Cov}[X, Y]}{\mathsf{Var}[X]}. \tag{4.3}$$

Substituting this value in (4.2) and simplifying, we find that the ratio of the variance of the optimally controlled estimator to that of the uncontrolled estimator is

$$\frac{\mathsf{Var}[\bar{Y} - b^*(\bar{X} - \mathsf{E}[X])]}{\mathsf{Var}[\bar{Y}]} = 1 - \rho_{XY}^2. \tag{4.4}$$

A few observations follow from this expression:

- With the optimal coefficient b^*, the effectiveness of a control variate, as measured by the variance reduction ratio (4.4), is determined by the strength of the correlation between the quantity of interest Y and the control X. The sign of the correlation is irrelevant because it is absorbed in b^*.
- If the computational effort per replication is roughly the same with and without a control variate, then (4.4) measures the computational speed-up resulting from the use of a control. More precisely, the number of replications of the Y_i required to achieve the same variance as n replications of the control variate estimator is $n/(1 - \rho_{XY}^2)$.
- The variance reduction factor $1/(1 - \rho_{XY}^2)$ increases very sharply as $|\rho_{XY}|$ approaches 1 and, accordingly, it drops off quickly as $|\rho_{XY}|$ decreases away from 1. For example, whereas a correlation of 0.95 produces a ten-fold speed-up, a correlation of 0.90 yields only a five-fold speed-up; at $|\rho_{XY}| = 0.70$ the speed-up drops to about a factor of two. This suggests that a rather high degree of correlation is needed for a control variate to yield substantial benefits.

These remarks and equation (4.4) apply if the optimal coefficient b^* is known. In practice, if $\mathsf{E}[Y]$ is unknown it is unlikely that σ_Y or ρ_{XY} would be known. However, we may still get most of the benefit of a control variate using an estimate of b^*. For example, replacing the population parameters in (4.3) with their sample counterparts yields the estimate

$$\hat{b}_n = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}. \tag{4.5}$$

Dividing numerator and denominator by n and applying the strong law of large numbers shows that $\hat{b}_n \to b^*$ with probability 1. This suggests using the estimator $\bar{Y}(\hat{b}_n)$, the sample mean of $Y_i(\hat{b}_n) = Y_i - \hat{b}_n(X_i - \mathsf{E}[X])$, $i = 1, \ldots, n$. Replacing b^* with $\hat{b}_n$ introduces some bias; we return to this point in Section 4.1.3.

The expression in (4.5) is the slope of the least-squares regression line through the points (X_i, Y_i), $i = 1, \ldots, n$. The link between control variates and regression is useful in the statistical analysis of control variate estimators and also permits a graphical interpretation of the method. Figure 4.1 shows a hypothetical scatter plot of simulation outputs (X_i, Y_i) and the estimated regression line for these points, which passes through the point $(\bar{X}, \bar{Y})$. In the figure, $\bar{X} < \mathsf{E}[X]$, indicating that the n replications have underestimated $\mathsf{E}[X]$. If the X_i and Y_i are positively correlated, this suggests that the simulation estimate $\bar{Y}$ likely underestimates $\mathsf{E}[Y]$. This further suggests that we should adjust the estimator upward. The regression line determines the magnitude of the adjustment; in particular, $\bar{Y}(\hat{b}_n)$ is the value fitted by the regression line at the point $\mathsf{E}[X]$.