Proceedings of the 2003 Winter Simulation Conference
S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.

CONTROL VARIATES TECHNIQUES FOR MONTE CARLO SIMULATION

Roberto Szechtman

Operations Research Department
Naval Postgraduate School
Monterey, CA 93943, U.S.A.

ABSTRACT

In this paper we present an overview of classical results about the variance reduction technique of control variates. We emphasize aspects of the theory that are of importance to the practitioner, as well as presenting relevant applications.

1 INTRODUCTION

The method of control variates is one of the most widely used variance reduction techniques. Its popularity rests on the ease of implementation, the availability of controls, and on the straight intuition of the underlying theory.

To keep the presentation simple, we frame the results for the case where the parameter to be estimated is a scalar in the setting of terminating simulations, although the theory extends to the multi-response setting and to the steady-state simulation context. The emphasis is on creating valid confidence intervals and on understanding the variance reduction achieved by the estimator.

For the remainder of Section 1 we outline the paper and discuss the relevant literature. In the second section we present the basic formulation of control variates, which includes finding the optimal control coefficient and creating an asymptotically valid confidence interval. Because the optimal control coefficient is generally unknown, we discuss the loss of variance reduction caused by its estimation and introduce the idea of loss factor.

Section 3 presents the relationship between control variates and the method of regressions. This relationship is useful to obtain an expression about the limiting variance of the control variate estimator that uses an estimate of the optimal control variates coefficient. We also show the relationship between least squares regression and control variates.

In Section 4, we discuss the method of batch means as a way to overcome the bias introduced in the control variates estimator when the optimal control variates coefficient needs to be estimated. We close that section by presenting an asymptotically valid confidence interval.

Section 5 deals with non-linear control variates. We show that these type of control schemes are no more efficient, in terms of variance reduction, than linear control variates.

In the last section we present some applications of control variates in the realm of finance. We make use of the examples to illustrate the more general problem of finding and selecting control variates.

The literature in the theory and applications of control variates is quite extensive, and we do not intend to provide an exhaustive list here. The paper by Nelson (1990) and the work of Loh (1995) contain a very complete list of relevant references. We also recommend the paper of Lavenberg and Welch (1981), which was the first to give a complete and rigorous exposition of control variates. At a more introductory level, the books by Bratley et al. (1987) and Law and Kelton (1991) provide the fundamentals of control variates.

2 PROBLEM FORMULATION AND BASIC RESULTS

We start by considering the problem of estimating by Monte Carlo simulation a scalar parameter α that can be expressed as the expectation of a random variable Y, that is $\alpha = EY$. Let Y_i be an output of the i'th iteration of the simulation, done in a way so that the replications $Y_1, Y_2, \ldots, Y_n$ obtained after n iterations are independent and identically distributed (i.i.d.) as the random variable Y. The natural point estimator of α is the average,

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

The method of control variates arises when the simulationist has available a random vector $\mathbf{C} \in \mathbb{R}^d$ (notation point: vectors are columns and $\cdot^T$ denotes transpose) with known mean $\mu_{\mathbf{c}}$ that is jointly distributed with Y as an additional simulation output. Let $\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_n$ be the sequence of i.i.d. observed outputs. The idea is to use the "error" $1/n \sum_{i=1}^{n} \mathbf{C}_i - \mu_{\mathbf{c}}$ to control $\bar{Y}_n$. Intuitively,

when Y and $\mathbf{C}$ are positively correlated, we should introduce $1/n\sum_{i=1}^{n}\mathbf{C}_i - \mu_{\mathbf{c}}$ in a manner so that we adjust $\bar{Y}_n$ upwards when $\bar{\mathbf{C}}_n < \mu_{\mathbf{c}}$ and downwards when $\bar{\mathbf{C}}_n > \mu_{\mathbf{c}}$. One manner to achieve that is via the linear transformation,

$$\bar{Y}_n(\boldsymbol{\lambda}) = \bar{Y}_n - \boldsymbol{\lambda}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{C}_i - \mu_{\mathbf{c}}\right), \tag{1}$$

where the vector $\boldsymbol{\lambda} \in \mathbb{R}^d$ is chosen to minimize $\operatorname{Var}\bar{Y}_n(\boldsymbol{\lambda})$. Assume that $E(Y^2 + \mathbf{C}^T\mathbf{C}) < \infty$, and let $\sigma_{y\mathbf{c}}$ be the d-dimensional vector whose elements are the covariances of Y with each of the d components of $\mathbf{C}$. Then, $\boldsymbol{\lambda}$ is chosen so that,

$$\boldsymbol{\lambda} = \arg\min\left\{\operatorname{Var} Y - 2\boldsymbol{\lambda}^T\sigma_{y\mathbf{c}} + \boldsymbol{\lambda}^T E\mathbf{C}\mathbf{C}^T\boldsymbol{\lambda}\right\}.$$

If we assume that $E\mathbf{C}\mathbf{C}^T$ is non-singular, then the first and second order optimality conditions of the minimization problem imply that there exists a unique optimal solution,

$$\boldsymbol{\lambda} = (E\mathbf{C}\mathbf{C}^T)^{-1}\sigma_{y\mathbf{c}}.$$

With this choice of $\boldsymbol{\lambda}$ the variance reduction achieved is

$$\frac{\operatorname{Var}(\bar{Y}_n(\boldsymbol{\lambda}))}{\operatorname{Var}(\bar{Y}_n)} = 1 - R^2_{Y\mathbf{C}}, \tag{2}$$

where $R^2_{Y\mathbf{C}} = \sigma^T_{y\mathbf{c}}(E\mathbf{C}\mathbf{C}^T)^{-1}\sigma_{y\mathbf{c}}/\operatorname{Var} Y$ is the square of the multiple correlation coefficients between Y and $\mathbf{C}$.

The following assumption will hold throughout the paper.

Assumption 1. (Functional Central Limit Theorem) Assume that the stochastic processes $Y = (Y(t) \in \mathbb{R} : t \geq 0)$ and $\mathbf{C} = (\mathbf{C}(t) \in \mathbb{R}^d : t \geq 0)$ are outputs of the simulation, and define $\mathbf{X} = (\mathbf{X}(t) = (Y(t), \mathbf{C}(t)) : t \geq 0)$. Let,

$$\bar{\mathbf{X}}_n(t) = \frac{1}{n}\int_0^{nt}\mathbf{X}(s)ds.$$

We assume that there exists a constant $\mu_{\mathbf{x}} \in \mathbb{R}^{d+1}$ and a positive-definite matrix $\Sigma \in \mathbb{R}^{(d+1)\times(d+1)}$ such that the following limit holds:

$$n^{1/2}(\bar{\mathbf{X}}_n(t) - \mu_{\mathbf{x}}t) \Rightarrow \Sigma^{1/2}B(t), \text{ for } 0 \leq t \leq 1,$$

as $n \to \infty$, where $B(\cdot)$ is a standard Brownian motion in $\mathbb{R}^{d+1}$, and the convergence is weak in the space $D[0,1]$ (a good reference on this topic is Billingsley 1999). In this paper we study the process $Y(t) = Y_{\lfloor t\rfloor}$ along with the vector valued control process $\mathbf{C}(t) = \mathbf{C}_{\lfloor t\rfloor}$.

One of the key issues of every simulation is to assess the accuracy of the final estimator via confidence intervals. The distribution of $(\bar{Y}_n(\boldsymbol{\lambda}) - \alpha)/\operatorname{Var}(\bar{Y}_n(\boldsymbol{\lambda}))^{1/2}$ is approximately Student's-t with $n-1$ degrees of freedom, and we can construct the confidence interval,

$$P(\bar{Y}_n(\boldsymbol{\lambda}) - t_{n-1}(1-\gamma/2)\operatorname{Var}(\bar{Y}_n(\boldsymbol{\lambda}))^{1/2} \leq \alpha \leq \bar{Y}_n(\boldsymbol{\lambda}) + t_{n-1}(1-\gamma/2)\operatorname{Var}(\bar{Y}_n(\boldsymbol{\lambda}))^{1/2}) \approx 1-\gamma,$$

where $t_{n-1}(1-\gamma/2)$ is the $1-\gamma/2$ quantile of the t-distribution with $n-1$ degrees of freedom for $0 < \gamma < 1$. When the random vector $(Y, \mathbf{C})$ is multivariate normal the last approximation is exact.

In general, however, the covariance structure of the random vector $(Y, \mathbf{C})$ is not fully known prior to the simulation. An efficient approach to overcome this difficulty is to estimate these moments from the already available samples $(Y_i, \mathbf{C}_i)$, $i = 1, \ldots, n$. The relevant unbiased statistics for the second moments are

$$S_{\mathbf{c}}(n) = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{C}_i - \bar{\mathbf{C}}_n)(\mathbf{C}_i - \bar{\mathbf{C}}_n)^T,$$

$$S_y(n) = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2,$$

and,

$$S_{y\mathbf{c}}(n) = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)(\mathbf{C}_i - \bar{\mathbf{C}}_n).$$

In terms of these statistics, the optimal control parameter $\boldsymbol{\lambda}_n$ is

$$\boldsymbol{\lambda}_n = S_{\mathbf{c}}(n)^{-1}S_{y\mathbf{c}}(n).$$

The modified point estimator for α now becomes

$$\bar{Y}_n(\boldsymbol{\lambda}_n) = \bar{Y}_n - \boldsymbol{\lambda}_n^T(\bar{\mathbf{C}}_n - \mu_{\mathbf{c}}).$$

One problem with this approach is that we cannot use the t-statistic to generate confidence intervals because the controlled output replicates $[Y_i - \boldsymbol{\lambda}_n^T(\mathbf{C}_i - \mu_{\mathbf{c}})]$, $i = 1, \ldots, n$, are, in general, dependent of each other. Under the additional assumption that $(Y, \mathbf{C})$ has a multivariate normal distribution in $\mathbb{R}^{d+1}$, however, it can be shown (see Lavenberg and Welch 1981 and Lavenberg et al. 1982) that $\bar{Y}_n(\boldsymbol{\lambda}_n)$ is unbiased and an expression for $\operatorname{Var}(\bar{Y}_n(\boldsymbol{\lambda}_n))$ can be obtained by standard regression techniques. It also can be shown that $(\bar{Y}_n(\boldsymbol{\lambda}_n) - \alpha)/\operatorname{Var}(\bar{Y}_n(\boldsymbol{\lambda}_n))^{1/2}$ has a t-distribution with $n-d-1$ degrees of freedom, which permits the creation of an exact $(1-\gamma)100\%$ confidence interval,

$$P(\bar{Y}_n(\boldsymbol{\lambda}_n) - t_{n-d-1}(1-\gamma/2)\operatorname{Var}(\bar{Y}_n(\boldsymbol{\lambda}_n))^{1/2} \leq \alpha \leq \bar{Y}_n(\boldsymbol{\lambda}_n) + t_{n-d-1}(1-\gamma/2)\operatorname{Var}(\bar{Y}_n(\boldsymbol{\lambda}_n))^{1/2}) = 1-\gamma.$$

The variance reduction reduction achieved in this setting is (more details in Lavenberg and Welch 1981 and Lavenberg et al. 1982)

$$\frac{\mathrm{Var}(\bar{Y}_n(\boldsymbol{\lambda}_n))}{\mathrm{Var}(\bar{Y}_n)} = \frac{n-2}{n-d-2}(1-R^2_{Y\mathbf{C}}).$$

The factor $(n-2)/(n-d-2) > 1$ determines the variance increase due to the estimation of the covariance structure of $(Y, \mathbf{C})$ (compare with equation (2)), and for this reason it is called the *loss factor*. We say that a sequence of random vectors $(\chi_n : n \geq 1)$ is $o_p(n^{-1/2})$ if $n^{1/2}\chi_n \Rightarrow 0$ as $n \to \infty$, and that $(\chi_n : n \geq 1)$ is $o(a_n)$ a.s. if $\chi_n/a_n \to 0$ a.s. as $n \to \infty$.

The following expansion of the loss factor,

$$\frac{\mathrm{Var}(\bar{Y}_n(\boldsymbol{\lambda}_n))}{\mathrm{Var}(\bar{Y}_n(\boldsymbol{\lambda}))} = 1 + \frac{d}{n} + o(n^{-1}),$$

asserts that the variance reduction loss caused by estimating $\boldsymbol{\lambda}$ converges to zero at a rate d/n. In addition, for any consistent estimator $\hat{\boldsymbol{\lambda}}_n$ of $\boldsymbol{\lambda}$, under further uniform integrability assumptions (see Loh 1995, p. 22) and Assumption 1, the following result holds:

$$\lim_{n\to\infty} \frac{\mathrm{Var}(\bar{Y}_n(\hat{\boldsymbol{\lambda}}_n))}{\mathrm{Var}(\bar{Y}_n(\boldsymbol{\lambda}))} = 1. \tag{3}$$

Equation (3) asserts that as long as the estimator of $\boldsymbol{\lambda}$ is consistent, we will obtain the $1 - R^2_{Y\mathbf{C}}$ variance reduction guaranteed by $\bar{Y}_n(\boldsymbol{\lambda})$.

The results of this section can be extended to the multiresponse setting, in which case Y and α are vectors in $\mathbb{R}^q$, $q > 1$; see Rubinstein and Marcus (1985) for details.

3 THE REGRESSION APPROACH TO CONTROL VARIATES

We mentioned in the last section that an expression for $\mathrm{Var}(\bar{Y}_n(\boldsymbol{\lambda}_n))$ can be obtained using regression techniques. In this section we make the connection between regression analysis and control variates in the bivariate normal setting (with just one control variable) more explicit.

Specifically, assume that the pairs $(Y_1, C_1), (Y_2, C_2), \ldots, (Y_n, C_n)$ are i.i.d. bivariate normal. Conditional on C_i we have $E(Y_i|C_i) = \alpha + \lambda(C_i - \mu_c)$ where $\lambda = -\mathrm{Cov}(Y, C)/\mathrm{Var}\, C$, and $\mathrm{Var}(Y_i|C_i) = \mathrm{Var}\, Y(1-\rho^2_{YC})$, where ρ_{YC} is the correlation coefficient between Y and C. Then we can express,

$$Y_i = \alpha + \lambda(C_i - \mu_c) + \epsilon_i, \text{ for } i = 1, \ldots, n,$$

where the ϵ_i's are i.i.d. normal with mean zero and variance $\mathrm{Var}\, Y(1-\rho^2_{YC})$, and independent of (Y_i, C_i) for $i = 1, \ldots, n$.

The minimum least squares problem is to find parameters $\hat{\alpha}$ and $\hat{\lambda}$ that

$$\text{minimize}_{\alpha,\lambda} \sum_{i=1}^{n} \epsilon_i^2.$$

Solving the problem above we obtain $\hat{\alpha} = \bar{Y}_n - \hat{\lambda}(\bar{C}_n - \mu_c)$ and $\hat{\lambda} = S_{yc}(n)/S_c(n)$, the least squares estimators of α and λ respectively. The variance of $\hat{\alpha}$ can be found to equal (see Lavenberg and Welch 1981)

$$\mathrm{Var}\,\hat{\alpha} = S^2(n)(n^{-1} + (n-1)^{-1}(\bar{C}_n - \mu_c)^2/S_c(n)),$$

where $S^2(n) = \frac{n-1}{n-2}(S_y(n) - S^2_{yc}(n)/S_c(n))$. This result can be extended to the multi-control case to provide an expression for $\mathrm{Var}(\bar{Y}_n(\boldsymbol{\lambda}_n))$.

Relaxing the normality assumption, getting the best linear fit of the pairs $(Y_1, C_1), (Y_2, C_2), \ldots, (Y_n, C_n)$ by minimizing $\sum_{i=1}^{n}(Y_i - \alpha - \lambda(C_i - \mu_c))^2$ in α and λ yields $\hat{\alpha} = \bar{Y}_n - \hat{\lambda}(\bar{C}_n - \mu_c)$ and $\hat{\lambda} = S_{yc}(n)/S_c(n)$ respectively. That is, the line $\hat{\alpha} - \hat{\lambda}(\bar{C}_n - \mu_c)$ passes through the points $(\bar{Y}_n, \bar{C}_n)$ and $(\hat{\alpha}, \mu_c)$, so that $\bar{Y}_n$ is adjusted to $\hat{\alpha}$.

4 BATCH MEANS

As already mentioned in Section 1, a problem with using the estimated optimal coefficient $\boldsymbol{\lambda}_n$ is that the controlled simulation outputs $(Y_1 - \boldsymbol{\lambda}_n(\mathbf{C}_1 - \mu_{\mathbf{c}}), Y_2 - \boldsymbol{\lambda}_n(\mathbf{C}_2 - \mu_{\mathbf{c}}), \ldots, Y_n - \boldsymbol{\lambda}_n(\mathbf{C}_n - \mu_{\mathbf{c}}))$ are no longer independent of each other. This precludes the computation of confidence intervals for α, and $\bar{Y}_n(\boldsymbol{\lambda}_n)$ is no longer unbiased. The same issue happens in the steady-state simulation context. The batch means method tackles this problem by splitting the output into a fixed number of m batches (with $mk = n$, the integer k being the batch size), and forming a batch means sequence with elements that are asymptotically i.i.d. normal as $n \to \infty$. As in Section 2, our main objective is to determine a valid confidence interval, and to find the variance reduction achieved by this method.

More precisely, let $\mathbf{X}_i = (Y_i, \mathbf{C}_i)$, for $i = 1, \ldots, n$. For the batches $i = 1, \ldots, m$, define the batch means by,

$$\bar{\mathbf{X}}(i, n) = \frac{1}{k} \sum_{j=(i-1)k+1}^{ik} \mathbf{X}_j,$$

with $\bar{Y}(i, n)$ (resp., $\bar{\mathbf{C}}(i, n)$) being the first (resp., second to $(d+1)$'th) component(s) of $\bar{\mathbf{X}}(i, n)$. We argue by induction

that the elements $(\bar{\mathbf{X}}(i,n) : i = 1,\ldots,m)$ are asymptotically i.i.d. normal. Consider the first batch,

$$\bar{\mathbf{X}}(1,n) = m\bar{\mathbf{X}}_n(1/m).$$

By Assumption 1,

$$m\bar{\mathbf{X}}_n(1/m) \Rightarrow \left(\frac{1}{m}\Sigma\right)^{1/2} N(\mu_{\mathbf{x}}, I),$$

where Σ is the matrix with first row $(\mathrm{Var}\, Y, \sigma_{y\mathbf{c}}^T)$ and second row $(\sigma_{y\mathbf{c}}, E\mathbf{C}\mathbf{C}^T)$, and I is the $d \times d$ identity matrix. Considering now $\bar{\mathbf{X}}((i+1),n)$ we have,

$$\bar{\mathbf{X}}((i+1),n) = m(\bar{\mathbf{X}}_n((i+1)/m) - \bar{\mathbf{X}}_n(i/m)).$$

Using Assumption 1,

$$\begin{aligned} m(\bar{\mathbf{X}}_n((i+1)/m) - \bar{\mathbf{X}}_n(i/m)) - \mu_{\mathbf{x}} \\ \Rightarrow \Sigma^{1/2}\left(B\left(\frac{i+1}{m}\right) - B\left(\frac{i}{m}\right)\right) \\ \sim \left(\frac{1}{m}\Sigma\right)^{1/2} N(0, I). \end{aligned}$$

The Brownian motion term above is, by the independent increments property of Brownian motion, independent of $\Sigma^{1/2}(B(i/m) - B((i-1)/m))$ (which is, by the induction hypothesis, the asymptotic distribution of $\bar{\mathbf{X}}(i,n)$). The asymptotic normality of the batch means ensures that we can construct confidence intervals that are asymptotically valid.

The relevant statistics in the batch means context are

$$S_{\mathbf{c}}(m,n) = \frac{1}{m-1}\sum_{i=1}^{m}(\bar{\mathbf{C}}(i,n) - \bar{\mathbf{C}}_n)(\bar{\mathbf{C}}(i,n) - \bar{\mathbf{C}}_n)^T,$$

and,

$$S_{y\mathbf{c}}(m,n) = \frac{1}{m-1}\sum_{i=1}^{m}(\bar{Y}(i,n) - \bar{Y}_n)(\bar{\mathbf{C}}(i,n) - \bar{\mathbf{C}}_n).$$

Letting $\boldsymbol{\lambda}(m,n) = S_{\mathbf{c}}(m,n)^{-1}S_{y\mathbf{c}}(m,n)$, the batch means controlled estimator in terms of $\boldsymbol{\lambda}(m,n)$ is

$$\begin{aligned} \bar{Y}_{m,n}(\boldsymbol{\lambda}(m,n)) = \frac{1}{m}\sum_{i=1}^{m}\bar{Y}(i,n) \\ - \boldsymbol{\lambda}(m,n)^T\frac{1}{m}\sum_{i=1}^{m}(\bar{\mathbf{C}}(i,n) - \mu_{\mathbf{c}}). \end{aligned}$$

We want to compare the variance reduction achieved by this estimator with that of the batch means controlled estimator in terms of $\boldsymbol{\lambda}$, which is

$$\bar{Y}_{m,n}(\boldsymbol{\lambda}) = \frac{1}{m}\sum_{i=1}^{m}\bar{Y}(i,n) - \boldsymbol{\lambda}^T\frac{1}{m}\sum_{i=1}^{m}(\bar{\mathbf{C}}(i,n) - \mu_{\mathbf{c}}).$$

Since the batch means $\bar{\mathbf{X}}(i,n)$ are asymptotically i.i.d. normal as $n \to \infty$, the number of batches m in the equation above is the analogous of the number of replications n of Equation (1). Indeed, under certain uniform integrability conditions on the sequences $(\bar{Y}_{m,n}(\boldsymbol{\lambda}(m,n)))_n$ and $(\bar{Y}_{m,n}(\boldsymbol{\lambda}))_n$, and Assumption 1, one can show (see Loh 1995, p. 37) that for $m > d+2$,

$$\frac{\mathrm{Var}\,\bar{Y}_{m,n}(\boldsymbol{\lambda}(m,n))}{\mathrm{Var}\,\bar{Y}_{m,n}(\boldsymbol{\lambda})} \to \frac{m-2}{m-d-2},$$

as $n \to \infty$. Consequently there is a tangible loss of variance reduction for the batch means estimator that uses an estimate of the optimal control coefficient $\boldsymbol{\lambda}$.

One of the advantages of the batch means approach is that it allows the creation of asymptotically exact confidence intervals. It can be argued (see Nelson 1990 and Loh 1995) that,

$$\frac{\bar{Y}_{m,n}(\boldsymbol{\lambda}(m,n)) - \alpha}{\mathrm{Var}(\bar{Y}_{m,n}(\boldsymbol{\lambda}(m,n)))^{1/2}},$$

converges in distribution to a Student's-t random variable with $m-d-1$ degrees of freedom as $n \to \infty$. So the batch means method decreases the number of degrees of freedom. A confidence interval can be generated,

$$\begin{aligned} P(\bar{Y}_{m,n}(\boldsymbol{\lambda}(m,n)) \\ - t_{m-d-1}(1-\gamma/2)\,\mathrm{Var}(\bar{Y}_{m,n}(\boldsymbol{\lambda}(m,n)))^{1/2} \le \alpha \\ \le \bar{Y}_{m,n}(\boldsymbol{\lambda}(m,n)) \\ + t_{m-d-1}(1-\gamma/2)\,\mathrm{Var}(\bar{Y}_{m,n}(\boldsymbol{\lambda}(m,n)))^{1/2}) \to 1-\gamma, \end{aligned}$$

as $n \to \infty$.

Our presentation of the method of batch means makes clear that selecting the appropriate number of batches m is an important decision for the analyst; this issue is well explained in Nelson (1990).

5 NON-LINEAR CONTROL VARIATES

In this section we consider the performance of control variates when they are related to $\bar{Y}_n$ in a non-linear way.

The results presented in this section are contained in Glynn and Whitt (1989). For example,

$$\bar{Y}_n \frac{\bar{C}_n}{\mu_c},$$

and,

$$\bar{Y}_n^{\frac{\bar{C}_n}{\mu_c}}.$$

would be two such schemes when $C \in \mathbb{R}$. More generally, we deal with a scalar function f with domain in $\mathbb{R}^{d+1}$. The function f of the last two examples is $f(y,c) = yc/\mu_c$ and $f(y,c) = y^{c/\mu_c}$ respectively, and satisfies $f(y,\mu_c) = y$. This last property ensures that $f(\bar{Y}_n, \bar{\mathbf{C}}_n) \Rightarrow \alpha$ if $(\bar{Y}_n, \bar{\mathbf{C}}_n) \Rightarrow (\alpha, \mu_\mathbf{c})$, so we only will consider such functions in the discussion that follows.

The variance reduction associated with any given function f will depend on the limiting variance of $n^{1/2}(f(\bar{Y}_n, \bar{\mathbf{C}}_n) - \alpha)$. Now, when f has continuous first partial derivatives in a neighborhood around $(\alpha, \mu_\mathbf{c})$, we can obtain via Taylor's theorem a first-order linear approximation of f around $f(\alpha, \mu_\mathbf{c})$,

$$f(\bar{Y}_n, \bar{\mathbf{C}}_n) = f(\alpha, \mu_\mathbf{c}) + (\bar{Y}_n - \alpha, \bar{\mathbf{C}}_n - \mu_\mathbf{c})^T \nabla f(\xi_n, \varepsilon_n),$$

where the random variable ξ_n and the random vector $\varepsilon_n \in \mathbb{R}^d$ lie on a segment with end-points $(\bar{Y}_n, \alpha)$ and $(\bar{\mathbf{C}}_n, \mu_\mathbf{c})$ respectively. Since $(\bar{Y}_n, \bar{\mathbf{C}}_n) \Rightarrow (\alpha, \mu_\mathbf{c})$, we also have $(\xi_n, \varepsilon_n) \Rightarrow (\alpha, \mu_\mathbf{c})$, and we can write

$$f(\bar{Y}_n, \bar{\mathbf{C}}_n) = \alpha + (\bar{Y}_n - \alpha, \bar{\mathbf{C}}_n - \mu_\mathbf{c})^T \nabla f(\alpha, \mu_\mathbf{c}) + o_p(n^{-1/2}).$$

Note that $\nabla f(\alpha, \mu_\mathbf{c}) = 1$, so that,

$$f(\bar{Y}_n, \bar{\mathbf{C}}_n) = \bar{Y}_n + (\bar{\mathbf{C}}_n - \mu_\mathbf{c})^T \nabla_\mathbf{c} f(\alpha, \mu_\mathbf{c}) + o_p(n^{-1/2}), \quad (4)$$

where $\nabla_\mathbf{c} f$ is the vector of partial derivatives of f with respect to the $\mathbf{C}$ components. Thus, the limiting distribution of $n^{1/2} f(\bar{Y}_n, \bar{\mathbf{C}}_n)$ is the same as that of the linear control $n^{1/2}(\bar{Y}_n - \boldsymbol{\lambda}^T(\bar{\mathbf{C}}_n - \mu_\mathbf{c}))$, with $\nabla_\mathbf{c} f(\alpha, \mu_\mathbf{c})$ standing in lieu of $-\boldsymbol{\lambda}$. This result implies that non-linear control variates cannot improve the variance reduction achieved by linear control variates, in the limit as $n \to \infty$. Indeed, equation (4) results in,

$$n^{1/2}(f(\bar{Y}_n, \bar{\mathbf{C}}_n) - \alpha) = n^{1/2}(\bar{Y}_n - \alpha) + n^{1/2}(\bar{\mathbf{C}}_n - \mu_\mathbf{c})^T \nabla_\mathbf{c} f(\alpha, \mu_\mathbf{c}) + o_p(1).$$

Sending $n \to \infty$, Assumption 1 and the converging-together lemma imply that,

$$n^{1/2}(f(\bar{Y}_n, \bar{\mathbf{C}}_n) - \alpha) \Rightarrow N(0, \sigma_f^2),$$

where,

$$\sigma_f^2 = \mathrm{Var}\, Y + 2\nabla_\mathbf{c} f(\alpha, \mu_\mathbf{c})^T \sigma_{y\mathbf{c}} + \nabla_\mathbf{c} f(\alpha, \mu_\mathbf{c})^T E(\mathbf{C}\mathbf{C}^T) \nabla_\mathbf{c} f(\alpha, \mu_\mathbf{c}).$$

Selecting f so that $\nabla_\mathbf{c} f(\alpha, \mu_\mathbf{c}) = -(E\mathbf{C}\mathbf{C}^T)^{-1}\sigma_{y\mathbf{c}}$ is, according to the discussion of Section 2, the variance minimizing function. This selection, in turn, implies that the optimal control variate function is linear with control coefficient given by $-(E\mathbf{C}\mathbf{C}^T)^{-1}\sigma_{y\mathbf{c}}$.

6 APPLICATIONS OF CONTROL VARIATES

In this section we present several applications of control variates in finance. Our first example uses what are called "internal" control variables, so called because the control variables are random variables, or functions of them, used as an input to the simulation model and are often easy to parameterize. One of the advantages of using internal control variables is that the additional computational cost incurred by adding them is usually small relative to the overall cost. Normal random variables are often used in finance to drive pricing models, suggesting the use of their known mean and variance as control variables.

Another example of internal control variables (borrowed from Szechtman and Glynn 2001) is provided by the computation of an Asian option via simulation under the risk-neutral measure (see Duffie 1996 for details). More specifically, assume that the price process $\xi = (\xi(s) : s \geq 0)$ of the underlying asset is geometric Brownian motion, and let k denote the strike price. Then, the price of the Asian option is given by the expectation of the random variable Y given by,

$$Y = \left(\int_0^t \xi(s) ds - k \right)^+,$$

where $(a)^+ = \max(a, 0)$ for a scalar a.

In this context, one can analytically find the expectation of the integrand in the last equation,

$$C = \int_0^t \xi(s) ds,$$

to be $EC = 2(\exp(1/2) - 1)$. Therefore C can be used as an internal control for Y.

External controls are controls that are jointly distributed with the replicates of the random variable whose expectation we wish to estimate, and that are generated in addition (often

as a separate model driven by the same random input as the main model) to the main model.

From Glasserman (2003), we show an example of external control variables. We want to price an option with expiration time T with strike price k and whose underlying asset price $S(t)$ has dynamics driven by,

$$\frac{dS(t)}{S(t)} = r\,dt + \sigma(t)\,dB(t),$$

where the volatility $\sigma(t)$ may be random or a function of $S(t)$. In order to simulate the price dynamics, we simulate S at discrete times $t_1, \ldots, t_n = T$ via the recursion,

$$\frac{S(t_i)}{S(t_{i-1})} = \exp([r - 1/2\sigma(t_{i-1})^2](t_i - t_{i-1}) + \sigma(t_{i-1})(t_i - t_{i-1})^{1/2} Z_i),$$

where the Z_i's are i.i.d. standard normal random variables and $\sigma(t_i)$ is driven by its own recursion. The idea is to run another simulation alongside with constant volatility $\hat{\sigma}$ and initial condition $\hat{S}(0) = S(0)$,

$$\frac{\hat{S}(t_i)}{\hat{S}(t_{i-1})} = \exp([r - 1/2\hat{\sigma}^2](t_i - t_{i-1}) + \hat{\sigma}(t_i - t_{i-1})^{1/2} Z_i),$$

where the Z_i's are the same (common random numbers) as in the model for S. If the price of the underlying asset follows a geometric Brownian motion, then we can use the Black-Scholes formula to find $E(\hat{S}(t_n) - k)^+$ analytically. With these assumptions, generate controlled replications,

$$(S(t_n) - k)^+ - \lambda\left((\hat{S}(t_n) - k)^+ - E(\hat{S}(t_n) - k)^+\right),$$

to form the usual control variates estimator.

The efficiency of this approach will depend, among other factors, on judiciously choosing the constant volatility $\hat{\sigma}$.

REFERENCES

Billingsley, P. 1999. *Convergence of Probability Measures.* 2nd ed. New York: Wiley.

Bratley, P, B. L. Fox and L. E. Schrage. 1987. *A Guide to Simulation*. New York: Springer-Verlag.

Duffie, D. 1996. *Dynamic Asset Pricing theory*. 2nd ed. Princeton: Princeton University Press.

Glasserman, P. 2003. *Monte Carlo Methods in Financial Engineering*. New York: Springer-Verlag.

Glynn, P. W., and W. Whitt. 1989. Indirect estimation via $L = \lambda W$. *Operations Research* 37:82-103.

Lavenberg, S. S., and P. D. Welch. 1981. A perspective on the use of control variables to increase the efficiency of Monte Carlo simulations. *Management Science* 27:322-335.

Lavenberg, S. S., T. L. Moeller and P. D. Welch. 1982. Statistical results on control variates with application to queueing network simulation. *Operations Research* 30:182-202.

Law, A. M., and W. D. Kelton. 1991. *Simulation Modeling and Analysis.* 2nd ed. New York: McGraw-Hill.

Loh, W. W. 1995. On the Method of Control Variates. *Ph. D. Thesis.* Department of Operations Research. Stanford University.

Nelson, B. L. 1990. Control variate remedies. *Operations Research* 38:974-992.

Rubinstein, R. Y., and R. Marcus. 1985. Efficiency of multivariate control variates in Monte Carlo simulation. *Operations Research* 33:661-677.

Szechtman, R., and P. W. Glynn. 2001. Constrained Monte Carlo and the method of control variates. In *Proceedings of the 2001 Winter Simulation Conference*, ed. B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, 394-400. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers.

AUTHOR BIOGRAPHY

ROBERTO SZECHTMAN is Assistant Professor in the Operations Research Department at the Naval Postgraduate School since 2003. His research interests include applied probability, simulation theory, and supply chain management.