

Control-variate estimation using estimated control means

RAGHU PASUPATHY¹, BRUCE W. SCHMEISER², MICHAEL R. TAAFFE^{1,*} and JIN WANG³

¹Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA 24061, USA
E-mail: taaffe@vt.edu

²School of Industrial Engineering, Purdue University, West Lafayette, IN 47907, USA

³Department of Mathematics and Computer Science, Valdosta State University, Valdosta, GA 31698, USA

Received May 2011 and accepted July 2011

This article studies control-variate estimation where the control mean itself is estimated. Control-variate estimation in simulation experiments can significantly increase sampling efficiency and has traditionally been restricted to cases where the control has a known mean. In a previous paper the current authors generalized the idea of control variate estimation to the case where the control mean is only approximated. The result is a biased but possibly useful estimator. For that case, a mean square error optimal estimator was provided and its properties were discussed. This article generalizes classical control variate estimation to the case of Control Variates using Estimated Means (CVEMs). CVEMs replace the control mean with an estimated value for the control mean obtained from a prior simulation experiment. Although the resulting control-variate estimator is unbiased, it does introduce additional sampling error and so its properties are not the same as those of the standard control-variate estimator. A CVEM estimator is developed that minimizes the overall estimator variance. Both biased control variates and CVEMs can be used to improve the efficiency of stochastic simulation experiments. Their main appeal is that the restriction of having to know (deterministically) the exact value of the control mean is eliminated; thus, the space of possible controls is greatly increased.

Keywords: Simulation, Monte Carlo, variance reduction

1. Introduction

In an earlier paper we introduced the idea of generalizing classical control-variate estimation by eliminating the requirement that the control mean be known deterministically (Schmeiser, Taaffe, and Wang (STW), 2001). In that paper we generalized control variates to the case where the control mean can be approximated. We called the resulting estimators *Biased Control Variates* (BCVs). In this article we again generalize classical control variates but now to the case where the control mean is estimated. To be consistent with the notation of the BCV paper we paraphrase the following two paragraphs from that paper.

In a stochastic simulation experiment we compute a point estimator $\hat{\theta}$ of a performance measure θ for a model of interest. Following the notation of STW, we refer to the model of interest as the *principal model*. When the principal model can be simplified to obtain an *approximation model* with performance measure θ^a and when θ^a is known we can often greatly increase the efficiency of the simulation experiment by constructing a control-variate estimator. The point estimator $\hat{\theta}$ and the approximation measure θ^a can

be combined via the classical control-variate estimator:

$$\hat{\theta}_O^c(\beta) \equiv \hat{\theta} - \beta(\hat{\theta}^a - \theta^a), \quad (1)$$

where $\hat{\theta}$ is the *direct-simulation* estimator with unknown mean θ , $\hat{\theta}^a$ is the *control-simulation* estimator with known mean θ^a and β is the rate of change in $\hat{\theta}_O^c(\beta)$ per unit change in $\hat{\theta}^a$. Hammersley and Handscomb (1964), Lavenberg and Welch (1981), Wilson (1984), Nelson and Schmeiser (1986), Nelson (1989, 1990), Swain and Schmeiser (1989), and Law and Kelton (2000), for example, discuss classical control variates, including estimation of the minimal-variance control weight β^* and extensions to higher dimensions. The two key requirements for choosing the approximation model are that the mean θ^a is known and that $\hat{\theta}$ and $\hat{\theta}^a$ are highly correlated.

Consider the case where θ^a is unknown. Perhaps we can closely approximate θ^a analytically or numerically, or perhaps from a prior simulation experiment we have an estimate of θ^a . In classical control-variate estimation we could not make use of these approximate/estimated values of θ^a ; thus, we could not take advantage of the efficiencies of a control-variate estimation.

In STW we generalized the ideas of control variates to the case where the mean of the control, θ^a , can be approximated. We called that generalization of control-variate

*Corresponding author

estimation BCVs and explored the statistical properties of such estimators.

In a similar spirit, we now generalize classical control variates to the case where we have available an estimate, $\hat{\theta}^a$, of the control mean. The estimate may be available from a prior simulation experiment. We call this generalization of classical control variates *Control Variates using Estimated Means* (CVEMs). Unlike BCVs, CVEMs are unbiased; however, CVEMs do introduce an additional source of sampling variability. Both BCVs and CVEMs eliminate the restriction of deterministically knowing the control mean (as is the case with classical control variates) and both BCVs and CVEMs can greatly increase the efficiency of simulation experiments.

In Section 2 we investigate statistical properties of CVEMs. Section 3 contains implementation comments, as well as the summary and conclusions.

2. Control variates using estimated control means $\tilde{\theta}^a$

Our purpose is to estimate the performance measure θ of the principal model. Our approach, as in STW, is to simulate the principal model and an approximation model to obtain a correlated pair of point estimators $(\hat{\theta}, \hat{\theta}^a)$. However, in the present context, we consider a new source for an estimation of the approximation-model mean.

Consider an independent prior auxiliary simulation experiment that provides $\tilde{\theta}^a$, an estimate of the approximation-model mean θ^a . Notice that the approximation model is simulated at two different times, providing two different and independent estimators of θ^a . One simulation experiment is an auxiliary experiment that is run prior to the direct-simulation and control-simulation experiments and provides estimator $\tilde{\theta}^a$. The control-simulation estimator, $\hat{\theta}^a$, provides an independent estimate for the same approximation-model performance measure, θ^a , but at a different time. Rather than being a deterministic approximation as in STW, $\tilde{\theta}^a$ is now a random variable with mean θ^a and variance $\text{var}[\tilde{\theta}_i^a]/k'$, where k' is the number of replications in the prior auxiliary experiment. The analysis error $\varepsilon = \tilde{\theta}^a - \theta^a$ now has mean zero and variance $\text{var}[\tilde{\theta}_i^a]/k'$, whereas in STW θ^a was a constant and thus had a non-zero mean and zero variance.

Because the control-variate experiment will use a particular realization of $\tilde{\theta}^a$ and therefore a particular ε , the results in this article are in terms of expected performance over instances from the earlier paper (STW). Unlike the previous paper, however, the practitioner now has—in the sample standard error—an objective measure of the quality of $\tilde{\theta}^a$.

By having a control-variate structure where one is allowed to use an estimated approximation-model mean, *any* model can be used as the approximation model, not just those whose means are known or can be well approximated.

The approximation model can then be chosen based solely on its potential correlation with the principal model and computational costs. We consider two computation costs, both of which are for generating $\hat{\theta}_i^a$, the point estimator from one replication of the approximation model, relative to computing cost for the direct simulation. As in STW, we denote the relative cost during the control simulation to be c . The relative cost during the prior auxiliary experiment we denote by c' .

Computing costs do not refer to computing *times*. If they did, then c and c' would be equal because both relate to simulating the approximation model. Rather, the computing costs are meant to reflect urgency or financial cost or other utility measure. For example, perhaps the auxiliary experiment can be run before the principal model is available, possibly overnight or in the background. Or perhaps a single approximation model can serve as a control for several principal models. Despite the difficulty in interpretation, we assume that reasonable values for c and c' can be specified and proceed with an analysis. The results based on these costs should be viewed as rough guidelines, in the spirit of inventory models that contain a parameter for customer goodwill.

A prior auxiliary experiment to estimate the approximation-model mean is reasonable to consider when c' is small compared to c . Inexpensive computation spent on the prior auxiliary experiment provides information about the principal model, replacing expensive computation on the principal model. For small values of c' (as would be appropriate if the prior auxiliary experiment were run on a personal computer over the weekend), the prior auxiliary experiment sample size k' should be large, resulting in a small standard error for $\tilde{\theta}^a$ and the CVEM estimator of this article behaving like the classical (unbiased) control-variate estimator. However, in practice the prior auxiliary run is finite and we wish to investigate the relationships among run lengths, costs, and statistical properties.

The following is a summary of the model type, performance measure, estimators of that performance measure from one single simulation replication, the relative cost of obtaining one simulation replication of the model, the estimator of that performance measure from a number of independent simulation replications, the expected value of that estimator from a number of independent simulation replications, and the variance of that estimator from a number of independent simulation replications.

The principal model and experiment:

- (a) performance measure $\equiv \theta$;
- (b) estimator from one simulation replication of the principal model $\equiv \hat{\theta}_i$;
- (c) scaled cost to obtain one simulation replication of the principal model $\equiv 1$;
- (d) estimator from k independent simulation replications of the principal model $\equiv \hat{\theta} = k^{-1} \sum_{i=1}^k \hat{\theta}_i$;

- (e) the expectation of the estimator from k independent simulation replications of the principal model $\equiv E[\hat{\theta}] = \theta$;
- (f) the variance of the estimator from k independent simulation replications of the principal model $\equiv \text{var}[\hat{\theta}] = k^{-1}\text{var}[\hat{\theta}_i]$.

The approximation model and control experiment:

- (a) performance measure $\equiv \theta^a$;
- (b) estimator from one simulation replication of the approximation model $\equiv \hat{\theta}_i^a$;
- (c) relative cost to obtain one simulation replication of the approximation model with respect to the cost to obtain one simulation replication of the principal model $\equiv c$;
- (d) estimator from k independent simulation replications of the approximation model $\equiv \hat{\theta}^a = k^{-1} \sum_{i=1}^k \hat{\theta}_i^a$;
- (e) the expectation of the estimator from k independent simulation replications of the approximation model $\equiv E[\hat{\theta}^a] = \theta^a$;
- (f) the variance of the estimator from k independent simulation replications of the approximation model $\equiv \text{var}[\hat{\theta}^a] = k^{-1}\text{var}[\hat{\theta}_i^a]$.

The approximation model and prior auxiliary experiment:

- (a) performance measure $\equiv \theta^a$;
- (b) estimator from one prior auxiliary simulation replication of the approximation model $\equiv \tilde{\theta}_i^a$;
- (c) relative cost to obtain one prior auxiliary simulation replication of the approximation model with respect to the cost to obtain one simulation replication of the principal model $\equiv c'$;
- (d) estimator from k' prior auxiliary independent simulation replications of the approximation model $\equiv \tilde{\theta}^a = k'^{-1} \sum_{i=1}^{k'} \tilde{\theta}_i^a$;
- (e) the expectation of the estimator from k' prior auxiliary independent simulation replications of the approximation model $\equiv E[\tilde{\theta}^a] = \theta^a$;
- (f) the variance of the estimator from k' prior auxiliary independent simulation replications of the approximation model $\equiv \text{var}[\tilde{\theta}^a] = k'^{-1}\text{var}[\tilde{\theta}_i^a]$.

The CVEM estimator expression is structurally the same as the usual control-variate estimator expression, with only the constant representing the known value of the approximation-model mean performance measure replaced by an estimate of the approximation-model mean performance measure obtained in a prior auxiliary experiment.

$$\begin{aligned} \text{BCV: } \hat{\theta}^c(\beta) &= \hat{\theta} - \beta(\hat{\theta}^a - \theta^a). \\ \text{CVEM: } \hat{\theta}^c(\beta) &= \hat{\theta} - \beta(\hat{\theta}^a - \tilde{\theta}^a). \end{aligned}$$

Recalling that $\epsilon = \tilde{\theta}^a - \theta^a$, and assuming that the direct and auxiliary simulation experiments are conducted independently, the variance and the bias of the conditional ran-

dom variable $\hat{\theta}^c(\beta)|\epsilon$ are

$$\begin{aligned} \text{var}[\hat{\theta}^c(\beta)|\epsilon] &= \frac{\text{var}[\hat{\theta}_i]}{k} + \beta^2 \frac{\text{var}[\hat{\theta}_i^a]}{k} - 2\beta \frac{\text{cov}[\hat{\theta}_i, \hat{\theta}_i^a]}{k}, \\ \text{bias}[\hat{\theta}^c(\beta)|\epsilon, \theta] &= \beta(\hat{\theta}^a - \theta^a) = \beta\epsilon. \end{aligned}$$

The Mean Squared Error (MSE) of the combined (prior auxiliary and control-variate) experiment is then

$$\begin{aligned} \text{mse}[\hat{\theta}^c(\beta), \theta] &= E_{\tilde{\theta}^a} \left[\text{var}[\hat{\theta}^c(\beta)|\epsilon] + (\text{bias}[\hat{\theta}^c(\beta)|\epsilon, \theta])^2 \right] \\ &= \frac{\text{var}[\hat{\theta}_i]}{k} + \beta^2 \frac{\text{var}[\hat{\theta}_i^a]}{k} - 2\beta \frac{\text{cov}[\hat{\theta}_i, \hat{\theta}_i^a]}{k} + E_{\tilde{\theta}^a} [\beta^2 \epsilon^2] \\ &= \frac{\text{var}[\hat{\theta}_i]}{k} + \beta^2 \frac{\text{var}[\hat{\theta}_i^a]}{k} - 2\beta \frac{\text{cov}[\hat{\theta}_i, \hat{\theta}_i^a]}{k} + \beta^2 \frac{\text{var}[\hat{\theta}_i^a]}{k'}. \end{aligned}$$

Setting the first derivative of $\text{mse}[\hat{\theta}^c(\beta), \theta]$ to zero yields the MSE-optimal control-variate weight:

$$\beta^*(k') = \frac{\text{cov}[\hat{\theta}_i, \hat{\theta}_i^a]}{\text{var}[\hat{\theta}_i^a]} \left(\frac{1}{1 + k/k'} \right), \tag{2}$$

whose corresponding optimal MSE is

$$\text{mse}[\hat{\theta}^c(\beta^*(k')), \theta] = \text{mse}[\hat{\theta}, \theta] \left[1 - \rho^2 \left(\frac{1}{1 + k/k'} \right) \right].$$

If $k' = 0$, then $\beta^*(k') = 0$ and the direct-simulation experiment is optimal. The MSE-optimal auxiliary sample size is $k' = \infty$, in which case $\beta^*(k')$ and $\text{mse}[\hat{\theta}^c(\beta^*(k')), \theta]$ correspond to the classical values.

To evaluate efficiency, we use Generalized-MSE (GMSE), the product of MSE and cost. The cost of the combined experiment is $c'k' + (1 + c)k$, which is not a function of β , and so $\beta^*(k')$ also produces the optimal GMSE.

For notational simplicity, define the experimental constant:

$$\alpha = \left[\left(\frac{\rho^2}{1 - \rho^2} \right) \left(\frac{1 + c}{c'} - 1 \right) \right]^{1/2}.$$

If $\beta^*(k')$ as given in Equation (2) is used, the GMSE-optimal auxiliary sample size k' is

$$k'^* = k \max\{0, \alpha - 1\}. \tag{3}$$

Not surprisingly, if ρ^2 is close to zero, or if c' is not much less than $1 + c$, the optimal auxiliary sample size k'^* is zero and the direct experiment is optimal. More specifically, CVEMs using the optimal parameters $\beta^*(k'^*)$ and k'^* have a smaller GMSE than the direct experiment if and only if $\alpha \geq 1$; that is,

$$\rho^2 \geq \frac{c'}{1 + c}.$$

The right-hand term $c'/(1 + c)$ is the ratio of the relative cost to obtain one auxiliary observation to the cost to obtain the real-time observations $(\hat{\theta}_i, \hat{\theta}_i^a)$.

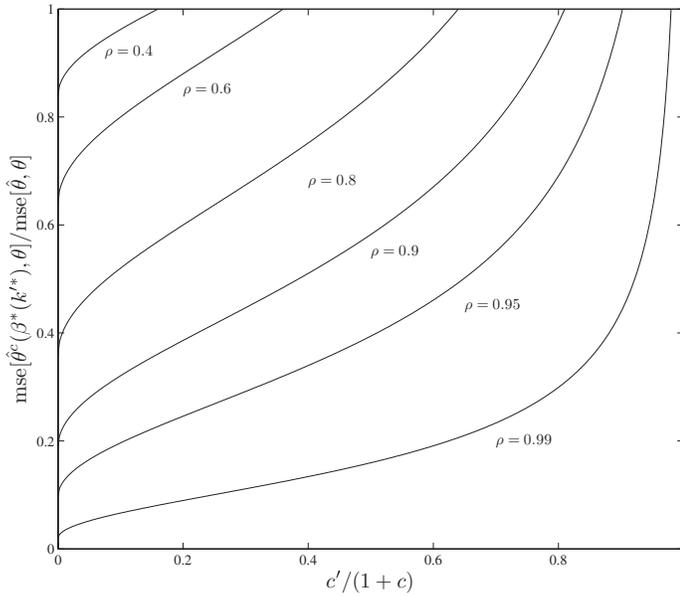


Fig. 1. Reduction in MSE due to using the GMSE-optimal sample size k^* and control-variate weight $\beta^*(k^*)$. The abscissa represents the ratio of the relative costs of obtaining the auxiliary and the real-time observations, and each curve corresponds to a value of the correlation (ρ^2) between the direct-simulation and control-simulation estimators. A reduction in MSE is guaranteed as long as the value of $c'/(1+c)$ does not exceed ρ^2 .

The MSE resulting from using the GMSE-optimal experiment parameters $\beta^*(k^*)$ and k^* is

$$\text{mse}[\widehat{\theta}^c(\beta^*(k^*)), \theta] = \text{mse}[\widehat{\theta}, \theta] \left[1 - \rho^2 \left(1 - \frac{1}{\alpha} \right) \right]. \quad (4)$$

The MSE reduction characterized by Equation (4) is depicted in Fig. 1, where the ratio $\text{mse}[\widehat{\theta}^c(\beta^*(k^*)), \theta]/\text{mse}[\widehat{\theta}, \theta]$ is plotted as a function of the ratio $c'/(1+c)$. In practice, the MSE reduction that is obtained will be smaller because the optimal experimental parameters are unknown.

Define the dimensionless ratio:

$$b \equiv \frac{\varepsilon}{\sqrt{\text{var}[\widehat{\theta}^a]}}$$

the analysis error in units of the standard error of the approximation-model point estimator. For classical control variates, $b = 0$ because $\varepsilon = 0$. For $|\varepsilon| > 0$, the ratio b is a function of either the numerator or the denominator. If the analysis error ε is held fixed, then b^2 is proportional to simulation run length, the number of macro-replications. If simulation run length is held fixed, then b^2 is simply a scaled analysis error. Also,

$$E[b^2] = \frac{\text{var}[\varepsilon]}{\text{var}[\widehat{\theta}^a]} = \frac{\text{var}[\widehat{\theta}_i^a]/k'}{\text{var}[\widehat{\theta}_i^a]/k} = \frac{k}{k'}$$

and if ε is normally distributed (which should usually be a good assumption since it is a sample mean of k'

observations):

$$\text{var}[b^2] = \frac{E[\varepsilon^4]}{\text{var}^2[\widehat{\theta}^a]} = \frac{3(\text{var}[\widehat{\theta}_i^a]/k')^2}{(\text{var}[\widehat{\theta}_i^a]/k)^2} = 3 \left(\frac{k}{k'} \right)^2,$$

implying that the coefficient of variation of b^2 is always $\sqrt{3}$. More specifically, normality of ε implies that the distribution of b^2 is a scaled chi-squared with one degree of freedom. Therefore, values of b^2 close to zero are quite likely, but the long tail will occasionally yield a control-variate experiment with much larger-than-average bias.

Occasionally a practitioner might wish to use a possibly non-optimal value of β ; for example, $\beta = 1$ is occasionally used because it is simple or because it is intuitive (e.g., when θ and θ^a have common units). The GMSE-optimal auxiliary sample size for an arbitrary value of β is

$$k^* = k \left\{ \left(\frac{c'}{1+c} \right) \left[\frac{\text{var}[\widehat{\theta}]}{\beta^2 \text{var}[\widehat{\theta}^a]} - \frac{2\rho}{\beta} \sqrt{\frac{\text{var}[\widehat{\theta}]}{\text{var}[\widehat{\theta}^a]} + 1} \right] \right\}^{-1/2},$$

which is obtained by setting the first derivative of GMSE with respect to k' equal to zero.

3. Implementation considerations and conclusions

For implementation, a practitioner needs to specify values for c, c', ρ and k . If the resulting k^* is positive, the auxiliary experiment is run with the approximation model to produce a sample mean $\widehat{\theta}^a$. If convenient, the k control-variate replications are also simulated to obtain $\{\widehat{\theta}_i^a; i = 1, \dots, k\}$ and the random-number seeds saved. (In this case, the values of c and c' would be equal and much less than one.) Then, when the principal model becomes available, the same seeds are used to obtain $\{\widehat{\theta}_i; i = 1, \dots, k\}$. Unless the practitioner has substantial confidence in the specified value of ρ , β^* should be estimated using the sample covariance and variance from the k pairs of observations $\{\widehat{\theta}_i, \widehat{\theta}_i^a; i = 1, \dots, k\}$. (The value of k should be chosen to be large enough to provide a good estimate, as discussed in Nelson (1989).)

Although values of c, c', ρ , and k are needed to design the experiment, of these, only k is controlled by the practitioner. Reasonable values of c' and c are obvious by comparing the complexity of the approximation and principal models and considering when and how the approximation runs (auxiliary and control) are to be made. The optimal value of k' increases with value of ρ , which is seldom known in advance. The increase is dramatic as ρ^2 approaches one, so the practitioner might want to investigate the sensitivity by trying more than one value of ρ while determining the auxiliary sample size k' . Having determined k' , the practitioner might well choose another value, because computation cost is often not linear in the number of replications. For example, if the auxiliary run is made overnight, then c' may be essentially zero until the practitioner returns to work in the

morning. Thus, typically the computer value of k' is simply an order-of-magnitude guideline.

References

- Hammersley, J.M. and Handscomb, D.C. (1964) *Monte Carlo Methods*, Methuen, London.
- Lavenberg, S.S. and Welch, P.D. (1981) A perspective on the use of control variates to increase the efficiency of Monte Carlo simulations. *Management Science*, **27**, 322–335.
- Law, A.W. and Kelton, W.D. (2000) *Simulation Modeling and Analysis*, third edition. McGraw-Hill, New York, NY.
- Nelson, B.L. (1989) Batch size effects on the efficiency of control variates in simulation. *European Journal of Operational Research*, **43**, 184–196.
- Nelson, B.L. (1990) Control variate remedies. *Operations Research*, **38**, 974–992.
- Nelson, B.L. and Schmeiser, B.W. (1986) Decomposition of some well-known variance reduction techniques. *Journal of Statistical Computation and Simulation*, **4**, 27–46.
- Schmeiser, B.W., Taaffe, M.R. and Wang, J. (2001) Biased control-variate estimation. *IIE Transactions*, **33**, 219–228.
- Swain, J.J. and Schmeiser, B.W. (1989) Control variates for Monte Carlo analysis of nonlinear statistical models, I: overview. *Communications in Statistics: B-Simulation and Computation*, **18**(3), 1011–1036.
- Wilson, J. (1984) Variance reduction techniques for digital simulation. *American Journal of Mathematical and Management Sciences*, **4**, 277–312.

Biographies

Raghu Pasupathy is an Associate Professor of Industrial and Systems Engineering at Virginia Tech. He received his Ph.D. from the School

of Industrial Engineering at Purdue University in 2005. His research interests lie broadly in Monte Carlo methods with a specific focus on simulation optimization, stochastic root finding, and input modeling. He is a member of INFORMS, IIE, and ASA and serves as an Associate Editor for ACM TOMACS and the *INFORMS Journal on Computing*. He is the Proceedings Chair for the 2013 Winter Simulation Conference.

Bruce W. Schmeiser is an Emeritus Professor in the School of Industrial Engineering at Purdue University. He received his Ph.D. from the School of Industrial and Systems Engineering at Georgia Tech. His primary research interests are the probabilistic and statistical aspects of digital-computer simulation. He is an active participant in the Winter Simulation Conference, including serving as Program Chair in 1983 and chairing the Board of Directors during 1988–1990. He is also active in the American Statistical Association, IIE, INFORMS, and Omega Rho, the Operations Research Honor Society.

Michael R. Taaffe is an Associate Professor at Industrial and Systems Engineering at Virginia Tech. He was previously on the faculties of the University of Minnesota and Purdue University. He received his Ph.D. from the Department of Industrial and Systems Engineering at The Ohio State University. His primary research interests are computational aspects of applied probability modeling and analysis, and combined use of simulation methods and computational probability methods. He has served as President of the INFORMS College on Simulation (1998–2000) and Director of the Operations Research Division of IIE (1997–1998).

Jin Wang is a Professor of Mathematics and Computer Science at Valdosta State University. He received his Ph.D. from the School of Industrial Engineering at Purdue University. His research interests include stochastic modeling and optimization, Monte Carlo methods, supply chain management, and applied probability and statistics.

Copyright of IIE Transactions is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.