Proceedings of the 1988 Winter Simulation Conference
M. Abrams, P. Haigh, and J. Comfort (eds.)

Multivariate estimation and variance reduction in terminating and steady-state simulation

Wei-Ning Yang
Department of Industrial and Systems Engineering
The Ohio State University
Columbus, Ohio 43210

Barry L. Nelson
Department of Industrial and Systems Engineering
The Ohio State University
Columbus, Ohio 43210

ABSTRACT

Research on the analysis of steady-state simulation experiments has concentrated on mitigating the effects of initial-condition bias and estimating the variance of the simulation point estimator, usually a sample mean. There has been little research on improving the precision of point estimators through variance reduction, especially in multivariate estimation problems. In fact, multivariate estimation procedures are rarely used in simulation output analysis.

We consider applying the non-overlaping batch means output analysis method in conjunction with the control-variates variance reduction technique to estimate a multivariate mean vector. The effect of the number of batches and the number of control variates on the multivariate point and region estimators and the univariate point and interval estimators are considered. Our results have implications for terminating simulations as well.

1. INTRODUCTION

Computer simulation is commonly employed for the analysis of stochastic systems. There are many situations in which we are interested in several performance measures of a stochastic system simultaneously, possibly of several different systems. However, multivariate estimation procedures are rarely used in simulation output analysis. Moreover, when we make simultaneous inferences on each individual response, difficulties arise from the fact that these response variables are often dependent.

Although simulation is frequently the only feasible method for estimating the parameters of a complex stochastic system, the computing cost for achieving acceptable precision can be a serious disadvantage. Variance reduction techniques can be used to reduce the population variance of estimators derived from the output of simulation experiments. Recent surveys of variance reduction include Nelson (1987) and Wilson (1984).

This paper examines the effect of applying the control-variate variance reduction technique to estimate a multivariate mean vector, in conjunction with batching to improve point and region-estimator performance.

Suppose we have an output process of the form $(\mathbf{Y}_i', \mathbf{C}_i')', i = 1, 2, \cdots, n$, where $\mathbf{Y}$ is a $p \times 1$ vector, $\mathbf{C}$ is a $q \times 1$ vector, and $'$ indicates the transpose of a matrix. Suppose that the output process is identically distributed and stationary. We are interested in estimating the p-variate mean vector $\boldsymbol{\Theta} = \mathrm{E}[\mathbf{Y}_i]$, when a q-variate control vector $\mathbf{C}_i$, with known expectation, $\boldsymbol{\mu}_{\mathbf{C}}$, can also be observed.

Such an output process can arise from either terminating or steady-state simulations. In terminating simulation, $(\mathbf{Y}_i', \mathbf{C}_i')'$ could be summary outputs from the *ith* replication, which are therefore independent. In steady-state simulation, after initial-condition effects have been removed (see e.g., Schruben 1981), $(\mathbf{Y}_i', \mathbf{C}_i')', i = 1, 2, \ldots, n$, could be the output of a single replication. Of course, to obtain a discrete-time process of the form assumed here we may have to transform the natural output process, possibly by batching by time. In either case, we want to form point and region estimators for $\boldsymbol{\Theta}$ using the control-variate variance reduction technique to improve the precision of the estimators.

Standard region estimation procedures require that the output process be independent and identically normally distributed. The assumption of normality is not necessarily true for the output process obtained from terminating simulations. The assumptions of normality and independence may be violated for the output process from steady-state simulations. Batching is an aid to improve both assumptions. More precisely, batching makes the output processes from both terminating and steady-state simulations closer to normality due to central limit theorem effects, and the output process from steady-state simulations less dependent for typical covariance structures.

The penalty for the improvement from batching is loss of degrees of freedom. The approach we take in this paper is to assume the conditions of independence and normality are actually satisfied, and then to study the potential penalty for further batching in terms of the effects on point and region estimator performance. We find that estimator performance is robust for all numbers of batches beyond a certain point. This result is useful for experimental design and analysis because it limits the range within which we need to search for an acceptable number of batches.

We first review both batching and control variates, and then examine batch-size effects on the variance of the point estimator and the expected volume of the joint confidence region. We also study the tradeoffs for using individual univariate estimation versus multivariate estimation procedures, in terms of, for example, the expected half width of the confidence intervals for individual univariate responses.

2. REVIEW OF BATCHING AND CONTROL VARIATES

To review the batch means method and control-variates variance reduction technique, let the output of the simulation experiment $(\mathbf{Y}, \mathbf{C})$ be as described above (we temporarily drop the subscript i). Let $\boldsymbol{\Theta} = (\theta_1, \theta_2, \ldots, \theta_p)' = \mathrm{E}[\mathbf{Y}]$ denote the unknown mean vector of interest. The variance-covariance matrix of $(\mathbf{Y}', \mathbf{C}')'$ can be represented by

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{YY}} & \boldsymbol{\Sigma}_{\mathbf{YC}} \\ \boldsymbol{\Sigma}_{\mathbf{CY}} & \boldsymbol{\Sigma}_{\mathbf{CC}} \end{pmatrix}$$

where $\boldsymbol{\Sigma}_{\mathbf{YY}}$ is the $p \times p$ matrix of $\mathrm{Cov}[\mathbf{Y}]$, $\boldsymbol{\Sigma}_{\mathbf{YC}}$ is the $p \times q$ matrix of $\mathrm{Cov}[\mathbf{Y}, \mathbf{C}]$, $\boldsymbol{\Sigma}_{\mathbf{CY}}$ is the $q \times p$ matrix of $\mathrm{Cov}[\mathbf{C}, \mathbf{Y}]$, and $\boldsymbol{\Sigma}_{\mathbf{CC}}$ is the $q \times q$ matrix of $\mathrm{Cov}[\mathbf{C}]$.

The idea behind control variates is to identify a q-variate control vector $\mathbf{C}$ which has known expectation, $\boldsymbol{\mu}_{\mathbf{C}}$, and is strongly correlated with the p-variate response, $\mathbf{Y}$. The deviation $\mathbf{C} - \boldsymbol{\mu}_{\mathbf{C}}$ is then used to counteract the unknown deviation $\mathbf{Y} - \boldsymbol{\Theta}$ by subtracting an appropriate linear transformation of $\mathbf{C} - \boldsymbol{\mu}_{\mathbf{C}}$ from the response. For any fixed $q \times p$ matrix of control coefficients, $\boldsymbol{\Phi}$, the control estimator of $\boldsymbol{\Theta}$ is

$$\hat{\boldsymbol{\Theta}}(\boldsymbol{\Phi}) = \mathbf{Y} - \boldsymbol{\Phi}'(\mathbf{C} - \boldsymbol{\mu}_{\mathbf{C}}).$$

Letting $|\cdot|$ denote the determinant of a matrix, the generalized variance of the control estimator is

$$|\mathrm{Cov}[\hat{\boldsymbol{\Theta}}(\boldsymbol{\Phi})]| = |\boldsymbol{\Sigma}_{\mathbf{YY}} - 2\boldsymbol{\Phi}'\boldsymbol{\Sigma}_{\mathbf{CY}} + \boldsymbol{\Phi}'\boldsymbol{\Sigma}_{\mathbf{CC}}\boldsymbol{\Phi}|$$

which is minimized by the optimal matrix of control coefficients

$$\boldsymbol{\Phi}^* = \boldsymbol{\Sigma}_{\mathbf{CC}}^{-1}\boldsymbol{\Sigma}_{\mathbf{CY}}$$

(Venkatraman and Wilson 1986). The minimum generalized variance is

$$|\mathrm{Cov}[\hat{\boldsymbol{\Theta}}(\boldsymbol{\Phi}^*)]| = |\mathrm{Cov}[\mathbf{Y}]| \cdot \prod_{j=1}^{\nu}(1 - \rho_j^2)$$

where $\nu = \mathrm{rank}(\boldsymbol{\Sigma}_{\mathbf{YC}})$, and ρ_js are the canonical correlations between the response $\mathbf{Y}$ and the control vector $\mathbf{C}$.

If we are only interested in the individual univariate responses in a multivariate estimation problem, then the trace of the covariance matrix of the control estimator is an important criterion. Consider the jth expected response, $\theta_j = \mathrm{E}[Y_j]$, where Y_j denotes the jth element of $\mathbf{Y}$. As a special case of the result above, the variance of the control estimator for θ_j *alone* is minimized by the optimal vector of control coefficients

$$\phi_j^* = \boldsymbol{\Sigma}_{\mathbf{CC}}^{-1}\boldsymbol{\Sigma}_{\mathbf{C}Y_j}$$

yielding the minimum variance

$$\mathrm{Var}[\hat{\theta}_j(\phi_j^*)] = [\boldsymbol{\Sigma}_{\mathbf{YY}}]_{jj} \cdot (1 - R_j^2),$$

where $\boldsymbol{\Sigma}_{\mathbf{C}Y_j}$ is the jth column of $\boldsymbol{\Sigma}_{\mathbf{CY}}$, $[\boldsymbol{\Sigma}_{\mathbf{YY}}]_{jj}$ is the jth diagonal element of $\boldsymbol{\Sigma}_{\mathbf{YY}}$, and R_j^2 is the squared multiple correlation coefficient between $\mathbf{C}$ and Y_j.

Since ϕ_j^* is the jth column of $\boldsymbol{\Phi}^*$, the control coefficient matrix that minimizes the generalized variance of the control estimator for $\boldsymbol{\Theta}$, $\boldsymbol{\Phi}^*$ also minimizes the trace of the covariance matrix of the control estimator. The minimum trace is

$$\mathrm{tr}(\mathrm{Cov}[\hat{\boldsymbol{\Theta}}(\boldsymbol{\Phi}^*)]) = \sum_{j=1}^{p}(1 - R_j^2)[\boldsymbol{\Sigma}_{\mathbf{YY}}]_{jj},$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. This result assumes that we use the same control vector $\mathbf{C}$ to estimate each univariate response. Later we discuss the possibility of using different controls.

In practice, $\boldsymbol{\Sigma}_{\mathbf{CY}}$ is unknown, so $\boldsymbol{\Phi}^*$ must be estimated. This results in an efficiency loss relative to the minimum generalized variance and trace. This efficiency loss was quantified by Venkatraman and Wilson (1986), and is discussed in the next section.

Batching, as we use the term, means to partition the output process into k nonoverlapping batches of size $b = n/k$ and to compute the batch-mean vectors $\bar{\mathbf{Y}}_j(k)$ and $\bar{\mathbf{C}}_j(k)$, where

$$\bar{\mathbf{Y}}_j(k) = \frac{1}{b}\sum_{i=(j-1)b+1}^{jb}\mathbf{Y}_i$$

$$\bar{\mathbf{C}}_j(k) = \frac{1}{b}\sum_{i=(j-1)b+1}^{jb}\mathbf{C}_i$$

for $j = 1, 2, \cdots, k$; b is called the batch size, k the number of batches, and we assume k divides n evenly.

In the case of terminating simulations, where the output process may be nonnormal, or in steady-state simulations, where the output process may be dependent and nonnormal, it is hoped that

$$\begin{pmatrix} \bar{\mathbf{Y}}_j(k) \\ \bar{\mathbf{C}}_j(k) \end{pmatrix} \overset{i.i.d.}{\sim} \mathbf{N}_{p+q}\left(\begin{pmatrix} \boldsymbol{\Theta} \\ \boldsymbol{\mu}_{\mathbf{C}} \end{pmatrix}, \quad \boldsymbol{\Sigma}(k)\right), \; j = 1, 2, \cdots, k$$

where

$$\boldsymbol{\Sigma}(k) = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{YY}}(k) & \boldsymbol{\Sigma}_{\mathbf{YC}}(k) \\ \boldsymbol{\Sigma}_{\mathbf{CY}}(k) & \boldsymbol{\Sigma}_{\mathbf{CC}}(k) \end{pmatrix}$$

is analogous to $\boldsymbol{\Sigma}$ for the original output process. The approx-

imation of independence and normality will tend to improve as k, the number of batches, decreases (b increases).

3. POINT ESTIMATOR

Let $\{(\bar{\mathbf{Y}}_j'(k), \bar{\mathbf{C}}_j'(k))', j = 1, \ldots, k\}$ denote the batch-mean vectors as defined above. Let $\bar{\bar{\mathbf{Y}}}$ and $\bar{\bar{\mathbf{C}}}$ denote the sample mean vectors of the response and the controls, respectively,

$$\bar{\bar{\mathbf{Y}}} = \frac{1}{k}\sum_{j=1}^{k} \bar{\mathbf{Y}}_j(k) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Y}_i$$

$$\bar{\bar{\mathbf{C}}} = \frac{1}{k}\sum_{j=1}^{k} \bar{\mathbf{C}}_j(k) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{C}_i.$$

Let $\hat{\mathbf{\Sigma}}_{\mathbf{YY}}(k)$, $\hat{\mathbf{\Sigma}}_{\mathbf{CY}}(k)$, and $\hat{\mathbf{\Sigma}}_{\mathbf{CC}}(k)$ denote, respectively, the sample analogues of $\mathbf{\Sigma}_{\mathbf{YY}}(k)$, $\mathbf{\Sigma}_{\mathbf{CY}}(k)$, and $\mathbf{\Sigma}_{\mathbf{CC}}(k)$, which are computed from the batch mean vectors as follows:

$$\hat{\mathbf{\Sigma}}_{\mathbf{YY}}(k) = \frac{1}{k-1}\sum_{j=1}^{k}(\bar{\mathbf{Y}}_j(k) - \bar{\bar{\mathbf{Y}}})(\bar{\mathbf{Y}}_j(k) - \bar{\bar{\mathbf{Y}}})'$$

$$\hat{\mathbf{\Sigma}}_{\mathbf{CY}}(k) = \frac{1}{k-1}\sum_{j=1}^{k}(\bar{\mathbf{C}}_j(k) - \bar{\bar{\mathbf{C}}})(\bar{\mathbf{Y}}_j(k) - \bar{\bar{\mathbf{Y}}})'$$

$$\hat{\mathbf{\Sigma}}_{\mathbf{CC}}(k) = \frac{1}{k-1}\sum_{j=1}^{k}(\bar{\mathbf{C}}_j(k) - \bar{\bar{\mathbf{C}}})(\bar{\mathbf{C}}_j(k) - \bar{\bar{\mathbf{C}}})'.$$

Then the optimal control coefficient can be estimated by

$$\hat{\mathbf{\Phi}}^* = \hat{\mathbf{\Sigma}}_{\mathbf{CC}}^{-1}(k)\hat{\mathbf{\Sigma}}_{\mathbf{CY}}(k)$$

and a control-variate point estimator of $\mathbf{\Theta}$ is

$$\hat{\mathbf{\Theta}}(k,p,q) = \bar{\bar{\mathbf{Y}}} - \hat{\mathbf{\Phi}}^{*\prime}(\bar{\bar{\mathbf{C}}} - \boldsymbol{\mu}_{\mathbf{C}}).$$

The following theorem establishes the basic properties of this estimator:

Theorem 1 (Venkatraman and Wilson 1986) *If* $(\mathbf{Y}_i', \mathbf{C}_i')'$, $i = 1, \ldots, n$ *are i.i.d. normal, then* $\mathrm{E}[\hat{\mathbf{\Theta}}(n,p,q)] = \mathbf{\Theta}$, *and*

$$\frac{|\mathrm{Cov}[\hat{\mathbf{\Theta}}(n,p,q)]|}{|\mathrm{Cov}[\bar{\bar{\mathbf{Y}}}]|} = \left(\frac{n-2}{n-q-2}\right)^p \cdot \prod_{j=1}^{\nu}(1-\rho_j^2)$$

where $\nu = \mathrm{Rank}(\mathbf{\Sigma}_{\mathbf{YC}}(k))$ *and* ρ_j*s are canonical correlations between* $\bar{\bar{\mathbf{Y}}}$ *and* $\bar{\bar{\mathbf{C}}}$.

Under the same assumptions as Theorem 1 we can show that

$$\mathrm{tr}(\mathrm{Cov}[\hat{\mathbf{\Theta}}(n,p,q)]) = \left(\frac{n-2}{n-q-2}\right)\cdot\frac{1}{n}\sum_{j=1}^{p}(1-R_j^2)[\mathbf{\Sigma}_{\mathbf{YY}}]_{jj}.$$

These results are for the case when there is no batching ($k = n$), and the original output process is i.i.d. normal. If the independence and normality assumptions are not valid, then we may batch the output process in hopes that for some number of batches k small enough (equivalently, some batch size b large enough), the batch means are approximately i.i.d. normal. For k in this range, $\mathbf{\Sigma}(k)/k = \mathrm{Cov}[(\bar{\bar{\mathbf{Y}}}', \bar{\bar{\mathbf{C}}}')']$, which equals $\mathbf{\Sigma}/n$ in the special case that the original output process is i.i.d. The following results, which are similar to Nelson (1986) for the case $p = 1$, assume that k is in the range such that the batch means are i.i.d. normal.

Theorem 2 *For fixed* p *and* q, *with* $q + 2 < k$,

$$\frac{|\mathrm{Cov}[\hat{\mathbf{\Theta}}(k,p,q)]|}{|\mathrm{Cov}[\bar{\bar{\mathbf{Y}}}]|} = \left(\frac{k-2}{k-q-2}\right)^p \cdot \prod_{j=1}^{\nu}(1-\rho_j^2)$$

and

$$\frac{\mathrm{tr}(\mathrm{Cov}[\hat{\mathbf{\Theta}}(k,p,q)])}{\mathrm{tr}(\mathrm{Cov}[\bar{\bar{\mathbf{Y}}}])} = \left(\frac{k-2}{k-q-2}\right)\cdot\frac{\sum_{j=1}^{p}(1-R_j^2)[\mathbf{\Sigma}_{\bar{\bar{\mathbf{Y}}}\bar{\bar{\mathbf{Y}}}}]_{jj}}{\sum_{j=1}^{p}[\mathbf{\Sigma}_{\bar{\bar{\mathbf{Y}}}\bar{\bar{\mathbf{Y}}}}]_{jj}}.$$

where $\mathbf{\Sigma}_{\bar{\bar{\mathbf{Y}}}\bar{\bar{\mathbf{Y}}}} = \mathrm{Cov}[\bar{\bar{\mathbf{Y}}}]$.

Notice that ρ_j^2 is a function of p and q but not of k, since $|\mathrm{Cov}[\bar{\bar{\mathbf{Y}}}] - \mathrm{Cov}[\bar{\bar{\mathbf{Y}}}, \bar{\bar{\mathbf{C}}}]\{\mathrm{Cov}[\bar{\bar{\mathbf{C}}}]\}^{-1}\mathrm{Cov}[\bar{\bar{\mathbf{C}}}, \bar{\bar{\mathbf{Y}}}]| = |\mathrm{Cov}[\bar{\bar{\mathbf{Y}}}]|\cdot[\prod_{j=1}^{\nu}(1-\rho_j^2)]$ under our assumptions. Similiarly, R_j^2 is a function of q only.

Theorem 3 *For fixed* p, q, *and* $q + 2 < k_1 < k_2$,

$$\frac{|\mathrm{Cov}[\hat{\mathbf{\Theta}}(k_1,p,q)]|}{|\mathrm{Cov}[\hat{\mathbf{\Theta}}(k_2,p,q)]|} = \left[\frac{(k_1-2)(k_2-q-1)}{(k_2-2)(k_1-q-2)}\right]^p > 1,$$

and

$$\frac{\mathrm{tr}(\mathrm{Cov}[\hat{\mathbf{\Theta}}(k_1,p,q)])}{\mathrm{tr}(\mathrm{Cov}[\hat{\mathbf{\Theta}}(k_2,p,q)])} = \frac{(k_1-2)(k_2-q-1)}{(k_2-2)(k_1-q-1)} > 1.$$

Theorems 2 and 3 compare the control-variate point estimator to the sample mean, and to itself for different numbers of batches, but always in the range such that the batch means are i.i.d. normal. For fixed p and q, increasing k decreases the generalized variance, especially for larger p, meaning that having a larger number of batches is more important when estimating more parameters. Similarly, increasing k decreases the trace of the covariance matrix, but the number of responses has no effect on this ratio. Figure 1 shows the loss ratio $((k-2)/(k-q-2))^p$ for different values of k and p when $q = 5$ controls. The curve $p = 1$ covers both the generalized variance for a single response, and the trace for any number of responses. The number of responses has a dramatic effect on the loss ratio when k is small, but little when $k \geq 80$.

For different numbers of batches, more batches is better (the ratios in Theorem 3 are greater than 1) as would be expected. However, it is important to notice that the ratios are nearly 1 if $k_1 \geq 80$, $p \leq 5$ and $q \leq 5$, no matter how large k_2 is. This means that the improvement from a larger number of batches is negligible beyond, say, 80, unless p is quite large.

We are considering different batch sizes when the total number of outputs, n, is fixed. It is interesting to constrast the batch-size effects with the effect of additional sampling. Suppose that, under the same assumptions as Theorem 3, $\hat{\mathbf{\Theta}}(k_1,p,q)$ is formed from k_1 batches of size b_1, while $\hat{\mathbf{\Theta}}(k_2,p,q)$ is formed

from k_2 batches of the *same size* b_1. Thus, $\hat{\Theta}(k_2, p, q)$ is based on a larger total sample. The result for generalized variance in Theorem 3 is valid if the right-hand side is multiplied by

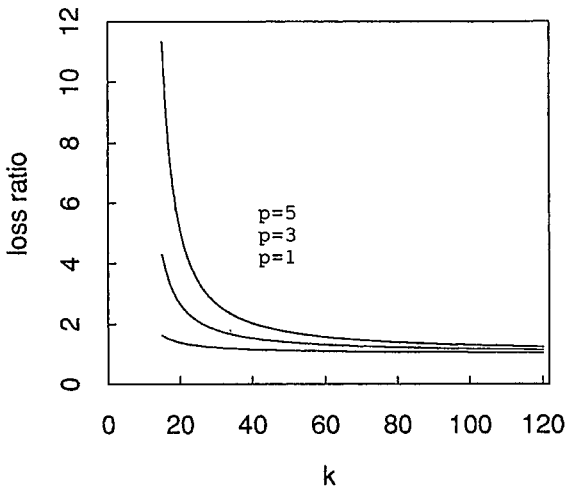

Figure 1: Loss Ratio for Control Variates with $q = 5$ Controls

$(k_2/k_1)^p$. Thus, in the case of additional sampling, the improvement from more batches is magnified by $(k_2/k_1)^p$ compared to the case where n is fixed.

Consider two different sets of control variates containing q_1 and q_2 controls. We add an argument (q) to ρ^2 and R^2 to emphasize their dependence on the particular controls variates. Then

Theorem 4 *For fixed k and p,*

$$|\mathrm{Cov}[\hat{\Theta}(k,p,q_2)]| < |\mathrm{Cov}[\hat{\Theta}(k,p,q_1)]|$$

if and only if

$$\frac{[\prod_{j=1}^{\nu_2}(1-\rho_j^2(q_2))]}{[\prod_{j=1}^{\nu_1}(1-\rho_j^2(q_1))]} < \left(\frac{k-q_2-2}{k-q_1-2}\right)^p,$$

and

$$\mathrm{tr}\left(\mathrm{Cov}\left[\hat{\Theta}(k,p,q_2)\right]\right) < \mathrm{tr}\left(\mathrm{Cov}\left[\hat{\Theta}(k,p,q_1)\right]\right)$$

if and only if

$$\frac{\sum_{i=1}^{p}(1-R_i^2(q_2))[\mathbf{\Sigma_{YY}}]_{ii}}{\sum_{i=1}^{p}(1-R_i^2(q_1))[\mathbf{\Sigma_{YY}}]_{ii}} < \frac{k-q_2-1}{k-q_1-1}.$$

A special case of Theorem 4 is adding control variates to a fixed set of q_1 controls. Since $[\prod_{j=1}^{\nu}(1-\rho_j^2(q))]$ and $\sum_{i=1}^{p}(1-R_i^2(q))$ are nonincreasing in q, Theorem 4 gives upper bounds on the decreases in $[\prod_{j=1}^{\nu}(1-\rho_j^2(q))]$ and $\sum_{i=1}^{p}(1-R_i^2(q))$ necessary to insure that adding control variates leads to a generalized variance and trace reduction, respectively. For generalized variance, this upper bound decreases as p increases, meaning that adding control variates is more likely to degrade the generalized variance when we are estimating more parameters. There is no such effect on trace. Figures 2 and 3 illustrate this point by plotting the marginal improvement ratio $((k-q_2-2)/(k-q_1-2))^p$ for various values of k, p and q.

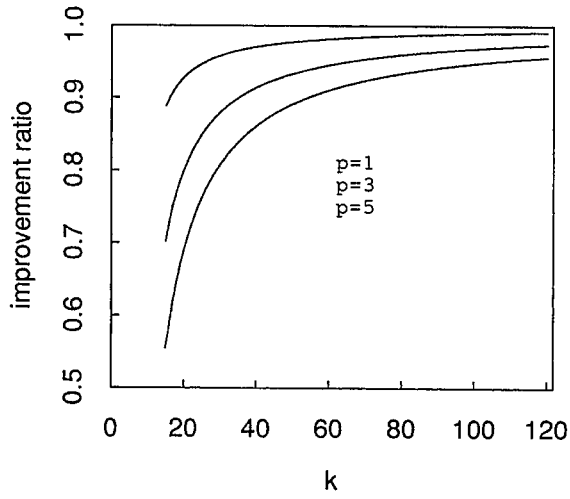

Figure 2: Marginal Improvement Ratio as a Function of k for $q_1 = 4$, $q_2 = 5$ Controls

4. REGION ESTIMATOR

A joint confidence region for Θ that incorporates control variates can be constructed using the batch mean vectors. Let

$$\begin{aligned}\mathbf{\Sigma}_{\bar{\mathbf{Y}}\cdot\bar{\mathbf{C}}} &= \mathrm{Cov}[\bar{\mathbf{Y}}] - \mathrm{Cov}[\bar{\mathbf{Y}},\bar{\mathbf{C}}]\{\mathrm{Cov}[\bar{\mathbf{C}}]\}^{-1}\mathrm{Cov}[\bar{\mathbf{C}},\bar{\mathbf{Y}}] \\ &= \mathbf{\Sigma}_{\bar{\mathbf{Y}}\bar{\mathbf{Y}}} - \mathbf{\Sigma}_{\bar{\mathbf{Y}}\bar{\mathbf{C}}}\mathbf{\Sigma}_{\bar{\mathbf{C}}\bar{\mathbf{C}}}^{-1}\mathbf{\Sigma}_{\bar{\mathbf{C}}\bar{\mathbf{Y}}}\end{aligned}$$

and

$$\begin{aligned}\mathbf{G} &= (k-1)\hat{\mathbf{\Sigma}}_{\bar{\mathbf{Y}}\cdot\bar{\mathbf{C}}} \\ &= (k-1)\{\frac{1}{k}[\hat{\mathbf{\Sigma}}_{\mathbf{YY}}(k) - \hat{\mathbf{\Sigma}}_{\mathbf{YC}}(k)\hat{\mathbf{\Sigma}}_{\mathbf{CC}}^{-1}(k)\hat{\mathbf{\Sigma}}_{\mathbf{CY}}(k)]\}.\end{aligned}$$

Then if the batch means are actually i.i.d. normal, a $(1-\alpha)100\%$ confidence region for Θ is (Wilson 1984)

$$\{\Theta : [\hat{\Theta}(k,p,q) - \Theta]'\mathbf{G}^{-1}[\hat{\Theta}(k,p,q) - \Theta] \leq \frac{p}{k-p-q}\cdot \mathrm{F}_\alpha(p, k-p-q)\cdot(1+T_q^2)\}$$

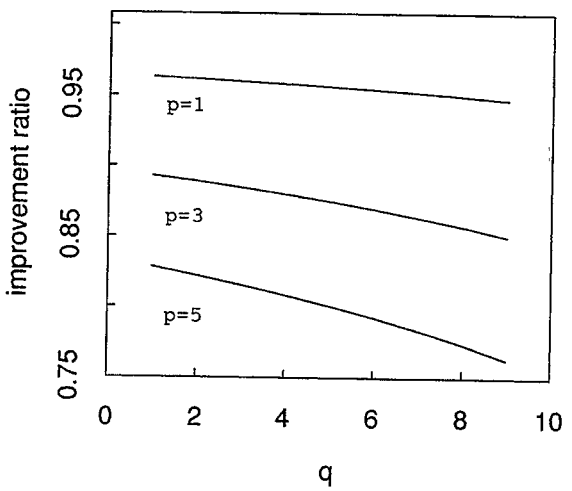

Figure 3: Marginal Improvement Ratio as a Function of q for $k = 30$ Batches

where $(k-1)T_q^2 = (\bar{\tilde{C}} - \mu_C)'(\hat{\Sigma}_{\bar{\tilde{C}}\bar{\tilde{C}}})^{-1}(\bar{\tilde{C}} - \mu_C)$, $\hat{\Sigma}_{\bar{\tilde{C}}\bar{\tilde{C}}} = k^{-1}\hat{\Sigma}_{CC}(k)$, and $\mathbf{F}_\alpha(p, k-p-q)$ is the $(1-\alpha)$ quantile of the $\mathbf{F}$ distribution with p and $k-p-q$ degrees of freedom.

4.1 Expected Volume of the Confidence Region

If $V_\alpha[\hat{\Theta}(k,p,q)]$ is the volume of the $(1-\alpha)100\%$ confidence region for Θ based on k batches and q control variates, then the expected volume is

$$\begin{aligned} \mathrm{E}\left[V_\alpha[\hat{\Theta}(k,p,q)]\right] &= D(k,p,q)\frac{\Gamma(\frac{k}{2})\Gamma(\frac{k-p-q}{2})}{\Gamma(\frac{k-q}{2})\Gamma(\frac{k-p}{2})} \\ &\times \prod_{i=1}^{p}[\sqrt{2}\frac{\Gamma(\frac{k-q-i+1}{2})}{\Gamma(\frac{k-q-i}{2})}]\cdot\sqrt{|\Sigma_{\bar{\tilde{Y}}\cdot\bar{\tilde{C}}}|} \end{aligned}$$

where

$$D(k,p,q) = \frac{2\pi^{p/2}}{p\Gamma(\frac{p}{2})}\left[\frac{p}{k-p-q}\mathbf{F}_\alpha(p,k-p-q)\right]^{p/2},$$

and $\Gamma(\cdot)$ is the Gamma function.

Suppose the number of responses, p, and the number of controls, q, are fixed but the number of batches, k, varies. The effect of number of batches, when the batch means are actually i.i.d. normal, can be summarized as follows:

1. As k increases, the expected volume of the joint confidence region decreases but at a decreasing rate, meaning that the gain from more batches decreases as the number of batches increases.
2. For larger p, decreases in the expected volume of the joint confidence region are still significant at larger values of k; in other words, having k large is more valuable when estimating more parameters.
3. For larger q, decrease in the expected volume of the joint confidence region is still significant at larger values of k, meaning that having k large is more valuable when more control variates are used.
4. With respect to the expected volume of the joint confidence region, there is little benefit from increasing the number of batches beyond $k = 80$ when $p \leq 5$ and $q \leq 5$, since the gain from more batches is insignificant.

Table 1 shows the number of batches k such that the marginal benefit in terms of reduced expected confidence region volume from 5 additional batches is just less than 5%. This is one way to define the number of batches at which increasing k further, with n fixed, has little additional benefit. Rows in the table show the effect of number of responses, while columns show the effect of number of controls.

5. CONFIDENCE INTERVALS

Even though constructing a joint confidence region for a multivariate response is important, practioners often need to make inferences on each univariate response, which leads to simultaneous inference or multiple comparisons. Bonferroni's procedure and Scheffé's projection procedure are two approaches for obtaining multiple univariate confidence intervals. Both of these procedures are conservative in the sense that the actual confidence level may be greater than what is prespecified. This section considers batch-size effects on the efficiency of these procedures when simultaneously applying control variates. In particular, expected half width of the univariate confidence interval will be given.

Table 1: Number of Batches k such that the Marginal Reduction in Expected Volume from 5 Additional Batches is Less Than 5%

	p	1	2	3	4	5
	0	14	22	29	37	44
	1	17	26	33	41	48
q	2	20	28	38	44	51
	3	22	32	40	48	55
	4	24	34	42	51	58
	5	26	36	45	54	62

5.1 Expected Half Width of Univariate Confidence Intervals

Let H_j denote the half width of the confidence interval for θ_j, the jth element of Θ, with $1-\alpha$ overall confidence level, when simultaneously applying control variates and batching. When the Bonferroni inequality is used to construct confidence intervals for individual univariate responses the expected half width of the confidence interval for each univariate response is

$$\begin{aligned}\mathrm{E}[H_j(B)] &= \frac{1}{2}\mathrm{E}\{V_{\alpha/p}[\hat{\Theta}(k,1,q)]\} \\ &= [\tfrac{2}{k-q-1}F_{\alpha/p}(1,k-q-1)]^{1/2}\tfrac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k-1}{2})}\cdot\sqrt{[\Sigma_{\bar{\mathbf{Y}}.\bar{\mathbf{C}}}]_{jj}},\end{aligned}$$

where $[\Sigma_{\bar{\mathbf{Y}}.\bar{\mathbf{C}}}]_{jj}$ is the jth diagonal element of $\Sigma_{\bar{\mathbf{Y}}.\bar{\mathbf{C}}}$, and B stands for Bonferroni.

Scheffé's projection procedure is used to construct confidence intervals for any linear combination of the mean vector and still achieve the overall confidence level. This projection procedure is very conservative when only the confidence intervals for each univariate mean is constructed.

The confidence interval for θ_j using Scheffé's projection procedure is

$$\{\theta_j : \mathbf{1}_j'[\hat{\Theta}(k,p,q)-\Theta]'\mathbf{G}^{-1}[\hat{\Theta}(k,p,q)-\Theta]\mathbf{1}_j \le \frac{p}{k-p-q}\cdot \mathrm{F}_\alpha(p,k-p-q)\cdot(1+T_q^2)\},$$

where $\mathbf{1}_j$ is a $p\times 1$ vector with 1 as the jth element and 0's elsewhere. The expected half width of the confidence interval for θ_j, with overall confidence level $1-\alpha$, using Scheffé's projection procedure, is

$$\mathrm{E}[H_j(S)] = \left[\frac{2p}{k-p-q}F_\alpha(p,k-p-q)\right]^{1/2}\frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k-1}{2})}\cdot\sqrt{[\Sigma_{\bar{\mathbf{Y}}.\bar{\mathbf{C}}}]_{jj}}.$$

5.2 Comparison of Bonferroni and Scheffé's Procedures

The ratio of the expected half width obtained by the Bonferroni procedure to that of Scheffé's projection procedure can be expressed as

$$\frac{\mathrm{E}[H_j(B)]}{\mathrm{E}[H_j(S)]} = \left[\frac{F_{\alpha/p}(1,k-q-1)}{F_\alpha(p,k-p-q)}\cdot\frac{k-p-q}{p(k-q-1)}\right]^{1/2}.$$

The results can be summarized as follows:

1. For fixed p and q, the Bonferroni procedure dominates Scheffé's projection procedure in the sense that the ratio is less than 1. The ratio increases as the number of batches, k, increases, meaning that the Bonferroni procedure is more sensitive to the number of batches.

2. For fixed k and p, the ratio decreases at an increasing rate as q increases, meaning that Scheffé's projection procedure is more conservative than the Bonferroni procedure when more control variables are applied. For larger k there is no significant decrease in the ratio as q increases.

3. For fixed k and q, the ratio decreases as p increases, meaning that Scheffé's projection procedure is more conservative than the Bonferroni procedure when estimating more parameters.

6. DISCUSSION

The results in this paper apply to both terminating and steady-state simulations when batching is used to improve the assumptions of independence and normality. Although not specifically examined here, batching can also improve the performance of the control-variate-point estimator in terms of bias, since unbiasedness depends on the normality of the output process.

The number of batches at which the departure from independence and normality is insignificant is usually unknown. Keeping the number of batches small improves the approximations of independence and normality, but if the number of batches is not too small then little is sacrificed in estimator performance due to the loss of degrees of freedom.

When the multivariate normality assumption is satisfied and we use the same q control variates for estimating each individual response, multivariate estimation procedures are more appropriate than the Bonferroni procedure since the former constructs a smaller joint confidence region and a confidence interval for any linear combination of the responses can be constructed based on Scheffé's procedure. If desired, the Bonferroni intervals can be constructed from the results of the multivariate estimation procedure.

However, using the same control variates to estimate each parameter may not be optimal if we are only interested in each parameter individually. For example, suppose $p=q=2$ and $(\mathbf{Y}',\mathbf{C}')'$ has the covariance structure below:

$$\Sigma = \begin{pmatrix}\Sigma_{\mathbf{YY}} & \Sigma_{\mathbf{YC}}\\ \Sigma_{\mathbf{CY}} & \Sigma_{\mathbf{CC}}\end{pmatrix} = \begin{pmatrix}1 & 0 & \rho & 0\\ 0 & 1 & 0 & \rho\\ \rho & 0 & 1 & 0\\ 0 & \rho & 0 & 1\end{pmatrix}.$$

Clearly, only the first control variate, if any, should be used to estimate the first response mean θ_1, since the second control and the first response are uncorrelated. However, the generalized variance of the control estimator for Θ is minimized by using both control variates if and only if

$$\left(\frac{n-2}{n-4}\right)^2(1-\rho^2)^2 < 1$$

and

$$\left(\frac{n-3}{n-4}\right)^2(1-\rho^2) < 1,$$

which is equivalent to

$$\rho^2 > 1 - \left(\frac{n-4}{n-3}\right)^2 .$$

Thus, the generalized variance and volume of the joint confidence region are minimized by using both controls, while the variance of the individual point estimators and the lengths of the joint Bonferroni intervals are minimized by using different controls for each estimator. This seems to suggest that there are benefits from selecting controls individually for each parameter. A similar argument can be made for batching each response variable individually to gain additional degrees of freedom where there is less dependence. Unfortunately, control-variate selection is a difficult problem in any case (see Bauer 1987 for the first systematic study), and multivariate batching algorithms proposed to date batch all outputs together (Añonuevo and Nelson 1988, Chen and Seila 1987). The value of our results is that they show that, beyond a certain number of batches, estimator performance is robust to the number of batches or control variates selected, and this number of batches is not very large. Thus, the benefits from individually selecting control variates or batch sizes are negligible in this range.

ACKNOWLEDGEMENT

This research was partially supported by National Science Foundation Grant No. ECS-8707634.

REFERENCES

Añonuevo, R. and Nelson B.L. (1988). Automated estimation and variance reduction via control variates for infinite-horizon simulations. *Computers & Operations Research*, in press.

Bauer, K.W. (1987). Control variate selection for multiresponse simulation. Unpublished Ph.D. dissertation. School of Industrial Engineering, Purdue University.

Chen, D.R. and Seila, A.F. (1987). Multivariate inference in stationary simulation using batch means. In: *Proceedings of the 1987 Winter Simuation Conference* (A. Thesen, H. Grant, and W.D. Kelton, eds.). Institute of Electrical and Electronics Engineers, Washington, D.C., 302-304.

Nelson, B.L. (1986). Batch size effects on the efficiency of control variates in simulation. Working Paper Series No. 1986-001. Department of Industrial and Systems Engineering, The Ohio State University.

Nelson, B.L. (1987). A perspective on variance reduction in dynamic simulation experiments. *Communications in Statistics* **B16**, 385-426.

Schruben, L. (1981). Control of initialization bias in multivariate simulation response. *Communications of the ACM* **24**, 246-252.

Venkatraman, S. and Wilson, J.R. (1986). The efficiency of control variates in multiresponse simulation. *Operations Research Letters* **5**, 37-42.

Wilson, J.R. (1984). Variance reduction techniques for digital simulation. *American Journal of Mathematical and Management Sciences* 4, 277-312.

AUTHORS' BIOGRAPHIES

WEI-NING YANG is a Ph.D. student in the Department of Industrial and Systems Engineering at The Ohio State University. He received a B.S. in Industrial Engineering from The National Tsing Hua University (Taiwan) in 1980, and an M.S. in Industrial and Systems Engineering from The Ohio State University in 1985. His research interests include multivariate variance reduction and output analysis, and ranking and selection.

Wei-Ning Yang
Department of Industrial and Systems Engineering
1971 Neil Avenue
The Ohio State University
Columbus, OH 43210
(614) 292-7820

BARRY L. NELSON is an assistant professor in the Department of Industrial and Systems Engineering at The Ohio State University. His Ph.D. is from the School of Industrial Engineering at Purdue University. His research interests are design and analysis of simulation experiments, particularly methods for statistically efficient simulation. He teaches courses in simulation and stochastic processes. Continuing memberships include ASA, IIE, ORSA, and TIMS. Dr. Nelson is an active member of the TIMS College on Simulation, and is currently its Secretary-Treasurer.

Barry L. Nelson
Department of Industrial and Systems Engineering
The Ohio State University
Columbus, OH 43210, USA
(614) 292-0610