The Optimal Linear Combination of Control Variates in the Presence of Asymptotically Negligible Bias

Peter W. Glynn and Donald L. Iglehart
Stanford University, Stanford, California 94305

The optimal linear combination of control variates is well known when the controls are assumed to be unbiased. We derive here the optimal linear combination of controls in the situation where asymptotically negligible bias is present. The small-sample linear control which minimizes the mean square error (MSE) is derived. When the optimal asymptotic linear control is used rather than the optimal small-sample control, the degradation in MSE is c/n^3, where n is the sample size and c is a known constant. This analysis is particularly relevant to the small-sample theory for control variates as applied to the steady-state estimation problem. Results for the method of multiple estimates are also given.

1. INTRODUCTION

The method of control variates has been extensively studied as a technique for obtaining variance reductions for complex simulations. The method basically requires that the practitioner be able to identify processes for which the exact or asymptotic mean is known; the knowledge of these means is then used to obtain a variance reduction.

Our goal here is to study a specific aspect of the small-sample theory for control variates. Our particular interest focuses on the loss of efficiency incurred when only the asymptotic mean is known, as opposed to the true (small-sample) mean. The results obtained here have implications for the application of control variates to the steady-state estimation problem. Specifically, in many steady-state simulations, only the asymptotic means of the control variates are known; see, for example, Section 8 of Glynn and Whitt [6] in which the arrival process to a queue is used as a control.

The results obtained here complement other small-sample studies on control variates (where it is assumed that the exact mean is known) in which the focus is on the degradation in performance caused by estimation of the optimal control coefficients; see, for example, Lavenberg, Moeller, and Welch [10], Rubinstein and Marcus [12], and Venkatraman and Wilson [13].

Our methods can also be used to study small-sample properties of the method of multiple estimates; see Section 6. Concluding remarks are stated in Section 7.

2. BACKGROUND ON CONTROL VARIATES

Suppose that one wishes to estimate a parameter r from a simulation. Assume that it is possible to generate variables (X_1,Y_1), (X_2,Y_2), . . . ,$(X_i \in \mathbb{R}^1, Y_i \in \mathbb{R}^d, d \geq 1)$ such that

Naval Research Logistics, Vol. 36, pp. 683–692 (1989)

CCC 0028-1441/89/050683-10$04.00.

$$\overline{X}_n \equiv \frac{1}{n} \sum_{i=1}^{n} X_i \Rightarrow r, \tag{1}$$

$$\overline{Y}_n \equiv \frac{1}{n} \sum_{i=1}^{n} Y_i \Rightarrow \mu \tag{2}$$

($\Rightarrow$ denotes weak convergence) as $n \to \infty$, where μ is known. Clearly, the estimator $\overline{X}_n$ is consistent for r under (1). The main application we have in mind is the steady-state simulation problem in which r is a parameter associated with the steady-state distribution of the process $\{X_n: n \geqslant 0\}$. We do not assume the sequence $\{(X_n, Y_n): n \geqslant 0\}$ is independent, identically distributed, nor that $E(Y_n) = \mu$. Also we note that similar results could be obtained for processes with a continuous, rather than discrete, time parameter.

In this section we summarize the standard approach to the asymptotic efficiency for control variates in the steady-state simulation problem. We later relate this approach to our small-sample results.

Equations (1) and (2) together imply that for $\lambda \in \mathbb{R}^d$,

$$\overline{U}_n(\lambda) \equiv \overline{X}_n - \lambda^t(\overline{Y}_n - \mu) \Rightarrow r$$

as $n \to \infty$, so $\overline{U}_n(\lambda)$ is also consistent for r. (We adopt here the convention that all elements of $\mathbb{R}^d$ are represented as column vectors; a^t denotes the transpose of $a \in \mathbb{R}^d$.) Since μ is known, $\overline{U}_n(\lambda)$ is an estimator which can legitimately be constructed from the simulated data. Furthermore, λ is at our disposal, so that λ may be chosen so as to maximize the efficiency of the estimator $U_n(\lambda)$.

To maximize the asymptotic efficiency of $U_m(\lambda)$, it is common to assume a strengthened form of (1) and (2), namely,

$$n^{1/2}(\overline{X}_n - r, \overline{Y}_n - \mu) \Rightarrow N(0,C) \tag{3}$$

as $n \to \infty$, where $N(0,C)$ is a multivariate normal distribution with mean vector zero and $(d + 1) \times (d + 1)$ covariance matrix

$$C = \begin{bmatrix} \sigma_x^2 & c_{xy}^t \\ c_{xy} & c_{yy} \end{bmatrix}$$

(c_{xy} and c_{yy} are $d \times 1$ and $d \times d$ matrices, respectively.) Given (3), the continuous mapping lemma (Billingsley [2, p. 31]) shows that

$$n^{1/2}(\overline{U}_n(\lambda) - r) \Rightarrow \sigma^2(\lambda)\, N(0,1) \tag{4}$$

as $n \to \infty$, where

$$\sigma^2(\lambda) = \sigma_x^2 - \lambda^t c_{xy} - c_{xy}^t \lambda + \lambda^t c_{yy} \lambda.$$

To optimize the asymptotic efficiency of $\overline{U}_n(\lambda)$, (4) suggests that one should choose

λ so as to minimize $\sigma^2(\lambda)$. Assuming that c_{yy} is positive definite (and hence nonsingular), the value λ^* which minimizes $\sigma^2(\lambda)$ is given by

$$\lambda^* = c_{yy}^{-1}c_{xy} \tag{5}$$

(see p. 31) of Anderson [1]; the corresponding value of $\sigma^2(\lambda)$ is then

$$\sigma^2(\lambda^*) = \sigma_x^2 - c_{xy}^t c_{yy}^{-1} c_{xy}. \tag{6}$$

3. THE OPTIMAL SOLUTION IN THE PRESENCE OF BIAS

The development of the formulas (5) and (6) discussed in Section 2 relied heavily on the asymptotic limit theory for the estimator $\overline{U}_n(\lambda)$. A somewhat different approach, which permits study of small-sample behavior, can also be taken. As we shall see, the two viewpoints coincide, under appropriate regularity conditions, in the limit.

A reasonable critrion (when small-sample bias is present) for choosing the control coefficient vector λ is to choose λ so that the mean square error (MSE) of $U_n(\lambda)$ is minimized. For this criterion to make sense, assume that

$$E(X_n^2 + Y_n^t Y_n) < \infty \text{ for } n \geq 1. \tag{7}$$

Let the biases in $\overline{X}_n$ and $\overline{Y}_n$ be denoted by

$$\begin{aligned} b_x(n) &= E(\overline{X}_n - r), \\ b_y(n) &= E(\overline{Y}_n - \mu). \end{aligned}$$

Also let the second order moments be denoted by

$$\begin{aligned} \sigma_x^2(n) &= \text{var}(\overline{X}_n), \\ c_{xy}(n) &= E(\overline{X}_n\overline{Y}_n) - E(\overline{X}_n)E(\overline{Y}_n), \\ c_{yy}(n) &= E(\overline{Y}_n\overline{Y}_n^t) - E(\overline{Y}_n)E(\overline{Y}_n)^t, \end{aligned}$$

and let

$$\text{MSE}_n(\lambda) = E(U_n(\lambda) - r)^2.$$

Then,

$$\begin{aligned} \text{MSE}_n(\lambda) &= \text{var}(U_n(\lambda)) + (E(U_n(\lambda) - r))^2 \\ &= \sigma_x^2(n) - \lambda^t c_{xy}(n) - c_{xy}(n)^t\lambda \\ &\quad + \lambda^t c_{yy}(n)\lambda + b_x^2(n) \\ &\quad + b_x(n)\lambda^t b_y(n) + b_x(n)b_y(n)^t\lambda \\ &\quad + \lambda^t b_y(n)b_y(n)^t\lambda. \end{aligned} \tag{8}$$

This quadratic form in λ has precisely the same structure as does $\sigma^2(\lambda)$, so that the minimizer λ_n^* of $\text{MSE}_n(\lambda)$ is given by

$$\lambda_n^* = A(n)^{-1}d(n), \tag{9}$$

where

$$\begin{aligned} A(n) &= c_{yy}(n) + b_y(n)b_y(n)^t \\ d(n) &= c_{xy}(n) + b_x(n)b_y(n). \end{aligned}$$

Of course, we again require that $c_{yy}(n)$ be positive definite. Note that this implies $A(n)$ is positive definite. The corresponding minimal value of $\mathrm{MSE}_n(\lambda)$ is then

$$\mathrm{MSE}_n\,(\lambda_n^*) = \sigma_x^2(n) + b_x^2(n) - d(n)^t A(n)^{-1} d(n). \tag{10}$$

We summarize our discussion thus far with the following proposition.

PROPOSITION 1: Assume (7). Then, if $c_{yy}(n)$ is positive definite, the minimizer λ_n^* of $\mathrm{MSE}_n(\lambda)$ is given by (9) and the minimizing value of $\mathrm{MSE}_n(\lambda)$ is given by (10).

As promised earlier, we will now show that the MSE criterion used in this section coincides with the asymptotic efficiency criterion used in Section 2. We will require the additional regularity condition:

$$\{n(\overline{X}_n - r)^2 + n(\overline{Y}_n - \mu)^t(\overline{Y}_n - \mu): n \geqslant 1\} \tag{11}$$

is uniformly integrable. This assumption allows us to pass expectations through the limit theorem (3), thereby yielding

$$\begin{aligned} n^{1/2} b_x(n) &\to 0, \\ n^{1/2} b_y(n) &\to 0, \\ n\, \sigma_x^2(n) &\to \sigma_x^2, \\ n\, c_{xy}(n) &\to c_{xy}, \\ n\, c_{yy}(n) &= c_{yy}, \end{aligned} \tag{12}$$

as $n \to \infty$. We may therefore conclude that

$$n\, A(n) \to c_{yy},$$

and

$$n\, d(n) \to c_{xy}.$$

So that if c_{yy} is positive definite, then

$$\begin{aligned} \lambda_n^* &= (n\, A(n))^{-1}(n\, d(n)) \\ &\to c_{yy}^{-1} c_{xy} = \lambda^*. \end{aligned}$$

Similarly,

$$n\, \mathrm{MSE}_n(\lambda_n^*) \to \sigma^2(\lambda^*),$$

thereby yielding the following result.

PROPOSITION 2: Assume (3) and (11). [Equation (11) implies (7).] If c_{yy} is positive definite, then $\lambda_n^* \to \lambda^*$ and $n\, \mathrm{MSE}_n(\lambda_n^*) \to \sigma^2(\lambda^*)$, as $n \to \infty$.

This proposition is a formal statement of the fact that the MSE and asymptotic efficiency criterion coincide as $n \to \infty$.

4. SMALL-SAMPLE THEORY FOR STEADY-STATE CONTROL VARIATE SCHEMES

Suppose that r is a steady-state parameter of a stochastic system, and the $(X_1,Y_1),(X_2, Y_2), \ldots$ represent observations gathered during the time evolution of the system. Our primary goal here is to obtain an asymptotic expansion for λ_n^* and $\mathrm{MSE}_n(\lambda_n^*)$.

We will need to assume that both (3) and (11) are valid for our steady-state simulation. One condition which guarantees this is to require that $\{(X_n,Y_n) : n \geq 1\}$ be a nondelayed regenerative sequence with regeneration times $T_0 = 0, T_1,T_2, \ldots$. If one imposes the moment condition

$$E\left\{\left(\sum_{k=0}^{T_1-1} (X_k^2 + Y_k^t Y_k + 1)\right)^2\right\} < \infty,$$

then (3) and (11) follow (see pp. 99–104 of CHUNG [3] for a proof in the Markov chain setting; the general case can be argued in precisely the same way).

We will further require that the bias terms take the form

$$\begin{aligned} b_x(n) &= \frac{1}{n} b_x + o\left(\frac{1}{n}\right) \\ b_y(n) &= \frac{1}{n} b_y + o\left(\frac{1}{n}\right) \end{aligned} \tag{13}$$

for some constants b_x and b_y. It has been shown in Glynn [5] that the bias terms have this form for a wide class of equilibrium processes. The assumption (13) is satisfied in a variety of steady-state contexts.

Suppose, for example, that

$$\sum_{n=1}^{\infty} |\mathrm{EX}_n - r| < \infty. \tag{14}$$

Then, if we set

$$b_x = \sum_{n=1}^{\infty} (\mathrm{EX}_n - r),$$

it follows that

$$\begin{aligned} b_x(n) &= \frac{1}{n} \sum_{i=1}^{n} (\mathrm{EX}_i - r) \\ &= \frac{1}{n} b_x - \frac{1}{n} \sum_{i=n+1}^{\infty} (\mathrm{EX}_i - r) \\ &= \frac{1}{n} b_x + o\left(\frac{1}{n}\right). \end{aligned}$$

A similar analysis for $b_y(n)$ shows that

$$\sum_{n=1}^{\infty} |\mathrm{EY}_n - \mu| < \infty \tag{15}$$

is a sufficient condition for the second bias expansion. The absolute convergence of the sums in (14) and (15) occurs automatically if the expectations converge geometrically fast

$$\begin{aligned} EX_n &= r + o(\rho^n), \\ EY_n &= \mu + o(\rho^n), \end{aligned} \tag{16}$$

for some ρ satisfying $|\rho| < 1$. The geometric convergence dictated by (16) is frequently satisfied in a Markov process context, for example. In particular, many aperiodic Markov chains satisfy (16) see Lemma 7.2, p. 224, of Doob [4] and pp. 75–101 of Kemeny and Snell [9].

Assuming now that (3), (11), and (13) are in force, observe that

$$\begin{aligned} \lambda_n^* &= A(n)^{-1}\, d(n) \\ &= (n\, A(n))^{-1}\, (n\, d(n)). \end{aligned}$$

From (12) [this is implied by (11) and (13)] it is evident that

$$\begin{aligned} n\, A(n) &= c_{yy} + \frac{1}{n}\, b_y b_y^t + o\left(\frac{1}{n}\right) \\ n\, d(n) &= c_{xy} + \frac{1}{n}\, b_x b_y + o\left(\frac{1}{n}\right). \end{aligned}$$

Hence, assuming c_{yy} is positive definite,

$$\begin{aligned} (n\, A(n))^{-1} &= \left(c_{yy} \cdot \left(I + \frac{1}{n}\, c_{yy}^{-1} b_y b_y^t + o\left(\frac{1}{n}\right)\right)\right)^{-1} \\ &= \left(I + \frac{1}{n}\, c_{yy}^{-1} b_y b_y^t + o\left(\frac{1}{n}\right)\right)^{-1} c_{yy}^{-1}. \end{aligned}$$

Now for n large enough, the matrix

$$F(n) = \frac{1}{n}\, c_{yy}^{-1} b_y b_{yy}^t + o\left(\frac{1}{n}\right).$$

has a special radius less than one since all elements converge to zero, so that

$$\begin{aligned} (I + F(n))^{-1} &= I - F(n) + F(n)^2 - F(n)^3 + \cdots \\ &= I - \frac{1}{n}\, c_{yy}^{-1} b_y b_y^t + o\left(\frac{1}{n}\right). \end{aligned}$$

Consequently

$$\begin{aligned} \lambda_n^* &= \left(I - \frac{1}{n}\, c_{yy}^{-1} b_y b_y^t + o\left(\frac{1}{n}\right)\right) \cdot c_{yy}^{-1} \cdot \left(c_{xy} + \frac{1}{n}\, b_x b_y + o\left(\frac{1}{n}\right)\right) \\ &= c_{yy}^{-1} c_{xy} - \frac{1}{n}\, c_{yy}^{-1} b_y b_y^t\, c_{yy}^{-1} c_{xy} \\ &\quad + \frac{1}{n}\, c_{yy}^{-1} b_x b_y + o\left(\frac{1}{n}\right). \end{aligned}$$

Similarly, we find that

$$\begin{aligned} \text{n MSE}(\lambda_n^*) &= \sigma_x^2 - c_{xy} c_{yy}^{-1} c_{xy} \\ &- \frac{1}{n} b_x b_y^t c_{yy}^{-1} c_{xy} + \frac{1}{n} b_x^2 \\ &- \frac{1}{n} c_{xy}^t c_{yy}^{-1} b_x b_y \\ &+ \frac{1}{n} c_{xy}^t c_{yy}^{-1} b_y b_y^t c_{yy}^{-1} c_{xy} \\ &+ o\left(\frac{1}{n}\right). \end{aligned}$$

We therefore obtain the following result.

THEOREM 1: Assume (3), (11), and (13). If c_{yy} is positive definite, then

$$\begin{aligned} \lambda_n^* &= \lambda^* - \frac{1}{n} c_{yy}^{-1} b_y b_y^t c_{yy}^{-1} c_{xy} \\ &+ \frac{1}{n} c_{yy}^{-1} b_x b_y + o\left(\frac{1}{n}\right) \end{aligned}$$

and

$$\begin{aligned} n \, \text{MSE}_n(\lambda_n^*) &= \sigma^2(\lambda^*) - \frac{2}{n} b_x b_y^t c_{yy}^{-1} c_{xy} \\ &+ \frac{1}{n} c_{xy}^t c_{yy}^{-1} b_y b_y^t c_{yy}^{-1} c_{xy} \\ &+ \frac{1}{n} b_x^2 + o\left(\frac{1}{n}\right). \end{aligned}$$

5. DEGRADATION USING ASYMPTOTIC CONTROL VECTOR

It is of some interest to examine the degradation in MSE of the control variate scheme when the asymptotic control vector λ^* is used, rather than the small-sample optimal vector λ_n^*.

Let $M_n(\lambda) = n \, \text{MSE}_n(\lambda)$. It is easily verified from (8) that for arbitrary λ and λ_0,

$$\begin{aligned} M_n(\lambda) &= M_n(\lambda_0) - \nabla M_n(\lambda_0)^t \cdot (\lambda - \lambda_0) \\ &+ (\lambda - \lambda_0)^t H_n(\lambda - \lambda_0), \end{aligned} \tag{17}$$

where

$$\begin{aligned} \nabla M_n(\lambda_0) &= 2n(c_{yy}(n) + b_y(n) b_y(n)^t) \lambda_0 \\ &- 2n c_{xy}(n) + 2n \, b_x(n) b_y(n) \end{aligned}$$

and

$$H_n = n \, c_{yy}(n) + n \, b_y(n) \, b_y^t(n).$$

[This is just a Taylor expansion of $M_n(\lambda)$ around $\lambda = \lambda_0$.] Setting $\lambda = \lambda^*$ and $\lambda_0 = \lambda_n^*$, we observe that $\nabla M_n(\lambda_n^*) = 0$, so that (17) becomes

$$M_n(\lambda^*) = M_n(\lambda_n^*) + (\lambda^* - \lambda_n^*)^t H_n(\lambda^* - \lambda_n^*). \tag{18}$$

Letting

$$d = c_{yy}^{-1} b_y b_y^t c_{yy}^{-1} c_{xy} - c_{yy}^{-1} b_x b_y,$$

Theorem 1 shows that $\lambda_n^* = \lambda^* - d/n + o(1/n)$, whereas (12) yields $H_n \rightarrow c_{yy}$ as $n \rightarrow \infty$. It follows from (18) that

$$M_n(\lambda^*) = M_n(\lambda_n^*) + \frac{1}{n^2} d^t c_{yy} d + o\left(\frac{1}{n^2}\right).$$

As a consequence, we obtain the following result.

PROPOSITION 3: Assume (3), (11), (13), and that c_{yy} is positive definite. Then, the degradation in $\mathrm{MSE}_n(\lambda)$ caused by using λ^* rather than λ_n^* is given by $n^{-3} d^t c_{yy} d + o(1/n^3)$. (Since c_{yy} is positive definite, $d^t c_{yy} d$ is always non-negative.)

Thus, the degradation in MSE is of small order, since it decreases as the reciprocal of the cube of the sample size. However, in certain small-sample situations, the degradation could be significant. In such a situation, Theorem 1 provides a possible key to improving the performance of the control scheme.

Let $\hat{\lambda}_n = \lambda^* - d/n$. Noting that $\hat{\lambda}_n = \lambda_n^* + o(1/n)$, it follows from (18) that $M_n(\hat{\lambda}_n) = M_n(\lambda_n^*) + o(1/n^2)$. Thus, using $\hat{\lambda}_n$ as the control vector is "almost" as good as using the optimal vector λ_n^*.

Clearly, in order to obtain optimal asymptotic efficiency from the control variate scheme, $\lambda^* = c_{yy}^{-1} c_{xy}$ must be estimated. Ordinarily, this will require consistent estimation of both c_{yy} and c_{xy}; see Iglehart and Lewis [8] for details in the regenerative case. Thus, if the simulation can estimate the quantities b_x and b_y, $\hat{\lambda}_n$ can be obtained and used to improve performance. The terms b_x and b_y can be estimated in the regenerative case using the results contained in [5]. Note that even if the estimators are not particularly accurate, their influence "washes out" fairly rapidly since $\hat{\lambda}_n \rightarrow \lambda^*$ as $n \rightarrow \infty$. Thus, one should never lose too much efficiency, even with poor estimators.

6. SMALL-SAMPLE THEORY FOR STEADY-STATE MULTIPLE ESTIMATE SCHEMES

Our goal here is to establish small-sample results, analogous to those obtained in Section 4, for the method of multiple estimates. Given a steady-state parameter $r \in \mathbb{R}$ and the vector $e \in \mathbb{R}^d$ with all components 1, suppose that one can generate an $\mathbb{R}^d$-valued sequence $Z_1, Z_2, \ldots$ such that

$$n^{1/2}(\bar{Z}_n - r\,e) \Rightarrow N(0,C) \tag{19}$$

as $n \rightarrow \infty$, where $\bar{Z}_n = (Z_1 + \cdots + Z_n)/n$, and C is a $d \times d$ covariance matrix. The idea behind the method of multiple estimates is that for any vector α such that $\alpha^t e = 1$, (19) implies that

$$\alpha^t \bar{Z}_n \Rightarrow r;$$

one now chooses α so as to maximize efficiency. Heidelberger [7] explored this technique in the context of Markov chains, and showed how one can generate Z_i's with property (20).

It is worth pointing out that the method of multiple estimates can be viewed as a special case of control variates. Let

$$X_n = e'Z_n/d.$$

Then,

$$\alpha' \bar{Z}_n = \bar{X}_n - \lambda' \bar{Y}_n,$$

where $Y_n = Z_n$ and $\lambda = e/d - \alpha$. The constraint $\alpha' e = 1$ translates into choosing λ so that $\lambda' e = 0$. With the (X_n, Y_n)'s defined in this way, we are again in the setting of Section 2 through 4. Although it would be possible to derive all the asymptotic theory for multiple estimates by appealing to the previously developed results for control variates, it seems easier to obtain them directly.

Note that the continuous mapping lemma, as applied to (19), yields

$$n^{1/2} (\alpha' \bar{Z}_n - r) \Rightarrow \sigma^2(\alpha) N(0,1),$$

where $\sigma^2(\alpha) = \alpha' C \alpha$. The minimizer of $\sigma^2(\alpha)$ subject to $\alpha' e = 1$ is given by

$$\alpha^* = C^{-1}e/e'C^{-1}e,$$

provided that C is positive definite (see p. 60 of Rao [11] . The minimal value of $\sigma^2(\alpha)$ is then given by $\sigma^2(\alpha^*) = (e'C^{-1}e)^{-1}$. The following theorem summarizes the situation.

THEOREM 2: Assume that (19) holds with C positive definite. Then (i) $\sigma^2(\alpha)$ has minimal value $(e'C^{-1}e)^{-1}$, and is minimized at $\alpha^* = C^{-1}e/(e'C^{-1}e)$.

If, in addition, $\{n(\bar{Z}_n - re)' (\bar{Z}_n - re) : n \geq 1\}$ is uniformly integrable and if $E\bar{Z}_n = re + b/n + o(1/n)$ for some $b \in \mathbb{R}^d$, then: (ii) $\text{MSE}_n(\alpha) \equiv E(\alpha' \bar{Z}_n - r)^2$ has minimal value $(e'(C(n) + b(n)b(n)')^{-1}e)^{-1}$ and is minimized at

$$\alpha_n^* = \frac{(C(n) + b(n)b(n)')^{-1}e}{e'(C(n) + b(n)b(n)')^{-1}e},$$

where $C(n) = E(Z_n Z_n') - (EZ_n)(EZ_n)'$ and $b(n) = E\bar{Z}_n - re$.

(iii)
$$\alpha_n^* = \frac{C^{-1}e}{(e'C^{-1}e)} \left(1 + \frac{1}{n} \frac{e'C^{-1}bb'C^{-1}e}{(e'C^{-1}e)}\right) - \frac{1}{n} \frac{C^{-1}bb'C^{-1}e}{(e'C^{-1}e)} + o\left(\frac{1}{n}\right)$$

and

$$\text{MSE}(\alpha_n^*) = \frac{1}{e'C^{-1}e} \left(1 + \frac{e'C^{-1}bb'C^{-1}e}{n\, e'C^{-1}e}\right) + o\left(\frac{1}{n}\right)$$

(iv) $\text{MSE}(\alpha^*) = \text{MSE}(\alpha_n^*) + n^{-3} d'Cd + o(n^{-3})$, where

$$d = \left(\frac{e'C^{-1}bb'C^{-1}e}{(e'C^{-1}e)^2}\right) C^{-1}e - \frac{C^{-1}bb'C^{-1}e}{e'C^{-1}e}$$

Thus, the results obtained for the method of multiple estimates are qualitatively similar to those obtained for control variates.

7. CONCLUSIONS

Using the MSE criterion, we have shown that under rather general conditions, the small-sample optimal control coefficients λ_n^*, for steady-state simulations, differ from the asymptotically optimal control coefficients λ^* by a factor of order n^{-1}. The first-order error term involves only the asymptotic covariance structure, and the first-order bias terms; the (exact) small-sample covariance structure plays no role, even in the case of a nonstationary steady-state simulation. The loss, in MSE efficiency, created by using the asymptotically optimal λ^*, rather than the small-sample optimal λ_n^*, is of order n^{-3}. Thus, the loss in efficiency is of small order. Similar results hold for the method of multiple estimates.

ACKNOWLEDGMENTS

This work was partially supported by U.S. Army Research Contract DAAG29-84-K-0030 and National Science Foundation Grants MCS-8203483 and DCR-85-09668. We are grateful for the referees comments which helped improve the exposition of this article.

REFERENCES

[1] Anderson, T. W., *An Introduction to Multivariate Statistical Analysis*, 2nd Ed., John Wiley, New York, 1984.
[2] Billingsley, P., *Convergence of Probability Measures*, John Wiley, New York, 1968.
[3] Chung, K. L., *Markov Chains with Stationary Transition Probabilities*, Springer-Verlag, New York, 1967.
[4] Doob, J. L., *Stochastic Processes*, John Wiley, New York, 1953.
[5] Glynn, P. W., "A low bias steady-state estimator for equilibrium processes," *Mathematics of Operations Research* (to be published).
[6] Glynn, P. W. and Whitt, W., "Indirect estimation via $L = \lambda W$" (to be published).
[7] Heidelberger, P., "Variance reduction techniques for the simulation of Markov processes. I: Multiple estimates," *IBM Journal of Research and Development*, **24**, 570–581 (1980).
[8] Iglehart, D. L. and Lewis, P. A. W., "Regenerative simulation with internal controls," *Journal of the ACM*, **26**, 271–282 (1979).
[9] Kemeny, J. L. and Snell, J. L., *Finite Markov Chains*, Van Nostrand, Princeton, NJ, 1960.
[10] Lavenberg, S., Moeller, S. T. L., and Welch, P. D., "Statistical results on control variables with application to queueing network simulation," *Operations Research*, **30**, 182–202 (1982).
[11] Rao, C. R., *Linear Statistical Inference and its Applications*, John Wiley, New York, 1973.
[12] Rubinstein, R. V. and Marcus, R., "Efficiency of multivariate control variates and Monte Carlo simulation," *Operational Research*, **33**, 661–677 (1985).
[13] Venkratraman, S. and Wilson, J. R., "Efficiency of control variates in multiresponse simulation," Technical Report, School of Industrial Engineering, Purdue University, West Lafayette, IN, 1985.

Manuscript received March 25, 1986
Revised manuscript received August 1988
Accepted August 18, 1988