

# INTEGRATED VARIANCE REDUCTION STRATEGIES FOR SIMULATION

ATHANASSIOS N. AVRAMIDIS  
*SABRE Decision Technologies,  
Dallas/Fort Worth Airport, Texas*

JAMES R. WILSON  
*North Carolina State University,  
Raleigh, North Carolina*

We develop strategies for integrated use of certain well-known variance reduction techniques to estimate a mean response in a finite-horizon simulation experiment. The building blocks for these integrated variance reduction strategies are the techniques of conditional expectation, correlation induction (including antithetic variates and Latin hypercube sampling), and control variates; and all pairings of these techniques are examined. For each integrated strategy, we establish sufficient conditions under which that strategy will yield a smaller response variance than its constituent variance reduction techniques will yield individually. We also provide asymptotic variance comparisons between many of the methods discussed, with emphasis on integrated strategies that incorporate Latin hypercube sampling. An experimental performance evaluation reveals that in the simulation of stochastic activity networks, substantial variance reductions can be achieved with these integrated strategies. Both the theoretical and experimental results indicate that superior performance is obtained via joint application of the techniques of conditional expectation and Latin hypercube sampling.

*Subject classifications:* Simulation, efficiency: conditioning, control variates, correlation induction. Simulation, design of experiments: antithetic variates, Latin hypercube sampling. Simulation, statistical analysis: combined Monte Carlo methods.

*Area of review:* SIMULATION.

For many complex stochastic systems, purely mathematical methods of analysis are unavailable, and deterministic numerical methods have extremely limited utility. By making it feasible to analyze the performance of such systems, simulation has become one of the most widely applied tools of operations research. Direct simulation, however, may require excessive run lengths (or replication counts) to yield estimators with acceptable precision, thus becoming prohibitively expensive. A diversity of variance reduction techniques (VRTs) have been developed to improve the efficiency of simulations—that is, to reduce the computing effort necessary to obtain some specified precision. For a survey of variance reduction techniques, see Wilson (1984), Nelson (1987), and L’Ecuyer (1994).

Relatively little work has been directed toward integrating the most widely used variance reduction techniques into an overall strategy that can exploit various sources of efficiency improvement simultaneously. Kleijnen (1975) combined the techniques of antithetic variates (AV) and common random numbers (CRN) to estimate the mean difference between the responses of two systems. He showed by simple examples that some implementations of the combined technique (AV+CRN) may be inferior to either antithetic variates or common random numbers used alone. For each of the variance reduction techniques under consideration (AV, CRN, and AV+CRN), Kleijnen proposed a scheme for optimally allocating replications to the two systems; and when using these schemes, he found experimentally that his combined technique (AV+CRN) was superior to antithetic variates or common random numbers used alone. Burt, Gaver, and Perlas (1970) combined the techniques of antithetic variates and control variates (CV) for the simulation of activity networks, and experimentally they found the combined technique (AV+CV) to provide more precise results than either antithetic variates or control variates used alone. Loulou and Beale (1976) observed similar improvements in efficiency when they combined antithetic variates with a version of systematic sampling for the simulation of activity networks.

Few attempts have been made either to quantify the efficiency improvements resulting from integrated variance reduction strategies or to establish general conditions under which those integrated strategies are preferable to direct simulation or standard variance reduction techniques used alone. Schruben and Margolin (1978) considered the estimation of a simulation metamodel—that is, a linear regression model of a simulation-generated performance measure expressed in terms of a vector of design variables for the target system. Schruben and Margolin provided conditions under which the techniques of antithetic variates and common random numbers jointly yield guaranteed efficiency improvements compared to (a) using common random number streams at all design points in a simulation experiment, and (b) using independent sets of random number streams at different design

points. Extending this work, Tew and Wilson (1994) incorporated control variates into the Schruben-Margolin scheme and established conditions under which their combined approach is superior to (i) the original Schruben-Margolin scheme, (ii) control variates used alone, and (iii) direct simulation using independent random number streams at each design point.

In this paper, we develop strategies for jointly applying certain well-known variance reduction techniques to estimate the expected value of a univariate response in a finite-horizon simulation experiment. These strategies incorporate the following variance reduction techniques in pairs: conditional expectation, control variates, and correlation induction. Although our general definition of correlation induction encompasses several related variance reduction techniques such as common random numbers and antithetic variates, we focus much of our analysis on a special case of correlation induction known as Latin hypercube sampling (LHS). For each integrated variance reduction strategy, we formulate and justify conditions on the structure and operation of the simulation under which that strategy will yield a smaller response variance than its constituent variance reduction techniques will yield individually. We also derive asymptotic variance comparisons between many of the methods that are discussed, with emphasis on integrated strategies that incorporate Latin hypercube sampling. Finally, we present Monte Carlo evidence that large efficiency gains can be achieved by applying these integrated variance reduction strategies to the simulation of stochastic activity networks.

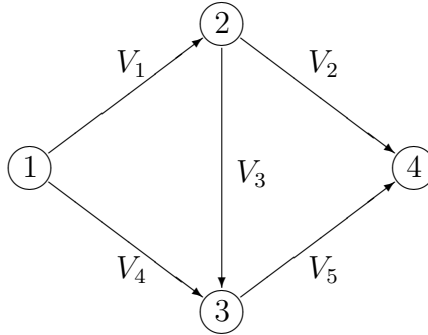
This paper is organized as follows. In Section 1 we define our notation and review the basic variance reduction techniques that will be used as building blocks for the integrated variance reduction strategies. In Section 2 we formulate and analyze the integrated strategies. In Section 3 we provide asymptotic variance comparisons for all of the strategies that involve Latin hypercube sampling. Applications to simulation of stochastic activity networks are detailed in Section 4, including implementation of the integrated strategies and validation of their underlying assumptions; moreover, Section 4 contains an experimental performance evaluation of all of the methods previously discussed. In Section 5 we summarize the main findings of this work, and we recommend directions for future research. Although this paper is based on Avramidis (1993), some of our results were also presented in Avramidis and Wilson (1993b).

## 1. NOTATION AND BACKGROUND

The basic problem is to estimate the expected value  $\theta$  of a stochastic simulation response  $Y$ . For an appropriate choice of  $Y$ , this problem includes estimating noncentral moments and

probabilities; but it does not include estimating, for example, central moments or quantiles. We assume throughout this paper that  $E[Y^2] < \infty$  so that  $\theta \equiv E[Y]$  and  $\sigma_Y^2 \equiv \text{Var}[Y]$  are both finite. The response is assumed to have the form  $Y = f(V_1, \dots, V_p)$ , where the function  $f(\cdot)$  has a fixed number of inputs; and the *input random variates*  $\{V_1, \dots, V_p\}$  have a known probabilistic structure. By this we merely mean that we have a way of generating the random vector  $\mathbf{V} \equiv (V_1, \dots, V_p)$  so that it has the correct distribution. The input random variates are generated as  $\mathbf{V} = \mathcal{H}(\mathbf{U})$ , where the random vector  $\mathbf{U} \equiv (U_1, \dots, U_d)$  with fixed dimension  $d$  is composed of independent *random numbers* that are uniformly distributed on the unit interval  $(0, 1)$ ; and  $\mathcal{H}$  is a *sampling plan* that describes the variate-generation scheme used in the simulation. At some points in this paper, it is convenient to view  $Y$  as a function of the vector  $\mathbf{V}$  of input random variates, whereas elsewhere we prefer to view  $Y$  as a function of the vector  $\mathbf{U}$  of input random numbers. In the latter situation, we write  $Y = f[\mathcal{H}(\mathbf{U})] \equiv y(\mathbf{U})$ . Throughout this paper, the word *function* will mean a Borel measurable function, taking either real scalar values or real vector values. While bold symbols will usually be used to represent vectors and matrices, nonbold symbols will usually be used to represent functions and scalar quantities. All vectors will be row vectors unless otherwise stated.

**Example 1.** Figure 1 depicts a stochastic activity network with source node 1 and sink node 4. The input random variates are  $\{V_1, \dots, V_5\}$ , where  $V_j$  is the duration of activity



**Figure 1.** A stochastic activity network.

(arc)  $j$  for  $j = 1, \dots, 5$ . Thus  $p = 5$  in this example. Let  $T$  denote the longest directed path from the source node to the sink node so that

$$T = \max\{V_1 + V_2, V_1 + V_3 + V_5, V_4 + V_5\}; \quad (1)$$

and suppose the objective is to estimate  $\theta = \Pr\{T \leq t\}$  for a given cutoff time  $t$ . The corresponding response is  $Y = \mathbf{1}\{T \leq t\}$ , the indicator function of the event  $\{T \leq t\}$ . Suppose that the random variates  $V_1$ ,  $V_4$ , and  $V_5$  are mutually independent with known

distributions and that the random vector  $(V_2, V_3)$  is independent of  $V_1$ ,  $V_4$ , and  $V_5$  with a known bivariate distribution. Consider the sampling plan  $\mathcal{H}$  defined by

$$V_1 = H_1(U_1, U_2), \quad V_2 = H_2(U_3), \quad V_3 = H_3(U_3, U_4), \quad V_4 = H_4(U_5), \quad \text{and} \quad V_5 = H_5(U_6), \quad (2)$$

where  $\{U_1, \dots, U_6\}$  are independent random numbers and  $H_1(\cdot), \dots, H_5(\cdot)$  are given functions that can be evaluated readily. Here we do not use the method of inversion to generate all input variates; instead  $V_1$  is generated by some other method that requires two random numbers. Moreover,  $V_3$  is generated conditional on  $V_2$ ; and thus  $V_3$  is also a function of two random numbers. With the sampling plan (2), we have  $d = 6$ ; and the response function  $y(\cdot)$  has the form

$$y(u_1, \dots, u_6) = \begin{cases} 1, & \text{if } \max \{H_1(u_1, u_2) + H_2(u_3), \\ & H_1(u_1, u_2) + H_3(u_3, u_4) + H_5(u_6), H_4(u_5) + H_5(u_6)\} \leq t, \\ 0, & \text{otherwise.} \end{cases} \quad \blacksquare$$

In a direct-simulation experiment, we perform  $n$  independent replications that yield independent identically distributed (i.i.d.) observations of the response  $\{Y_i : i = 1, \dots, n\}$ . The direct-simulation estimator is the corresponding sample mean  $\bar{Y}(n)$ , which is unbiased and has variance  $n^{-1}\sigma_Y^2$ . The aim of variance reduction techniques is to identify an alternative estimator  $\hat{\theta}(n)$  based on  $n$  replications (which are not necessarily i.i.d.) such that

$$\mathbb{E}[\hat{\theta}(n)] = \theta \quad \text{and} \quad \text{Var}[\hat{\theta}(n)] < \text{Var}[\bar{Y}(n)].$$

Sometimes we also consider biased estimators; in this situation, the reason for preferring  $\hat{\theta}(n)$  over  $\bar{Y}(n)$  is usually a reduction in mean square error,

$$\text{MSE}[\hat{\theta}(n)] \equiv \mathbb{E}\left\{\left[\hat{\theta}(n) - \theta\right]^2\right\} < \text{Var}[\bar{Y}(n)].$$

In addition to analyzing the behavior of a new estimator  $\hat{\theta}(n)$  for fixed values of the sample size  $n$ , we will analyze also the asymptotic behavior of  $\hat{\theta}(n)$  as  $n$  tends to infinity. Even when  $\hat{\theta}(n)$  is based on  $n$  dependent simulation runs, typically a central limit theorem (CLT) holds so that

$$n^{1/2}[\hat{\theta}(n) - \theta] \xrightarrow{\mathcal{D}} N(0, \sigma^2) \quad \text{as } n \rightarrow \infty, \quad (3)$$

where  $\xrightarrow{\mathcal{D}}$  denotes convergence in distribution (Wolff 1989, p. 43) and  $N(\mu, \sigma^2)$  denotes a normal random variable with mean  $\mu$  and variance  $\sigma^2$ . We then say that  $\hat{\theta}(n)$  has *asymptotic*

mean  $\theta$  and asymptotic variance parameter  $\sigma^2$ . Suppose that we have two estimators  $\hat{\theta}_1(n)$  and  $\hat{\theta}_2(n)$  satisfying CLTs of the form (3) with respective variance parameters  $\sigma_1^2$  and  $\sigma_2^2$  such that  $\sigma_1^2 \leq \sigma_2^2$ . We then say that  $\hat{\theta}_1$  *asymptotically dominates*  $\hat{\theta}_2$ . For any given finite sample size  $n$ , this does not guarantee that either the bias or the variance of  $\hat{\theta}_1(n)$  has smaller magnitude than the corresponding characteristic of  $\hat{\theta}_2(n)$ . However, asymptotic dominance is a reasonable criterion for comparing estimators when it is difficult to obtain exact expressions for the bias and variance of each estimator at each sample size  $n$ . For simplicity, we will occasionally suppress the argument  $n$  in the discussion of alternative simulation-based estimators  $\hat{\theta}_1$  and  $\hat{\theta}_2$  when no confusion can result from this usage.

We choose to compare estimators in terms of their finite-sample variances and their asymptotic variance parameters without considering their computing costs. This is done mainly because the computing cost associated with each estimator is hard to quantify, since it depends on the specific configuration of hardware and software that is used as well as other machine-dependent characteristics. Fortunately, in many complex simulations (for which variance reduction techniques are most needed), the computing cost is fairly insensitive to the type of estimator used; and therefore such a comparison is appropriate (Fishman 1989). Next we review the VRTs that will be used as the building blocks for our integrated variance reduction strategies.

### 1.1. Conditional Expectation (CE)

Suppose we can identify an auxiliary random vector  $\mathbf{X}$  generated on each replication of the simulation such that we can evaluate the conditional expectation  $\zeta(\mathbf{x}) \equiv E[Y|\mathbf{X} = \mathbf{x}]$  analytically or numerically for each possible value of  $\mathbf{x}$ . Thus the random variable  $Z \equiv \zeta(\mathbf{X})$  is an alternative estimator of  $\theta$  based on a single replication. From the results  $\{\mathbf{X}_i : i = 1, \dots, n\}$  of  $n$  independent replications, we compute the corresponding random sample  $\{Z_i = \zeta(\mathbf{X}_i) : i = 1, \dots, n\}$ ; and the *conditional-expectation* (CE) estimator of  $\theta$  is

$$\hat{\theta}_{\text{CE}}(n) \equiv n^{-1} \sum_{i=1}^n Z_i.$$

The double expectation theorem (Bickel and Doksum 1977, p. 6) and the conditional variance relation (Wolff 1989, p. 34) respectively imply that

$$E[Z] = E[E(Y|\mathbf{X})] = \theta \quad \text{and} \quad \sigma_Z^2 \equiv \text{Var}[Z] = \sigma_Y^2 - E[\text{Var}(Y|\mathbf{X})]. \quad (4)$$

It follows immediately that  $\hat{\theta}_{\text{CE}}(n)$  is an unbiased estimator of  $\theta$ . It also follows from (4) that  $\text{Var}[\hat{\theta}_{\text{CE}}(n)] \leq \text{Var}[\bar{Y}(n)]$ , with equality holding if and only if  $Y$  is a function of  $\mathbf{X}$  alone so that  $\text{Var}(Y|\mathbf{X})$  vanishes with probability one.

**Example 1 (continued).** Suppose that the following cumulative distribution functions (c.d.f.'s) can be evaluated analytically or numerically:  $F_{2,3}(\cdot, \cdot)$  is the joint c.d.f. of the input random vector  $(V_2, V_3)$ , and  $F_4(\cdot)$  is the c.d.f. of  $V_4$ . Then we can readily compute the conditional expectation of  $Y = \mathbf{1}\{T \leq t\}$  given  $\mathbf{X} \equiv (X_1, X_2) \equiv (V_1, V_5)$ . We have

$$\begin{aligned}\zeta(x_1, x_2) &= \Pr\{T \leq t | X_1 \equiv V_1 = x_1, X_2 \equiv V_5 = x_2\} \\ &= \Pr\{V_2 \leq t - x_1, V_3 \leq t - x_1 - x_2, V_4 \leq t - x_2\} \\ &= F_{2,3}(t - x_1, t - x_1 - x_2) \cdot F_4(t - x_2)\end{aligned}$$

so that

$$\begin{aligned}Z \equiv \zeta(\mathbf{X}) &= F_{2,3}(t - X_1, t - X_1 - X_2) \cdot F_4(t - X_2) \\ &= F_{2,3}(t - V_1, t - V_1 - V_5) \cdot F_4(t - V_5). \quad \blacksquare\end{aligned}\tag{5}$$

## 1.2. Correlation Induction (CI)

To estimate the expected response of a single simulated system, we often use correlation-induction methods such as antithetic variates to obtain negatively correlated replications of the response, thereby reducing the variance of the sample mean response. On the other hand to compare several simulated systems with respect to their expected responses, we often use the correlation-induction method of common random numbers to obtain positively correlated responses from each simulated system, thereby reducing the variance of the difference between each pair of responses. Since this paper is limited to the analysis of a single system, we concentrate our attention on correlation-induction techniques that yield negatively correlated responses.

To achieve maximum generality for our results on integrated variance reduction strategies that are presented in Sections 2.1 and 2.2, here we describe a method for obtaining negatively correlated replications of an arbitrary random output  $W$  observed in the simulation. (Since we will apply the following development not only to the original response  $Y$  but also to other simulation-generated outputs, we let the symbol  $W$  denote a generic simulation output to which a correlation-induction strategy will be applied.) We view  $W$  as a function of the input random numbers,

$$W = w(U_j : j \in I_W),$$

where  $I_W$  is a subset of  $\{1, \dots, d\}$ , and the function  $w(\cdot)$  is defined by the simulation code.

A useful condition that often guarantees negative induced correlation is based on the notion of *negative quadrant dependence* defined by Lehmann (1966). We say that the distri-

bution of the bivariate random vector  $(A_1, A_2)^T$  is negatively quadrant dependent (n.q.d.) if

$$\Pr\{A_1 \leq a_1, A_2 \leq a_2\} \leq \Pr\{A_1 \leq a_1\} \cdot \Pr\{A_2 \leq a_2\} \quad \text{for all } a_1, a_2.$$

(Throughout this paper, the roman superscript  $^T$  denotes the transpose of a vector or matrix, whereas the italic letter  $T$  denotes a network completion time like (1); and since  $T$  is never used as a superscript, no confusion should arise from our use of these symbols.) We will exploit the concept of negative quadrant dependence in Result 1 below to provide the desired sufficient condition for negatively correlated simulation responses. Moreover, we use this concept to define a special class  $\mathcal{G}$  of distributions for the random-number inputs. Every distribution  $G \in \mathcal{G}$  must have the following correlation-induction properties:

CI<sub>1</sub> For some  $k \geq 2$ , the distribution  $G$  is  $k$ -variate with univariate marginals that are uniform on the unit interval  $(0, 1)$ .

CI<sub>2</sub> Each bivariate marginal of  $G$  is n.q.d.

CI<sub>3</sub> All bivariate marginals of  $G$  are the same.

When it is desirable to indicate explicitly that a distribution in  $\mathcal{G}$  is  $k$ -variate, we will write “ $G^{(k)} \in \mathcal{G}$ ” rather than “ $G \in \mathcal{G}$ .” Throughout this paper, we let  $G_{\text{IR}}^{(k)}$  denote the distribution of  $k$  independent random numbers. It is clear that  $G_{\text{IR}}^{(k)}$  has properties CI<sub>1</sub>–CI<sub>3</sub> so that  $G_{\text{IR}}^{(k)} \in \mathcal{G}$ .

First we describe a general scheme for obtaining some stochastic dependence between replications of the response  $W = w(U_j : j \in I_W)$ ; and then we state conditions on the function  $w(\cdot)$  under which this scheme yields negatively correlated replications. To generate  $k$  dependent replications of the simulation output  $W$ , we choose a  $k$ -variate distribution  $G^{(k)} \in \mathcal{G}$  and a set  $L_W \subseteq I_W$  consisting of the indices of the arguments of  $w(\cdot)$  that will be used for inducing dependence. Let  $U_j^{(i)}$  denote the  $j$ th input random number used in the  $i$ th replication, where  $i = 1, \dots, k$  and  $j \in I_W$ . We obtain  $k$  dependent replications,

$$W^{(i)} = w(U_j^{(i)} : j \in I_W) \quad \text{for } i = 1, \dots, k, \quad (6)$$

by sampling the column vectors of input random numbers,

$$\mathcal{U}_j \equiv [U_j^{(1)}, \dots, U_j^{(k)}]^T \quad \text{for } j \in I_W, \quad (7)$$

according to a scheme satisfying the following conditions—

SC<sub>1</sub> For every index  $j \in L_W$ , the random vector  $\mathcal{U}_j$  has distribution  $G^{(k)}$ .



SC<sub>2</sub> For every index  $j \in I_W - L_W$ , the random vector  $\mathcal{U}_j$  has distribution  $G_{\text{IR}}^{(k)}$ .

SC<sub>3</sub> The column vectors  $\mathcal{U}_1, \dots, \mathcal{U}_d$  are mutually independent.

Sampling condition SC<sub>1</sub> specifies that we induce dependence between the outputs  $\{W^{(i)} : i = 1, \dots, k\}$  by arranging a negative quadrant dependence between the  $j$ th random numbers sampled on each pair of replications, provided  $j \in L_W$ . Sampling condition SC<sub>2</sub> specifies that for each  $j \notin L_W$ , the  $j$ th random number should be sampled independently on different replications. Finally sampling condition SC<sub>3</sub> requires mutual independence of the random numbers used within the  $i$ th replication to generate the output  $W^{(i)}$ ; and this guarantees that each  $W^{(i)}$  has the correct distribution.

**Example 1 (continued).** Suppose that the output of immediate interest is  $W = V_1 + V_2$ , the duration of the first path from node 1 to node 4 in the definition (1) of the response  $T$ . The sampling plan (2) implies

$$W = w(U_1, U_2, U_3) = H_1(U_1, U_2) + H_2(U_3)$$

so that  $I_W = \{1, 2, 3\}$ . Suppose moreover that we wish to obtain 4 dependent replications of  $W$  by using some quadrivariate distribution  $G^{(4)} \in \mathcal{G}$  to generate 4 dependent replications of  $V_1$  while using  $G_{\text{IR}}^{(4)}$  to generate 4 independent replications of  $V_2$ ; thus we have  $L_W = \{1, 2\}$ . For simplicity in this illustration of the method of correlation induction, we ignore the other activities in the network and their corresponding random-number inputs; and then the simulation experiment would involve sampling an array of 12 random numbers as depicted below.

$$\begin{array}{c}
 \text{independent columns} \\
 \hline
 \begin{array}{ccc}
 \mathcal{U}_1 & \mathcal{U}_2 & \mathcal{U}_3 \\
 \wr & \wr & \wr \\
 G^{(4)} & G^{(4)} & G_{\text{IR}}^{(4)} \\
 \downarrow & \downarrow & \downarrow
 \end{array} \\
 \left[ \begin{array}{ccc}
 U_1^{(1)} & U_2^{(1)} & U_3^{(1)} \\
 U_1^{(2)} & U_2^{(2)} & U_3^{(2)} \\
 U_1^{(3)} & U_2^{(3)} & U_3^{(3)} \\
 U_1^{(4)} & U_2^{(4)} & U_3^{(4)}
 \end{array} \right] \rightarrow \left. \begin{array}{l} W^{(1)} \\ W^{(2)} \\ W^{(3)} \\ W^{(4)} \end{array} \right\} \begin{array}{l} \text{dependent rows yield} \\ \text{dependent responses} \end{array}
 \end{array}$$

In this correlation-induction scheme, we observe the following characteristics: (a) each row of the array  $[U_j^{(i)}]$  corresponds to a replication of the simulation; (b) the first two columns ( $\mathcal{U}_1$

and  $\mathcal{U}_2$ ) of the array are sampled from the distribution  $G^{(4)}$  so that each of these columns consists of negatively quadrant dependent random numbers; (c) the third column ( $\mathcal{U}_3$ ) is sampled from the distribution  $G_{\text{IR}}^{(4)}$  so that it consists of independent random numbers; and (d) the three columns are mutually independent random vectors. The effect of this scheme is that the resulting replications of the response  $W$  are dependent. ■

Having generated  $k$  dependent replications of  $W$  as in (6), we define the average

$$W_{\text{CI}}(G^{(k)}, L_W) \equiv k^{-1} \sum_{i=1}^k W^{(i)}, \quad (8)$$

where we make explicit the dependence of  $W_{\text{CI}}$  on the distribution  $G^{(k)}$  and the index set  $L_W$  where  $G^{(k)}$  is applied. The mean and variance of the estimator  $W_{\text{CI}}(G^{(k)}, L_W)$  are easily derived. Clearly, for any  $G^{(k)} \in \mathcal{G}$  and  $L_W \subseteq I_W$ ,

$$\mathbb{E}[W_{\text{CI}}(G^{(k)}, L_W)] = \mathbb{E}[W], \quad (9)$$

and

$$\begin{aligned} \text{Var}[W_{\text{CI}}(G^{(k)}, L_W)] &= k^{-2} \left\{ \sum_{i=1}^k \text{Var}[W^{(i)}] + 2 \sum_{i=1}^{k-1} \sum_{\ell=i+1}^k \text{Cov}[W^{(i)}, W^{(\ell)}] \right\} \\ &= k^{-1} \text{Var}(W) [1 + (k-1) \rho_W(G^{(k)}, L_W)], \end{aligned} \quad (10)$$

where

$$\rho_W(G^{(k)}, L_W) \equiv \text{Corr}[W^{(1)}, W^{(2)}]. \quad (11)$$

Now we state conditions on the response function (6) under which the dependence-induction scheme (7) is guaranteed to yield negatively correlated replications. The next result follows easily from Theorem 1(iii) and Lemma 3 of Lehmann (1966). It will be used often in the remainder of the paper.

**Result 1.** *If  $G^{(k)}$  has property  $\text{CI}_2$ , if  $\{W^{(i)} : i = 1, \dots, k\}$  is generated as in (6) and (7) subject to conditions  $\text{SC}_1$ – $\text{SC}_3$ , and if  $w(\cdot)$  is a monotone function of each argument with index in  $L_W$ , then for  $i, \ell = 1, \dots, k$  and  $i \neq \ell$ , we have  $\text{Cov}[W^{(i)}, W^{(\ell)}] \leq 0$ , with equality holding if and only if  $W^{(i)}$  and  $W^{(\ell)}$  are independent.*

Thus  $W_{\text{CI}}(G^{(k)}, L_W)$  has no larger variance than  $\bar{W}(k)$ , the average of  $k$  independent replications of  $W$ , whenever  $w(\cdot)$  is a monotone function of each random-number input  $U_j$  with index  $j \in L_W$ ; no assumption is needed with respect to the behavior of  $w(\cdot)$  as a function of  $U_j$  for  $j \notin L_W$ .

**Remark 1.** Property  $\text{CI}_3$  was used merely to obtain the second equality in display (10). Assuming only that  $G^{(k)}$  has properties  $\text{CI}_1$  and  $\text{CI}_2$ , we still obtain (9); and it follows from Result 1 and the first equality in display (10) that  $\text{Var}[W_{\text{CI}}(G^{(k)}, L_W)] \leq \text{Var}[\bar{W}(k)]$ . We chose to include property  $\text{CI}_3$  in the definition of  $\mathcal{G}$  because this property frequently holds in practice and because it yields a simplified variance expression in the second equality of display (10).

Finally we formulate  $\hat{\theta}_{\text{CI}}(G^{(k)}, n)$ , the *correlation-induction* (CI) estimator of  $\theta$  based on the  $k$ -variate distribution  $G^{(k)}$  and  $n$  simulation runs. Specifically,  $\hat{\theta}_{\text{CI}}(G^{(k)}, n)$  is obtained by averaging  $m = n/k$  i.i.d. replications of the statistic  $Y_{\text{CI}}(G^{(k)}, L_Y)$ , where we take  $L_Y = I_Y = \{1, \dots, d\}$ ; and to obtain a single observation of  $Y_{\text{CI}}(G^{(k)}, \{1, \dots, d\})$ , we average  $k$  negatively correlated responses, where all  $d$  random-number inputs are used for correlation induction. In terms of the notation in (6) and (8), the correlation-induction estimator of  $\theta$  is defined formally as

$$\hat{\theta}_{\text{CI}}(G^{(k)}, n) \equiv m^{-1} \sum_{\ell=1}^m \tilde{Y}_{\ell}, \text{ where } m = n/k \text{ and } \{\tilde{Y}_{\ell}\}_{\ell=1}^m \stackrel{\text{i.i.d.}}{\sim} Y_{\text{CI}}(G^{(k)}, \{1, \dots, d\}). \quad (12)$$

To simplify the exposition, we assume throughout this paper that  $n$  is an integral multiple of  $k$ . Next we review two important special cases of the method of correlation induction.

### 1.2.1. Antithetic Variates (AV)

To generate  $k = 2$  correlated replications by the method of antithetic variates, we sample the random numbers  $\{U_j^* : j = 1, \dots, d\}$  independently and compute the column vectors of (7) according to the relation

$$\mathcal{U}_j = (U_j^*, 1 - U_j^*)^T \text{ for } j = 1, 2, \dots, d. \quad (13)$$

We let  $G_{\text{AV}}^{(2)}$  denote the distribution of  $\mathcal{U}_j$ . It is straightforward to check that  $G_{\text{AV}}^{(2)}$  satisfies conditions  $\text{CI}_1$ – $\text{CI}_3$  so that  $G_{\text{AV}}^{(2)} \in \mathcal{G}$ . Since  $L_Y = I_Y$ , the method of antithetic variates is clearly a special case of the general correlation-induction scheme described by (6) and (7). If the simulation response  $y(\cdot)$  is a monotone function of each of its random-number inputs, then it follows that

$$\text{Var}[\hat{\theta}_{\text{CI}}(G_{\text{AV}}^{(2)}, n)] \leq \text{Var}[\bar{Y}(n)]. \quad (14)$$

More generally, relation (14) holds if the method of antithetic variates is applied only to the random-number inputs on which  $y(\cdot)$  depends monotonically while all other random-number

inputs are sampled independently—thus it is not necessary for the simulation response to be a monotonic function of all of its random-number inputs in order to achieve a variance reduction using this correlation-induction technique. For example in a queueing network simulation, the method of antithetic variates could be applied to the random-number inputs driving the service times at different stations that are generated by monotonic transformations (such as inversion), while all other random-number inputs (such as those driving the selection of routes) could be sampled independently; and then Result 1 guarantees reduced variance for the estimator of mean flowtime. Of course a similar remark applies to all other correlation-induction techniques discussed in this paper.

### 1.2.2. Latin Hypercube Sampling (LHS)

To generate  $k$  correlated replications via Latin hypercube sampling, we compute the input random numbers according to the relation

$$U_j^{(i)} = \frac{\pi_j(i) - 1 + U_{ij}^*}{k} \quad \text{for } i = 1, \dots, k \text{ and } j = 1, \dots, d, \quad (15)$$

where

- (a)  $\pi_1(\cdot), \dots, \pi_d(\cdot)$  are permutations of the integers  $\{1, \dots, k\}$  that are randomly sampled with replacement from the set of  $k!$  such permutations, with  $\pi_j(i)$  denoting the  $i$ th element in the  $j$ th randomly sampled permutation; and
- (b)  $\{U_{ij}^* : j = 1, \dots, d, i = 1, \dots, k\}$  are random numbers sampled independently of each other and of the permutations  $\pi_1(\cdot), \dots, \pi_d(\cdot)$ .

We let  $G_{\text{LH}}^{(k)}$  denote the distribution of each  $k$ -dimensional column vector of input random numbers generated in this way so that

$$\mathcal{U}_j \sim G_{\text{LH}}^{(k)} \quad \text{if } \mathcal{U}_j = [U_j^{(1)}, \dots, U_j^{(k)}]^\text{T} \text{ is generated according to (15).}$$

The key property of LHS is that for each  $j$  ( $j = 1, \dots, d$ ), the components of the column vector  $\mathcal{U}_j$  form a stratified sample of size  $k$  from the uniform distribution on the unit interval  $(0, 1)$  such that there is a single observation in each stratum and the observations within the sample are negatively quadrant dependent; moreover, different stratified samples of size  $k$  are independent. Since  $\pi_j(\cdot)$  is a random permutation of the integers  $\{1, \dots, k\}$ , each element  $\pi_j(i)$  for  $i = 1, \dots, k$  has the discrete uniform distribution on the set  $\{1, \dots, k\}$ ; and thus in the definition (15), the variate  $\pi_j(i)$  randomly indexes a subinterval (stratum) of the form

$\left((\ell-1)/k, \ell/k\right]$  for some  $\ell \in \{1, \dots, k\}$ . Since  $U_{ij}^*$  is a random number sampled independently of  $\pi_j(i)$ , we see that  $U_j^{(i)}$  is uniformly distributed in the subinterval indexed by  $\pi_j(i)$ ; and it follows that  $U_j^{(i)}$  is uniformly distributed on the unit interval  $(0, 1)$ . Moreover, since  $\pi_j(\cdot)$  is a permutation of  $\{1, \dots, k\}$ , every subinterval (stratum) of the form  $\left((\ell-1)/k, \ell/k\right]$  for  $\ell = 1, \dots, k$  contains exactly one of the negatively quadrant dependent random numbers  $\{U_j^{(i)} : i = 1, \dots, k\}$  so that the components of  $\mathcal{U}_j$  constitute a stratified sample of the uniform distribution on  $(0, 1)$ . Finally, we notice that the column vectors  $\mathcal{U}_1, \dots, \mathcal{U}_d$  are independent since the random permutations  $\{\pi_j(\cdot) : j = 1, \dots, d\}$  and the random numbers  $\{U_{ij}^* : i = 1, \dots, k; j = 1, \dots, d\}$  are all generated independently.

McKay, Beckman, and Conover (1979) invented LHS and showed that if  $y(\cdot)$  is a monotone function of each of its arguments, then  $\text{Var}[\hat{\theta}_{\text{CI}}(G_{\text{LH}}^{(k)}, n)] \leq \text{Var}[\bar{Y}(n)]$ . A slightly simpler proof of their result can be obtained by showing that  $G_{\text{LH}}^{(k)} \in \mathcal{G}$  and then applying Result 1. These authors assumed that the input random variates are mutually independent and that each input random variate is generated by inversion; and in follow-up work on the large-sample properties of LHS, Stein (1987) and Owen (1992) adopted the same assumptions.

We define a more general LHS estimator in which a specified subset  $\{V_j : j \in J\}$  of the input random variates must be sampled by inversion, where  $J$  is a (possibly empty) subset of  $\{1, \dots, p\}$ ; and *any* valid sampling scheme may be used to generate the remaining input variates  $\{V_j : j \in \{1, \dots, p\} - J\}$ . Suppose that for a given index set  $J \subseteq \{1, \dots, p\}$ , the following independence properties hold:

IP<sub>1</sub> The input random variates  $\{V_j : j \in J\}$  are mutually independent.

IP<sub>2</sub> The vectors  $(V_j : j \in J)$  and  $(V_j : j \in J')$  are independent, where  $J' \equiv \{1, \dots, p\} - J$ .

Given an index set  $J$  with properties IP<sub>1</sub> and IP<sub>2</sub>, we define the *Latin hypercube sampling* estimator  $\hat{\theta}_{\text{LH}}(\mathcal{H}_J, n)$  to be the correlation-induction estimator  $\hat{\theta}_{\text{CI}}(G_{\text{LH}}^{(n)}, n)$  based on a sampling plan  $\mathcal{H}_J$  of the form

$$V_j = \begin{cases} F_j^{-1}(U_j), & j \in J, \\ H_j(U_\tau : \tau \in \{1, \dots, d\} - J), & j \in J', \end{cases} \quad (16)$$

where  $F_j^{-1}(\cdot)$  is the inverse c.d.f. of  $V_j$  for  $j \in J$ ; and the remaining part of the sampling plan consists of functions  $\{H_j(\cdot) : j \in J'\}$  that are selected by the user to yield the correct joint distribution for the random vector  $(V_j : j \in J')$ .

If the input random variates are independent with readily evaluated inverse c.d.f.'s, then we can take  $J = J_p \equiv \{1, \dots, p\}$  (and  $J' = \emptyset$ ) in (16) to obtain the sampling plan  $\mathcal{H}_{J_p}$ ;

and the resulting estimator  $\hat{\theta}_{\text{LH}}(\mathcal{H}_{J_p}, n)$  is the usual Latin hypercube sampling estimator studied by McKay, Beckman, and Conover (1979), Stein (1987), and Owen (1992). On the other hand, irrespective of the joint distribution of the input random variates we can always take  $J = \emptyset$  (and  $J' = J_p$ ) in (16) to obtain the sampling plan  $\mathcal{H}_\emptyset$ ; and this yields the most general formulation of Latin hypercube sampling considered in this paper. Clearly for any  $J \subseteq \{1, \dots, p\}$ , the estimator  $\hat{\theta}_{\text{LH}}(\mathcal{H}_J, n)$  is a special case of the estimator  $\hat{\theta}_{\text{LH}}(\mathcal{H}_\emptyset, n)$ .

### 1.3. Control Variates (CV)

Suppose we can identify a  $1 \times q$  vector of concomitant random variables  $\mathbf{C} = (C_1, \dots, C_q)$  that are generated by the simulation and that have a known, finite expectation  $\boldsymbol{\mu}_{\mathbf{C}} \equiv \mathbb{E}[\mathbf{C}]$  as well as a strong linear association with  $Y$ . When using the method of control variates, we try to predict the unknown deviation  $Y - \theta$  as a linear combination of the known deviation  $\mathbf{C} - \boldsymbol{\mu}_{\mathbf{C}}$  in order to adjust the response accordingly; thus for an appropriate  $1 \times q$  vector  $\mathbf{b}$  of control coefficients, we have the controlled response

$$Y_{\text{CV}} \equiv Y - \mathbf{b}(\mathbf{C} - \boldsymbol{\mu}_{\mathbf{C}})^{\text{T}}.$$

If  $\mathbf{b}$  is constant, then  $Y_{\text{CV}}$  is an unbiased estimator of  $\theta$ . Let  $\boldsymbol{\sigma}_{Y\mathbf{C}} \equiv \text{Cov}(Y, \mathbf{c})$  denote the  $1 \times q$  vector of covariances  $[\text{Cov}(Y, C_1), \dots, \text{Cov}(Y, C_q)]$  and let  $\boldsymbol{\Sigma}_{\mathbf{C}} \equiv \text{Var}(\mathbf{C})$  denote the  $q \times q$  variance-covariance matrix of  $\mathbf{C}$ , where we assume that  $\boldsymbol{\Sigma}_{\mathbf{C}}$  is positive definite.

The variance of  $Y_{\text{CV}}$  is minimized by the optimal control coefficient vector

$$\mathbf{b}^* = \boldsymbol{\sigma}_{Y\mathbf{C}} \boldsymbol{\Sigma}_{\mathbf{C}}^{-1} \quad (17)$$

(Lavenberg, Moeller, and Welch 1982). Although in some applications  $\boldsymbol{\Sigma}_{\mathbf{C}}$  may be known,  $\boldsymbol{\sigma}_{Y\mathbf{C}}$  is almost always unknown, and therefore  $\mathbf{b}^*$  must be estimated. Suppose we have available  $n$  i.i.d. observations  $\{(Y_i, \mathbf{C}_i) : i = 1, \dots, n\}$ . The most commonly used control coefficient vector is the sample analog of  $\mathbf{b}^*$ ,

$$\hat{\mathbf{b}} = \mathbf{S}_{Y\mathbf{C}} \mathbf{S}_{\mathbf{C}}^{-1},$$

where  $\mathbf{S}_{Y\mathbf{C}}$  is the  $1 \times q$  vector of sample covariances  $[\widehat{\text{Cov}}(Y, C_1), \dots, \widehat{\text{Cov}}(Y, C_q)]$  and  $\mathbf{S}_{\mathbf{C}}$  is the sample variance-covariance matrix of  $\mathbf{C}$ .

The *control-variates* (CV) estimator based on the sample  $\{(Y_i, \mathbf{C}_i) : i = 1, \dots, n\}$  is then defined as

$$\hat{\theta}_{\text{CV}}(n) \equiv \bar{Y} - \hat{\mathbf{b}}(\bar{\mathbf{C}} - \boldsymbol{\mu}_{\mathbf{C}})^{\text{T}},$$

where  $\bar{Y}$  and  $\bar{\mathbf{C}}$  are the sample means of  $\{Y_i : i = 1, \dots, n\}$  and  $\{\mathbf{C}_i : i = 1, \dots, n\}$ , respectively. If  $(Y, \mathbf{C})$  has a multivariate normal distribution, then  $\hat{\theta}_{\text{CV}}(n)$  is an unbiased estimator of  $\theta$  with variance

$$\text{Var}[\hat{\theta}_{\text{CV}}(n)] = n^{-1} \sigma_Y^2 (1 - R_{Y\mathbf{C}}^2) \frac{n-2}{n-q-2},$$

where  $R_{Y\mathbf{C}}^2 = \boldsymbol{\sigma}_{Y\mathbf{C}} \boldsymbol{\Sigma}_{\mathbf{C}}^{-1} \boldsymbol{\sigma}_{Y\mathbf{C}}^T / \sigma_Y^2$  is the squared coefficient of multiple correlation between  $Y$  and  $\mathbf{C}$  (Lavenberg, Moeller, and Welch 1982). Without some additional assumptions about the distribution of  $(Y, \mathbf{C})$ , there is no guarantee that  $\hat{\theta}_{\text{CV}}(n)$  is unbiased or that it has smaller MSE or smaller variance than  $\bar{Y}(n)$ . However, Nelson (1990) pointed out that irrespective of the distribution of  $(Y, \mathbf{C})$ , we have the following central limit theorem for the method of control variates:

$$n^{1/2} [\hat{\theta}_{\text{CV}}(n) - \theta] \xrightarrow{\mathcal{D}} N[0, \sigma_Y^2 (1 - R_{Y\mathbf{C}}^2)] \quad \text{as } n \rightarrow \infty. \quad (18)$$

Thus  $\hat{\theta}_{\text{CV}}(n)$  asymptotically dominates  $\bar{Y}(n)$ . Observe that the asymptotic behavior of  $\hat{\theta}_{\text{CV}}(n)$  depends on the joint distribution of  $Y$  and  $\mathbf{C}$  only through  $R_{Y\mathbf{C}}$ . The asymptotic property (18) is the main guarantee when using the method of control variates, since in some applications the normality assumption is not even approximately satisfied. Tew and Wilson (1992, pp. 91–92) describe a practical method for checking the assumption of multivariate normality. Nelson (1990) provides a comprehensive discussion of remedies for the problems arising in applications of the method of control variates. See also Avramidis and Wilson (1993a).

## 2. INTEGRATED VARIANCE REDUCTION STRATEGIES

Building on the individual VRTs reviewed in Section 1, we formulate and analyze integrated variance reduction strategies that are based on joint application of the following pairs of individual VRTs: (a) conditional expectation and correlation induction (Section 2.1); (b) correlation induction and control variates (Section 2.2); and (c) conditional expectation and control variates (Section 2.3).

### 2.1. Conditional Expectation and Correlation Induction (CE+CI)

We begin by expressing the conditioning vector  $\mathbf{X}$  as a function of the input random numbers,

$$\mathbf{X} = x(U_j : j \in I_{\mathbf{X}}) \quad \text{for some } I_{\mathbf{X}} \subseteq \{1, \dots, d\}, \quad (19)$$

where  $I_{\mathbf{X}}$  is the set of indices of the random numbers on which  $\mathbf{X}$  depends. As seen in Section 1.1, the random variable  $Z \equiv E[Y|\mathbf{X}] \equiv \zeta(\mathbf{X})$  is another unbiased estimator of  $\theta$  based on a single replication of the simulation; and the variance of  $Z$  does not exceed the variance of  $Y$ . Thus we may view  $Z$  as the new response of interest, and we seek an even more precise estimator by applying the technique of correlation induction to the new response. For this purpose, we express  $Z$  as a function of the input random numbers,

$$Z = \zeta(\mathbf{X}) = \zeta[x(U_j : j \in I_{\mathbf{X}})] \equiv z(U_j : j \in I_{\mathbf{X}}). \quad (20)$$

**Example 1 (continued).** In Section 1.1 we took  $\mathbf{X} = (V_1, V_5)$ . In view of (2), we have  $I_{\mathbf{X}} = \{1, 2, 6\}$ . Moreover, (2) and (5) imply

$$\begin{aligned} Z &= z(U_1, U_2, U_6) \\ &\equiv F_{2,3}[t - H_1(U_1, U_2), t - H_1(U_1, U_2) - H_5(U_6)] \cdot F_4[t - H_5(U_6)]. \quad \blacksquare \end{aligned}$$

Starting from the new response (20), we formulate the *conditional expectation–correlation induction* (CE+CI) estimator  $\hat{\theta}_{\text{CE+CI}}(G^{(k)}, n)$  based on the distribution  $G^{(k)}$  and  $n$  simulation runs. Specifically,  $\hat{\theta}_{\text{CE+CI}}(G^{(k)}, n)$  is obtained by averaging  $m = n/k$  i.i.d. replications of the statistic  $Z_{\text{CI}}(G^{(k)}, I_{\mathbf{X}})$ ; and to obtain a single observation of  $Z_{\text{CI}}(G^{(k)}, I_{\mathbf{X}})$ , we average  $k$  negatively correlated responses of the form (20), where the random numbers with indices in  $I_{\mathbf{X}}$  are used for correlation induction. Notice that the random numbers with indices in  $\{1, \dots, d\} - I_{\mathbf{X}}$  need not be sampled, since  $Z$  does not depend on them. In terms of the notation in (6) and (8), the CE+CI estimator is defined formally as

$$\hat{\theta}_{\text{CE+CI}}(G^{(k)}, n) \equiv m^{-1} \sum_{\ell=1}^m \tilde{Z}_{\ell} \quad \text{where } m = n/k \quad \text{and} \quad \{\tilde{Z}_{\ell}\}_{\ell=1}^m \stackrel{\text{i.i.d.}}{\sim} Z_{\text{CI}}(G^{(k)}, I_{\mathbf{X}}). \quad (21)$$

It follows immediately from (9) and (10) that for any  $G^{(k)} \in \mathcal{G}$ , the statistic  $\hat{\theta}_{\text{CE+CI}}(G^{(k)}, n)$  is an unbiased estimator of  $\theta$ , and

$$\text{Var}[\hat{\theta}_{\text{CE+CI}}(G^{(k)}, n)] = n^{-1} \sigma_Z^2 [1 + (k-1) \rho_Z(G^{(k)}, I_{\mathbf{X}})],$$

where  $\rho_Z(G^{(k)}, I_{\mathbf{X}})$  is defined as in (11). If  $z(\cdot)$  is a monotone function of each of its arguments, then  $\rho_Z(G^{(k)}, I_{\mathbf{X}}) \leq 0$  by Result 1 so that  $\text{Var}[\hat{\theta}_{\text{CE+CI}}(G^{(k)}, n)] \leq \text{Var}[\hat{\theta}_{\text{CE}}(n)]$ .

It is also interesting to compare  $\hat{\theta}_{\text{CE+CI}}(G^{(k)}, n)$  with  $\hat{\theta}_{\text{CI}}(G^{(k)}, n)$ , the pure correlation-induction estimator based on the same distribution  $G^{(k)}$  for inducing correlation. To carry out this comparison, we assume that (a) the conditioning vector  $\mathbf{X}$  is a subvector of the full



vector  $\mathbf{V}$  of input random variates so that with an appropriate reindexing of the components of  $\mathbf{V}$ , we may write

$$\mathbf{V} = (\mathbf{V}_*, \mathbf{V}_{**}) = (\mathbf{X}, \mathbf{V}_{**});$$

and (b) the conditioning vector  $\mathbf{X}$  is independent of the vector  $\mathbf{V}_{**}$  of remaining input random variates. Formally this assumption about the form of  $\mathbf{X}$  is stated as

$$\left. \begin{array}{l} \mathbf{X} \equiv (V_j : j \in J_{\mathbf{X}}) \text{ and } \mathbf{V}_{**} \equiv (V_j : j \in J'_{\mathbf{X}}) \text{ are independent,} \\ \text{where } J_{\mathbf{X}} \subset \{1, \dots, p\}, J_{\mathbf{X}} \neq \emptyset, \text{ and } J'_{\mathbf{X}} \equiv \{1, \dots, p\} - J_{\mathbf{X}} \neq \emptyset. \end{array} \right\} \quad (22)$$

It follows from Lemma 2.7 of Whitt (1976) that the full vector  $(\mathbf{X}, \mathbf{V}_{**})$  of input random variates could in principle be sampled as a function of a *single* random number; and under such a sampling scheme, we could not make the desired comparison between the CE+CI and CI estimators. Throughout this paper we assume that when (22) holds, the vectors  $\mathbf{X}$  and  $\mathbf{V}_{**}$  are generated as functions of *disjoint* sets of the random-number inputs so that the sampling plan has the form

$$\left. \begin{array}{l} \mathbf{X} = x(U_j : j \in I_{\mathbf{X}}) \text{ and } \mathbf{V}_{**} = v_{**}(U_j : j \in I'_{\mathbf{X}}), \text{ where} \\ I_{\mathbf{X}} \subset \{1, \dots, d\}, I_{\mathbf{X}} \neq \emptyset, \text{ and } I'_{\mathbf{X}} \equiv \{1, \dots, d\} - I_{\mathbf{X}} \neq \emptyset. \end{array} \right\} \quad (23)$$

Condition (23) is a reasonable assumption about the form of the sampling plan—in fact, (23) is a natural approach to generating the random vectors  $\mathbf{X}$  and  $\mathbf{V}_{**}$  under assumption (22). Additional discussion of the significance of assumption (22) is given after the statement of Theorem 1 below.

We compare the estimators  $\hat{\theta}_{\text{CI}}(G^{(k)}, n)$  and  $\hat{\theta}_{\text{CE+CI}}(G^{(k)}, n)$  by analyzing the behavior of all relevant stochastic quantities when we generate the pure correlation-induction estimator  $\hat{\theta}_{\text{CI}}(G^{(k)}, n)$ . For  $i = 1, \dots, k$ , let  $Y^{(i)}$ ,  $\mathbf{X}^{(i)}$ , and  $\mathbf{V}_{**}^{(i)}$  respectively denote the  $i$ th correlated replication of  $Y$ ,  $\mathbf{X}$ , and  $\mathbf{V}_{**}$  when we use the general correlation-induction scheme (6)–(7) with  $W = Y$  and  $L_W = I_W = \{1, \dots, d\}$ . Let  $\Xi$  denote the matrix of overall conditioning information,

$$\Xi \equiv \left[ \left\{ \mathbf{X}^{(1)} \right\}^T, \dots, \left\{ \mathbf{X}^{(k)} \right\}^T \right]^T.$$

Exploiting the properties of this scheme for generating  $\hat{\theta}_{\text{CI}}(G^{(k)}, n)$ , we will show that

$$\mathbb{E} \left[ Y_{\text{CI}}(G^{(k)}, \{1, \dots, d\}) \mid \Xi \right] = k^{-1} \sum_{i=1}^k \mathbb{E} \left[ f(\mathbf{X}^{(i)}, \mathbf{V}_{**}^{(i)}) \mid \Xi \right] \quad (24)$$

$$= k^{-1} \sum_{i=1}^k \mathbb{E} \left[ f(\mathbf{X}^{(i)}, \mathbf{V}_{**}^{(i)}) \mid \mathbf{X}^{(i)} \right] \quad (25)$$

$$= k^{-1} \sum_{i=1}^k \zeta[\mathbf{X}^{(i)}] \quad (26)$$

$$= k^{-1} \sum_{i=1}^k z \left[ U_j^{(i)} : j \in I_{\mathbf{X}} \right] \quad (27)$$

$$\sim Z_{\text{CI}}(G^{(k)}, I_{\mathbf{X}}) . \quad (28)$$

Equation (24) follows directly from the assumption (22) about the form of the full vector of input random variates; and the key to the rest of this argument is the observation that  $\mathbf{V}_{**}^{(i)}$  depends only on the column vectors  $\{\mathcal{U}_j : j \in I'_{\mathbf{X}}\}$ , while  $\Xi$  depends only on the column vectors  $\{\mathcal{U}_j : j \in I_{\mathbf{X}}\}$ . Sampling condition  $\text{SC}_3$  prescribes that the column vectors  $\{\mathcal{U}_j : j = 1, \dots, d\}$  defined by (7) are sampled independently; and thus assumption (23) about the form of the sampling plan implies that  $\mathbf{V}_{**}^{(i)}$  and  $\Xi$  are independent. Therefore the conditional distribution of  $[\mathbf{X}^{(i)}, \mathbf{V}_{**}^{(i)}]$  is the same whether we condition on  $\Xi$  or  $\mathbf{X}^{(i)}$ ; and (25) follows immediately. Display (26) follows from assumption (22) and the definition of  $\zeta(\cdot)$ ; and display (27) follows from the definition (20). Now we notice that the same joint distribution for the correlated replications  $\{\mathbf{X}^{(i)} : i = 1, \dots, k\}$  would be obtained by taking  $W = \mathbf{X}$  and  $L_W = I_W = I_{\mathbf{X}}$  in the general correlation-induction scheme (6)–(7); and this latter approach is used to generate the statistic  $Z_{\text{CI}}(G^{(k)}, I_{\mathbf{X}})$  on which the conditional expectation–correlation induction estimator  $\hat{\theta}_{\text{CE+CI}}(G^{(k)}, n)$  is based. Thus relation (28) follows immediately.

To complete the comparison of  $\hat{\theta}_{\text{CI}}(G^{(k)}, n)$  and  $\hat{\theta}_{\text{CE+CI}}(G^{(k)}, n)$ , we apply the conditional variance relation (4) to  $Y_{\text{CI}}(G^{(k)}, \{1, \dots, d\})$  when conditioning on  $\Xi$ ; and then we see that

$$\text{Var} \left[ Y_{\text{CI}}(G^{(k)}, \{1, \dots, d\}) \right] \geq \text{Var} \left[ \mathbb{E} \left[ Y_{\text{CI}}(G^{(k)}, \{1, \dots, d\}) \mid \Xi \right] \right] = \text{Var} \left[ Z_{\text{CI}}(G^{(k)}, I_{\mathbf{X}}) \right], \quad (29)$$

where the last equality follows from (28). Display (29) together with the definitions of  $\hat{\theta}_{\text{CI}}(G^{(k)}, n)$  and  $\hat{\theta}_{\text{CE+CI}}(G^{(k)}, n)$  given respectively by (12) and (21) imply that

$$\text{Var} \left[ \hat{\theta}_{\text{CE+CI}}(G^{(k)}, n) \right] \leq \text{Var} \left[ \hat{\theta}_{\text{CI}}(G^{(k)}, n) \right], \quad (30)$$

and equality holds in (30) if and only if  $Y_{\text{CI}}(G^{(k)}, \{1, \dots, d\})$  is a function of  $\Xi$  alone. Notice that if the original simulation response  $Y$  depends only on the conditioning vector  $\mathbf{X}$  and not on the vector  $\mathbf{V}_{**}$  of remaining input random variates, then it follows immediately that  $Y_{\text{CI}}(G^{(k)}, \{1, \dots, d\})$  is a function of  $\Xi$  alone so that equality holds in (30).

The preceding development formalizes the intuitive result that “solving the problem analytically” on a subset of the random-number inputs and using a correlation-induction technique on the rest of the random-number inputs is at least as good a variance-reduction strategy as using the same correlation-induction technique on all of the random-number inputs. We summarize this discussion in the following result.

**Theorem 1.** *Suppose that  $G^{(k)} \in \mathcal{G}$ . Then  $\hat{\theta}_{\text{CE+CI}}(G^{(k)}, n)$  is an unbiased estimator of  $\theta$ . If the conditioning vector  $\mathbf{X}$  has the form (22) and if the sampling plan has the form (23), then*

$$\text{Var}[\hat{\theta}_{\text{CE+CI}}(G^{(k)}, n)] \leq \text{Var}[\hat{\theta}_{\text{CI}}(G^{(k)}, n)],$$

*with equality holding if and only if  $Y_{\text{CI}}(G^{(k)}, \{1, \dots, d\})$  is a function of  $\Xi$  alone. Moreover, irrespective of the form of  $\mathbf{X}$ , if  $z(\cdot)$  is a monotone function of each of its arguments, then*

$$\text{Var}[\hat{\theta}_{\text{CE+CI}}(G^{(k)}, n)] \leq \text{Var}[\hat{\theta}_{\text{CE}}(n)].$$

Some additional comments should be made about Theorem 1. The term “input random vector” is commonly interpreted to mean a vector of fundamental simulation inputs that are directly observable in the real system and have immediate physical meaning. Under this interpretation, the assumption (22) about the form of the input random vector  $\mathbf{V}$  in terms of the conditioning vector  $\mathbf{X}$  may appear to be unduly restrictive. However as discussed in the first paragraph of Section 1, our definition of the term “input random vector” is more general than the common interpretation of this term—an input random vector is *any* random vector  $\mathbf{V}$  observed in the simulation such that the response  $Y$  can be expressed as a function of  $\mathbf{V}$  alone. To clarify this point, suppose that the activity connecting nodes 1 and 2 in Example 1 is composed of two basic tasks that must be performed sequentially. Let  $V_{1,1}$  and  $V_{1,2}$  respectively denote the random durations of these two tasks. According to the usual interpretation, the input random vector is  $(V_{1,1}, V_{1,2}, V_2, V_3, V_4, V_5)$ . Now set  $V_1 \equiv V_{1,1} + V_{1,2}$  and let  $\mathbf{V} \equiv (V_1, V_2, V_3, V_4, V_5)$ . According to our interpretation,  $\mathbf{V}$  is a valid input random vector, since the longest directed path from node 1 to node 4 can be expressed as a function of  $\mathbf{V}$  alone according to (1).

Some comments should also be made about the monotonicity assumption in Theorem 1. For a general conditioning vector  $\mathbf{X}$  not necessarily having the form (22), no monotonicity properties are guaranteed for the conditional-expectation function  $\zeta(\cdot)$  even if the original response function  $f(\cdot)$  is monotone in each of its arguments. However, if  $\mathbf{X}$  has the form (22) and  $f(\cdot)$  is a monotone function of each component of  $\mathbf{X}$ , then  $\zeta(\cdot)$  is a monotone function of each of its arguments. In the latter situation, it is desirable to use a monotone sampling scheme  $x(\cdot)$  to generate  $\mathbf{X}$  because this will guarantee that  $z(\cdot)$  is a monotone function of each of its arguments; and as shown in Theorem 1, the integrated CE+CI strategy will then be superior to both the CE and CI techniques.

Finally, we point out that Carson (1985) obtained a result similar to Theorem 1 for the special case in which  $G^{(k)} = G_{\text{AV}}^{(2)}$  and  $Z$  is a specific conditional-expectation estimator due to Burt and Garman (1971).

## 2.2. Correlation Induction and Control Variates (CI+CV)

Our approach to the joint application of the methods of correlation induction and control variates is based on the observation that the control vector  $\mathbf{C}$  usually depends only on a proper subset  $\{U_j : j \in I_{\mathbf{C}}\}$  of the input random numbers  $\{U_j : j = 1, \dots, d\}$ , so that we may write

$$\mathbf{C} = c(U_j : j \in I_{\mathbf{C}}) \text{ for some } I_{\mathbf{C}} \subset \{1, \dots, d\}, \text{ where } I'_{\mathbf{C}} \equiv \{1, \dots, d\} - I_{\mathbf{C}} \neq \emptyset. \quad (31)$$

**Example 1 (continued).** Suppose that the control variate of interest is  $\mathbf{C} = V_4 + V_5$ , the duration of the third path from node 1 to node 4 in the definition (1) of  $T$ . The sampling plan (2) implies

$$c(u_5, u_6) = H_4(u_5) + H_5(u_6)$$

so that  $I_{\mathbf{C}} = \{5, 6\}$  and  $I'_{\mathbf{C}} = \{1, 2, 3, 4\}$ . ■

Our development is in the same spirit as the approach of Tew and Wilson (1994) for integrating the Schruben-Margolin strategy with the method of control variates. The key idea is to induce the desired negative correlation between the responses by sampling dependently between replications only the input random numbers that do not affect the control vector, thus preserving the dependency structure between the response and the control vector on each simulation run.

Given an arbitrary  $k$ -dimensional distribution  $G^{(k)} \in \mathcal{G}$ , we perform  $k$  dependent replications of the simulation using the distribution  $G^{(k)}$  to sample the random numbers with indices in  $I'_{\mathbf{C}}$ . The random numbers with indices in  $I_{\mathbf{C}}$  are sampled independently according to  $G_{\text{IR}}^{(k)}$ . Following the notation in (8), we define the auxiliary quantities

$$\tilde{Y} \equiv Y_{\text{CI}}(G^{(k)}, I'_{\mathbf{C}}) \text{ and } \tilde{\mathbf{C}} \equiv [C_{1,\text{CI}}(G^{(k)}, I'_{\mathbf{C}}), \dots, C_{q,\text{CI}}(G^{(k)}, I'_{\mathbf{C}})]. \quad (32)$$

Since all the random numbers that affect the control vector are sampled independently according to  $G_{\text{IR}}^{(k)}$ , it is clear that  $\tilde{\mathbf{C}}$  is the average of  $k$  independent replications of  $\mathbf{C}$ , although the subscript  $\text{CI}$  appended to each of the components of  $\tilde{\mathbf{C}}$  in (32) might suggest that correlation induction is applied to the control vector. Thus

$$\mathbb{E}[\tilde{\mathbf{C}}] = \boldsymbol{\mu}_{\mathbf{C}} \text{ and } \text{Var}[\tilde{\mathbf{C}}] \equiv \boldsymbol{\Sigma}_{\tilde{\mathbf{C}}} = k^{-1} \boldsymbol{\Sigma}_{\mathbf{C}}. \quad (33)$$

On the other hand,  $\tilde{Y}$  is the average of  $k$  dependent replications of  $Y$ . To simplify the notation throughout the rest of this section, we will let

$$\rho_Y \equiv \rho_Y(G^{(k)}, I'_{\mathbf{C}})$$

denote the induced correlation between any pair of replications of the response  $Y$ . By (9) and (10), we have

$$\mathbb{E}[\tilde{Y}] = \theta \quad \text{and} \quad \text{Var}[\tilde{Y}] \equiv \sigma_{\tilde{Y}}^2 = k^{-1}\sigma_Y^2[1 + (k-1)\rho_Y]. \quad (34)$$

We view the statistics  $\tilde{Y}$  and  $\tilde{\mathbf{C}}$  as an aggregated response and an aggregated control vector, respectively; and we use the control-variates technique to further reduce the variance of  $\tilde{Y}$ . Let  $m = n/k$  and let

$$\{(\tilde{Y}_\ell, \tilde{\mathbf{C}}_\ell)\}_{\ell=1}^m \stackrel{\text{i.i.d.}}{\sim} (\tilde{Y}, \tilde{\mathbf{C}}) \quad \text{as in display (32)}. \quad (35)$$

We define  $\hat{\theta}_{\text{CI+CV}}(G^{(k)}, n)$ , the *correlation induction-control variates* (CI+CV) estimator based on the distribution  $G^{(k)}$  and  $n$  replications, as the control-variates estimator

$$\hat{\theta}_{\text{CI+CV}}(G^{(k)}, n) \equiv \bar{\tilde{Y}} - \hat{\mathbf{b}} \left( \bar{\tilde{\mathbf{C}}} - \boldsymbol{\mu}_{\mathbf{C}} \right)^{\text{T}},$$

where:  $\bar{\tilde{Y}}$  and  $\bar{\tilde{\mathbf{C}}}$  are the sample means of  $\{\tilde{Y}_\ell : \ell = 1, \dots, m\}$  and  $\{\tilde{\mathbf{C}}_\ell : \ell = 1, \dots, m\}$ , respectively;

$$\hat{\mathbf{b}} = \mathbf{S}_{\tilde{Y}\tilde{\mathbf{C}}} \mathbf{S}_{\tilde{\mathbf{C}}}^{-1};$$

$\mathbf{S}_{\tilde{Y}\tilde{\mathbf{C}}}$  denotes the  $1 \times q$  vector of sample covariances between  $\tilde{Y}$  and the components of  $\tilde{\mathbf{C}}$  in (32); and  $\mathbf{S}_{\tilde{\mathbf{C}}}$  denotes the  $q \times q$  sample variance-covariance matrix of  $\tilde{\mathbf{C}}$ . Notice that the sample covariances  $\mathbf{S}_{\tilde{Y}\tilde{\mathbf{C}}}$  and  $\mathbf{S}_{\tilde{\mathbf{C}}}$  (and thus the estimated control coefficient vector  $\hat{\mathbf{b}}$ ) are based on the random sample (35) of size  $m = n/k$  rather than the original sample of size  $n$  used in the conventional control-variates method.

To compute the mean and variance of  $\hat{\theta}_{\text{CI+CV}}(G^{(k)}, n)$ , we assume that  $(\tilde{Y}, \tilde{\mathbf{C}})$  has a multivariate normal distribution. Then  $\hat{\theta}_{\text{CI+CV}}(G^{(k)}, n)$  is an unbiased estimator of  $\theta$ , and its variance is

$$\begin{aligned} \text{Var}[\hat{\theta}_{\text{CI+CV}}(G^{(k)}, n)] &= m^{-1}\sigma_{\tilde{Y}}^2 \left(1 - R_{\tilde{Y}\tilde{\mathbf{C}}}^2\right) \frac{m-2}{m-q-2} \\ &= n^{-1}\sigma_Y^2[1 + (k-1)\rho_Y] \left(1 - R_{\tilde{Y}\tilde{\mathbf{C}}}^2\right) \frac{n-2k}{n-(q+2)k}, \end{aligned} \quad (36)$$

where  $R_{\tilde{Y}\tilde{\mathbf{C}}}$  is the coefficient of multiple correlation between  $\tilde{Y}$  and  $\tilde{\mathbf{C}}$ . To express  $R_{\tilde{Y}\tilde{\mathbf{C}}}$  as a function of  $R_{Y\mathbf{C}}$  and  $\rho_Y$ , we observe that

$$\boldsymbol{\sigma}_{\tilde{Y}\tilde{\mathbf{C}}} \equiv \text{Cov}(\tilde{Y}, \tilde{\mathbf{C}}) = \mathbb{E}[\tilde{Y}\tilde{\mathbf{C}}] - \mathbb{E}[\tilde{Y}]\mathbb{E}[\tilde{\mathbf{C}}]$$

$$\begin{aligned}
&= k^{-2} \left\{ \mathbb{E} \left[ \sum_{i=1}^k Y^{(i)} \sum_{\ell=1}^k \mathbf{C}^{(\ell)} \right] \right\} - \theta \boldsymbol{\mu}_{\mathbf{C}} \\
&= k^{-2} \left\{ k \mathbb{E} \left[ Y^{(1)} \mathbf{C}^{(1)} \right] + k(k-1) \theta \boldsymbol{\mu}_{\mathbf{C}} \right\} - \theta \boldsymbol{\mu}_{\mathbf{C}} \\
&= k^{-1} \boldsymbol{\sigma}_{Y\mathbf{C}}.
\end{aligned} \tag{37}$$

Using (33), (34), and (37), we have

$$R_{\tilde{Y}\tilde{\mathbf{C}}}^2 = \frac{\boldsymbol{\sigma}_{\tilde{Y}\tilde{\mathbf{C}}} \boldsymbol{\Sigma}_{\tilde{\mathbf{C}}}^{-1} \boldsymbol{\sigma}_{\tilde{Y}\tilde{\mathbf{C}}}^T}{\sigma_{\tilde{Y}}^2} = \frac{\boldsymbol{\sigma}_{Y\mathbf{C}} \boldsymbol{\Sigma}_{\mathbf{C}}^{-1} \boldsymbol{\sigma}_{Y\mathbf{C}}^T}{\sigma_Y^2 [1 + (k-1)\rho_Y]} = \frac{R_{Y\mathbf{C}}^2}{1 + (k-1)\rho_Y}.$$

Substituting this last result into (36), we obtain

$$\text{Var} \left[ \hat{\theta}_{\text{CI+CV}} \left( G^{(k)}, n \right) \right] = n^{-1} \sigma_Y^2 \left[ 1 + (k-1)\rho_Y - R_{Y\mathbf{C}}^2 \right] \frac{n-2k}{n - (q+2)k}. \tag{38}$$

Observe the additive effect of the two sources of variance reduction in equation (38): the application of correlation induction contributes the term  $(k-1)\rho_Y$ , and the application of control variates contributes the term  $-R_{Y\mathbf{C}}^2$ . The loss factor  $(n-2k)/[n - (q+2)k]$  is larger than the conventional control-variate loss factor  $(n-2)/(n-q-2)$  since in the CI+CV strategy the method of control variates is applied to a random sample of size  $m = n/k$  rather than to a sample of size  $n$ .

To compare the variance of  $\hat{\theta}_{\text{CI+CV}} \left( G^{(k)}, n \right)$  with the variance of the standard control-variate estimator  $\hat{\theta}_{\text{CV}}(n)$ , we also assume that  $(Y, \mathbf{C})$  has a multivariate normal distribution. Then the ratio of variances is

$$\frac{\text{Var} \left[ \hat{\theta}_{\text{CI+CV}} \left( G^{(k)}, n \right) \right]}{\text{Var} \left[ \hat{\theta}_{\text{CV}}(n) \right]} = \left[ 1 + \frac{(k-1)\rho_Y}{1 - R_{Y\mathbf{C}}^2} \right] \left[ \frac{(n-2k)(n - (q+2))}{(n-2)(n - (q+2)k)} \right]. \tag{39}$$

For a fixed  $G^{(k)}$ , the second factor in square brackets on the right-hand side of (39) converges to 1 as  $n \rightarrow \infty$ , so the first factor in square brackets becomes critical. If  $y(\cdot)$  is a monotone function of each argument whose index belongs to  $I'_{\mathbf{C}}$ , then  $\rho_Y \leq 0$  by Result 1; and if  $\rho_Y < 0$ , then  $\text{Var} \left[ \hat{\theta}_{\text{CI+CV}} \left( G^{(k)}, n \right) \right] < \text{Var} \left[ \hat{\theta}_{\text{CV}}(n) \right]$  for  $n$  sufficiently large.

To determine the asymptotic distribution of  $\hat{\theta}_{\text{CI+CV}} \left( G^{(k)}, n \right)$ , we relax the assumptions of joint normality for  $(Y, \mathbf{C})$  and  $(\tilde{Y}, \tilde{\mathbf{C}})$  made above. For a fixed  $k$ -dimensional distribution  $G^{(k)} \in \mathcal{G}$ , and irrespective of the distributions of  $(Y, \mathbf{C})$  and  $(\tilde{Y}, \tilde{\mathbf{C}})$ , we obtain the following central limit theorem analogous to (18) for the CI+CV estimator:

$$n^{1/2} \left[ \hat{\theta}_{\text{CI+CV}} \left( G^{(k)}, n \right) - \theta \right] \xrightarrow{\mathcal{D}} N \left\{ 0, \sigma_Y^2 \left[ 1 + (k-1)\rho_Y - R_{Y\mathbf{C}}^2 \right] \right\} \text{ as } n \rightarrow \infty.$$

Coupled with (18), this last result shows that for any  $G^{(k)} \in \mathcal{G}$ , the ratio  $\sigma_{\text{CI+CV}}^2/\sigma_{\text{CV}}^2$  of asymptotic variance parameters is given by the first factor in square brackets on the right-hand side of (39). We summarize this discussion in the following result.

**Theorem 2.** *Suppose that  $\mathbf{C}$  has the form (31),  $G^{(k)} \in \mathcal{G}$ , and  $y(\cdot)$  is a monotone function of each argument with index in  $I_{\mathbf{C}}'$ . If each of  $(Y, \mathbf{C})$  and  $(\tilde{Y}, \tilde{\mathbf{C}})$  has a multivariate normal distribution, then  $\hat{\theta}_{\text{CI+CV}}(G^{(k)}, n)$  is an unbiased estimator of  $\theta$ ; and ignoring terms of the form  $O(1/n)$ , we have*

$$\frac{\text{Var}[\hat{\theta}_{\text{CI+CV}}(G^{(k)}, n)]}{\text{Var}[\hat{\theta}_{\text{CV}}(n)]} = 1 + \frac{(k-1)\rho_Y}{1 - R_{Y\mathbf{C}}^2} \leq 1.$$

Moreover, irrespective of the distributions of  $(Y, \mathbf{C})$  and  $(\tilde{Y}, \tilde{\mathbf{C}})$ , the estimator  $\hat{\theta}_{\text{CI+CV}}$  asymptotically dominates  $\hat{\theta}_{\text{CV}}$ .

**Example 2.** We give a simple example in which  $(Y, \mathbf{C})$  and  $(\tilde{Y}, \tilde{\mathbf{C}})$  are both multivariate normal, and combining antithetic variates and control variates yields a particularly effective variance reduction strategy. Suppose that  $\mathbf{W} \sim N_{d-q}(\boldsymbol{\mu}_{\mathbf{W}}, \boldsymbol{\Sigma}_{\mathbf{W}})$  and  $\mathbf{C} \sim N_q(\boldsymbol{\mu}_{\mathbf{C}}, \boldsymbol{\Sigma}_{\mathbf{C}})$  are independent normal row vectors of dimension  $d-q$  and  $q$ , respectively ( $1 \leq q < d$ ). Moreover, suppose that the response of interest is  $Y = \boldsymbol{\alpha}\mathbf{W}^T + \boldsymbol{\beta}\mathbf{C}^T$ , where  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  are constant row vectors consisting respectively of  $d-q$  and  $q$  elements. Using the Cholesky decomposition  $\boldsymbol{\Sigma}_{\mathbf{W}} = \boldsymbol{\Theta}_{\mathbf{W}}\boldsymbol{\Theta}_{\mathbf{W}}^T$  (Anderson 1984, p. 586), we generate the random vector  $\mathbf{W}$  according to the scheme

$$\mathbf{W} = \boldsymbol{\mu}_{\mathbf{W}} + [\Phi^{-1}(U_1), \dots, \Phi^{-1}(U_{d-q})] \boldsymbol{\Theta}_{\mathbf{W}}^T,$$

where  $\Phi^{-1}(\cdot)$  is the inverse standard normal c.d.f. and the random numbers  $(U_1, \dots, U_{d-q})$  are independent. The control vector  $\mathbf{C}$  is generated by a similar scheme using the independently generated random numbers  $(U_{d-q+1}, \dots, U_d)$ . As a practical matter this scheme for generating  $\mathbf{W}$  and  $\mathbf{C}$  would involve using, for example, the *Applied Statistics* algorithm AS 111 (Griffiths and Hill 1985) to approximate  $\Phi^{-1}(\cdot)$ .

Since the standard normal distribution is symmetric about the origin, using the method of antithetic variates (13) to induce correlation between two replications of the vector  $(U_1, \dots, U_{d-q})$  yields  $\tilde{Y} = \boldsymbol{\alpha}\boldsymbol{\mu}_{\mathbf{W}}^T + \boldsymbol{\beta}\tilde{\mathbf{C}}^T$ . Since  $\tilde{\mathbf{C}}$  is the sum of independent normal vectors,  $\tilde{\mathbf{C}}$  is itself multivariate normal; and thus  $(\tilde{Y}, \tilde{\mathbf{C}})$  is (singular) multivariate normal, with  $R_{\tilde{Y}\tilde{\mathbf{C}}}^2 = 1$ . In this extreme example, the method of antithetic variates induces a linear relationship between  $\tilde{Y}$  and  $\tilde{\mathbf{C}}$ ; and thus the integrated AV+CV strategy achieves 100% variance reduction in the estimation of  $\theta$ . ■

Unfortunately, a variance comparison between  $\hat{\theta}_{\text{CI+CV}}(G^{(k)}, n)$  and  $\hat{\theta}_{\text{CI}}(G^{(k)}, n)$  is not possible in general. For  $\hat{\theta}_{\text{CI}}(G^{(k)}, n)$  we sample independently between replications all the random-number inputs, whereas for  $\hat{\theta}_{\text{CI+CV}}(G^{(k)}, n)$  we sample independently between replications only the random-number inputs with indices in  $I'_C$ ; and the extra degree of induced correlation achieved by  $\hat{\theta}_{\text{CI}}(G^{(k)}, n)$  might outweigh the benefit of using control variates to form the estimator  $\hat{\theta}_{\text{CI+CV}}(G^{(k)}, n)$ . For an elaboration of this point and relevant experimental results, see Kwon and Tew (1994).

### 2.3. Conditional Expectation and Control Variates (CE+CV)

To combine the methods of conditional expectation and control variates, we must select a control vector  $\mathbf{C}$  and a conditioning vector  $\mathbf{X}$  such that we can evaluate the conditional expectations  $\zeta(\mathbf{x}) \equiv \mathbb{E}[Y|\mathbf{X} = \mathbf{x}]$  and  $\delta(\mathbf{x}) \equiv \mathbb{E}[\mathbf{C}|\mathbf{X} = \mathbf{x}]$  analytically or numerically for every possible value of  $\mathbf{x}$ . We define the auxiliary random vectors  $Z \equiv \zeta(\mathbf{X})$  and  $\mathbf{D} \equiv \delta(\mathbf{X})$ , and we observe that  $\mathbb{E}[Z] = \theta$  and  $\mathbb{E}[\mathbf{D}] = \boldsymbol{\mu}_C$ .

**Example 1 (continued).** With the previously defined vectors  $\mathbf{X} \equiv (X_1, X_2) \equiv (V_1, V_5)$  and  $\mathbf{C} = V_4 + V_5$ , we have

$$\delta(x_1, x_2) = \mathbb{E}[V_4 + V_5 | X_1 \equiv V_1 = x_1, X_2 \equiv V_5 = x_2] = \mathbb{E}[V_4] + x_2$$

so that

$$\mathbf{D} = \delta(\mathbf{X}) = \mathbb{E}[V_4] + X_2 = \mathbb{E}[V_4] + V_5. \quad \blacksquare$$

We view  $Z$  and  $\mathbf{D}$  as the new response and control vector respectively, and we use the control-variates technique to reduce further the variance of  $Z$ . By analogy with the standard control-variates methodology, we assume that the vector  $\mathbf{D}$  has a positive definite (p.d.) variance-covariance matrix  $\boldsymbol{\Sigma}_D$ ; and we will see this implies that  $\boldsymbol{\Sigma}_C$  is also p.d. Let  $\{\mathbf{X}_i : i = 1, \dots, n\}$  be i.i.d. observations of  $\mathbf{X}$ . In terms of the auxiliary observations

$$Z_i = \zeta(\mathbf{X}_i) \quad \text{and} \quad \mathbf{D}_i = \delta(\mathbf{X}_i) \quad \text{for } i = 1, \dots, n, \quad (40)$$

we define  $\hat{\theta}_{\text{CE+CV}}(n)$ , the *conditional expectation-control variates* (CE+CV) estimator based on  $n$  replications, as the control-variates estimator

$$\hat{\theta}_{\text{CE+CV}}(n) \equiv \bar{Z} - \hat{\mathbf{b}} (\bar{\mathbf{D}} - \boldsymbol{\mu}_C)^T,$$

where:  $\bar{Z}$  and  $\bar{\mathbf{D}}$  are the sample means of  $\{Z_i : i = 1, \dots, n\}$  and  $\{\mathbf{D}_i : i = 1, \dots, n\}$ , respectively;

$$\hat{\mathbf{b}} \equiv \mathbf{S}_{ZD} \mathbf{S}_D^{-1};$$



$\mathbf{S}_{Z\mathbf{D}}$  denotes the  $1 \times q$  vector of sample covariances between  $Z$  and the components of  $\mathbf{D}$ ; and  $\mathbf{S}_{\mathbf{D}}$  denotes the  $q \times q$  sample variance-covariance matrix of  $\mathbf{D}$ .

To compute the mean and variance of  $\hat{\theta}_{\text{CE+CV}}(n)$ , we assume that  $(Z, \mathbf{D})$  has a multivariate normal distribution. Then  $\hat{\theta}_{\text{CE+CV}}(n)$  is an unbiased estimator of  $\theta$ , and its variance is

$$\text{Var}[\hat{\theta}_{\text{CE+CV}}(n)] = n^{-1} \sigma_Z^2 (1 - R_{Z\mathbf{D}}^2) \frac{n-2}{n-q-2},$$

where  $R_{Z\mathbf{D}}$  is the coefficient of multiple correlation between  $Z$  and  $\mathbf{D}$ . The remarks following (18) are again applicable so that irrespective of the distribution of  $(Z, \mathbf{D})$ , the combined estimator  $\hat{\theta}_{\text{CE+CV}}$  asymptotically dominates  $\hat{\theta}_{\text{CE}}$ .

To compare the variance of  $\hat{\theta}_{\text{CE+CV}}(n)$  with the variance of the standard control-variates estimator  $\hat{\theta}_{\text{CV}}(n)$ , we also assume that  $(Y, \mathbf{C})$  has a multivariate normal distribution and that the following condition holds:

$$\left. \begin{array}{l} \text{For every constant } 1 \times q \text{ vector } \boldsymbol{\beta} \neq \mathbf{0}, \text{ the random} \\ \text{variable } \boldsymbol{\beta} \mathbf{C}^T \text{ is not a function of } \mathbf{X} \text{ alone.} \end{array} \right\} \quad (41)$$

Adjusted to eliminate the effects of sample size, the difference between the variance of  $\hat{\theta}_{\text{CV}}(n)$  and the variance of  $\hat{\theta}_{\text{CE+CV}}(n)$  is

$$\frac{n(n-q-2)}{n-2} \left\{ \text{Var}[\hat{\theta}_{\text{CV}}(n)] - \text{Var}[\hat{\theta}_{\text{CE+CV}}(n)] \right\} = \sigma_Y^2 (1 - R_{Y\mathbf{C}}^2) - \sigma_Z^2 (1 - R_{Z\mathbf{D}}^2). \quad (42)$$

We now state the key result permitting the comparison between the CE+CV strategy and the standard control-variates technique. The proof of this result is given in the Appendix.

**Lemma 1.** *If  $\Sigma_{\mathbf{D}}$  is positive definite and (41) holds, then  $\sigma_Y^2 (1 - R_{Y\mathbf{C}}^2) - \sigma_Z^2 (1 - R_{Z\mathbf{D}}^2) \geq 0$ .*

Under the assumptions of Lemma 1, it follows immediately from (42) that  $\text{Var}[\hat{\theta}_{\text{CE+CV}}(n)] \leq \text{Var}[\hat{\theta}_{\text{CV}}(n)]$  for all  $n$ .

To determine the asymptotic distribution of  $\hat{\theta}_{\text{CE+CV}}(n)$ , we relax the assumptions of joint normality for  $(Y, \mathbf{C})$  and  $(Z, \mathbf{D})$  made above. Irrespective of the distributions of  $(Y, \mathbf{C})$  and  $(Z, \mathbf{D})$ , we obtain the following central limit theorem analogous to (18) for the CE+CV estimator:

$$n^{1/2} [\hat{\theta}_{\text{CE+CV}}(n) - \theta] \xrightarrow{\mathcal{D}} N[0, \sigma_Z^2 (1 - R_{Z\mathbf{D}}^2)] \quad \text{as } n \rightarrow \infty.$$

Coupled with (18), this last result shows that under the assumptions of Lemma 1,  $\hat{\theta}_{\text{CE+CV}}$  asymptotically dominates  $\hat{\theta}_{\text{CV}}$ . We have proved the following result.

**Theorem 3.** *Suppose that  $\Sigma_{\mathbf{D}}$  is positive definite and (41) holds. If each of  $(Y, \mathbf{C})$  and  $(Z, \mathbf{D})$  has a  $(q+1)$ -variate normal distribution, then  $\hat{\theta}_{\text{CE+CV}}(n)$  is an unbiased estimator*

of  $\theta$ , and

$$\text{Var}[\hat{\theta}_{\text{CE}+\text{CV}}(n)] \leq \min\{\text{Var}[\hat{\theta}_{\text{CE}}(n)], \text{Var}[\hat{\theta}_{\text{CV}}(n)]\} \quad \text{for } n \geq q/R_{\mathbf{ZD}}^2 + 2.$$

Moreover, irrespective of the distributions of  $(Y, \mathbf{C})$  and  $(Z, \mathbf{D})$ , the estimator  $\hat{\theta}_{\text{CE}+\text{CV}}$  asymptotically dominates  $\hat{\theta}_{\text{CV}}$  and  $\hat{\theta}_{\text{CE}}$ .

**Example 3.** Suppose that  $(Y, \mathbf{C}, \mathbf{X})$  is nonsingular multivariate normal. Then Theorems 2.4.3, 2.5.1, and A.3.2 of Anderson (1984) imply that  $(Y, \mathbf{C})$  and  $(Z, \mathbf{D})$  are both multivariate normal,  $\Sigma_{\mathbf{D}}$  is positive definite, and (41) holds. ■

### 3. ASYMPTOTIC VARIANCE COMPARISONS INVOLVING LHS

In this section we show that Latin hypercube sampling is asymptotically more efficient than the method of control variates when a certain class of controls is used. We also establish some general conditions under which an integrated variance reduction strategy based on the methods of conditional expectation and Latin hypercube sampling is asymptotically more efficient than many of the other strategies discussed in this paper.

Suppose that each component of the control vector  $\mathbf{C}$  has the additive form

$$C_\tau = \sum_{j \in J_{\mathbf{C}}} \varphi_{\tau,j}(V_j) \quad \text{for } \tau = 1, \dots, q \quad \text{and for some } J_{\mathbf{C}} \subseteq \{1, \dots, p\}, \quad (43)$$

subject to the following ancillary conditions:

AC<sub>1</sub> The input random variates  $\{V_j : j \in J_{\mathbf{C}}\}$  are mutually independent.

AC<sub>2</sub> The random vectors  $(V_j : j \in J_{\mathbf{C}})$  and  $(V_j : j \in J'_{\mathbf{C}})$  are independent, where  $J'_{\mathbf{C}} \equiv \{1, \dots, p\} - J_{\mathbf{C}}$ .

AC<sub>3</sub> For  $\tau = 1, \dots, q$  and for  $j \in J_{\mathbf{C}}$ , the symbol  $\varphi_{\tau,j}(\cdot)$  denotes an arbitrary univariate function.

In other words, each component  $C_\tau$  of the control vector  $\mathbf{C}$  is a *separable* function of a set of independent input random variates; and although the remaining set of input random variates may be stochastically interdependent, the latter set is independent of the former set. This situation often occurs in practice since many input variates are generated independently of each other, and control variates are often taken to be sums of selected input variates. For example, in queueing simulations, sums or averages of service times observed at selected service centers are frequently used as controls (Wilson 1984); and in simulations of stochastic

activity networks, sums of activity times along selected paths are often used as controls (see Section 4.1.1 below).

**Example 1 (continued).** Using again  $\mathbf{C} = V_4 + V_5$ , we see that (43) holds with  $J_{\mathbf{C}} = \{4, 5\}$ ,  $q = 1$ , and  $\varphi_{1,4}(v) \equiv \varphi_{1,5}(v) \equiv v$  for all real  $v$ . ■

Using the Latin hypercube sampling estimator  $\hat{\theta}_{\text{LH}}(\mathcal{H}_J, n)$  as defined in Section 1.2.2 with  $J = J_{\mathbf{C}}$ , we obtain the following result comparing  $\hat{\theta}_{\text{LH}}(\mathcal{H}_{J_{\mathbf{C}}}, n)$  and the control-variates estimator  $\hat{\theta}_{\text{CV}}(n)$ . The proof of this result is given in the Appendix.

**Theorem 4.** *If the response  $Y$  is bounded and the control vector  $\mathbf{C}$  has components of the form (43), then  $\hat{\theta}_{\text{LH}}(\mathcal{H}_{J_{\mathbf{C}}}, n)$  asymptotically dominates  $\hat{\theta}_{\text{CV}}(n)$ .*

In the rest of this section we examine the asymptotic efficiency of an integrated variance reduction strategy based on the methods of conditional expectation and Latin hypercube sampling. Our development requires assumptions (22) and (23), which were elaborated in Section 2.1. We reiterate that  $I_{\mathbf{X}}$  (respectively,  $J_{\mathbf{X}}$ ) is a set of indices of the input random numbers (respectively, the input random variates) upon which  $\mathbf{X}$  depends. For any subset  $J$  of  $\{1, \dots, p\}$  that satisfies conditions  $\text{IP}_1$  and  $\text{IP}_2$  in the general definition of Latin hypercube sampling as stated in Section 1.2.2, we define the *conditional expectation–Latin hypercube sampling* estimator  $\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_J, n)$  to be  $\hat{\theta}_{\text{CE+CI}}(G_{\text{LH}}^{(n)}, n)$ , where the input random variates  $\{V_j : j \in J_{\mathbf{X}} \cap J\}$  are sampled by inversion. For concreteness, we state the sampling plan  $\mathcal{H}_J$  corresponding to  $\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_J, n)$ ,

$$V_j = \begin{cases} F_j^{-1}(U_j), & j \in J_{\mathbf{X}} \cap J, \\ H_j(U_\tau : \tau \in I_{\mathbf{X}} - J), & j \in J_{\mathbf{X}} - J, \end{cases} \quad (44)$$

where the functions  $\{H_j(\cdot) : j \in J_{\mathbf{X}} - J\}$  are selected by the user to yield the correct joint distribution for the random vector  $(V_j : j \in J_{\mathbf{X}} - J)$ .

The following result provides a consolidated statement of our asymptotic comparisons of  $\hat{\theta}_{\text{CE+LH}}$  with many of the other estimators described in this paper. The proof of Theorem 5 is given in the Appendix.

**Theorem 5.**

- (i) *If the response  $Y$  is bounded, then  $\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_J, n)$  asymptotically dominates  $\hat{\theta}_{\text{CE}}(n)$  for any  $J \subseteq \{1, \dots, p\}$  that satisfies conditions  $\text{IP}_1$  and  $\text{IP}_2$ .*
- (ii) *If the response  $Y$  is bounded and the conditioning vector  $\mathbf{X}$  satisfies assumptions (22) and (23), then  $\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_J, n)$  asymptotically dominates  $\hat{\theta}_{\text{LH}}(\mathcal{H}_J, n)$  for any  $J \subseteq \{1, \dots, p\}$  that satisfies conditions  $\text{IP}_1$  and  $\text{IP}_2$ .*

(iii) If the response  $Y$  is bounded, if the conditioning vector  $\mathbf{X}$  satisfies assumptions (22) and (23), and if the control vector  $\mathbf{C}$  has components of the form (43), then the estimator  $\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_C}, n)$  asymptotically dominates  $\hat{\theta}_{\text{CE+CV}}(n)$  and  $\hat{\theta}_{\text{CV}}(n)$ .

**Example 1 (continued).** In view of the conclusion of part (iii) of Theorem 5, we take  $J = J_{\mathbf{C}} = \{4, 5\}$ . Since  $\mathbf{X} = (V_1, V_5)$  and  $J_{\mathbf{X}} = \{1, 5\}$ , we must reindex some of the random-number inputs to conform to (44). We accomplish this reindexing by interchanging  $U_5$  and  $U_6$  in the original sampling plan (2), so that now  $V_4 = H_4(U_6)$  and  $V_5 = H_5(U_5)$ . Let  $F_5^{-1}(\cdot)$  denote the inverse c.d.f. of  $V_5$ . Thus the sampling plan  $\mathcal{H}_{J_C}$  is given by

$$V_1 = H_1(U_1, U_2) \quad \text{and} \quad V_5 = F_5^{-1}(U_5).$$

Recalling the form of the conditional-expectation estimator in (5), we generate  $n$  dependent replications of this estimator,

$$Z^{(i)} = F_{2,3} \left[ t - H_1(U_1^{(i)}, U_2^{(i)}), t - H_1(U_1^{(i)}, U_2^{(i)}) - F_5^{-1}(U_5^{(i)}) \right] \cdot F_4 \left[ t - F_5^{-1}(U_5^{(i)}) \right]$$

for  $i = 1, \dots, n$ , where  $\{U_j^{(i)} : j = 1, 2, 5\}$  are sampled under LHS as in (15) with  $k = n$ . The resulting CE+LH estimator is

$$\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_C}, n) = n^{-1} \sum_{i=1}^n Z^{(i)}. \quad \blacksquare$$

## 4. APPLICATION TO STOCHASTIC ACTIVITY NETWORKS

We illustrate the application of our integrated variance reduction strategies to simulation of stochastic activity networks (SANs). In Section 4.1 we discuss how we implemented the standard VRTs of Section 1 and the integrated strategies of Section 2 for simulation of SANs, and we explain how we validated the assumptions underlying the integrated strategies. In Section 4.2 we describe the specific simulation experiments that were performed, and in Section 4.3 we summarize the results of those experiments.

### 4.1. Implementation and Validation Issues

#### 4.1.1. Setup for Simulating Stochastic Activity Networks

The setup for describing and simulating an arbitrary stochastic activity network is a straightforward extension of Example 1. The graph-theoretic structure of a stochastic activity network is described by the pair  $(\mathcal{N}, \mathcal{A})$ , where the nodes in the network constitute the set

$\mathcal{N} \equiv \{1, \dots, \nu\}$  and the activities in the network constitute the set

$$\mathcal{A} \equiv \{(a_j, b_j) : \text{activity } j \text{ has start node } a_j \in \mathcal{N} \text{ and end node } b_j \in \mathcal{N}, j = 1, \dots, p\}.$$

The network is assumed to be acyclic with a source node and a sink node in  $\mathcal{N}$ . Each activity  $j$  has a random duration  $V_j$ , so the input random variates are  $\{V_j : j = 1, \dots, p\}$ ; and the probabilistic structure of the network is described by the joint distribution of the random vector  $(V_1, \dots, V_p)$ .

The objective of simulating the network is to estimate the distribution of the time to realize the sink node—that is, the time to complete the network when all of the precedence relations between the activities in  $\mathcal{A}$  are taken into account. Let  $\xi$  denote the number of directed paths from source to sink, and let

$$A(\omega) \equiv \{j : \text{activity } j \text{ is on } \omega\text{th source-to-sink path}\} \quad \text{for } \omega = 1, \dots, \xi.$$

The duration of the  $\omega$ th path is the random variable

$$P_\omega \equiv \sum_{j \in A(\omega)} V_j \quad \text{for } \omega = 1, \dots, \xi; \quad (45)$$

and paralleling (1) is the *network completion time*

$$T \equiv \max\{P_1, \dots, P_\xi\}. \quad (46)$$

We seek to estimate the c.d.f.  $F_T(\cdot)$  of  $T$  at each cutoff value in a given set  $\mathcal{T}$ . For a selected  $t \in \mathcal{T}$ , the response of interest is

$$Y = f(V_1, \dots, V_p) \equiv \mathbf{1}\{T \leq t\}. \quad (47)$$

Here we view the overall estimation problem as a set of univariate estimation problems so that each value in  $\mathcal{T}$  corresponds to a single estimand of interest.

Our sampling plan is based on the assumption that each activity duration  $V_j$  ( $j = 1, \dots, p$ ) is independent of all other activity durations and has a known c.d.f.  $F_j(\cdot)$ . We use the method of inversion to generate all random variates, so the sampling plan  $\mathcal{H}_{J_p}$  is given by

$$V_j = F_j^{-1}(U_j) \quad \text{for } j \in J_p \equiv \{1, \dots, p\}. \quad (48)$$

Thus in terms of the notation established in Section 1, we have  $d = p$ . We observe that several numerical-analysis libraries—such as IMSL (International Mathematical and Statistical Library 1987) and the *Applied Statistics* algorithms (Griffiths and Hill 1985)—provide

inversion routines for most of the probability distributions that are commonly used in system simulation. Moreover, inversion is the *only* variate-generation scheme used in the INSIGHT simulation language (Roberts 1983); and in the SIMAN simulation language (Systems Modeling Corporation 1989, pp. 377–387), inversion is used to sample every distribution except the multiphase Erlang distribution, for which a combination of inversion and composition is used. Thus as a practical matter, we believe that use of the sampling plan (48) is not overly restrictive.

The variance reduction techniques discussed in Sections 1 and 2 are applied as follows. To implement the method of conditional expectation, we adapt an estimator originally developed for stochastic shortest route problems by Sigal, Pritsker, and Solberg (1980). A *uniformly directed cutset*  $\mathcal{L}$  is a set of activities such that each directed path from source to sink contains exactly one activity in  $\mathcal{L}$ . Following Sigal, Pritsker, and Solberg, we use a conditioning vector of the form

$$\mathbf{X} = (V_j : j \in J_{\mathbf{X}}), \quad \text{where } J_{\mathbf{X}} = \{1, \dots, p\} - \mathcal{L} \quad (49)$$

and  $\mathcal{L}$  is the uniformly directed cutset that is chosen. See Sigal, Pritsker, and Solberg for properties of a uniformly directed cutset and for a derivation of the associated conditional-expectation estimator of  $E[Y]$ . See Provan and Kulkarni (1989) for an efficient algorithm to identify a “good” uniformly directed cutset.

To implement the method of control variates, we use the approach of Avramidis, Bauer, and Wilson (1991). Ranking the directed paths from source to sink in decreasing order of expected duration, we let  $\omega_1$ ,  $\omega_2$ , and  $\omega_3$  denote the indices of the paths ranked first, second, and third, respectively. We use as control variables the durations of these three paths so that

$$C_{\tau} \equiv P_{\omega_{\tau}} = \sum_{j \in A(\omega_{\tau})} V_j \quad \text{for } \tau = 1, 2, \text{ and } 3. \quad (50)$$

Observe that for each  $\tau$  we can easily compute  $E[C_{\tau}]$  as the sum of the corresponding mean activity durations; and each mean activity duration  $E[V_j]$  is either known at the outset, or it can be evaluated using the known c.d.f.  $F_j(\cdot)$ . Also observe that the controls are of the form (43), with  $J_{\mathbf{C}} = \cup_{\tau=1}^3 A(\omega_{\tau})$  and

$$\varphi_{\tau,j}(v) \equiv \begin{cases} v, & \text{if } j \in A(\omega_{\tau}), \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } \tau = 1, 2, 3 \text{ and } j \in J_{\mathbf{C}}.$$

Finally, we have implemented the correlation-induction techniques of antithetic variates and Latin hypercube sampling along the lines detailed in Sections 1.2.1 and 1.2.2, respectively.

#### 4.1.2. Validation of the Integrated Variance Reduction Strategies

Given this setup for simulating stochastic activity networks, we can assess the validity of each of the assumptions underlying the development of Sections 2 and 3. First we consider the assumptions of Theorems 1 and 2. In view of (45) through (47), it is clear that  $f(\cdot)$  is a monotone function of each of its arguments. Moreover, the sampling plan (48) ensures that the conditioning vector in (49) satisfies conditions (22) and (23) as required in Theorem 1. From the discussion of monotonicity following Theorem 1 (that is, in the next-to-last paragraph of Section 2.1), it follows that  $\zeta(\cdot)$  is a monotone function of each of its arguments. Therefore *any* monotone sampling plan will guarantee the monotonicity of both  $y(\cdot)$  and  $z(\cdot)$ . Finally for all of the SANs used in our simulation experiments, the controls defined by (50) are readily seen to have the form (31) as required in Theorem 2.

Next we examine the validity of the assumption that (41) holds and  $\Sigma_{\mathbf{D}}$  is positive definite; this assumption is required in Theorem 3 to obtain the variance comparisons between  $\hat{\theta}_{\text{CE}+\text{CV}}$ ,  $\hat{\theta}_{\text{CV}}$ , and  $\hat{\theta}_{\text{CE}}$ . We will show that the following continuity properties of the controls  $\mathbf{C}$  and  $\mathbf{D}$  are sufficient to satisfy this assumption:

CP<sub>1</sub> Given  $\mathbf{X}$ , the conditional distribution of  $\mathbf{C}$  has a density or is of mixed type.

CP<sub>2</sub> The distribution of  $\mathbf{D} = \mathbf{E}[\mathbf{C}|\mathbf{X}] = \delta(\mathbf{X})$  has a density or is of mixed type.

(A distribution of mixed type has both discrete and absolutely continuous parts; see Neuts 1973, p. 170.) Proposition 1 of Porta Nova and Wilson (1989) is easily extended to cover distributions of mixed type and not only distributions having a density; then applying this extended result twice, we see that property CP<sub>1</sub> implies condition (41), and property CP<sub>2</sub> implies  $\Sigma_{\mathbf{D}}$  is p.d. For the SANs used in our simulation experiments, each distribution  $F_j(\cdot)$  has a density or is of mixed type,  $\mathbf{C}$  has the form (50), and  $\mathbf{X}$  has the form (49). Exploiting this structure, we reach the following conclusions about our SAN simulations:

- (a) The definition of a uniformly directed cutset ensures that property CP<sub>1</sub> holds.
- (b) Since  $\mathbf{D} = \delta(\mathbf{X})$  is a smooth (continuously differentiable) function of  $\mathbf{X}$ , the change-of-variables formula (Bickel and Doksum 1977, Theorem 1.2.2) ensures that property CP<sub>2</sub> holds.

In summary, properties CP<sub>1</sub> and CP<sub>2</sub> provide a convenient means of checking that  $\Sigma_{\mathbf{D}}$  is p.d. and condition (41) holds.

The assumption of normality necessary for the small-sample comparisons in Theorems 2 and 3 is clearly violated for the response (47) used in our SAN simulations. However, since

standard control-variates estimators have proven to be effective in a variety of settings where the normality assumption was violated (Avramidis, Bauer, and Wilson 1991), we expect that  $\hat{\theta}_{\text{CI+CV}}$  and  $\hat{\theta}_{\text{CE+CV}}$  will have a similar behavior.

In Theorems 4 and 5, the asymptotic variance comparisons between  $\hat{\theta}_{\text{CE}}$ ,  $\hat{\theta}_{\text{CV}}$ ,  $\hat{\theta}_{\text{LH}}$ ,  $\hat{\theta}_{\text{CE+CV}}$ , and  $\hat{\theta}_{\text{CE+LH}}$  require that the response  $Y$  is bounded. Clearly this boundedness assumption is satisfied for responses of the form (47). Since any simulation model must ultimately be executed on a computer with a finite word length, all variables in the simulation must be bounded; and thus the boundedness assumption does not seem to have much practical importance for *any* applications of the integrated variance reduction strategies discussed in this paper.

## 4.2. Description of the Simulation Experiments

The experimental performance evaluation was based on two SANs. For each activity duration  $V_j$  in a given network, the associated distribution was taken to be either (a) a normal distribution with a nominal mean  $\mu_j$  and standard deviation  $\sigma_j = \mu_j/4$  whose probability mass below the origin has been relocated to the origin; or (b) an exponential distribution with a specified mean  $\mu_j$ . We chose the exponential distribution as the nonnormal alternative for reasons elaborated in Avramidis, Bauer, and Wilson (1991). The first SAN was taken from Elmaghraby (1977, p. 275); and it is depicted in Figure 2. The set of activities with “adjusted” normal durations as in (a) was taken to be  $\{(1, 2), (1, 3), (2, 4), (6, 9), (7, 8)\}$ ; all other activities were taken to be exponentially distributed as in (b). As a uniformly directed cutset, we chose  $\mathcal{L} = \{(3, 6), (2, 6), (5, 6), (5, 8), (4, 7)\}$ .

The second SAN was taken from Antill and Woodhead (1982, Figure 8.5(b), p. 190); and it is depicted in Figure 3. Here the set of activities with “adjusted” normal durations was taken to be  $\{(1, 3), (2, 6), (2, 4), (8, 11), (10, 13), (12, 18), (16, 17), (17, 21), (17, 23), (17, 19), (18, 19), (23, 24)\}$ . The uniformly directed cutset for this network was  $\mathcal{L} = \{(2, 9), (4, 7), (5, 7), (1, 6), (3, 6), (3, 8)\}$ .

The experiments on networks 1 and 2 were executed with a general simulation program for stochastic activity networks. Random numbers were generated by the IMSL routine `RNUNF`. To implement the sampling plan (48), we used the *Applied Statistics* algorithm AS 111 to approximate the inverse standard normal c.d.f. To generate random permutations of the integers  $\{1, \dots, n\}$  as required by estimation procedures involving Latin hypercube sampling, we used the IMSL routine `RNPER`. Our simulation program is available upon request.

The purpose of the Monte Carlo study was to estimate the variance reductions achieved by



**Figure 2.** Network 1 used in the Monte Carlo study.

**Figure 3.** Network 2 used in the Monte Carlo study.

the following estimators: (i) the conditional-expectation estimator  $\hat{\theta}_{\text{CE}}(n)$ ; (ii) the antithetic-variates estimator  $\hat{\theta}_{\text{AV}}(n) \equiv \hat{\theta}_{\text{CI}}(G_{\text{AV}}^{(2)}, n)$ ; (iii) the Latin hypercube sampling estimator  $\hat{\theta}_{\text{LH}}(\mathcal{H}_{J_p}, n)$ ; (iv) the control-variates estimator  $\hat{\theta}_{\text{CV}}(n)$ ; (v) the conditional expectation–control variates estimator  $\hat{\theta}_{\text{CE+CV}}(n)$ ; and (vi) the conditional expectation–Latin hypercube sampling estimator  $\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_p}, n)$ . In the experimental performance evaluation, we used the following sample sizes:  $n = 32, 64$ , and  $128$ .

For each selected combination of a stochastic activity network, a sample size  $n$ , and an estimator  $\hat{\theta}$ , we used the following protocol to approximate  $\text{Var}[\hat{\theta}]$ . We conducted a set of  $\mathcal{M}$  independent *macroreplications* of the estimation procedure that yields  $\hat{\theta}$ , where each macroreplication consisted of  $n$  (possibly correlated) replications of the network simulation that are required to compute a single observation of  $\hat{\theta}$ . Then  $\text{Var}[\hat{\theta}]$  was estimated as the sample variance of  $\mathcal{M}$  i.i.d. observations of  $\hat{\theta}$ . Table I shows the number of macroreplications used to approximate  $\text{Var}[\hat{\theta}]$  for each estimator  $\hat{\theta}$  that we studied, including the direct-simulation estimator  $\bar{Y}(n)$ .

**Table I**  
Number of Independent Macroreplications  $\mathcal{M}$  of  $\hat{\theta}$   
Used to Approximate  $\text{Var}[\hat{\theta}]$  for Each Estimator  $\hat{\theta}$

Estimator	$\mathcal{M}$
$\bar{Y}(n)$	4096
$\hat{\theta}_{\text{CE}}(32)$	4096
$\hat{\theta}_{\text{AV}}(32)$	4096
$\hat{\theta}_{\text{LH}}(\mathcal{H}_{J_p}, 32)$	4096
$\hat{\theta}_{\text{CV}}(32)$	4096
$\hat{\theta}_{\text{CE+CV}}(32)$	4096
$\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_p}, 32)$	4096
$\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_p}, 64)$	2048
$\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_p}, 128)$	1024

### 4.3. Experimental Results

For networks 1 and 2 respectively, Tables II and III contain our approximations to the variance ratio  $\text{Var}[\bar{Y}(n)] / \text{Var}[\hat{\theta}]$  for each selected estimator  $\hat{\theta}$ . Because of the large number of degrees of freedom (at least 1023) in the numerator and denominator of each estimated variance ratio, there is relatively little error associated with the entries of Tables II and III.

By independently replicating Tables II and III, we found that the largest standard error for any of the entries in these tables was less than 7% of the corresponding entry. To improve the readability of these tables, we omitted the standard errors associated with the estimated variance ratios. We remark that based on  $n = 4096$  independent replications of network 1, the direct-simulation estimates of  $F_T(30)$ ,  $F_T(50)$ ,  $F_T(70)$ , and  $F_T(90)$  are 0.044, 0.44, 0.78, and 0.92, respectively; and for network 2, the direct-simulation estimates of  $F_T(500)$ ,  $F_T(600)$ ,  $F_T(700)$ , and  $F_T(900)$  are 0.18, 0.47, 0.72, and 0.94, respectively.

For estimating the c.d.f. of the network completion time in the given stochastic activity networks, the method of conditional expectation appears to be the most effective of the individual VRTs, followed by Latin hypercube sampling and control variates. Moreover, Tables II and III clearly reveal the synergism that results from the joint application of two VRTs in an integrated variance reduction strategy. For example, if the efficiency improvements due to successive application of the methods of conditional expectation and Latin hypercube sampling were independent effects, then the variance ratio for  $\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_p}, n)$  would equal the product of the variance ratios for  $\hat{\theta}_{\text{CE}}(n)$  and  $\hat{\theta}_{\text{LH}}(\mathcal{H}_{J_p}, n)$  at every sample size  $n$ ; however in Tables II and III the variance ratio for  $\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_p}, 32)$  substantially exceeds the product of the corresponding variance ratios for  $\hat{\theta}_{\text{CE}}(32)$  and  $\hat{\theta}_{\text{LH}}(\mathcal{H}_{J_p}, 32)$ . Notice also that  $\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_p}, 32)$  is more precise than  $\hat{\theta}_{\text{CE}}(32)$ ,  $\hat{\theta}_{\text{AV}}(32)$ ,  $\hat{\theta}_{\text{CV}}(32)$ ,  $\hat{\theta}_{\text{LH}}(\mathcal{H}_{J_p}, 32)$ , and  $\hat{\theta}_{\text{CE+CV}}(32)$ ; and this small-sample comparison is consistent with the asymptotic comparisons given in Section 3. Although asymptotic comparisons of  $\hat{\theta}_{\text{AV}}$  against the other estimators do not seem possible in general, our experimental results indicate that  $\hat{\theta}_{\text{AV}}$  is usually less precise than all of the other estimators. We concluded that for the simulation of stochastic activity networks, Latin hypercube sampling is the correlation-induction method of choice; moreover, integrated variance reduction strategies can yield large improvements in precision relative to individual VRTs.

Because of the superior performance of the conditional expectation–Latin hypercube sampling estimator  $\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_p}, n)$ , we investigated the effect on this estimator of increasing the sample size  $n$ . If the response  $Y$  is bounded and the conditioning vector  $\mathbf{X}$  has the form (22), then Corollary 1 of Stein (1987) implies that

$$\lim_{n \rightarrow \infty} \frac{\text{Var}[\bar{Y}(n)]}{\text{Var}[\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_p}, n)]} = \frac{\sigma_Y^2}{\sigma_{\text{CE+LH}}^2(\mathcal{H}_{J_p})} \geq 1, \quad (51)$$

where  $\sigma_{\text{CE+LH}}^2(\mathcal{H}_{J_p})$  is the asymptotic variance parameter for the CE+LH estimator with the sampling plan  $\mathcal{H}_{J_p}$  defined by (48). Although we cannot give a complete characterization of the way in which the variance ratio in (51) approaches the limiting value

**Table II**

Estimates of the Variance Ratio  $\text{Var}[\bar{Y}(n)] / \text{Var}[\hat{\theta}]$   
for Different Estimators  $\hat{\theta}$  and Sample Sizes  $n$  in Network 1

Estimator	Estimand			
	$F_T(30)$	$F_T(50)$	$F_T(70)$	$F_T(90)$
$\hat{\theta}_{\text{CE}}(32)$	11.2	4.7	4.17	4.31
$\hat{\theta}_{\text{AV}}(32)$	1.05	1.22	1.08	1.02
$\hat{\theta}_{\text{LH}}(\mathcal{H}_{J_p}, 32)$	1.19	2.26	2.92	2.18
$\hat{\theta}_{\text{CV}}(32)$	1.12	1.35	1.33	1.20
$\hat{\theta}_{\text{CE+CV}}(32)$	16.7	19.8	10.9	7.62
$\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_p}, 32)$	42.3	72.6	52.2	18.3
$\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_p}, 64)$	42.4	83.2	70.7	47.1
$\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_p}, 128)$	44.6	87.4	78.1	56.3

**Table III**

Estimates of the Variance Ratio  $\text{Var}[\bar{Y}(n)] / \text{Var}[\hat{\theta}]$   
for Different Estimators  $\hat{\theta}$  and Sample Sizes  $n$  in Network 2

Estimator	Estimand			
	$F_T(500)$	$F_T(600)$	$F_T(700)$	$F_T(900)$
$\hat{\theta}_{\text{CE}}(32)$	3.50	2.88	2.66	2.37
$\hat{\theta}_{\text{AV}}(32)$	1.17	1.18	1.11	1.02
$\hat{\theta}_{\text{LH}}(\mathcal{H}_{J_p}, 32)$	1.50	2.15	2.29	1.65
$\hat{\theta}_{\text{CV}}(32)$	1.15	1.27	1.26	1.19
$\hat{\theta}_{\text{CE+CV}}(32)$	5.64	6.48	6.14	3.86
$\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_p}, 32)$	11.2	17.9	17.5	7.99
$\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_p}, 64)$	11.0	19.5	23.0	11.2
$\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_p}, 128)$	11.2	20.9	25.4	14.2

$\sigma_Y^2/\sigma_{\text{CE+LH}}^2(\mathcal{H}_{J_p})$ , the experimental results in Tables II and III as well as other empirical results not reported here support the conclusion that this variance ratio is generally an *increasing* function of the sample size—that is, the precision of  $\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_p}, n)$  relative to direct simulation seems to *improve* as  $n$  increases. The point at which the variance ratio  $\text{Var}[\bar{Y}(n)]/\text{Var}[\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_p}, n)]$  levels off seems to vary with the application.

## 5. CONCLUSIONS AND RECOMMENDATIONS

We believe that the development presented in this paper provides a framework for effective application of integrated variance reduction strategies in many contexts. Beyond the theoretical comparisons of the various integrated strategies, the experimental results for moderately complex stochastic activity networks provide substantial evidence of the practical value of using these techniques to improve the efficiency of large-scale simulations. In particular, we observed a synergism due to joint application of variance reduction techniques so that the variance ratio (relative to direct simulation) for each integrated variance reduction strategy substantially exceeded the product of the variance ratios for the corresponding individual variance reduction techniques.

Although follow-up work is needed in a number of areas, perhaps the most immediate need is for more extensive experimentation. A major unresolved issue concerns the performance of the integrated variance reduction strategies when the assumptions underlying those strategies are violated. Moreover, it is unclear whether the large efficiency improvements observed by integrating the methods of conditional expectation and Latin hypercube sampling are typical of the gains that can be anticipated in practice. In the spirit of Nelson (1990) and Avramidis, Bauer, and Wilson (1991), a comprehensive experimental evaluation is required for the integrated variance reduction strategies developed in this paper.

Follow-up work is also required to extend the theoretical development to cover a larger class of simulation experiments. Although our development is limited to simulations where the dimension  $d$  of the vector of random-number inputs is fixed, we believe that much of this analysis can ultimately be extended to simulations where  $d$  is random. Such a complication naturally arises in the following situations: (a) a finite-horizon simulation with a sampling plan involving, for example, the acceptance-rejection method, so that  $p$  is fixed but  $d$  is random; and (b) an infinite-horizon simulation in which both  $p$  and  $d$  are random. All of the results presented in this paper are limited to independent replications of a univariate simulation response. These results should be generalized to multiresponse simulations. Moreover, the integrated variance reduction strategies should be adapted to responses generated within

a single prolonged replication of a simulation model in steady-state operation. In light of the observed effectiveness of the joint application of Latin hypercube sampling and the method of conditional expectation, we believe that in future research emphasis should be given to integrated strategies involving these techniques. Finally, the integrated variance reduction strategies formulated in this paper should be extended to accommodate joint application of three or more basic variance reduction techniques; and the potential for synergism between these techniques should be investigated.

## APPENDIX

### Proof of Lemma 1

We express the adjusted difference in variances (42) as

$$\sigma_Y^2(1 - R_{Y\mathbf{C}}^2) - \sigma_Z^2(1 - R_{Z\mathbf{D}}^2) = \sigma_Y^2 - \boldsymbol{\sigma}_{Y\mathbf{C}}\boldsymbol{\Sigma}_{\mathbf{C}}^{-1}\boldsymbol{\sigma}_{Y\mathbf{C}}^T - \sigma_Z^2 + \boldsymbol{\sigma}_{Z\mathbf{D}}\boldsymbol{\Sigma}_{\mathbf{D}}^{-1}\boldsymbol{\sigma}_{Z\mathbf{D}}^T, \quad (52)$$

where we have partitioned the variance-covariance matrix of  $(Z, \mathbf{D})$  as

$$\text{Var}[(Z, \mathbf{D})] = \begin{bmatrix} \sigma_Z^2 & \boldsymbol{\sigma}_{Z\mathbf{D}} \\ \boldsymbol{\sigma}_{Z\mathbf{D}}^T & \boldsymbol{\Sigma}_{\mathbf{D}} \end{bmatrix}.$$

Similarly, we partition the expectation of the conditional variance-covariance matrix of  $(Y, \mathbf{C})$  given  $\mathbf{X}$  as

$$\mathbf{P} \equiv \text{E}\{\text{Var}[(Y, \mathbf{C}) | \mathbf{X}]\} = \begin{bmatrix} P_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{12}^T & \mathbf{P}_{22} \end{bmatrix}, \quad (53)$$

where  $P_{11}$  is a scalar,  $\mathbf{P}_{12}$  is a  $1 \times q$  vector, and  $\mathbf{P}_{22}$  is a  $q \times q$  matrix. Exploiting the matrix version of the conditional variance relation

$$\begin{aligned} \text{Var}[(Y, \mathbf{C})] &= \text{E}\{\text{Var}[(Y, \mathbf{C}) | \mathbf{X}]\} + \text{Var}\{\text{E}[(Y, \mathbf{C}) | \mathbf{X}]\} \\ &= \mathbf{P} + \text{Var}[(Z, \mathbf{D})], \end{aligned}$$

we will repeatedly use the relations

$$P_{11} = \sigma_Y^2 - \sigma_Z^2, \quad \mathbf{P}_{12} = \boldsymbol{\sigma}_{Y\mathbf{C}} - \boldsymbol{\sigma}_{Z\mathbf{D}}, \quad \mathbf{P}_{22} = \boldsymbol{\Sigma}_{\mathbf{C}} - \boldsymbol{\Sigma}_{\mathbf{D}}. \quad (54)$$

To continue the analysis of (52), we show that  $\mathbf{P}_{22}$  and  $\boldsymbol{\Sigma}_{\mathbf{C}}$  are p.d. Notice that we require  $\boldsymbol{\Sigma}_{\mathbf{C}}$  to be p.d. in equation (52), the first step of this proof; and we will require  $\mathbf{P}_{22}$  to be

p.d. in equations (61) and (64) below. Clearly  $\mathbf{P}_{22} = \mathbb{E}[\text{Var}(\mathbf{C}|\mathbf{X})]$  is positive semidefinite (p.s.d.) since  $\text{Var}(\mathbf{C}|\mathbf{X} = \mathbf{x})$  is p.s.d. for every  $\mathbf{x}$ . Moreover, for any deterministic  $1 \times q$  vector  $\boldsymbol{\beta} \neq \mathbf{0}$ ,

$$\boldsymbol{\beta} \mathbf{P}_{22} \boldsymbol{\beta}^T = \mathbb{E}[\boldsymbol{\beta} \text{Var}(\mathbf{C}|\mathbf{X}) \boldsymbol{\beta}^T] = \mathbb{E}[\text{Var}(\boldsymbol{\beta} \mathbf{C}^T | \mathbf{X})] . \quad (55)$$

Since  $\boldsymbol{\beta} \mathbf{C}^T$  is not a function of  $\mathbf{X}$  alone, the event  $\{\text{Var}(\boldsymbol{\beta} \mathbf{C}^T | \mathbf{X}) > 0\}$  must have positive probability of occurring so that

$$\mathbb{E}[\text{Var}(\boldsymbol{\beta} \mathbf{C}^T | \mathbf{X})] > 0 . \quad (56)$$

It follows immediately from (55) and (56) that  $\mathbf{P}_{22}$  is p.d.; and (54) implies that  $\boldsymbol{\Sigma}_{\mathbf{C}} = \mathbf{P}_{22} + \boldsymbol{\Sigma}_{\mathbf{D}}$  is also p.d.

The next step in the analysis of (52) is to obtain expressions for  $\boldsymbol{\sigma}_{Y\mathbf{C}} \boldsymbol{\Sigma}_{\mathbf{C}}^{-1} \boldsymbol{\sigma}_{Y\mathbf{C}}^T$  and  $\boldsymbol{\sigma}_{Z\mathbf{D}} \boldsymbol{\Sigma}_{\mathbf{D}}^{-1} \boldsymbol{\sigma}_{Z\mathbf{D}}^T$  that can be combined conveniently. Using elementary properties of  $\text{tr}(\cdot)$ , the trace operator, we have

$$\begin{aligned} \boldsymbol{\sigma}_{Y\mathbf{C}} \boldsymbol{\Sigma}_{\mathbf{C}}^{-1} \boldsymbol{\sigma}_{Y\mathbf{C}}^T &= \text{tr}(\boldsymbol{\sigma}_{Y\mathbf{C}}^T \boldsymbol{\sigma}_{Y\mathbf{C}} \boldsymbol{\Sigma}_{\mathbf{C}}^{-1}) \\ &= \text{tr}[\boldsymbol{\sigma}_{Y\mathbf{C}}^T \boldsymbol{\sigma}_{Y\mathbf{C}} \boldsymbol{\Sigma}_{\mathbf{C}}^{-1} (\boldsymbol{\Sigma}_{\mathbf{C}} - \mathbf{P}_{22}) \boldsymbol{\Sigma}_{\mathbf{D}}^{-1}] \\ &= \text{tr}(\boldsymbol{\sigma}_{Y\mathbf{C}}^T \boldsymbol{\sigma}_{Y\mathbf{C}} \boldsymbol{\Sigma}_{\mathbf{D}}^{-1}) - \text{tr}(\boldsymbol{\sigma}_{Y\mathbf{C}}^T \boldsymbol{\sigma}_{Y\mathbf{C}} \boldsymbol{\Sigma}_{\mathbf{C}}^{-1} \mathbf{P}_{22} \boldsymbol{\Sigma}_{\mathbf{D}}^{-1}) . \end{aligned} \quad (57)$$

Similarly, we have

$$\begin{aligned} \boldsymbol{\sigma}_{Z\mathbf{D}} \boldsymbol{\Sigma}_{\mathbf{D}}^{-1} \boldsymbol{\sigma}_{Z\mathbf{D}}^T &= \text{tr}(\boldsymbol{\sigma}_{Z\mathbf{D}}^T \boldsymbol{\sigma}_{Z\mathbf{D}} \boldsymbol{\Sigma}_{\mathbf{D}}^{-1}) \\ &= \text{tr}[(\boldsymbol{\sigma}_{Y\mathbf{C}} - \mathbf{P}_{12})^T (\boldsymbol{\sigma}_{Y\mathbf{C}} - \mathbf{P}_{12}) \boldsymbol{\Sigma}_{\mathbf{D}}^{-1}] \\ &= \text{tr}(\boldsymbol{\sigma}_{Y\mathbf{C}}^T \boldsymbol{\sigma}_{Y\mathbf{C}} \boldsymbol{\Sigma}_{\mathbf{D}}^{-1}) + \text{tr}\left\{ \left[ -(\boldsymbol{\sigma}_{Y\mathbf{C}}^T \mathbf{P}_{12} + \mathbf{P}_{12}^T \boldsymbol{\sigma}_{Y\mathbf{C}}) + \mathbf{P}_{12}^T \mathbf{P}_{12} \right] \boldsymbol{\Sigma}_{\mathbf{D}}^{-1} \right\} . \end{aligned} \quad (58)$$

Now if  $\mathbf{A}$  and  $\mathbf{B}$  are square matrices of the same dimension and if  $\mathbf{B}$  is symmetric, then it is easy to show that

$$\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{A}^T \mathbf{B}) . \quad (59)$$

If we apply (59) to (58) with  $\mathbf{A} \equiv \boldsymbol{\sigma}_{Y\mathbf{C}}^T \mathbf{P}_{12}$  and  $\mathbf{B} \equiv \boldsymbol{\Sigma}_{\mathbf{D}}^{-1}$ , then we obtain

$$\boldsymbol{\sigma}_{Z\mathbf{D}} \boldsymbol{\Sigma}_{\mathbf{D}}^{-1} \boldsymbol{\sigma}_{Z\mathbf{D}}^T = \text{tr}(\boldsymbol{\sigma}_{Y\mathbf{C}}^T \boldsymbol{\sigma}_{Y\mathbf{C}} \boldsymbol{\Sigma}_{\mathbf{D}}^{-1}) + \text{tr}\left[ \left( -2\boldsymbol{\sigma}_{Y\mathbf{C}}^T \mathbf{P}_{12} + \mathbf{P}_{12}^T \mathbf{P}_{12} \right) \boldsymbol{\Sigma}_{\mathbf{D}}^{-1} \right] . \quad (60)$$

Subtracting (57) from (60), we get

$$\begin{aligned} & -\boldsymbol{\sigma}_{Y\mathbf{C}} \boldsymbol{\Sigma}_{\mathbf{C}}^{-1} \boldsymbol{\sigma}_{Y\mathbf{C}}^T + \boldsymbol{\sigma}_{Z\mathbf{D}} \boldsymbol{\Sigma}_{\mathbf{D}}^{-1} \boldsymbol{\sigma}_{Z\mathbf{D}}^T \\ &= \text{tr}\left[ \left( \boldsymbol{\sigma}_{Y\mathbf{C}}^T \boldsymbol{\sigma}_{Y\mathbf{C}} - 2\boldsymbol{\sigma}_{Y\mathbf{C}}^T \mathbf{P}_{12} \mathbf{P}_{22}^{-1} \boldsymbol{\Sigma}_{\mathbf{C}} + \mathbf{P}_{12}^T \mathbf{P}_{12} \mathbf{P}_{22}^{-1} \boldsymbol{\Sigma}_{\mathbf{C}} \right) (\boldsymbol{\Sigma}_{\mathbf{C}}^{-1} \mathbf{P}_{22} \boldsymbol{\Sigma}_{\mathbf{D}}^{-1}) \right] \\ &= \text{tr}\left\{ \left[ \boldsymbol{\sigma}_{Y\mathbf{C}}^T (\boldsymbol{\sigma}_{Y\mathbf{C}} - \mathbf{P}_{12} \mathbf{P}_{22}^{-1} \boldsymbol{\Sigma}_{\mathbf{C}}) - \boldsymbol{\Sigma}_{\mathbf{C}} \mathbf{P}_{22}^{-1} \mathbf{P}_{12}^T \boldsymbol{\sigma}_{Y\mathbf{C}} + \mathbf{P}_{12}^T \mathbf{P}_{12} \mathbf{P}_{22}^{-1} \boldsymbol{\Sigma}_{\mathbf{C}} \right] (\boldsymbol{\Sigma}_{\mathbf{C}}^{-1} \mathbf{P}_{22} \boldsymbol{\Sigma}_{\mathbf{D}}^{-1}) \right\} . \end{aligned} \quad (61)$$



In the last equality in (61) we used (59) again, this time with  $\mathbf{A} \equiv \boldsymbol{\sigma}_{YC}^T \mathbf{P}_{12} \mathbf{P}_{22}^{-1} \boldsymbol{\Sigma}_C$  and  $\mathbf{B} \equiv \boldsymbol{\Sigma}_C^{-1} \mathbf{P}_{22} \boldsymbol{\Sigma}_D^{-1} = \boldsymbol{\Sigma}_D^{-1} - \boldsymbol{\Sigma}_C^{-1}$ . If we define the auxiliary quantities

$$\boldsymbol{\alpha} \equiv \boldsymbol{\sigma}_{YC} - \mathbf{P}_{12} \mathbf{P}_{22}^{-1} \boldsymbol{\Sigma}_C \quad \text{and} \quad \boldsymbol{\Omega} \equiv \boldsymbol{\Sigma}_C^{-1} \mathbf{P}_{22} \boldsymbol{\Sigma}_D^{-1} = \boldsymbol{\Sigma}_D^{-1} - \boldsymbol{\Sigma}_C^{-1},$$

then we can rewrite (61) as

$$\begin{aligned} -\boldsymbol{\sigma}_{YC} \boldsymbol{\Sigma}_C^{-1} \boldsymbol{\sigma}_{YC}^T + \boldsymbol{\sigma}_{ZD} \boldsymbol{\Sigma}_D^{-1} \boldsymbol{\sigma}_{ZD}^T &= \text{tr} \left\{ \left[ \boldsymbol{\alpha}^T \boldsymbol{\alpha} - (\boldsymbol{\Sigma}_C \mathbf{P}_{22}^{-1} - \mathbf{I}_q) \mathbf{P}_{12}^T \mathbf{P}_{12} \mathbf{P}_{22}^{-1} \boldsymbol{\Sigma}_C \right] \boldsymbol{\Omega} \right\} \\ &= \text{tr} \left\{ \left[ \boldsymbol{\alpha}^T \boldsymbol{\alpha} - (\boldsymbol{\Sigma}_D \mathbf{P}_{22}^{-1}) \mathbf{P}_{12}^T \mathbf{P}_{12} \mathbf{P}_{22}^{-1} \boldsymbol{\Sigma}_C \right] \boldsymbol{\Omega} \right\} \\ &= \text{tr} \left\{ \left[ \boldsymbol{\alpha}^T \boldsymbol{\alpha} - \boldsymbol{\Sigma}_D (\mathbf{P}_{22}^{-1} \mathbf{P}_{12}^T \mathbf{P}_{12}) \mathbf{P}_{22}^{-1} \boldsymbol{\Sigma}_C \right] \boldsymbol{\Omega} \right\} \\ &= \text{tr} \left[ \boldsymbol{\alpha}^T \boldsymbol{\alpha} \boldsymbol{\Omega} - (\mathbf{P}_{22}^{-1} \mathbf{P}_{12}^T \mathbf{P}_{12}) (\mathbf{P}_{22}^{-1} \boldsymbol{\Sigma}_C \boldsymbol{\Omega} \boldsymbol{\Sigma}_D) \right] \\ &= \text{tr}(\boldsymbol{\alpha}^T \boldsymbol{\alpha} \boldsymbol{\Omega}) - \text{tr}(\mathbf{P}_{22}^{-1} \mathbf{P}_{12}^T \mathbf{P}_{12}) \\ &= \boldsymbol{\alpha} \boldsymbol{\Omega} \boldsymbol{\alpha}^T - \mathbf{P}_{12} \mathbf{P}_{22}^{-1} \mathbf{P}_{12}^T, \end{aligned} \tag{62}$$

where  $\mathbf{I}_q$  denotes the  $q \times q$  identity matrix. In view of (54) and (62), the difference (52) has the form

$$\sigma_Y^2(1 - R_{YC}^2) - \sigma_Z^2(1 - R_{ZD}^2) = \boldsymbol{\alpha} \boldsymbol{\Omega} \boldsymbol{\alpha}^T + (P_{11} - \mathbf{P}_{12} \mathbf{P}_{22}^{-1} \mathbf{P}_{12}^T). \tag{63}$$

To complete the proof of Lemma 1, we will establish that the two terms on the right-hand side of (63) are nonnegative. From the definition (53), it follows that  $\mathbf{P}$  is p.s.d. since  $\text{Var}[(Y, \mathbf{C}) | \mathbf{X} = \mathbf{x}]$  is p.s.d. for all  $\mathbf{x}$ ; and thus by a simple modification of Theorem A.6 of Arnold (1981), we see that the term in parentheses on the right-hand side of (63) is nonnegative. Finally we show that the matrix  $\boldsymbol{\Omega}$  is p.d. so that the first term on the right-hand side of (63) is nonnegative. There exists a nonsingular matrix  $\boldsymbol{\Theta}$  such that  $\boldsymbol{\Theta}^T \boldsymbol{\Sigma}_C \boldsymbol{\Theta}$  and  $\boldsymbol{\Theta}^T \boldsymbol{\Sigma}_D \boldsymbol{\Theta}$  are both diagonal matrices (Searle 1982, Theorem 3, p. 313)—say,

$$\boldsymbol{\Theta}^T \boldsymbol{\Sigma}_C \boldsymbol{\Theta} = \text{diag}(\varepsilon_1, \dots, \varepsilon_q) \quad \text{and} \quad \boldsymbol{\Theta}^T \boldsymbol{\Sigma}_D \boldsymbol{\Theta} = \text{diag}(\eta_1, \dots, \eta_q).$$

Since  $\mathbf{P}_{22} = \boldsymbol{\Sigma}_C - \boldsymbol{\Sigma}_D$  is p.d., we see that

$$\begin{aligned} \boldsymbol{\Theta}^T \mathbf{P}_{22} \boldsymbol{\Theta} &= \boldsymbol{\Theta}^T \boldsymbol{\Sigma}_C \boldsymbol{\Theta} - \boldsymbol{\Theta}^T \boldsymbol{\Sigma}_D \boldsymbol{\Theta} \\ &= \text{diag}(\varepsilon_1 - \eta_1, \dots, \varepsilon_q - \eta_q) \quad \text{is p.d.}; \end{aligned} \tag{64}$$

and thus we have  $\varepsilon_\tau > \eta_\tau > 0$  for  $\tau = 1, \dots, q$ . It follows that  $\eta_\tau^{-1} - \varepsilon_\tau^{-1} > 0$  for  $\tau = 1, \dots, q$  so that

$$\begin{aligned} \boldsymbol{\Theta}^{-1} \boldsymbol{\Omega} (\boldsymbol{\Theta}^T)^{-1} &= (\boldsymbol{\Theta}^T \boldsymbol{\Sigma}_D \boldsymbol{\Theta})^{-1} - (\boldsymbol{\Theta}^T \boldsymbol{\Sigma}_C \boldsymbol{\Theta})^{-1} \\ &= \text{diag}(\eta_1^{-1} - \varepsilon_1^{-1}, \dots, \eta_q^{-1} - \varepsilon_q^{-1}) \quad \text{is p.d.}; \end{aligned}$$

and thus we finally conclude that  $\boldsymbol{\Omega}$  is p.d.  $\blacksquare$

## Proof of Theorem 4

We have seen in (18) that the asymptotic variance parameter for  $\hat{\theta}_{\text{CV}}$  is

$$\sigma_{\text{CV}}^2 = \sigma_Y^2 - \boldsymbol{\sigma}_{Y\mathbf{C}} \boldsymbol{\Sigma}_{\mathbf{C}}^{-1} \boldsymbol{\sigma}_{Y\mathbf{C}}^T = \text{Var}(Y - \mathbf{b}^* \mathbf{C}^T), \quad (65)$$

where  $\mathbf{b}^*$  is defined in (17). Writing  $\mathbf{b}^* = (b_1^*, \dots, b_q^*)$  and using (43), we have

$$Y - \mathbf{b}^* \mathbf{C}^T = Y - \sum_{\tau=1}^q b_\tau^* C_\tau = Y - \sum_{j \in J_{\mathbf{C}}} \sum_{\tau=1}^q b_\tau^* \varphi_{\tau,j}(V_j). \quad (66)$$

Applying Theorem 1 of Owen (1992), we see that for any  $J \subseteq \{1, \dots, p\}$  satisfying conditions  $\text{IP}_1$  and  $\text{IP}_2$  with associated sampling plan  $\mathcal{H}_J$  defined by (16), the following CLT holds:

$$n^{1/2} [\hat{\theta}_{\text{LH}}(\mathcal{H}_J, n) - \theta] \xrightarrow{\mathcal{D}} N[0, \sigma_{\text{LH}}^2(\mathcal{H}_J)] \quad \text{as } n \rightarrow \infty, \quad (67)$$

where the corresponding variance parameter is

$$\sigma_{\text{LH}}^2(\mathcal{H}_J) = \text{Var} \left[ Y - \sum_{j=1}^d \mathbb{E}(Y|U_j) \right] \quad \text{with} \quad Y = f[\mathcal{H}_J(\mathbf{U})]. \quad (68)$$

Result (67) was originally established by Stein (1987), but Stein's proof has some gaps that were filled by Owen (1992). Next we observe that

$$\text{Var} \left[ Y - \sum_{j=1}^d \mathbb{E}(Y|U_j) \right] \leq \text{Var} \left[ Y - \sum_{j=1}^d \psi_j(U_j) \right] \quad (69)$$

for any univariate functions  $\{\psi_j(\cdot) : j \in \{1, \dots, d\}\}$ . The inequality (69) follows from Appendix B of Stein (1987). To compare the asymptotic variance parameters  $\sigma_{\text{CV}}^2$  and  $\sigma_{\text{LH}}^2(\mathcal{H}_{J_{\mathbf{C}}})$  when we take  $J = J_{\mathbf{C}}$  in (16) to obtain the sampling plan  $\mathcal{H}_{J_{\mathbf{C}}}$ , we choose  $\psi_j(u) \equiv \sum_{\tau=1}^q b_\tau^* \varphi_{\tau,j}[F_j^{-1}(u)]$  for  $j \in I_{\mathbf{C}}$  and  $\psi_j(u) \equiv 0$  for  $j \in I'_{\mathbf{C}}$  in the right-hand side of (69). Since  $I_{\mathbf{C}} = J_{\mathbf{C}}$  in the sampling plan  $\mathcal{H}_{J_{\mathbf{C}}}$ , we see that (65), (66), and (69) yield the desired result.  $\blacksquare$

Next we establish a lemma that will be used repeatedly in the proof of Theorem 5.

**Lemma 2.** *Let  $J$  be an arbitrary subset of  $\{1, \dots, p\}$  and define the random vectors  $\dot{\mathbf{V}} \equiv (V_j : j \in J)$  and  $\ddot{\mathbf{V}} \equiv (V_j : j \in \{1, \dots, p\} - J)$ . Let  $\dot{\mathbf{U}}$  denote the set of random numbers on which  $\dot{\mathbf{V}}$  depends so that  $\dot{\mathbf{V}} = \dot{\mathcal{H}}(\dot{\mathbf{U}})$ , where  $\dot{\mathcal{H}}(\cdot)$  is the appropriate part of the overall sampling plan  $\mathcal{H}$ . If  $\ddot{\mathbf{V}}$  does not depend on  $\dot{\mathbf{U}}$ , then for any response  $W$  that is a function of  $(\dot{\mathbf{V}}, \ddot{\mathbf{V}})$  alone, we have*

$$\mathbb{E}[W | \dot{\mathbf{V}}] = \mathbb{E}[W | \dot{\mathbf{U}}].$$

**Proof.** Since  $\ddot{\mathbf{V}}$  does not depend on  $\dot{\mathbf{U}}$ , the joint distribution of  $(\dot{\mathbf{V}}, \ddot{\mathbf{V}})$  is the same whether we condition on  $\dot{\mathbf{V}}$  or  $\dot{\mathbf{U}}$ ; and the conclusion of Lemma 2 follows immediately.  $\blacksquare$

## Proof of Theorem 5

If  $Y$  is bounded, then so is  $Z \equiv E[Y|\mathbf{X}]$ ; and applying Theorem 1 of Owen (1992), we see that for any  $J \subseteq \{1, \dots, p\}$  satisfying conditions  $\text{IP}_1$  and  $\text{IP}_2$  with associated sampling plan  $\mathcal{H}_J$  defined in (44), the asymptotic variance parameter for  $\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_J, n)$  is

$$\sigma_{\text{CE+LH}}^2(\mathcal{H}_J) = \text{Var} \left[ Z - \sum_{j \in I_{\mathbf{X}}} E(Z|U_j) \right], \quad (70)$$

where the dependence of  $Z$  on  $\mathcal{H}_J$  in (70) parallels the dependence of the response on the sampling plan that was shown explicitly in (68). Applying Appendix B of Stein (1987), we also have

$$\text{Var} \left[ Z - \sum_{j \in I_{\mathbf{X}}} E(Z|U_j) \right] \leq \text{Var} \left[ Z - \sum_{j \in I_{\mathbf{X}}} \psi_j(U_j) \right] \quad (71)$$

for any univariate functions  $\{\psi_j(\cdot) : j \in I_{\mathbf{X}}\}$ . By choosing  $\psi_j(u) \equiv 0$  for  $j \in I_{\mathbf{X}}$ , we obtain  $\text{Var}(Z) = \sigma_{\text{CE}}^2$  on the right-hand side of (71); and combining this result with (70), we obtain the conclusion of part (i).

To prove part (ii), we will show that for any  $J \subseteq \{1, \dots, p\}$  satisfying conditions  $\text{IP}_1$  and  $\text{IP}_2$ , the sampling plan  $\mathcal{H}_J$  defined by (44) yields the asymptotic variance parameter

$$\sigma_{\text{CE+LH}}^2(\mathcal{H}_J) = \text{Var} \left\{ E(Y|\mathbf{X}) - \sum_{j \in I_{\mathbf{X}}} E[E(Y|U_\tau : \tau \in I_{\mathbf{X}}) | U_j] \right\} \quad (72)$$

$$= \text{Var} \left[ E(Y|\mathbf{X}) - \sum_{j \in I_{\mathbf{X}}} E(Y|U_j) \right] \quad (73)$$

$$= \text{Var} \left\{ E(Y|\mathbf{X}) - \sum_{j \in I_{\mathbf{X}}} E[E(Y|U_j) | U_\tau : \tau \in I_{\mathbf{X}}] \right\} \quad (74)$$

$$= \text{Var} \left\{ E(Y|\mathbf{X}) - \sum_{j \in I_{\mathbf{X}}} E[E(Y|U_j) | \mathbf{X}] \right\} \quad (75)$$

$$= \text{Var} \left\{ E(Y|\mathbf{X}) - \sum_{j=1}^d E[E(Y|U_j) | \mathbf{X}] \right\} \quad (76)$$

$$= \text{Var} \left\{ E \left[ Y - \sum_{j=1}^d E(Y|U_j) \middle| \mathbf{X} \right] \right\} \quad (77)$$

$$\leq \text{Var} \left[ Y - \sum_{j=1}^d E(Y|U_j) \right] = \sigma_{\text{LH}}^2(\mathcal{H}_J). \quad (78)$$

Display (72) follows from (70), the definition of  $Z$ , and an application of Lemma 2 with  $\dot{\mathbf{V}} = \mathbf{X}$ ,  $\dot{\mathbf{U}} = (U_\tau : \tau \in I_{\mathbf{X}})$ , and  $W = Y$ . Display (73) follows from the law of total probability

for conditional expectations (Karlin and Taylor 1975, p. 246); and (74) follows from the substitution theorem for conditional expectations (Bickel and Doksum 1977, p. 5). Display (75) follows from repeated application of Lemma 2 with  $\dot{\mathbf{V}} = \mathbf{X}$ ,  $\dot{\mathbf{U}} = (U_\tau : \tau \in I_{\mathbf{X}})$  and  $W = E(Y|U_j)$  for each  $j \in I_{\mathbf{X}}$ . Display (76) follows by observing that for every  $j \in \{1, \dots, d\} - I_{\mathbf{X}}$ , the random vectors  $E(Y|U_j)$  and  $\mathbf{X}$  are independent so that  $E[E(Y|U_j) | \mathbf{X}] = E[E(Y|U_j)] = \theta$  by the double expectation theorem (Bickel and Doksum 1977, p. 6). Finally, the inequality in display (78) follows by applying the conditional variance relation (4) to  $Y - \sum_{j=1}^d E(Y|U_j)$  when conditioning on  $\mathbf{X}$ . We have shown that  $\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_J, n)$  asymptotically dominates  $\hat{\theta}_{\text{LH}}(\mathcal{H}_J, n)$  for any  $J \subseteq \{1, \dots, p\}$ ; and this completes the proof of part (ii).

To prove part (iii), we observe that  $\mathbf{D} \equiv E[\mathbf{C}|\mathbf{X}]$  has components

$$\begin{aligned} D_\tau \equiv E[C_\tau|\mathbf{X}] &= \sum_{j \in J_{\mathbf{C}}} E\left[\varphi_{\tau,j}(V_j) \mid (V_\tau : \tau \in J_{\mathbf{X}})\right] \\ &= \sum_{j \in J_{\mathbf{C}} \cap J_{\mathbf{X}}} \varphi_{\tau,j}(V_j) + \sum_{j \in J_{\mathbf{C}} - J_{\mathbf{X}}} E[\varphi_{\tau,j}(V_j)], \quad \tau = 1, \dots, q, \end{aligned}$$

where the last equality follows from conditions AC<sub>1</sub> and AC<sub>2</sub>; and thus the control vector  $\mathbf{D}$  has components of the form (43). Now Theorem 4 can be applied to the response  $Z$  and the control vector  $\mathbf{D}$ , showing that  $\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_{\mathbf{C}}}, n)$  asymptotically dominates  $\hat{\theta}_{\text{CE+CV}}(n)$ . It remains to show that  $\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_{\mathbf{C}}}, n)$  asymptotically dominates  $\hat{\theta}_{\text{CV}}(n)$ . Applying part (ii) of Theorem 5 with  $J = J_{\mathbf{C}}$ , we see that  $\hat{\theta}_{\text{CE+LH}}(\mathcal{H}_{J_{\mathbf{C}}}, n)$  asymptotically dominates  $\hat{\theta}_{\text{LH}}(\mathcal{H}_{J_{\mathbf{C}}}, n)$ . Finally, Theorem 4 ensures that  $\hat{\theta}_{\text{LH}}(\mathcal{H}_{J_{\mathbf{C}}}, n)$  asymptotically dominates  $\hat{\theta}_{\text{CV}}(n)$ ; and since asymptotic dominance is a transitive relation, we obtain the desired result. ■

## ACKNOWLEDGMENT

This work was partially supported by NSF Grant No. DMS-8717799 and by a David Ross Grant from the Purdue Research Foundation. The authors thank Bruce Schmeiser for many enlightening discussions on this paper. The authors also thank the Associate Editor and the referees for numerous suggestions that improved the readability of this paper.

## REFERENCES

- ANDERSON, T. W. 1984. *An Introduction to Multivariate Statistical Analysis*, 2nd ed. John Wiley, New York.
- ANTILL, J. M., AND R. W. WOODHEAD. 1982. *Critical Path Methods in Construction Practice*. John Wiley, New York.

- ARNOLD, S. F. 1981. *The Theory of Linear Models and Multivariate Analysis*. John Wiley, New York.
- AVRAMIDIS, A. N. 1993. Variance Reduction Techniques for Simulation with Applications to Stochastic Networks. Ph.D. Dissertation. School of Industrial Engineering, Purdue University, West Lafayette, Ind.
- AVRAMIDIS, A. N., K. W. BAUER, JR. AND J. R. WILSON. 1991. Simulation of Stochastic Activity Networks Using Path Control Variates. *Naval Research Logistics* **38**, 183–201.
- AVRAMIDIS, A. N., AND J. R. WILSON. 1993a. A Splitting Scheme for Control Variates. *Operations Research Letters* **14**, 187–198.
- AVRAMIDIS, A. N., AND J. R. WILSON. 1993b. Integrated Variance Reduction Strategies. In *Proceedings of the 1993 Winter Simulation Conference*, G. W. Evans, M. Mollaghasemi, E. C. Russell, and W. E. Biles (eds.). Institute of Electrical and Electronics Engineers, Piscataway, N.J., 445–454.
- BICKEL, P. J., AND K. A. DOKSUM. 1977. *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden-Day, San Francisco.
- BURT, J. M., JR., AND M. B. GARMAN. 1971. Conditional Monte Carlo: A Simulation Technique for Stochastic Network Analysis. *Management Science* **18**, 207–217.
- BURT, J. M., JR., D. P. GAVER AND M. PERLAS. 1970. Simple Stochastic Networks: Some Problems and Procedures. *Naval Research Logistics Quarterly* **17**, 439–459.
- CARSON, J. S. 1985. On the Efficiency of Antithetic Sampling and Conditional Monte Carlo in Simulations of Stochastic Activity Networks. Technical Report. School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Ga.
- ELMAGHRABY, S. 1977. *Activity Networks: Project Planning and Control by Network Models*. Wiley-Interscience, New York.
- FISHMAN, G. S. 1989. Monte Carlo, Control Variates, and Stochastic Ordering. *SIAM Journal on Scientific and Statistical Computing* **10**, 187–204.
- GRIFFITHS, P., AND I. D. HILL. 1985. *Applied Statistics Algorithms*. Ellis Horwood Limited, Chichester, England.
- International Mathematical and Statistical Library. 1987. *STAT/LIBRARY User's Manual*. IMSL, Inc., Houston.
- KARLIN, S., AND H. M. TAYLOR. 1975. *A First Course in Stochastic Processes*, 2nd ed. Academic Press, New York.
- KLEIJNEN, J. P. C. 1975. Antithetic Variates, Common Random Numbers and Optimal Computer Time Allocation in Simulation. *Management Science* **21**, 1176–1185.
- KWON, C., AND J. D. TEW. 1994. Strategies for Combining Antithetic Variates and

- Control Variates in Designed Simulation Experiments. *Management Science* **40**, 1021–1034.
- LAVENBERG, S. S., T. L. MOELLER AND P. D. WELCH. 1982. Statistical Results on Control Variables with Application to Queueing Network Simulation. *Operations Research* **30**, 182–202.
- L’ECUYER, P. 1994. Efficiency Improvement and Variance Reduction. In *Proceedings of the 1994 Winter Simulation Conference*, J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila (eds.). Institute of Electrical and Electronics Engineers, Piscataway, N.J., 122–132.
- LEHMANN, E. L. 1966. Some Concepts of Dependence. *Annals of Mathematical Statistics* **37**, 1137–1153.
- LOULOU, R., AND T. BEALE. 1976. A Comparison of Variance Reducing Techniques in PERT Simulations. *INFOR* **14**, 259–269.
- MCKAY, M. D., R. J. BECKMAN AND W. J. CONOVER. 1979. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics* **21**, 239–245.
- NELSON, B.L. 1987. A Perspective on Variance Reduction in Dynamic Simulation Experiments. *Communications in Statistics—Simulation and Computation* **B16**, 385–426.
- NELSON, B. L. 1990. Control Variate Remedies. *Operations Research* **38**, 974–992.
- NEUTS, M. F. 1973. *Probability*. Allyn and Bacon, Boston.
- OWEN, A. B. 1992. A Central Limit Theorem for Latin Hypercube Sampling. *Journal of the Royal Statistical Society* **B54**, 541–551.
- PORTA NOVA, A. M. DE O., AND J. R. WILSON. 1989. Estimation of Multiresponse Simulation Metamodels Using Control Variates. *Management Science* **35**, 1316–1333.
- PROVAN, J. S., AND V. G. KULKARNI. 1989. Exact Cuts in Networks. *Networks* **19**, 281–289.
- ROBERTS, S. D. 1983. *Simulation Modeling and Analysis with INSIGHT*. Regenstrief Institute, Indianapolis, Ind.
- SCHRUBEN L. W., AND B. H. MARGOLIN. 1978. Pseudorandom Number Assignment in Statistically Designed Simulation and Distribution Sampling Experiments. *Journal of the American Statistical Association* **73**, 504–525.
- SEARLE, S. R. 1982. *Matrix Algebra Useful for Statistics*. John Wiley, New York.
- SIGAL, C. E., A. A. B. PRITSKER AND J. J. SOLBERG. 1980. The Stochastic Shortest Route Problem. *Operations Research* **28**, 1122–1129.
- STEIN, M. 1987. Large Sample Properties of Simulations Using Latin Hypercube Sampling.

- Technometrics* **29**, 143–151.
- Systems Modeling Corporation. 1989. *The SIMAN IV Reference Guide*. Systems Modeling Corporation, Sewickley, Pa.
- TEW, J. D., AND J. R. WILSON. 1992. Validation of Simulation Analysis Methods for the Schruben-Margolin Correlation-Induction Strategy. *Operations Research* **40**, 87–103.
- TEW, J. D., AND J. R. WILSON. 1994. Estimating Simulation Metamodels Using Combined Correlation-Based Variance Reduction Techniques. *IIE Transactions* **26**, 2–16.
- WHITT, W. 1976. Bivariate Distributions with Given Marginals. *The Annals of Statistics* **4**, 1280–1289.
- WILSON, J. R. 1984. Variance Reduction Techniques for Digital Simulation. *American Journal of Mathematical and Management Sciences* **4**, 277–312.
- WOLFF, R. W. 1989. *Stochastic Modeling and the Theory of Queues*. Prentice Hall, Englewood Cliffs, N.J.