Proceedings of the 1990 Winter Simulation Conference
Osman Balci, Randall P. Sadowski, Richard E. Nance (eds.)

CONTROL VARIATES FOR STOCHASTIC NETWORK SIMULATION

Thanos N. Avramidis
James R. Wilson

School of Industrial Engineering
Purdue University
West Lafayette, Indiana 47907

ABSTRACT

In this paper we examine several procedures for using path control variates to improve the accuracy of simulation-based point and confidence-interval estimators of the mean completion time of a stochastic activity network (SAN). Because each path control variate is the duration of the corresponding directed path in the network from the source to the sink, the vector of selected path controls has both a known mean and a known covariance matrix. This information is incorporated into control-variates estimation procedures that do not require normally distributed responses. The simulation-generated observations are split into three groups, and control-variates procedures applied within each group are combined in such a way that the overall point estimator and the associated variance estimator are always unbiased. To evaluate the performance of these procedures experimentally, we examine the bias, variance, and mean square error of the controlled point estimators as well as the average half-length and coverage probability of the corresponding confidence-interval estimators for some selected SANs that are typical of large-scale PERT applications. The experimental results show that in comparison to standard linear control-variates procedures, the proposed procedures can yield substantial improvements in point-estimator accuracy and confidence-interval coverage while achieving almost the same magnitude of variance reduction.

1. INTRODUCTION

Stochastic activity networks (SANs) are widely used in the scheduling and management of large projects. However, the analysis of such networks is greatly complicated by stochastic dependencies among network components that arise, for example, when some activities are common to several paths or when several activity durations are correlated. Conventional analysis techniques are based on restrictive assumptions about the probability distributions of the activity durations or about the topology of the network [Grubbs 1962, Littlefield and Randolph 1987, MacCrimmon and Ryavec 1964, McBride and McClelland 1967]; and these assumptions generally yield approximations of unknown accuracy. Because of its ability to represent faithfully the dependencies among the components of a stochastic activity network and to yield estimates of desired performance measures with controllable accuracy, Monte Carlo simulation is frequently the method of choice for the analysis of such networks.

In the simulation of a SAN, the usual objective is to obtain point and confidence interval estimators for the mean completion time θ of the network. Let the random variable Y denote the completion time of a given SAN. Direct simulation simply computes the sample mean response $\overline{Y}$ from n independent replications of the network to yield an unbiased estimator of θ with $\mathrm{var}(\overline{Y}) = \mathrm{var}(Y)/n$. Since the variance of $\overline{Y}$ declines as the inverse of the sample size, a large number of replications will usually be required to achieve acceptable precision (for example, see Table 2 below). Computing costs can then become prohibitive, and we naturally seek to derive an alternative estimator $\hat{\theta}$ with $E(\hat{\theta}) \simeq \theta$ and $E[(\hat{\theta}-\theta)^2] < \mathrm{var}(\overline{Y})$.

Several variance reduction techniques have been proposed for improving the efficiency of activity network simulations, including conditional Monte Carlo [Burt and Garman 1971, Kulkarni and Provan 1985], stratified sampling [Burt et al. 1970, Loulou and Beale 1976], antithetic sampling [Sullivan et al. 1982], control variates [Burt et al. 1970], and combinations of these techniques [Burt et al. 1970, Loulou and Beale 1976]. Some of our recent work [Bauer et al. 1987, Venkatraman and Wilson 1985] has led us to the conclusion that in comparison to the other commonly used variance reduction techniques, the method of control variates is more easily adapted to a wide variety of network configurations and has greater potential to yield large efficiency increases in general applications. To estimate the target parameter θ using the method of control variates, we identify a set of auxiliary variables $C = (C_1, \ldots, C_q)'$ that are generated by the same stochastic system, have a known expectation μ_C, and are strongly correlated with the response Y. We then try to predict and counteract the unknown deviation $Y - \theta$ by subtracting from Y an appropriate linear function of the known deviation $C - \mu_C$. The objective then is to determine a vector of control coefficients $b \equiv (b_1, \ldots, b_q)'$ that will minimize the variance of the controlled estimator $Y(b) = Y - b'(C - \mu_C)$. Lavenberg, Moeller, and Welch [1982] presented a comprehensive analysis of the control variate method for univariate responses.

This paper is organized as follows. In Section 2 we summarize the necessary statistical framework for applying (a) some standard linear control-variate procedures based on the assumption of normal responses, and (b) distribution-free variants of the standard control-variates procedures based on splitting the simulation-generated sample into three subsamples. This splitting technique enables us to construct controlled point and confidence-interval estimation procedures that are robust against nonnormality. In Section 3 we summarize some basic results about the moment structure of path control variates in stochastic activity networks. Section 4 details the results of an extensive experimental investigation of the performance of the various estimation procedures based on path control variates. In Section 5 we discuss the significance of the experimental results, and we summarize the main findings of this research in Section 6. This paper is partially based on results that were originally presented in Venkatraman and Wilson [1985], Bauer et al. [1987], and Avramidis et al. [1990].

2. STATISTICAL FRAMEWORK

The following notation is used throughout this paper. Let $\sigma_Y^2 = E[(Y-\theta)^2]$ denote the variance of the completion time Y for the target network, let $\boldsymbol{\sigma}_{CY} = E[(\boldsymbol{C}-\boldsymbol{\mu}_C)(Y-\theta)]$ denote the $q \times 1$ vector of covariances between the control vector $\boldsymbol{C}$ and the response Y, and let $\boldsymbol{\Sigma}_C = E[(\boldsymbol{C}-\boldsymbol{\mu}_C)(\boldsymbol{C}-\boldsymbol{\mu}_C)']$ denote the $q \times q$ covariance matrix of the controls. The variance of the controlled response

$$\mathrm{var}[Y(\boldsymbol{b})] = \sigma_Y^2 - 2\boldsymbol{b}'\boldsymbol{\sigma}_{CY} + \boldsymbol{b}'\boldsymbol{\Sigma}_C\boldsymbol{b}$$

is minimized by the optimal vector of control coefficients

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}_C^{-1}\boldsymbol{\sigma}_{CY}, \tag{1}$$

yielding the minimum variance

$$\mathrm{var}[Y(\boldsymbol{\beta})] = \sigma_Y^2(1 - R_{YC}^2), \tag{2}$$

where R_{YC} is the coefficient of multiple correlation between Y and $\boldsymbol{C}$. In practice, $\boldsymbol{\beta}$ must be estimated because at least $\boldsymbol{\sigma}_{CY}$ is generally unknown; and in many applications $\boldsymbol{\Sigma}_C$ is also unknown so that both terms on the right-hand side of (1) must be estimated from simulation-generated data. Estimation of $\boldsymbol{\beta}$ results in some loss of precision for the controlled point estimators of θ described below.

2.1 Control-Variates Procedures for Normal Reponses

First we summarize the conventional method for applying control variates to the estimation of θ. Let $\{[Y_u, \boldsymbol{C}_u'] : u = 1, \ldots, n\}$ denote the results observed on n independent replications of a simulation of the target network. Let $\overline{Y}$ and $\overline{\boldsymbol{C}}$ respectively denote the sample means of the response and the control vector computed over all n replications; and let S_Y^2, $\boldsymbol{S}_{CY}$, and $\boldsymbol{S}_C$ respectively denote the sample analogs of σ_Y^2, $\boldsymbol{\sigma}_{CY}$, and $\boldsymbol{\Sigma}_C$. Specifically we compute

$$\boldsymbol{S}_C = (n-1)^{-1}\sum_{u=1}^{n}(\boldsymbol{C}_u - \overline{\boldsymbol{C}})(\boldsymbol{C}_u - \overline{\boldsymbol{C}})'$$

and

$$\boldsymbol{S}_{CY} = (n-1)^{-1}\sum_{u=1}^{n}(\boldsymbol{C}_u - \overline{\boldsymbol{C}})(Y_u - \overline{Y})$$

so that the sample analog of (1) is

$$\hat{\boldsymbol{\beta}} = \boldsymbol{S}_C^{-1}\boldsymbol{S}_{CY}; \tag{3}$$

and the conventional controlled point estimator of θ is

$$\overline{Y}(\hat{\boldsymbol{\beta}}) = \overline{Y} - \hat{\boldsymbol{\beta}}'(\overline{\boldsymbol{C}} - \boldsymbol{\mu}_C). \tag{4}$$

In general, $\overline{Y}(\hat{\boldsymbol{\beta}})$ is a biased estimator of θ because $\hat{\boldsymbol{\beta}}$ and $\overline{\boldsymbol{C}}$ are dependent so that $E[\hat{\boldsymbol{\beta}}'(\overline{\boldsymbol{C}} - \boldsymbol{\mu}_C)] \neq 0$. However, in many large-scale simulation experiments, the response and the controls are (approximately) jointly normal because these statistics are simultaneously accumulated over the duration of each run and thus are subject to a central-limit effect (see Cheng [1978]). Thus it is often reasonable to assume that Y and $\boldsymbol{C}$ have a joint multivariate normal distribution

$$\begin{bmatrix} Y \\ \boldsymbol{C} \end{bmatrix} \sim N_{q+1}\left(\begin{bmatrix} \theta \\ \boldsymbol{\mu}_C \end{bmatrix}, \begin{bmatrix} \sigma_Y^2 & \boldsymbol{\sigma}'_{CY} \\ \boldsymbol{\sigma}_{CY} & \boldsymbol{\Sigma}_C \end{bmatrix}\right). \tag{5}$$

If (5) holds, then $\overline{Y}(\hat{\boldsymbol{\beta}})$ is an unbiased estimator of θ; and an exact $100(1-\alpha)\%$ confidence interval for θ is given by

$$\overline{Y}(\hat{\boldsymbol{\beta}}) \pm t_{1-\alpha/2}(n-q-1)DS_{Y \cdot C}, \tag{6}$$

where

$$D^2 = n^{-1} + (n-1)^{-1}(\overline{\boldsymbol{C}} - \boldsymbol{\mu}_C)'\boldsymbol{S}_C^{-1}(\overline{\boldsymbol{C}} - \boldsymbol{\mu}_C),$$

$$S_{Y \cdot C}^2 = (n-q-1)^{-1}(n-1)(S_Y^2 - \boldsymbol{S}'_{CY}\boldsymbol{S}_C^{-1}\boldsymbol{S}_{CY}),$$

and $t_{1-\alpha/2}(n-q-1)$ is the quantile of order $1-\alpha/2$ for Student's t-distribution with $n-q-1$ degrees of freedom (see Lavenberg et al. [1982]).

As the basis for a similar statistical-estimation procedure that exploits the known covariance matrix of the control vector, we have previously proposed the following alternative estimator for the control coefficient vector

$$\ddot{\boldsymbol{\beta}} = \boldsymbol{\Sigma}_C^{-1}\boldsymbol{S}_{CY} \tag{7}$$

[Venkatraman and Wilson 1985]; and this leads to a controlled point estimator for θ with the form

$$\overline{Y}(\ddot{\boldsymbol{\beta}}) = \overline{Y} - \ddot{\boldsymbol{\beta}}'(\overline{\boldsymbol{C}} - \boldsymbol{\mu}_C). \tag{8}$$

Under the assumption of joint multivariate normality in (5), Bauer [2] proved that $\overline{Y}(\ddot{\boldsymbol{\beta}})$ is an unbiased estimator of θ. Furthermore, an approximate $100(1-\alpha)\%$ confidence interval for θ is given by

$$\overline{Y}(\ddot{\beta}) \pm t_{1-\alpha/2}(n-q-1)\left[\frac{q+1}{n(n-1)}S_Y^2 + \frac{n-2}{n(n-1)}S_{Y\cdot C}^2\right]^{1/2}. \quad (9)$$

A comprehensive analysis of the controlled estimator $\overline{Y}(\ddot{\beta})$ is given in Bauer [1987].

2.2 Control-Variates Procedures for Nonnormal Responses

In Avramidis et al. [1990] we observed that nonnormality of the network completion time and the path controls can induce a substantial bias in the point estimators (4) and (8), and this bias can lead to severe degradation of coverage probability for the associated confidence intervals (6) and (9). We also observed that although jackknifing alleviates these problems, most of the potential variance reduction is lost with this approach. As an alternative to jackknifing for nonnormal responses, we propose a variant of what is sometimes referred to as "splitting" [Nelson 1989]. The complete sample $\{[Y_u, C_u'] : u = 1, \ldots, n\}$ is split into three groups or subsamples; thus from the gth group ($g = 1, 2, 3$)

$$\Gamma_g \equiv \{\, [Y_u, C_u'] : u \in U_g \,\}$$

composed of observations whose indices belong to the index-set

$$U_g \equiv \{u : u = \lfloor (g-1)n/3 \rfloor + 1, \ldots, \lfloor gn/3 \rfloor\},$$

we compute the control coefficient estimates analogous to (3) and (7)

$$\hat{\beta}_g = \hat{\beta}_g(\Gamma_g) \text{ and } \ddot{\beta}_g = \ddot{\beta}_g(\Gamma_g).$$

Next we define the circular group-assignment function

$$\tau(g) \equiv g(\mathrm{mod}\,3)+1 \quad (g = 1, 2, 3)$$

so that for each group Γ_g, the control coefficient vector to be applied to the observations in that group is computed from the assigned group $\Gamma_{\tau(g)}$. Thus the uth controlled response in the gth group is given by

$$Y_u[\hat{\beta}_{\tau(g)}] = Y_u - \hat{\beta}'_{\tau(g)}(C_u - \mu_C) \quad \text{for } u \in U_g$$

and $g = 1, 2, 3$. Similarly we can define $Y_u[\ddot{\beta}_{\tau(g)}]$ for all u and g. Since simulation-generated observations from different groups are stochastically independent, the uth observation $[Y_u, C_u']$ in group Γ_g and the estimated control coefficient vector $\hat{\beta}_{\tau(g)}$ are stochastically independent. It follows that the controlled observations $\{Y_u[\hat{\beta}_{\tau(g)}] : u \in U_g\}$ are pairwise uncorrelated with expected value θ. A similar argument shows that the $\{Y_u[\ddot{\beta}_{\tau(g)}] : u \in U_g\}$ are also pairwise uncorrelated with expected value θ.

For simplicity in the following development, we only discuss our proposed extension of the standard linear control-variate estimator (4). From the properties cited in the preceding paragraph, it can be shown that the point estimators

$$\overline{Y}_g \equiv n_g^{-1} \sum_{u \in U_g} Y_u[\hat{\beta}_{\tau(g)}] \quad \text{with} \quad n_g \equiv |U_g|$$

and

$$S_g^2 \equiv (n_g - 1)^{-1} \sum_{u \in U_g} \left\{ Y_u[\hat{\beta}_{\tau(g)}] - \overline{Y}_g \right\}^2$$

are respectively unbiased estimators of θ and $\mathrm{var}\{Y_u[\hat{\beta}_{\tau(g)}]\}$. Thus the analogue of (4) based on splitting is

$$\overline{Y}_{\bullet}(\hat{\beta}) = \frac{1}{3}\sum_{g=1}^{3} \overline{Y}_g; \quad (10)$$

and the analogue of (6) is

$$\overline{Y}_{\bullet}(\hat{\beta}) \pm t_{1-\alpha/2}(\hat{\nu}_{\mathrm{eff}}) \frac{S_{\bullet}^2(\hat{\beta})}{\sqrt{n}}, \quad (11)$$

where: (a) the pooled variance estimator $S_{\bullet}^2(\hat{\beta})$ is the average of the three variance estimators $\{S_g^2 : g = 1, 2, 3\}$; and (b) the so-called "effective degrees of freedom" $\hat{\nu}_{\mathrm{eff}}$ in (11) is equal to two divided by the squared sample coefficient of variation of the $\{S_g^2\}$

$$\hat{\nu}_{\mathrm{eff}} = \frac{4\left[S_{\bullet}^2(\hat{\beta})\right]^2}{\sum_{g=1}^{3}\left[S_g^2 - S_{\bullet}^2(\hat{\beta})\right]^2}. \quad (12)$$

Note that the analogues of (10) and (11) for linear control-variate procedures based on a known Σ_C will be respectively denoted by

$$\overline{Y}_{\bullet}(\ddot{\beta}) \text{ and } \overline{Y}_{\bullet}(\ddot{\beta}) \pm t_{1-\alpha/2}(\ddot{\nu}_{\mathrm{eff}}) \frac{S_{\bullet}^2(\ddot{\beta})}{\sqrt{n}}, \quad (13)$$

where $\ddot{\nu}_{\mathrm{eff}}$ is given by the right-hand side of (12) when $S_{\bullet}^2(\hat{\beta})$ is replaced with $S_{\bullet}^2(\ddot{\beta})$.

3. ESTIMATION WITH PATH CONTROLS

The graph-theoretic structure of a given SAN is described by the pair (Ψ, Δ), where the set of all nodes (vertices) in the network is $\Psi \equiv \{1, 2, \ldots, \psi\}$, and the set of all activities (directed lines) in the network is $\Delta \equiv \{(u_i, v_i)$: activity i has start node $u_i \in \Psi$ and end node $v_i \in \Psi$, $i = 1, \ldots, \delta\}$. We assume that the network is acyclic with a single source node and a single sink node. The probabilistic structure of the network is described by the given joint distribution function $F(a_1, \ldots, a_\delta)$ of the random vector $(A_1, \ldots, A_\delta)$ whose ith element A_i is the duration of the ith activity $(u_i, v_i) \in \Delta$. Thus for $i = 1, 2, \ldots, \delta$, the activity duration A_i has a known marginal distribution $F_i(a_i)$ whose mean μ_i and variance σ_i^2 can at least be evaluated numerically. Moreover for $h, i = 1, \ldots, \delta$, the covariance σ_{hi} between the activity durations A_h and A_i is also known or can be evaluated numerically. (In many SANs the activity durations are assumed to be stochastically independent so that $\sigma_{hi} = 0$ for

$h \neq i$.)

The path controls and the overall network completion time depend on the path structure of the network as follows. Let ξ denote the number of directed paths from the source to the sink. Corresponding to the jth directed path π_j is the index-set of component arcs $I(j) \equiv \{i$: activity (u_i, v_i) is on path $\pi_j\}$ for $j = 1, \ldots, \xi$. The duration of path π_j is the random variable

$$P_j \equiv \sum_{i \in I(j)} A_i$$

with mean and variance

$$E(P_j) = \sum_{i \in I(j)} \mu_i \text{ and } \mathrm{var}(P_j) = \sum_{i \in I(j)} \sigma_i^2 + \sum_{\substack{h,\, i \in I(j) \\ h \neq i}} \sigma_{hi} \quad (14)$$

respectively. Note also that for $j, l = 1, \ldots, \xi$, the covariance between the path durations P_j and P_l is

$$\mathrm{Cov}(P_j, P_l) = \sum_{i \in I(j) \cap I(l)} \sigma_i^2 + \sum\sum_{\substack{h \in I(j) \\ i \in I(l) \\ h \neq i}} \sigma_{hi}\,. \quad (15)$$

The overall project completion time is $Y \equiv \max\{P_1, \ldots, P_\xi\}$, and the desired estimand is $\theta = E(Y)$.

From the set $\{P_j : j = 1, \ldots, \xi\}$ of all path durations for the stochastic activity network (Ψ, Δ), we must select the control vector $\boldsymbol{C}$ that is to be applied to point estimators of θ with the form (4), (8), (10), or (13). For simplicity in the experimental evaluation of all of these controlled estimators, we employed the following control-variate selection rule. Ranking the expected path durations $\{E(P_j) : j = 1, \ldots, \xi\}$ in ascending order so that

$$E[P_{(1)}] \le E[P_{(2)}] \le \cdots \le E[P_{(\xi)}], \quad (16)$$

we chose the last q paths in this list to build the control vector

$$\boldsymbol{C} = [P_{(\xi - q + 1)}, P_{(\xi - q + 2)}, \ldots, P_{(\xi - 1)}, P_{(\xi)}]'.$$

(In all of the experimentation described in the next two sections, we took $q = 3$ for reasons that are detailed in Avramidis et al. [1990].) The mean vector $\boldsymbol{\mu}_C$ and the dispersion matrix Σ_C of the resulting q-dimensional control vector were then computed from (14) and (15). Of course in general applications the user would need some guidance in determining q and in selecting the appropriate set of q controls from the set of available path controls, but this is a separate issue that is not addressed in this paper. For controlled estimators of the form (4) and (8), Bauer and Wilson [1989] have devised some control-variate selection criteria that are based on minimizing the mean square volume of the delivered confidence region and that appear to be effective when the normality assumption (5) holds; however a more extensive Monte Carlo study is required to support any general conclusions about the performance of these selection procedures.

4. EXPERIMENTAL EVALUATION

We conducted an extensive Monte Carlo study to evaluate the performance of the following controlled estimators for θ: $\overline{Y}(\hat{\beta})$ (estimator 1, also denoted $\hat{\theta}(1)$); $\overline{Y}(\ddot{\beta})$ (estimator 2, also denoted $\hat{\theta}(2)$); $\overline{Y}_*(\hat{\beta})$ (estimator 3, also denoted $\hat{\theta}(3)$); and $\overline{Y}_*(\ddot{\beta})$ (estimator 4, also denoted $\hat{\theta}(4)$). This study involved the simulation of a set of three SANs in which the following characteristics were systematically varied: (a) the size of the network (the number of nodes and activities); (b) the topology of the network; (c) the percentage of activities with exponentially distributed durations; and (d) the relative dominance (criticality index) of the critical path. All efficiency gains were reported relative to the direct simulation estimator $\overline{Y}$ (estimator 0, also denoted $\hat{\theta}(0)$). Table 1 shows the range in the number of nodes and activities for the three networks used in the study. Note that (a) network 1 was taken from page 275 of Elmaghraby [1977], (b) network 2 was taken from page 190 of Antill and Woodhead [1982], and (c) network 3 was taken from page 324 of McKenney and Rosenbloom [1969].

Table 1. Stochastic Activity Networks Used in the Experiments

Network	Nodes	Activities
1	9	13
2	24	42
3	51	65

For each activity duration A_i in a selected network, the associated distribution $F_i(a_i)$ was taken to be either (a) a normal distribution with the specified mean μ_i and standard deviation $\sigma_i = \mu_i/4$ whose tail was truncated below the cutoff value zero; or (b) an exponential distribution with mean μ_i. We chose the exponential distribution as the nonnormal alternative because it has a higher coefficient of variation (equal to 1) than the beta and triangular distributions commonly used in the simulation of SANs [Loulou and Beale 1976], and this property partially counteracts the central-limit effect described in Section 2. Sullivan, Hayya, and Schaul [1982] used a similar approach in their experimental study. For each of the three networks, we varied the percentage of exponentially distributed activity durations over the five levels $\{0\%, 25\%, 50\%, 75\%, 100\%\}$. For each network and for each specified percentage of exponentially distributed activities, we assigned appropriate exponential distributions to the activities in the network according to a series of independent Bernoulli trials with success probability equal to the specified percentage of exponential activities; moreover, this assignment was made prior to performing any simulations of the network.

Relative dominance of a given path π_j is defined as the probability $\Pr\{Y = P_j\}$ that path π_j has the longest duration in a single realization of the network. For each network and for each selected percentage of exponentially distributed activities, we rescaled the expected duration of each activity on the so-called "critical path" (that is, the path with the longest expected duration) to achieve a prespecified level of relative dominance for that path. In view of (16), the critical path is $\pi_{(\xi)}$ with mean duration $E[P_{(\xi)}]$; and the associated index-set of activities on the critical path is denoted $I[(\xi)] = \{i : \text{activity } (u_i, v_i) \text{ is on path } \pi_{(\xi)}\}$. For every $i \in I[(\xi)]$, a common scale factor ϕ was multiplied by the nominal mean μ_i° to yield the actual mean $\mu_i = \phi\mu_i^\circ$ that was used when sampling A_i. For each network to be simulated, we determined empirically values of ϕ that achieved levels of relative dominance for the critical path in the following ranges: $\{20\%-40\%, 50\%-70\%, 80\%-100\%\}$.

For a given configuration of a target activity network (that is, for a selected SAN with a given level of relative dominance and a given percentage of exponentially distributed activities), we determined the corresponding mean completion time θ to within $\pm 0.2\%$ of its true value by a preliminary Monte Carlo experiment involving direct simulation of the target network; and the final number of replications N^* in this preliminary experiment was determined by a relative-precision stopping rule. To estimate θ by a $100(1-\alpha)\%$ confidence interval of the form $\overline{Y} \pm \gamma|\overline{Y}|$ that is asympotically consistent and efficient as the percentage error tolerance $\gamma \to 0$, we employed a variant of Nádas's [1969] sequential confidence-interval estimation procedure that was proposed by Law, Kelton, and Koenig [1981]. For a fixed number of replications n of the target network, let $\overline{Y}(n)$ and $S_Y(n)$ denote the corresponding sample mean and standard deviation of the observed network completion times. Given prespecified values of the percentage error tolerance γ, the confidence coefficient α, and the preliminary sample size n_0, we determined the final number of replications according to the stopping rule

$$N^* = \min\left\{ n : n \geq n_0,\ n = 0 \,(\mathrm{mod}\ 10),\ S_Y(n) > 0, \text{ and } t_{1-\alpha/2}(n-1)\frac{S_Y(n)}{\sqrt{n}} \leq \gamma|\overline{Y}(n)| \right\}, \tag{17}$$

and the "true" value of θ was taken to be $\overline{Y}(N^*)$. In the preliminary experimentation with the selected networks, we took $n_0 = 1000$, $\gamma = 0.002$, and $\alpha = 0.01$. Table 2 summarizes the results of the preliminary experimentation to determine θ for each network configuration in which 50% of the activities were assigned an exponentially distributed duration. The results for other percentages of exponentially distributed activity durations are not displayed because they were not used in the final analysis of the controlled estimation procedures as explained in Section 5 below. To simulate each network configuration as efficiently as possible, we developed a general simulation program for stochastic activity networks that uses the discrete-event component of the SLAM II simulation language [Pritsker 1986] and that is available from the authors on request.

Inspection of Table 2 reveals that the true value of θ for each selected network configuration has been determined to at least two significant figures; but for some configurations, the third significant figure is in question. Throughout the rest of this paper, the estimates of θ in Table 2 are taken to be the "true" values of θ for each selected network configuration.

In the main experimental evaluation of the proposed estimation procedures for each selected network configuration, we conducted a *metaexperiment* composed of $m = 2048$ basic experiments that were executed separately and independently. A basic experiment involved $n = 96$ simulation runs (independent replications) of the target activity network; and from each basic experiment we computed point and confidence-interval estimators of θ using $q = 3$ selected path controls. To provide a fair assessment of the efficiency gains achieved in the simulation of each network, we estimated the bias, variance, and mean square error for each of the four controlled point estimators; moreover, for each of the four controlled confidence-interval estimators, we estimated the actual coverage probability of a nominal 90 percent confidence interval as well as the percentage reduction in confidence-interval half-length relative to direct simulation.

Properties of the controlled point estimators of θ were evaluated as follows. From the n simulation runs comprising the wth basic experiment ($w = 1, \ldots, m$), we computed the kth point estimator $\hat{\theta}_w(k)$ ($k = 0, \ldots, 4$). Across the entire metaexperiment, we computed the grand mean $\hat{\hat{\theta}}(k)$ and the sample variance $\hat{\hat{V}}(k)$ of the replicates $\{\hat{\theta}_w(k) : w = 1, \ldots, m\}$

$$\left.\begin{aligned} \hat{\hat{\theta}}(k) &= m^{-1}\sum_{w=1}^{m}\hat{\theta}_w(k) \\ \hat{\hat{V}}(k) &= (m-1)^{-1}\sum_{w=1}^{m}\left[\hat{\theta}_w(k) - \hat{\hat{\theta}}(k)\right]^2 \end{aligned}\right\}. \tag{18}$$

Thus for the kth controlled point estimator $\hat{\theta}_w(k)$ of the mean completion time θ, the corresponding bias was estimated by $\hat{\hat{\theta}}(k) - \theta$, the variance was estimated by $\hat{\hat{V}}(k)$, and the mean square error was estimated by

$$\hat{\widehat{\mathrm{MSE}}}(k) = m^{-1}\sum_{w=1}^{m}\left[\hat{\theta}_w(k) - \theta\right]^2. \tag{19}$$

Properties of the controlled confidence-interval estimators of θ were evaluated as follows. From the n simulation runs comprising the wth basic experiment ($w = 1, \ldots, m$), we computed not only $\hat{\theta}_w(k)$ but also $\hat{V}_w(k)$, the estimator of $\mathrm{var}[\hat{\theta}_w(k)]$; see (22) below. Thus in the wth basic experiment,

Table 2. Estimates of θ with Relative Precision $\pm 0.2\%$ at the 99% Confidence Level

Network	Relative Dominance of $\pi_{(\xi)}$	Final Estimate of θ	Final Sample Size N^*
	20% – 40%	57.00	232970
1	50% – 70%	88.31	287410
	80% – 100%	164.78	362310
	20% – 40%	981.37	194060
2	50% – 70%	1616.98	252640
	80% – 100%	2733.76	271200
	20% – 40%	68.39	122230
3	50% – 70%	81.43	93590
	80% – 100%	141.79	81960

the kth controlled confidence-interval estimator of θ has the general form

$$\hat{\Theta}_w(k) \equiv \left[\hat{\theta}_w(k) - \hat{H}_w(k),\ \hat{\theta}_w(k) + \hat{H}_w(k)\right] \tag{20}$$

with half-length

$$\hat{H}_w(k) \equiv t_{1-\alpha/2}[\nu(k)] \cdot \hat{V}_w^{1/2}(k)\,, \tag{21}$$

where: $\nu(k) \equiv n-1$ if $k = 0$, 3, or 4; $\nu(k) \equiv n-q-1$ if $k = 1$ or 2; and

$$\hat{V}_w(k) \equiv \begin{cases} S_Y^2/n & \text{if } k=0 \\ D^2 S_{Y\cdot C}^2 & \text{if } k=1 \\ [(q+1)S_Y^2 + (n-2)S_{Y\cdot C}^2]/[n(n-1)] & \text{if } k=2\,. \\ S_\cdot^2(\hat{\beta})/n & \text{if } k=3 \\ S_\cdot^2(\ddot{\beta})/n & \text{if } k=4 \end{cases} \tag{22}$$

Note that for $k = 1$, 2, 3, and 4, display (20) specializes to the forms (6), (9), (11), and (13) respectively. The average half-length of the kth confidence-interval estimator $\hat{\Theta}(k)$ computed across all m basic experiments is

$$\hat{\hat{H}}(k) \equiv m^{-1} \sum_{w=1}^{m} \hat{H}_w(k) \tag{23}$$

for $k = 0, 1, 2, 3, 4$; and the percentage reduction in half-length for $\hat{\Theta}(k)$ relative to direct simulation is estimated by

$$100[\hat{\hat{H}}(0) - \hat{\hat{H}}(k)]/\hat{\hat{H}}(0)\,.$$

Finally we consider the estimation of confidence-interval coverage probabilities. For the wth basic experiment, let

$$\hat{I}_w(k) \equiv \begin{cases} 1 & \text{if } \theta \in \hat{\Theta}_w(k) \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

for $k = 0, 1, 2, 3, 4$, and $w = 1, \ldots, m$. The actual coverage percentage for $\hat{\Theta}(k)$ is then given by $100\hat{\hat{I}}(k)$, where

$$\hat{\hat{I}}(k) \equiv m^{-1} \sum_{w=1}^{m} \hat{I}_w(k) \quad \text{for} \quad k = 0, 1, 2, 3, 4. \tag{25}$$

5. EXPERIMENTAL RESULTS

From previous experimentation [Avramidis et al. 1990] with the given networks, we observed marked departures from the joint normality assumption (5). This motivated our development of the controlled estimators (10), (11), and (13) based on splitting. We also observed that the percentage of exponentially distributed activities had no effect on confidence interval coverage, but it did affect the half-length of the confidence intervals. Since the worst-case behavior was observed with the percentage of exponentially distributed activities in the range 25% – 50%, we fixed this factor at the 50% level throughout the rest of this study.

For each controlled point estimator of θ, the corresponding sample bias is summarized in Table 3 for each network configuration with 50% exponential activity durations. As detailed in Section 4, the results in Table 3 and in all subsequent tables were computed from metaexperiments composed of $m = 2048$ basic experiments, and each basic experiment consisted of $n = 96$ independent simulation runs. From Table 3 we see that the estimators $\overline{Y}(\hat{\beta})$ and $\overline{Y}(\ddot{\beta})$ designed for normal responses (that is, $\hat{\theta}(1)$ and $\hat{\theta}(2)$ respectively) possess a marked negative bias; and the bias of $\overline{Y}(\ddot{\beta})$ appears to be an order of magnitude greater in absolute value than the bias of the conventional controlled estimator $\overline{Y}(\hat{\beta})$. By itself, this bias may not have much practical significance since it is less than 3% of the estimand θ for each network configuration studied; however, this bias may significantly affect the performance of the corresponding confidence-interval estimators. This issue will be elaborated in the discussion given below on confidence-interval estimators. Finally, we note that the estimators $\overline{Y}_\cdot(\hat{\beta})$ and $\overline{Y}_\cdot(\ddot{\beta})$ based on splitting (that is, $\hat{\theta}(3)$ and $\hat{\theta}(4)$ respectively) have negligible bias.

Table 4 summarizes the sample variance of each of the four controlled point estimators of θ. The estimator $\overline{Y}(\ddot{\beta})$ based on the known covariance structure of the controls consistently displayed more variability than the conventional controlled estimator $\overline{Y}(\hat{\beta})$. It is interesting to note that the point estimators $\overline{Y}_{\bullet}(\hat{\beta})$ and $\overline{Y}_{\bullet}(\ddot{\beta})$ frequently displayed nearly the same variability as their conventional counterparts.

For each controlled point estimator of θ, the sample mean square error defined by equation (19) and displayed in Table 5 provides a single figure of merit that incorporates both the bias and variance of that estimator. In all cases the conventional point estimator $\overline{Y}(\hat{\beta})$ dominated the point estimator $\overline{Y}(\ddot{\beta})$ based on the known covariance structure of the controls. Moreover, the estimator $\overline{Y}_{\bullet}(\hat{\beta})$ was clearly superior to $\overline{Y}_{\bullet}(\ddot{\beta})$ with respect to mean square error. These conclusions are to some extent surprising since $\overline{Y}(\ddot{\beta})$ and $\overline{Y}_{\bullet}(\ddot{\beta})$ were specifically designed to exploit extra analytical information about the distribution of the control vector $\boldsymbol{C}$.

Finally we consider the performance of the various controlled confidence-interval estimation procedures. Table 6 summarizes the percentage reduction (relative to direct simulation) of the average half-length of nominal 90% confidence intervals generated by each procedure for each network configuration with 50% exponential activities. Inspection of this table reveals that relative dominance of the critical path was a highly significant factor. The percentage reduction in average confidence-interval half-length generally increased with increasing levels of relative dominance of $\pi_{(\xi)}$. This is to be expected for the same reasons discussed at the beginning of this section: at higher levels of relative dominance, the control $P_{(\xi)}$ for the critical path was more highly correlated with the overall completion time Y.

Table 7 summarizes the percentage of confidence intervals that actually covered the estimand θ for each controlled estimation procedure and for each network configuration with 50% exponential activities. Since a total of m = 2048 confidence

Table 3. Estimated Bias of Controlled Point Estimators

Network	Relative Dominance of $\pi_{(\xi)}$	$\hat{\hat{\theta}}(k)-\theta$				
		$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
	20% – 40%	-0.008	-0.250	-0.782	-0.079	-0.104
1	50% – 70%	0.097	-0.239	-1.34	-0.052	-0.050
	80% – 100%	0.201	-0.217	-2.51	-0.117	-0.115
	20% – 40%	-0.889	-1.75	-5.78	-1.05	-0.983
2	50% – 70%	-2.98	-2.65	-11.1	-2.35	-2.79
	80% – 100%	-7.91	-4.16	-18.7	-3.99	-3.49
	20% – 40%	-0.093	-0.171	-0.369	-0.116	-0.110
3	50% – 70%	-0.027	-0.158	-0.463	-0.089	-0.070
	80% – 100%	-0.011	-0.135	-0.782	-0.107	-0.112

Table 4. Estimated Variance of Controlled Point Estimators

Network	Relative Dominance of $\pi_{(\xi)}$	$\hat{\hat{V}}(k)$				
		$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
	20% – 40%	4.77	1.84	2.38	2.10	3.43
1	50% – 70%	14.7	1.27	3.79	1.53	6.20
	80% – 100%	63.6	0.52	12.3	0.69	20.7
	20% – 40%	1100.	104.	221.	113.	388.
2	50% – 70%	3961.	43.1	447.	48.4	1007.
	80% – 100%	12030.	15.4	1321.	17.4	2910.
	20% – 40%	3.35	2.58	2.69	2.75	2.93
3	50% – 70%	3.68	1.68	1.90	1.83	2.34
	80% – 100%	10.2	0.26	1.24	0.30	2.54

Table 5. Estimated Mean Square Error of Controlled Point Estimators

Network	Relative Dominance of $\pi_{(\xi)}$	$\hat{\hat{\mathrm{MSE}}}(k)$				
		$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$
1	20% – 40%	4.76	1.90	2.99	2.11	3.43
	50% – 70%	14.7	1.32	5.58	1.53	6.20
	80% – 100%	63.6	0.57	18.6	0.71	20.7
2	20% – 40%	1100.	107.	254.	115.	389.
	50% – 70%	3968.	50.1	570.	53.9	1014.
	80% – 100%	12087.	32.7	1670.	33.3	2921.
3	20% – 40%	3.36	2.61	2.82	2.76	2.94
	50% – 70%	3.68	1.70	2.11	1.84	2.34
	80% – 100%	10.1	0.28	1.85	0.31	2.55

Table 6. Percentage Reduction in Average Confidence-Interval Half-Length for Nominal 90% Confidence Intervals

Network	Relative Dominance of $\pi_{(\xi)}$	$\hat{\hat{H}}(0)$	$100[\hat{\hat{H}}(0) - \hat{\hat{H}}(k)]/\hat{\hat{H}}(0)$			
			$k = 1$	$k = 2$	$k = 3$	$k = 4$
1	20% – 40%	3.60	38.7	36.6	29.0	5.1
	50% – 70%	6.22	70.8	64.7	64.0	15.9
	80% – 100%	13.0	91.4	77.6	87.1	19.4
2	20% – 40%	56.3	71.0	64.8	67.3	31.9
	50% – 70%	106.	90.0	77.2	89.8	35.8
	80% – 100%	186.	96.7	79.2	95.4	37.6
3	20% – 40%	3.09	13.1	12.5	-2.9	-6.9
	50% – 70%	3.23	34.5	32.6	17.3	6.3
	80% – 100%	5.31	87.5	75.0	72.8	32.5

Table 7. Actual Coverage Percentages for Nominal 90% Confidence Intervals

Network	Relative Dominance of $\pi_{(\xi)}$	$100\hat{\hat{I}}(k)$				
		$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$
1	20% – 40%	90.0	86.5	79.4	89.6	91.4
	50% – 70%	89.1	85.8	71.4	90.4	93.1
	80% – 100%	89.4	77.8	66.8	86.3	95.5
2	20% – 40%	90.4	86.2	81.3	89.0	92.3
	50% – 70%	90.2	81.0	80.2	86.1	95.3
	80% – 100%	90.3	59.8	78.6	70.7	95.3
3	20% – 40%	90.0	88.1	86.0	91.5	91.5
	50% – 70%	90.0	86.8	81.9	90.3	92.1
	80% – 100%	89.6	63.3	72.3	77.2	94.9

intervals were generated by each estimation procedure for each network configuration, the standard error of the estimated coverage probability $\hat{\hat{I}}(k)$ in Table 7 is

$$\mathrm{SE}[\hat{\hat{I}}(k)] = \left\{\frac{E[\hat{I}_w(k)]\{1 - E[\hat{I}_w(k)]\}}{m}\right\}^{1/2}$$
$$\leq \left[\frac{(0.5)^2}{2048}\right]^{1/2} \simeq 0.011. \qquad (26)$$

Examination of Table 7 reveals that the level of relative dominance of the critical path had a more complex effect on coverage probability than on confidence-interval half-length. The conventional control-variate estimators (6) and (9) (that is, $\hat{\Theta}(1)$ and $\hat{\Theta}(3)$ respectively) performed poorly at high levels of relative dominance of $\pi_{(\xi)}$; note particularly the performance in networks 2 and 3 with relative dominance of $\pi_{(\xi)}$ in the range 80% – 100%. In this latter case, the critical-path control $P_{(\xi)}$ virtually always coincided with the overall completion time Y so that the resulting confidence intervals of the form (6) and (9) were either very small or degenerate; and when this phenomenon was coupled with the bias induced by the marked degree of nonnormality in the response and the controls, the net effect was a substantial loss of confidence-interval coverage. With respect to the confidence-interval estimation procedures based on the known covariance matrix of the controls, estimator (9) (that is, $\hat{\Theta}(2)$) also suffered from this effect while its analogue based on splitting (13) (that is, $\hat{\Theta}(4)$) did not.

Several conclusions emerged from our comparison of the performance of the four controlled confidence-interval procedures. The conventional estimator (6) and its splitting version (11) yielded consistently larger half-length reductions than the corresponding estimators (9) and (13) based on the known covariance structure of the controls. Half-length reductions of up to 97% were observed with (6) and (11). However, such reductions were not realized without cost. The actual coverage probability achieved by estimators (6) and (9) fell more than two standard errors below the nominal level 0.90 in every case reported in Table 7. On the other hand, the splitting estimator (13) achieved at least the nominal coverage level 0.90 for every network configuration on which it was tested. Of course the reductions in confidence-interval half-length achieved by (13) were much more modest than the reductions achieved by the other controlled estimation procedures.

6. CONCLUSIONS

In this paper we have examined four control-variate estimation procedures to be used in lieu of the conventional direct-simulation analysis of stochastic activity networks. All of these procedures are designed to improve upon the performance of direct simulation with respect to the accuracy of both point and confidence-interval estimators of mean completion time. As a fundamental principle for evaluating the performance improvements yielded by each of these procedures, we believe that confidence-interval coverage must be maintained at its nominal level while optimizing the accuracy of the corresponding point and confidence-interval estimators; thus it is unacceptable to achieve a large improvement in point-estimator accuracy at the expense of a significant loss of confidence-interval coverage. The second basic consideration in evaluating the performance of these control-variate estimation procedures is the additional computational overhead that they incur. In the simulation study reported here, the computational cost of these controlled estimation procedures was negligible compared to the cost of simulating the stochastic activity networks; moreover the cost of the controlled estimation procedures was very insensitive to the size of the selected networks. Thus considerable efficiency gains can be realized by the use of path control variates in large-scale applications.

ACKNOWLEDGMENTS

This research was partially supported by the National Science Foundation under Grant No. DMS-8717799 and by a David Ross Grant from the Purdue Research Foundation.

REFERENCES

Antill, J.M. and R.W. Woodhead (1982), *Critical Path Methods in Construction Practice*, John Wiley & Sons, New York, NY.

Avramidis, T.N., K.W. Bauer, and J.R. Wilson (1990), "Simulation of Stochastic Activity Networks Using Path Control Variates," *Naval Research Logistics* (to appear).

Bauer, K.W. (1987), "Control Variate Selection for Multiresponse Simulation," Ph.D. Thesis, School of Industrial Engineering, Purdue University, West Lafayette, IN.

Bauer, K.W., S. Venkatraman, and J.R. Wilson (1987), "Estimation Procedures Based on Control Variates with Known Covariance Matrix," In *Proceedings of the 1987 Winter Simulation Conference*, A. Thesen, F.H. Grant and W.D. Kelton, Eds. IEEE, Atlanta, GA, 334–341.

Bauer, K.W. and J.R. Wilson (1989), "Control-Variate Selection Criteria," Technical Report SMS 89-29, School of Industrial Engineering, Purdue University, West Lafayette, IN.

Bratley, P., B.L. Fox, and L. Schrage (1987), *A Guide to Simulation*, Second Edition, Springer-Verlag, New York, NY.

Burt, J.M. and M.B. Garman (1971), "Conditional Monte Carlo: A Simulation Technique for Stochastic Network Analysis," *Management Science 18*, 207–217.

Burt, J.M., D.P. Gaver, and M. Perlas (1970), "Simple Stochastic

Networks: Some Problems and Procedures," *Naval Research Logistics Quarterly 17*, 439–459.

Cheng, R.C.H. (1978), "The Analysis of Simulation Experiments under Normality Assumptions," *Journal of the Operational Research Society 29*, 493–497.

Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics, Philadelphia, PA.

Elmaghraby, S. (1977), *Activity Networks*, Wiley-Interscience, New York, NY.

Grubbs, F.E. (1962), "Attempts to Validate Certain PERT Statistics or 'Picking on PERT'," *Operations Research 10*, 912–915.

Kulkarni, V.G. and J.S. Provan (1985), "An Improved Implementation of Conditional Monte Carlo Estimation of Path Lengths in Stochastic Networks," *Operations Research 33*, 1389–1393.

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.

Law, A.M., W.D. Kelton, and L.W. Koenig (1981), "Relative Width Confidence Intervals for the Mean," *Communications in Statistics: Simulation and Computation B10*, 29–39.

Littlefield, T.K. and P.H. Randolph (1987), "Reply: An Answer to Sasieni's Question on PERT Times," *Management Science 33*, 1357–1359.

Loulou, R. and T. Beale (1976), "A Comparison of Variance Reducing Techniques in PERT Simulations," *INFOR 14*, 259–269.

MacCrimmon, K.R. and C.A. Ryavec (1964), "An Analytical Study of the PERT Assumptions," *Operations Research 12*, 16–37.

McBride, W. and C. McClelland (1967), "PERT and the Beta Distribution," *IEEE Transactions on Engineering Management EM-14*, 166–169.

McKenney, J.L. and R.S. Rosenbloom (1969), *Cases in Operations Management*, John Wiley & Sons, New York, NY.

Nádas, A. (1969), "An Extension of a Theorem of Chow and Robbins on Sequential Confidence Intervals for the Mean," *Annals of Mathematical Statistics 40*, 667–671.

Nelson, B.L. (1989), "Control-Variate Remedies," *Operations Research* (to appear).

Pritsker, A.A.B. (1986), *Introduction to Simulation and SLAM II*, Third Edition, Halsted Press, New York, NY.

Sullivan, R.S., J.C. Hayya, and R. Schaul (1982), "Efficiency of the Antithetic Variate Method for Simulating Stochastic Networks," *Management Science 28*, 563–572.

Tew, J.D. and J.R. Wilson (1990), "Validation of Simulation Analysis Methods for the Schruben-Margolin Correlation-Induction Strategy," *Operations Research*, to appear.

Venkatraman, S., and J.R. Wilson (1985), "Using Path Control Variates in Activity Network Simulation," In *Proceedings of the 1985 Winter Simulation Conference*, D.T. Gantz, S.L. Solomon, and G.C. Blais, Eds. IEEE, Piscataway, NJ, 217–222.