Control-variate Models of Common Random Numbers for Multiple Comparisons with the Best

Barry L. Nelson • Jason C. Hsu
Department of Industrial and Systems Engineering, Ohio State University, Columbus, Ohio 43210
Department of Statistics, Ohio State University, Columbus, Ohio 43210

Using common random numbers (CRN) in simulation experiment design is known to reduce the variance of estimators of differences in system performance. However, when more than two systems are compared, exact simultaneous statistical inference in conjunction with CRN is typically impossible. We introduce control-variate models of CRN that permit exact statistical inference, specifically multiple comparisons with the best. These models explain the effect of CRN via a linear regression of the simulation output on "control variates" that are functions of the simulation inputs. We establish theoretically, and illustrate empirically, that the control-variate models lead to sharper statistical inference in the sense that the probability of detecting differences in systems' performance is increased.
(*Simulation; Variance Reduction; Multiple Comparisons*)

1. Introduction

Stochastic simulation experiments are frequently used to compare the performance of two or more systems, often with the goal of selecting the best system. Simulators know that estimators of differences in expected performance are improved by inducing positive correlation across simulation responses via "common random numbers" (CRN). Even when simulators do not know this, simulation languages make CRN the default experiment design by initializing all experiments with the same random number seeds (careful experiment design can further enhance the effect of CRN; see Bratley et al. 1987, Chapter 2, and Law and Kelton 1991, Chapter 11).

The beneficial effect of CRN on point estimators of differences can be realized under fairly general conditions. This paper considers the associated statistical inference about those differences, where inference means determining whether or not observed differences are due to actual differences in expected performance, or are due instead to random error. Simulators usually do not take advantage of CRN in their statistical analyses because the induced correlations are unknown and possibly unequal. Ideally, CRN should sharpen statistical inference (e.g., reduce the widths of confidence intervals), so that smaller differences in expected performance can be discerned. We propose models that permit exact statistical inference under CRN.

We restrict our attention to the statistical-inference methods called *multiple comparisons*, and do not consider the related methods of ranking and selection. See Clark and Yang (1986) for a ranking procedure that incorporates CRN.

The paper is organized as follows: We first review methods for incorporating CRN into simultaneous statistical inference. Then we establish conditions under which multiple-comparison procedures can be derived. These conditions are satisfied by models of CRN that are introduced in §4. Some numerical examples are given in §5, which is followed by concluding remarks.

0025-1909/93/3908/0989$01.25

2. Review

This section briefly reviews some existing methods for statistical inference under CRN, and sets the stage for our method. Suppose that the goal of the simulation experiment is to compare the performance of r systems. Let $\theta = (\theta_1, \theta_2, \ldots, \theta_r)'$ be the $r \times 1$ vector whose ith element is the expected performance of system i, and let the simulation output from the jth replication across all r systems be $\mathbf{Y}_j = (Y_{1j}, Y_{2j}, \ldots, Y_{rj})'$, for $j = 1, 2, \ldots, n$, where $'$ indicates the transpose of a vector or matrix. We assume that $\theta = E[\mathbf{Y}_j]$ for all j. In simulation experiments designed for multiple replications, Y_{ij} is the response of system i on replication j; in simulation experiments designed for a single replication of a stationary process, Y_{ij} could be the jth batch mean from system i or an appropriate quantity accumulated over the jth regenerative cycle of system i.

Throughout this paper responses *within* a system ($Y_{i1}, Y_{i2}, \ldots, Y_{in}$) are independent and identically distributed (i.i.d.), but responses *across* systems ($Y_{1j}, Y_{2j}, \ldots, Y_{rj}$) may be dependent due to CRN. Stated differently, CRN induces a joint distribution on the elements of $\mathbf{Y}_j$, but $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ remain i.i.d.

A point estimator of θ is the sample mean

$$\bar{\mathbf{Y}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{Y}_j = \begin{bmatrix}\bar{Y}_1\\ \bar{Y}_2\\ \vdots\\ \bar{Y}_r\end{bmatrix}.$$

When systems i and l are compared in terms of the difference $\theta_l - \theta_i$, then the natural unbiased point estimator is $\bar{Y}_l - \bar{Y}_i$, which has variance

$$\mathrm{Var}[\bar{Y}_l - \bar{Y}_i] = \frac{1}{n}(\sigma_l^2 + \sigma_i^2 - 2\rho_{il}\sigma_l\sigma_i) \qquad (1)$$

where $\sigma_i^2 = \mathrm{Var}[Y_{ij}]$ and $\rho_{il} = \mathrm{Corr}[Y_{ij}, Y_{lj}]$. The goal of CRN is to induce $\rho_{il} > 0$.

Inference about the actual difference $\theta_l - \theta_i$ often takes the form of a confidence interval. For example, if $r = 2$ and the Y_{ij} are normally distributed, then with probability $1 - \alpha$,

$$\theta_1 - \theta_2 \in \bar{Y}_1 - \bar{Y}_2 \pm t_{1-\alpha/2,n-1}S_D/\sqrt{n} \qquad (2)$$

where S_D^2 is the sample variance of $D_j = Y_{1j} - Y_{2j}$, $j = 1, 2, \ldots, n$, and $t_{1-\alpha/2,n-1}$ is the $1 - \alpha/2$ quantile of the t distribution with $n - 1$ degrees of freedom.

Nelson (1987) compared the performance of the "paired-t" interval (2) and the "pooled-t" interval

$$\theta_1 - \theta_2 \in \bar{Y}_1 - \bar{Y}_2 \pm t_{1-\alpha/2,2n-2}S_P\sqrt{2/n} \qquad (3)$$

that is valid when $\rho_{12} = 0$ (no CRN) and $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (S_P^2 is the usual pooled estimator of σ^2). Surprisingly, the inference provided by (3) without CRN is sharper than the inference provided by (2) with CRN unless n and ρ_{12} are sufficiently large (here sharper inference means a larger probability of detecting a difference between θ_1 and θ_2).

In this paper we are interested in comparisons among more than two systems, but the case $r = 2$ illustrates three important issues:

- CRN can reduce point-estimator variance without sharpening the associated inference. Although point-estimator performance is perhaps most important, the associated inference indicates whether we can have confidence in our conclusions.
- The dependence induced by CRN is typically unknown. The paired-t procedure circumvents this problem by transforming the data so that $D_1, D_2, \ldots, D_n$ are i.i.d. Valid statistical analysis under CRN must account for the induced dependence in some way.
- The inference should be exact under the assumptions of the procedure. The paired-t procedure is exact, but extensions of it to more than two systems typically employ conservative inequalities to obtain simultaneous inference, leading to wide and inconclusive confidence intervals that have true confidence level much larger than $1 - \alpha$. See, for example, Bratley et al. (1987, pp. 84–85) for a procedure based on the Bonferroni inequality.

The most thoroughly studied problem of statistical analysis under CRN is that of estimating the $p \times 1$-vector parameter $\underline{\gamma}$ in the model

$$\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n \sim \text{i.i.d. } N(\mathbf{x}\underline{\gamma}, \sigma^2\Xi) \qquad (4)$$

where $\mathbf{x}$ is a fixed $r \times p$ design matrix with first column all 1s, and $\Xi = (\rho_{il})$ is a correlation matrix. Estimating expected differences in performance, such as $\theta_l - \theta_i$, can be formulated in this way, but more generally (4) defines a metamodel estimation problem. The usual

point estimator is the ordinary-least-squares estimator $\hat{\underline{\gamma}} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{Y}$, where

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \quad \text{and} \quad \mathbf{G} = \begin{pmatrix} \mathbf{x} \\ \mathbf{x} \\ \vdots \\ \mathbf{x} \end{pmatrix}.$$

If the r systems (more commonly called "design points" in this context) are simulated independently, then $\Xi = \mathbf{I}_{r\times r}$, the $r \times r$ identity matrix. Schruben and Margolin (1978) investigated the use of CRN and antithetic variates (AV) to impart advantageous structure to Ξ in the special case when $\mathbf{x}$ is orthogonally blockable into two blocks (AV induces negative correlation between design points). Their analysis was based on assuming a regular correlation structure: $\rho_{il} = \rho^+ > 0$ if design points i and l are simulated using CRN, and $\rho_{il} = -\rho^- < 0$ if design points i and l are simulated using AV, where $\rho^- \le \rho^+$ (Tew and Wilson (1992a) showed that the results apply under even weaker conditions on ρ^-).

Let $\mathbf{1}_r$ be an $r \times 1$ column vector of ones. If CRN is used across all design points in a replication, then the assumptions of Schruben and Margolin imply that

$$\Xi = (1 - \rho^+)\mathbf{I}_{r\times r} + \rho^+\mathbf{1}_r\mathbf{1}_r'. \tag{5}$$

Combining CRN and AV according to an "assignment rule" proposed by Schruben and Margolin leads to the correlation structure

$$\Xi = (1 - \rho^+)\mathbf{I}_{r\times r} + \left(\frac{\rho^+ + \rho^-}{2}\right)\mathbf{U}\mathbf{U}' + \left(\frac{\rho^+ - \rho^-}{2}\right)\mathbf{1}_r\mathbf{1}_r' \tag{6}$$

where $\mathbf{U} = (\mathbf{1}_{r_1}', -\mathbf{1}_{r_2}')'$ and r_1 and r_2 are the block sizes.

Let γ_0 be the first element, and $\underline{\gamma}_1$ the remaining $p - 1$ elements, of $\underline{\gamma}$; and let $\hat{\gamma}_0$ and $\hat{\underline{\gamma}}_1$ be their respective least-squares estimators. Schruben and Margolin showed that CRN reduces the generalized variance (determinant of the variance-covariance matrix) of $\hat{\underline{\gamma}}_1$, but at the expense of increasing the variance of $\hat{\gamma}_0$, when compared to simulating the design points independently. Under certain conditions the assignment rule reduces the generalized variance of $\hat{\underline{\gamma}}$ relative to both independent simulation and CRN.

Nozari et al. (1987) exploited the regular structure of (6) to derive statistical inference for $\underline{\gamma}$, including $(1 - \alpha)100\%$ simultaneous confidence intervals for γ_0, for $\mathbf{l}'\underline{\gamma}_1$, for all $\mathbf{l} \in \Re^{p-1}$, and for $\mathbf{l}'\underline{\gamma}$, for all $\mathbf{l} \in \Re^p$. Their procedures are exact for the advertised inference, but they are conservative when only a small number of simple comparisons, such as differences, are desired rather than all linear combinations of the parameters. Of course, they depend on being able to assume the structure of (6). Tew and Wilson (1992a) provide hypothesis tests to verify the assumptions of Schruben and Margolin.

Kleijnen (1988, 1992) considered a generalization of (4): estimate $\underline{\gamma}$ for the model

$$\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n \sim \text{i.i.d. } N(\mathbf{x}\underline{\gamma}, \Sigma) \tag{7}$$

where $\Sigma = \text{Var}[\mathbf{Y}_i]$. When CRN is used, he proposes point estimators and inference based on an estimate of Σ, say $\hat{\Sigma}$. For example, he suggests using the estimated-generalized-least-squares estimator,

$$\hat{\underline{\gamma}} = (\mathbf{x}'\hat{\Sigma}^{-1}\mathbf{x})^{-1}\mathbf{x}'\hat{\Sigma}^{-1}\bar{\mathbf{Y}},$$

to estimate $\underline{\gamma}$.

Kleijnen's approach is at the other extreme from Schruben and Margolin, in the sense that it makes no assumption about the structure of Σ under CRN. However, his statistical analysis is necessarily approximate due to substituting $\hat{\Sigma}$ for Σ, and conservative inequalities are employed for simultaneous inference. Our approach is between these two: we assume a regular structure with some unknown parameters; then we estimate those parameters.

3. Multiple Comparisons with the Best

Throughout the remainder of this paper we assume that the goal of the simulation experiment is to compare elements of the vector parameter $\theta = (\theta_1, \theta_2, \ldots, \theta_r)'$ defined in the previous section. To be specific, suppose that larger expected performance implies a better system. For system i, the parameter $\theta_i - \max_{l\neq i}\theta_l$ can be termed *system i performance minus the best of the other systems' performances*. In optimization problems, the parameters $\theta_i - \max_{l\neq i}\theta_l$, for $i = 1, \ldots, r$ are often the parameters of primary interest. This can be seen as follows.

If $\theta_i - \max_{l \neq i}\theta_l > 0$, then system i is the best, because its performance parameter is better (larger) than the best of the other systems' parameters. Similarly, if $\theta_i - \max_{l \neq i}\theta_l < 0$, then system i is not the best, since there is another system with larger performance parameter. However, even when system i is not the best, if $\theta_i - \max_{l \neq i}\theta_l > -\epsilon$, where ϵ is a positive number, then system i is within ϵ of the best. Simultaneous statistical inference (i.e., confidence intervals) on $\theta_i - \max_{l \neq i}\theta_l$, for $i = 1, \ldots, r$, is termed *multiple comparisons with the best* (MCB).

We focus on MCB because of its close connection to optimization, but many of our results apply to all-pairwise multiple comparisons ($\theta_i - \theta_l$, for all $i \neq l$) and multiple comparisons with a control ($\theta_i - \theta_r$, for all $i \neq r$) as well. When optimization is the goal of the experiment, MCB inference is typically sharper than all-pairwise inference because fewer confidence intervals are required to be simultaneously correct (r versus $r(r-1)/2$). In addition, MCB implies the inference of both the indifference zone and the subset selection methodologies of ranking and selection (Hsu and Nelson 1988).

Below we describe sufficient conditions for deriving MCB intervals. The models proposed in §4 lead to estimators that satisfy these conditions under CRN.

Suppose that $\hat{\theta}$ is a point estimator of θ with the property that

$$\hat{\theta} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \vdots \\ \hat{\theta}_r \end{bmatrix} \sim N(\theta, \tau^2\Delta) \tag{8}$$

where Δ is known. The $r \times r$ matrix $\Delta = (\delta_{il})$ need not be a correlation matrix, but together $\tau^2\Delta$ must be a positive definite variance-covariance matrix. In addition, suppose there is an estimator $\hat{\tau}^2$ of τ^2 such that $\hat{\tau}^2 \sim \tau^2\chi_\nu^2/\nu$ and is independent of $\hat{\theta}$, where χ_ν^2 denotes the chi-squared distribution with ν degrees of freedom.

Critical to the derivation of MCB intervals is the distribution of

$$\mathbf{D}^{(i)}\hat{\theta} = \begin{bmatrix} \hat{\theta}_1 - \hat{\theta}_i \\ \hat{\theta}_2 - \hat{\theta}_i \\ \vdots \\ \hat{\theta}_{i-1} - \hat{\theta}_i \\ \hat{\theta}_{i+1} - \hat{\theta}_i \\ \vdots \\ \hat{\theta}_r - \hat{\theta}_i \end{bmatrix}$$

where $\mathbf{D}^{(i)}$ is the $(r-1) \times r$ matrix obtained by inserting the column $-\mathbf{1}_{r-1}$ between the $(i-1)$st and the ith columns of $\mathbf{I}_{r-1 \times r-1}$. Assumption (8) implies that

$$\mathbf{D}^{(i)}\hat{\theta} \sim N\left(\begin{bmatrix} \theta_1 - \theta_i \\ \theta_2 - \theta_i \\ \vdots \\ \theta_{i-1} - \theta_i \\ \theta_{i+1} - \theta_i \\ \vdots \\ \theta_r - \theta_i \end{bmatrix}, \tau^2\Omega^{(i)} \right) \tag{9}$$

where $\Omega^{(i)} = \mathbf{D}^{(i)}\Delta\mathbf{D}^{(i)\prime}$.

Let $\omega_{ll}^{(i)}$ be the diagonal element of $\Omega^{(i)}$ corresponding to $\hat{\theta}_l - \hat{\theta}_i$, and define $x^- = -\min\{x, 0\}$ and $x^+ = \max\{x, 0\}$. Under model (8), and given appropriate critical values $d_{1-\alpha}^{(i)}$, Chang and Hsu (1991) showed that with probability $1 - \alpha$

$$\theta_i - \max_{l \neq i}\theta_l \in [D_i^-, D_i^+] \tag{10}$$

for $i = 1, 2, \ldots, r$, where

$$D_i^+ = \left(\min_{l \neq i}\{\hat{\theta}_i - \hat{\theta}_l + d_{1-\alpha}^{(i)}\hat{\tau}\sqrt{\omega_{ll}^{(i)}}\}\right)^+,$$

$$\mathcal{G} = \{i : \min_{l \neq i}(\hat{\theta}_i - \hat{\theta}_l + d_{1-\alpha}^{(i)}\hat{\tau}\sqrt{\omega_{ll}^{(i)}}) > 0\},$$

$$D_i^- = \begin{cases} 0 & \text{if } \mathcal{G} = \{i\}, \\ -\left(\min\limits_{\substack{l \in \mathcal{G} \\ l \neq i}}\{\hat{\theta}_i - \hat{\theta}_l - d_{1-\alpha}^{(l)}\hat{\tau}\sqrt{\omega_{ii}^{(l)}}\}\right)^- & \text{otherwise.} \end{cases}$$

Let $\Xi^{(i)}$ be the correlation matrix of $\mathbf{D}^{(i)}\hat{\theta}$. The critical value $d_{1-\alpha}^{(i)}$ is the constant that satisfies the equation

$$\Pr\left\{\frac{\hat{\theta}_l - \hat{\theta}_i - (\theta_l - \theta_i)}{\hat{\tau}\sqrt{\omega_{ll}^{(i)}}} \leq d_{1-\alpha}^{(i)} \quad \text{for all} \quad l \neq i\right\}$$

$$= \Pr\{T_l \leq d_{1-\alpha}^{(i)}, l = 1, 2, \ldots, r-1\} = 1 - \alpha \tag{11}$$

where $\mathbf{T}' = (T_1, T_2, \ldots, T_{r-1})$ is a multivariate-t random variable with correlation matrix $\Xi^{(i)}$ and ν degrees of freedom. In general, calculation of $d_{1-\alpha}^{(i)}$ requires an r-dimensional numerical integration to evaluate the left-hand side of (11) for each candidate value of $d_{1-\alpha}^{(i)}$, and a search to find the value that satisfies the equality. The calculation is computationally prohibitive unless $\Xi^{(i)}$ has structure l (Tong 1980, p. 13):

$$\Xi^{(i)} = \begin{bmatrix} 1 & \lambda_1^{(i)}\lambda_2^{(i)} & \cdots & \lambda_1^{(i)}\lambda_{r-1}^{(i)} \\ \lambda_2^{(i)}\lambda_1^{(i)} & 1 & \cdots & \lambda_2^{(i)}\lambda_{r-1}^{(i)} \\ \lambda_{r-1}^{(i)}\lambda_1^{(i)} & \lambda_{r-1}^{(i)}\lambda_2^{(i)} & \cdots & 1 \end{bmatrix} \tag{12}$$

where $\lambda_l^{(i)} \in (-1, 1)$. Given structure (12), probability (11) can be rewritten as

$$\int_0^{+\infty} \int_{-\infty}^{+\infty} \prod_{l=1}^{r-1} \Phi\left(\frac{\lambda_l^{(i)} z + d_{1-\alpha}^{(i)} u}{\sqrt{1-(\lambda_l^{(i)})^2}}\right) d\Phi(z) dF(u) = 1 - \alpha \tag{13}$$

where Φ is the standard normal distribution function and F is the distribution function of $\hat{\tau}/\tau$ (Hochberg and Tamhane 1987, pp. 366–367). This integral can be approximated efficiently by double Gaussian quadrature, so that the critical value $d_{1-\alpha}^{(i)}$ can be obtained using a root-finding algorithm.

A model for the simulation responses that satisfies these conditions is the one-way model

$$\mathbf{Y}_j = \boldsymbol{\theta} + \boldsymbol{\epsilon}_j \tag{14}$$

where $\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \ldots, \boldsymbol{\epsilon}_n$ are i.i.d. $N(\mathbf{0}, \sigma^2 \mathbf{I}_{r\times r})$ random variables, i.e., the systems are simulated independently, without CRN. Under (14), $\bar{\mathbf{Y}} \sim N(\boldsymbol{\theta}, \tau^2 \Delta)$, where $\tau^2 = \sigma^2$ and $\Delta = \mathbf{I}_{r\times r}/n$. For all i, the correlation matrix $\Xi^{(i)}$ of $\mathbf{D}^{(i)}\bar{\mathbf{Y}}$ is

$$\Xi^{(i)} = \begin{bmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{2} \\ \frac{1}{2} & 1 & \cdots & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \cdots & 1 \end{bmatrix}.$$

Thus, $\Xi^{(i)}$ has structure l with $\lambda_l^{(i)} = 1/\sqrt{2}$.

If CRN is employed, then $\Delta \neq \mathbf{I}_{r\times r}/n$ in general, and the off-diagonal elements are unknown. Even if we are willing to substitute an estimate for Δ, the implied correlation matrices $\Xi^{(i)}$ may not have structure l. Thus, CRN can prohibit exact statistical inference.

In the next section we introduce models that imply estimators $\mathbf{D}^{(i)}\hat{\boldsymbol{\theta}}$ whose correlation matrices $\Xi^{(i)}$ have structure l under CRN. Since it is more natural to think in terms of the distribution of $\hat{\boldsymbol{\theta}}$ rather than the distribution of $\mathbf{D}^{(i)}\hat{\boldsymbol{\theta}}$, we mention that $\Xi^{(i)}$ will have structure l if Δ is a diagonal matrix, or if all of the off-diagonal elements of Δ are equal. These two structures are the ones encountered in the proposed models.

4. Control-variate Models of CRN

As an experiment-design strategy, CRN is implemented at the level of the pseudorandom numbers in the simulation experiment. "CRN" means assigning the same pseudorandom numbers, for the same purpose, to the simulation of different systems.

Random-variate generation is the process of transforming pseudorandom numbers into observations from prespecified input distributions; examples of inputs include the service-time and interarrival-time random variables in queueing simulations, the demand and lead-time random variables in inventory simulations, and the activity-duration random variables in stochastic-activity-network simulations. For CRN to be effective, the dependence induced at the pseudorandom-number level must be transmitted to the simulation inputs, and then to the simulation responses (outputs) of interest.

Our approach is to approximate the relationship between the simulation inputs and outputs by a linear model with unknown parameters. Specifically, let $\mathbf{C}_{ij}$ be a $q \times 1$ vector of (possibly functions of) simulation input random variables from the jth replication of system i, and let $\boldsymbol{\mu}_i = E[\mathbf{C}_{ij}]$; we assume that $\boldsymbol{\mu}_i$ is known since the simulator specifies the distribution of the simulation inputs. We propose the model

$$Y_{ij} = \theta_i + (\mathbf{C}_{ij} - \boldsymbol{\mu}_i)'\boldsymbol{\beta}_i + \eta_{ij} \tag{15}$$

where $\boldsymbol{\beta}_i$ is a $q \times 1$ vector of unknown parameters and the residuals η_{ij} are i.i.d. $N(0, \tau^2)$ random variables for all i and j. The random variables $\mathbf{C}_{ij}$ are called *control variates* in the variance-reduction literature (e.g., Nelson 1990).

Let $\boldsymbol{\eta}_j = (\eta_{1j}, \eta_{2j}, \ldots, \eta_{rj})'$ denote the residuals across systems on replication j. The relationship (15) between the response and control variates on replication j can also be represented as

$$\mathbf{Y}_j = \mathbf{X}_j \underline{\gamma} + \boldsymbol{\eta}_j \tag{16}$$

for $j = 1, 2, \ldots, n$, where $\boldsymbol{\theta}$ forms the first r elements of the column vector $\underline{\gamma}$; the definition of $\mathbf{X}_j$ and the remainder of $\underline{\gamma}$ depends on additional assumptions discussed in the next two subsections.

The control-variate point estimator of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}}$, the first r elements of the least-squares estimator $\hat{\underline{\gamma}} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{Y}$, where

$$\mathbf{G} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{pmatrix}.$$

The generic estimator of τ^2 is $\hat{\tau}^2 = \|\mathbf{Y} - \mathbf{G}\hat{\underline{\gamma}}\|^2/\nu$, where the degrees of freedom ν, depend on the case considered.

Assuming model (16) pertains, $\hat{\theta}$ has the following properties under CRN:

(1) The conditional distribution $\hat{\theta}|\mathbf{C} \sim N(\theta, \tau^2\Delta)$, where $\Delta = \Delta(\mathbf{C})$ is a function of $\mathbf{C}$ alone, and $\mathbf{C}$ is the set of all control variates $\mathbf{C}_{ij}$, for all i, j;

(2) The conditional correlation matrix $\Xi^{(i)}$ of $\mathbf{D}^{(i)}\hat{\theta}$ has structure l;

(3) The estimator $\hat{\tau}^2$ is conditionally independent of $\hat{\theta}$ and is distributed $\tau^2\chi_\nu^2/\nu$, and ν is known.

Given properties (1)–(3), Chang and Hsu (1991) implies that

$$\Pr\{\theta_i - \max_{l \neq i} \theta_l \in [D_i^-, D_i^+] \quad \text{for all} \quad i|\mathbf{C}\} = 1 - \alpha \tag{17}$$

where D_i^- and D_i^+ are defined by (10). Since the probability on the right-hand side of (17) does not depend on $\mathbf{C}$, the intervals $[D_i^-, D_i^+]$ are unconditionally $(1 - \alpha)100\%$ simultaneous confidence intervals for $\theta_i - \max_{l \neq i}\theta_l$, $i = 1, 2, \ldots, r$.

MCB inference derived under model (16) is exact. However, assuming a linear relationship between the simulation inputs and outputs is nearly always an approximation (although less so than assuming such a relationship between the pseudorandom numbers and the responses). Nelson (1990) showed that inference based on control-variate estimators is robust to nonlinearity, provided the sample size is not too small.

Some additional comments regarding model (16) are in order:

- Model (16) assumes that $\eta_1, \eta_2, \ldots, \eta_n$ are i.i.d. $(\mathbf{0}, \tau^2\mathbf{I}_{r\times r})$ random variables, which implies that all of the dependence due to CRN is explained by the control variates $\mathbf{C}_{ij}$. This is in contrast to the control-variate model of Tew and Wilson (1992b), in which the control variates are independent and all the dependence due to CRN is explained by the residuals.
- The residuals η_{ij} are assumed to be normally distributed. We do not assume that the responses Y_{ij} are normally distributed, as is typical for multiple-comparison procedures.
- The residuals η_{ij} are assumed to have common variance τ^2. We do not assume that the responses Y_{ij} have common variance, as in the one-way model (14).

All of these assumptions are approximations, but notice that they can be checked empirically using standard tools for examining the residuals from a least-squares regression.

In the subsections that follow we introduce two specific control-variate models and examine their advantages and disadvantages. The proofs of all theorems are given in the Appendix and Nelson and Hsu (1990).

4.1. The Model of Yang and Nelson

In model (16), let

$$\underline{\gamma} = \begin{bmatrix} \theta \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_r \end{bmatrix} \tag{18}$$

be an $r(q + 1) \times 1$ vector of unknown parameters, and let $\mathbf{X}_j = [\mathbf{I}_{r\times r}, \mathbf{C}_j^{\text{yn}}]$, where

$$\mathbf{C}_j^{\text{yn}} = \begin{bmatrix} (\mathbf{C}_{1j} - \mu_1)' & & \mathbf{0}' \\ & \ddots & \\ \mathbf{0}' & & (\mathbf{C}_{rj} - \mu_r)' \end{bmatrix} \tag{19}$$

is an $r \times rq$ matrix of control variates. This model, which generalizes the model in Yang and Nelson (1991), is appropriate if different control variates are available in each system, or if the relationship between the response and control variates may be different across systems (i.e., $\beta_i \neq \beta_l$ for $i \neq l$). Although we have assumed that there are the same number of control variates q, associated with each system, this is only for convenience of exposition; see Nelson and Hsu (1990) for the generalization to different numbers of control variates.

THEOREM 4.1. *If model* (16) *with* $\underline{\gamma}$ *defined by* (18) *and* $\mathbf{X}_j$ *defined by* (19) *pertains, then* $\hat{\theta}|\mathbf{C} \sim N(\theta, \tau^2\Delta)$ *and* $\Delta = \text{diag}(\delta_{11}, \delta_{22}, \ldots, \delta_{rr})$, *where*

$$\delta_{ii} = \frac{1}{n} + \frac{1}{n-1}(\bar{\mathbf{C}}_i - \mu_i)'\mathbf{S}_{\mathbf{C}_i}^{-1}(\bar{\mathbf{C}}_i - \mu_i), \tag{20}$$

with $\bar{\mathbf{C}}_i$ *the sample mean, and* $\mathbf{S}_{\mathbf{C}_i}$ *the sample variance-covariance matrix, of the control variates from system* i. *In addition, conditional on* $\mathbf{C}$, $\nu\hat{\tau}^2/\tau^2$ *has a chi-squared distribution with* $\nu = r(n - q - 1)$ *degrees of freedom and is independent of* $\hat{\theta}$.

Theorem 4.1 implies that the conditional correlation matrix $\Xi^{(i)}$ of $\mathbf{D}^{(i)}\hat{\theta}$ has structure l with

$$\lambda_l^{(i)} = \sqrt{\frac{\delta_{ii}}{\delta_{ii} + \delta_{ll}}}. \tag{21}$$

The form of (21) does not depend on the use of CRN. However, properties of $\lambda_l^{(i)}$, and thus certain properties of the resulting intervals, do depend on it. We investigate those properties in a later subsection.

4.2. The Model of Nozari, Arnold and Pegden

In model (16), let

$$\underline{\gamma} = \begin{bmatrix} \theta \\ \beta \end{bmatrix} \tag{22}$$

be an $(r + q) \times 1$ vector of unknown parameters, and let $\mathbf{X}_j = [\mathbf{I}_{r\times r}, \mathbf{C}_j^{\text{nap}}]$, where

$$\mathbf{C}_j^{\text{nap}} = \begin{bmatrix} (\mathbf{C}_{1j} - \mu_1)' \\ (\mathbf{C}_{2j} - \mu_2)' \\ \vdots \\ (\mathbf{C}_{rj} - \mu_r)' \end{bmatrix}. \tag{23}$$

This is model (15) with $\beta_i = \beta$ for all i. In addition, we assume that the $\mathbf{C}_{ij}$ have a common distribution for all i and j (implying common expectation μ). This model is a special case of the model considered by Nozari et al. (1984), and it is appropriate when the same control variates are available in each system and it is believed that the relationship between the response and control variates is the same across all systems. The inventory model in Subsection 5.1 is an example for which this model is plausible.

Nozari et al. (1984) analyzed $\hat{\theta}$ when the control variates are independent across systems. The conditional correlation matrix of $\mathbf{D}^{(i)}\hat{\theta}$ does not have structure l in that case. However, if CRN causes the control variates to take identical values across systems, then properties (1)–(3) are satisfied.

THEOREM 4.2. *If model* (16) *with* $\underline{\gamma}$ *defined by* (22) *and* $\mathbf{X}_j$ *defined by* (23) *pertains, and CRN causes the control variates to take identical values across systems, then* $\hat{\theta}|\mathbf{C} \sim N(\theta, \tau^2\Delta)$ *and*

$$\Delta = \frac{1}{n}\mathbf{I}_{r\times r} + \frac{1}{r(n-1)}(\bar{\mathbf{C}} - \mu)'\mathbf{S}_{\mathbf{C}}^{-1}(\bar{\mathbf{C}} - \mu)\mathbf{1}_r\mathbf{1}_r', \tag{24}$$

with $\bar{\mathbf{C}}$ *the sample mean, and* $\mathbf{S}_{\mathbf{C}}$ *the sample variance-covariance matrix, of all of the control variates. In addition, conditional on* $\mathbf{C}$, $\nu\hat{\tau}^2/\tau^2$ *has a chi-squared distribution with* $\nu = rn - r - q$ *degrees of freedom and is independent of* $\hat{\theta}$.

It is easy to show that the conditional correlation matrix $\Xi^{(i)}$ of $\mathbf{D}^{(i)}\hat{\theta}$ has structure l with $\lambda_l^{(i)} = 1/\sqrt{2}$ for all i.

4.3. Comparison of the Models

The previous subsections introduced two models of the dependence induced via CRN; we refer to them as YN, for Yang and Nelson, and NAP, for Nozari, Arnold and Pegden. This subsection provides summary comparisons.

Point-estimator performance is at least as important as inference, so it is worthwhile to compare the variance of the control-variate point estimator to the sample mean, both under CRN. We compare the pairwise differences $\hat{\theta}_i - \hat{\theta}_l$ and $\bar{Y}_i - \bar{Y}_l$, rather than the MCB differences $\hat{\theta}_i - \max_{l\neq i}\hat{\theta}_l$ and $\bar{Y}_i - \max_{l\neq i}\bar{Y}_l$, because they illustrate the effect of CRN while avoiding the complications introduced by the maximum operator. We also add the assumption that $(Y_{ij}, \mathbf{C}_{ij}')$ are jointly normal so that we can obtain easily comparable expressions.

THEOREM 4.3. *If model* (15) *pertains, then the variance of* $\bar{Y}_i - \bar{Y}_l$ *is minimized when* $\beta_i = \beta$, *for all* i, *and the control variates are identical across systems under CRN. In that case*

$$\text{Var}[\mathbf{D}^{(i)}\bar{\mathbf{Y}}] = \frac{\tau^2}{n}[\mathbf{I}_{r-1\times r-1} + \mathbf{1}_{r-1}\mathbf{1}_{r-1}']$$

which implies that

$$\text{Var}[\bar{Y}_i - \bar{Y}_l] = \frac{2\tau^2}{n}.$$

Theorem 4.3 establishes the case of model (15) for which CRN is the most effective in reducing the variance of the sample-mean differences. The following theorems give the corresponding variances of the control-variate estimators for the same case.

THEOREM 4.4. *Under the same conditions as Theorem* 4.1, *and assuming* $(Y_{ij}, \mathbf{C}_{ij}')$ *are jointly normal,*

$$\text{Var}[\mathbf{D}^{(i)}\hat{\theta}] = \left(\frac{n-2}{n-q-2}\right)\frac{\tau^2}{n}[\mathbf{I}_{r-1\times r-1} + \mathbf{1}_{r-1}\mathbf{1}_{r-1}']$$

which implies that

$$\text{Var}[\hat{\theta}_i - \hat{\theta}_l] = \left(\frac{n-2}{n-q-2}\right)\frac{2\tau^2}{n}.$$

Thus, under CRN the sample mean may be more

precise than the control-variate point estimator when YN pertains, but the difference is negligible when n is large.

THEOREM 4.5. *Under the same conditions as Theorem 4.2, and assuming* $(Y_{ij}, \mathbf{C}'_{ij})$ *are jointly normal,*

$$\text{Var}[\mathbf{D}^{(i)}\hat{\boldsymbol{\theta}}] = \frac{\tau^2}{n}[\mathbf{I}_{r-1\times r-1} + \mathbf{1}_{r-1}\mathbf{1}'_{r-1}]$$

which implies that

$$\text{Var}[\hat{\theta}_i - \hat{\theta}_l] = \frac{2\tau^2}{n}.$$

Thus, we completely recover the variance reduction achieved by the sample mean (Theorem 4.3) when NAP pertains and we make use of that knowledge. This is true because having common control variates and common $\boldsymbol{\beta}$ imply a common control-variate effect across systems, and this effect cancels when we take differences. In fact, we can show that $\hat{\theta}_i - \hat{\theta}_l = \bar{Y}_i - \bar{Y}_l$ for this case.

For completeness, we mention that under the conditions of Theorem 4.2, but without CRN,

$$\text{Var}[\hat{\theta}_i - \hat{\theta}_l] = \left(\frac{rn - r - 1}{rn - r - q - 1}\right)\frac{2\tau^2}{n} \qquad (25)$$

(Nozari et al. 1984). Comparing (25) to Theorem 4.5 shows that CRN yields a variance reduction under NAP; notice that CRN has no effect on point-estimator variance under YN.

In summary, when both models YN and NAP pertain, NAP leads to a more precise point estimator and greater degrees of freedom for MCB inference than YN, so it is clearly superior. In large samples both models lead to point estimators with precision that is competitive with the sample mean under CRN, while still permitting MCB inference.

Of course, the assumptions behind NAP are more stringent than YN: common control variates across systems, common relationship between the response and control variates across systems, and identical values of the control variates under CRN (in addition to the linearity, normality and common residual variance assumptions that both models must satisfy). YN is more general, since it allows for different control variates—even different numbers of control variates—across systems, and it does not require the use of CRN to derive MCB inference. However, there is a secondary benefit from using CRN in conjunction with YN, as we show in the next subsection.

4.4. The Effect of CRN

Yang and Nelson (1991) studied model YN, defined by (18) and (19), in the special case when the distribution of the control variate $\mathbf{C}_{ij}$ does not depend on the system i. They assumed that the control variates take identical values across all r systems under CRN, but the relationship between the response and the control variates may be different across systems (i.e., $\boldsymbol{\beta}_i \neq \boldsymbol{\beta}_l$ for $i \neq l$). These assumptions imply that $\delta_{ii} = \delta_{ll}$, for all i and l, and thus $\lambda_l^{(i)} = 1/\sqrt{2}$ for all l.

Yang and Nelson compared, theoretically and empirically, the inference derived from this special case of YN (with CRN) and the inference derived from the one-way model (14) (without CRN); recall that exact inference is not possible under the one-way model with CRN. They showed that, as the sample size n increases, inference under YN yields a larger probability of identifying differences in system performance. These results generalize directly to inference under the general YN model and to NAP. Thus, the control-variate procedures provide sharper inference than the one-way model.

Theorem 4.1 extends the results of Yang and Nelson to include control variates that may be different across systems, and may be either independent or dependent. When the control variates are different across systems, $\lambda_l^{(i)}$ is a function of the control variates, and thus is a random variable; therefore, $d_{1-\alpha}^{(i)}$ is also a random variable. Asymptotically, however, the value of $d_{1-\alpha}^{(i)}$ is the same as the case of identical control variates.

THEOREM 4.6. *Under the same conditions as Theorem 4.1,* $\lambda_l^{(i)} \xrightarrow{P} 1/\sqrt{2}$ *as* $n \to \infty$ *where* $\xrightarrow{P}$ *denotes convergence in probability. This result holds even if the number of control variates differs across systems.*

Apparently there is little direct benefit from using CRN in conjunction with YN: the variance of the point estimator is the same with or without CRN; the degrees of freedom for inference is not changed by CRN; and the critical value $d_{1-\alpha}^{(i)}$ is the same with or without CRN in large samples.

There are, however, some secondary benefits from CRN when the control variates can be made identical. Most importantly, $d_{1-\alpha}^{(i)}$ is a constant, which eliminates

a source of variability in the confidence-interval width. In addition, $d_{1-\alpha}^{(i)}$ is the same for all i, and is the $1-\alpha$ quantile of the maximum of an $(r-1)$-dimensional multivariate-t random variable with common correlation $1/2$ and $r(n-q-1)$ degrees of freedom. Quantiles of this random variable can be found in tables (e.g., Hochberg and Tamhane 1987, Appendix 3, Table 4; Nelson 1992a, Table 102.5).

In the general case, the critical values cannot be precomputed and the value of $d_{1-\alpha}^{(i)}$ that satisfies (13) must be found numerically. Since software for computing these critical values is not yet generally available, we present a conservative approximation:

THEOREM 4.7. *Let $\iota^{(i)}$ be the minimum element of $\Xi^{(i)}$, for $i = 1, 2, \ldots, r$. For $i = 1, 2, \ldots, r$, set $d_{1-\alpha}^{(i)}$ equal to the $1-\alpha$ quantile of the maximum of an $(r-1)$-dimensional multivariate-t random variable with common correlation $\iota^{(i)}$. Then the resulting MCB intervals are conservative.*

The conservative approximation is only needed for the YN model when the controls are not identical across systems. In that case

$$\iota^{(i)} = \min_{l \neq k \neq i} \frac{\delta_{ii}}{\sqrt{(\delta_{ii} + \delta_{ll})(\delta_{ii} + \delta_{kk})}}.$$

Hochberg and Tamhane (1987, Appendix 3, Table 4) give critical values for the maximum of a multivariate-t random variable with common correlation 0.1, 0.3, 0.5 or 0.7. To use these tables, or any others, the largest correlation that is less than or equal to $\iota^{(i)}$ is selected.

5. Examples

Following some preliminaries, this section presents two simple examples that illustrate the potential benefits, and shortcomings, of the models of CRN proposed above. In both cases the true system performance parameters θ_i, $i = 1, 2, \ldots, r$, are known, so the performance of the MCB procedures can be evaluated.

Selecting control variates is a critical decision. A sufficient condition for the linear relationship (15) to hold is that $(Y_{ij}, \mathbf{C}'_{ij})$ are jointly normal. In some simulation experiments it is possible to insure that this condition holds asymptotically (as the length of the simulation goes to infinity) by the choice of control variate. Wilson and Pritsker (1984) showed that a standardized average of the input processes in queueing simulations is asymptotically normal with mean 0 and variance 1. Specifically, if $Z_{ij1}, Z_{ij2}, \ldots, Z_{ijm}$ is a sequence of i.i.d. random variables from the jth replication of system i, each with expectation μ and variance σ^2, then the standardized average is

$$C_{ij} = \frac{\sum_{k=1}^{m} (Z_{ijk} - \mu)}{\sigma\sqrt{m}}.$$

If Y_{ij} is itself an average accumulated over the jth replication, then it is plausible that $(Y_{ij}, \mathbf{C}'_{ij})$ are approximately jointly normal, and thus linearly related. We use standardized averages of input processes as the control variates in our examples.

5.1. Inventory System Example

Consider an (s, S) inventory system in which some discrete item is periodically reviewed. If the inventory level is found to be below s units, then an order is issued to bring the inventory level up to S units; otherwise no additional items are ordered. Different (s, S) inventory policies result in different inventory systems. Koenig and Law (1985) used this example to illustrate a subset selection procedure; see their paper for a detailed description of the model.

The only input process in the simulation is the demand for inventory in each period, which is assumed to be a sequence of i.i.d. Poisson random variables with common mean 25; we use a standardized average of the demands as a control variate.

Five (s, S) inventory policies are considered. We simulate and apply MCB to determine which policy has the minimum expected cost per period for 30 periods (minimization problems are addressed by considering the parameters $\theta_i - \min_{l \neq i} \theta_l$, for $i = 1, 2, \ldots, r$). An experiment consists of $n = 60$ replications of each policy, each replication 30 periods in length.

5.2. Machine-repair Example

Consider a manufacturing system consisting of ten machines and four spares, where the machines are subject to failure. Failed machines are repaired by one of s repairmen who each work at rate μ machines repaired/unit time. Suppose that a fixed overall repair rate, $s\mu = 12$, can be achieved through different combinations of equipment and personnel. Iglehart (1977) used this example to illustrate a ranking and selection procedure; see his paper for a detailed description of the model.

There are two input processes in the simulation: the

sequence of machine times-to-failure—a sequence of i.i.d. exponential random variables with common mean 1—and the sequence of repair times—a sequence of i.i.d. exponential random variables with common mean $1/\mu$. We use standardized averages of these two processes as control variates.

The five (s, μ) combinations (1, 12), (2, 6), (3, 4), (4, 3) and (6, 2) were considered. We simulate and apply MCB to determine which (s, μ) combination maximizes the steady-state expected number of machines in use.

An experiment consists of $n = 60$ replications of each system, each replication 300 machine failures in length, with statistics cleared after the first 100 failures to reduce initial-condition bias.

5.3. Experiment Results

For both examples we estimated the probability of correct inference $\mathcal{C}$, and correct and useful inference $\mathcal{C} \cap \mathcal{U}$, for nominal 95% MCB confidence intervals. If the procedure is performing as desired, then $\Pr\{\mathcal{C}\} \approx 0.95$; this is the probability that the intervals simultaneously contain all r parameters $\theta_i - \max_{l \neq i} \theta_l$, for $i = 1, 2, \ldots, r$. The event $\mathcal{C} \cap \mathcal{U}$ is the event that the intervals simultaneously contain all the parameters but do not contain 0 when $\theta_i - \max_{l \neq i} \theta_l \neq 0$. While MCB always bounds the differences $\theta_i - \max_{l \neq i} \theta_l \neq 0$, a correct and useful outcome means that we declare conclusively (and correctly) which system is the best. The probabilities we report were estimated by replicating each experiment 1,000 times, both with and without CRN.

The $\Pr\{\mathcal{C} \cap \mathcal{U}\}$ is a measure of the sharpness of the inference, and we would like it to be as large as possible. Unfortunately, it is always the case that $\Pr\{\mathcal{C} \cap \mathcal{U}\} \leq \Pr\{\mathcal{C}\}$. Designing an experiment (i.e., choosing the number of replications n) to achieve a particular $\Pr\{\mathcal{C} \cap \mathcal{U}\}$ is an open problem. Our examples illustrate that CRN in conjunction with the control-variate models can significantly increase $\Pr\{\mathcal{C} \cap \mathcal{U}\}$ relative to using the one-way model with or without CRN.

Experiment results for the inventory simulation are displayed in Table 1. When the five inventory policies are simulated independently, both the one-way model and the YN model are appropriate (NAP assumes that the control variates take on identical values across systems). Both models approximately achieve the nominal 95% coverage probability, but the control-variate procedure increases the probability of correct and useful inference from 0.45 to 0.63.

If the five inventory policies are simulated using CRN on the single input process then the NAP model is also appropriate. Notice that the one-way model becomes conservative ($\Pr\{\mathcal{C}\} \approx 1$). The control-variate procedures improve their performance under CRN.

Experiment results for the machine-repair simulation are displayed in Table 2. When the five (s, μ) combinations are simulated independently, both the one-way model and the YN model approximately achieve the nominal 95% coverage probability, but the control-variate procedure nearly triples the probability of correct and useful inference.

Because the simulation run length is a fixed number of machine failures, CRN causes the time-to-failure random variables to be identical across the systems. The experiment "CRN-failures" employs CRN only on the failure times, and samples the repair times independently across systems. The models "YN-failures" and NAP make use of only the single failure-time control variate, while the model "YN-all" uses both the failure-time and repair-time control variates. All four models are conservative ($\Pr\{\mathcal{C}\} > 0.95$), and the use of CRN on the failure times does not significantly increase the

Table 1 **Estimated Probability of Correct, and Correct and Useful Inference for the Inventory-system Example**

	Model					
	one way		YN		NAP	
Data	$\Pr\{\mathcal{C}\}$	$\Pr\{\mathcal{C} \cap \mathcal{U}\}$	$\Pr\{\mathcal{C}\}$	$\Pr\{\mathcal{C} \cap \mathcal{U}\}$	$\Pr\{\mathcal{C}\}$	$\Pr\{\mathcal{C} \cap \mathcal{U}\}$
independent	0.94	0.45	0.94	0.63		
CRN	1.00	0.43	0.95	0.77	0.95	0.74

Table 2 Estimated Probability of Correct, and Correct and Useful Inference for the Machine-repair Example

	Model							
	one way		YN-failures		YN-all		NAP	
Data	$\Pr\{\mathcal{C}\}$	$\Pr\{\mathcal{C} \cap \mathcal{U}\}$	$\Pr\{\mathcal{C}\}$	$\Pr\{\mathcal{C} \cap \mathcal{U}\}$	$\Pr\{\mathcal{C}\}$	$\Pr\{\mathcal{C} \cap \mathcal{U}\}$	$\Pr\{\mathcal{C}\}$	$\Pr\{\mathcal{C} \cap \mathcal{U}\}$
independent	0.96	0.16			0.94	0.46		
CRN-failures	1.00	0.10	0.97	0.22	0.98	0.48	0.97	0.22
CRN-all	1.00	0.00			1.00	0.46		

probability of correct and useful inference for YN-all. When CRN is used on both input processes ("CRN-all"), the probability of correct and useful inference drops dramatically for the one-way model. This occurs because CRN reduces the variance of the point estimator $(\bar{Y}_i - \max_{l \neq i} \bar{Y}_l)$, so that the confidence intervals are more nearly centered on the true parameters, but it does not reduce the length of the intervals (in fact, they have the same expected length with or without CRN for the one-way model).

Why are the control-variate models effective for the inventory example but not as effective for the machine-repair example? A partial answer can be obtained by looking at correlation matrices.

The matrix (26) is an estimate of the correlation matrix of $\mathbf{Y}_j = (Y_{1j}, Y_{2j}, \ldots, Y_{5j})'$ for the inventory example under CRN (i.e., the correlations across systems induced by CRN). The correlations are all positive, as desired.

$$\begin{bmatrix} 1.00 & 0.49 & 0.85 & 0.41 & 0.65 \\ & 1.00 & 0.46 & 0.93 & 0.42 \\ & & 1.00 & 0.38 & 0.59 \\ & & & 1.00 & 0.37 \\ & & & & 1.00 \end{bmatrix} \quad (26)$$

The matrix (27) is an estimate of the correlation matrix of $\boldsymbol{\eta}_j = (\eta_{1j}, \eta_{2j}, \ldots, \eta_{5j})'$ after fitting the YN model. Notice that in most cases the correlations are greatly reduced, meaning that the model is doing an adequate job of explaining the dependence induced via CRN.

$$\begin{bmatrix} 1.00 & -0.10 & 0.50 & -0.03 & 0.20 \\ & 1.00 & -0.06 & 0.91 & 0.04 \\ & & 1.00 & -0.03 & 0.13 \\ & & & 1.00 & 0.06 \\ & & & & 1.00 \end{bmatrix} \quad (27)$$

Similarly, the matrix (28) is an estimate of the correlation matrix of $\mathbf{Y}_j = (Y_{1j}, Y_{2j}, \ldots, Y_{5j})'$ for the machine-repair example under CRN. The correlations are all positive and nearly 1.

$$\begin{bmatrix} 1.00 & 0.99 & 0.97 & 0.95 & 0.91 \\ & 1.00 & 0.99 & 0.97 & 0.92 \\ & & 1.00 & 0.99 & 0.94 \\ & & & 1.00 & 0.97 \\ & & & & 1.00 \end{bmatrix} \quad (28)$$

The matrix (29) is an estimate of the correlation matrix of $\boldsymbol{\eta}_j = (\eta_{1j}, \eta_{2j}, \ldots, \eta_{5j})'$ after fitting the YN-all model. The correlations are still quite large. Apparently the control-variate model does not do an adequate job of explaining the dependence due to CRN in this example. Although a different choice of control variates might improve performance, the alternative choice is not obvious. However, we can take some comfort in the fact that the control-variate models performed better than the one-way model, and that all the models yielded conservative results.

$$\begin{bmatrix} 1.00 & 0.97 & 0.93 & 0.88 & 0.78 \\ & 1.00 & 0.96 & 0.92 & 0.81 \\ & & 1.00 & 0.96 & 0.85 \\ & & & 1.00 & 0.90 \\ & & & & 1.00 \end{bmatrix} \quad (29)$$

Our experience is that the machine-repair example—in which the control-variate models were able to explain very little of the dependence due to CRN—is an extreme case. In practice, the simulator can determine whether or not the control-variate model has been effective by examining the residual dependence, as we did here. We are currently working on methods for modeling the residual dependence, as well.

6. Concluding Remarks

Yang and Nelson (1991) established the first exact multiple-comparison procedures under CRN, where

"exact" means that no conservative probabilistic inequalities are required. We have generalized their results to allow the systems to have different control variates (i.e., different input processes), so that the method is potentially applicable in any stochastic simulation experiment where comparisons among a small, finite number of systems are desired. We have also explored the effect of assuming even more structure than Yang and Nelson—that the relationship between the response and the control variates is common across systems—and found benefits, specifically reduced point-estimator variance and increased degrees of freedom for inference.

Of course, our MCB inference is exact only if the assumed control-variate model is correct. However, our examples show that we can expect some benefit from the control-variate models, in terms of sharper inference, even when the models are not entirely adequate. We hope eventually to prove that MCB based on the control-variate model is conservative when the control variates do not entirely explain the effect of CRN.

Nelson (1992b) gives instructions for implementing the control-variate MCB procedure, but realistically the approach will not be widely used until it is embedded in commercial simulation software and is transparent to the user. We feel that this is an important next step.[1]

[1] This material is based upon work supported by National Science Foundation Grant No. DDM-8922721. The authors acknowledge the careful, helpful review of the Department Editor and Associate Editor.

Appendix

The proofs of Theorems 4.1–4.5 are straightforward but tedious, and are given in Nelson and Hsu (1990).

PROOF OF THEOREM 4.6. From Theorem 5.2.3 in Anderson (1984), $n\bar{\mathbf{C}}_l'\mathbf{S}_{\mathbf{C}_l}^{-1}\bar{\mathbf{C}}_l \xrightarrow{\mathcal{D}} \chi_q^2$, where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution. Thus

$$\left(\frac{n}{n-1}\right)\bar{\mathbf{C}}_l'\mathbf{S}_{\mathbf{C}_l}^{-1}\bar{\mathbf{C}}_l \xrightarrow{\mathcal{P}} 0$$

by Slutsky's theorem and the fact that convergence in distribution to a constant implies convergence in probability as well. Thus, $n\delta_{ll} \xrightarrow{\mathcal{P}} 1$. The result then follows by application of Slutsky's theorem for the ratio of random variables.

Notice that the result does not depend on having the same number of control variates from each system. □

PROOF OF THEOREM 4.7. Dropping the superscript (i) for convenience, let

$$\mathbf{\Lambda} = (1-\iota)\mathbf{I}_{r-1\times r-1} + \iota\mathbf{1}_{r-1}\mathbf{1}_{r-1}',$$

where ι is the minimum element of Ξ. Let $d_{1-\alpha}$ be the critical value such that

$$\Pr_{\Lambda}\{T_l \le d_{1-\alpha}, l = 1, \ldots, r-1\} = 1-\alpha$$

where $(T_1, T_2, \ldots, T_{r-1})$ is a multivariate-t random variable with correlation matrix indicated by the subscript on Pr. Since $\Lambda \ge \Xi$ element by element, Theorems 3.1.1 and 2.1.1 in Tong (1980) imply that

$$\Pr_{\Xi}\{T_l \le d_{1-\alpha}, l = 1, \ldots, r-1\} \ge \Pr_{\Lambda}\{T_l \le d_{1-\alpha}, l = 1, \ldots, r-1\}.$$

Thus, the critical value $d_{1-\alpha}$ is conservative. □

References

Anderson, T., *An Introduction to Multivariate Statistical Analysis*, Second Ed., Wiley, New York, 1984.

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

Chang, J. and J. Hsu, "Optimal Designs for Multiple Comparisons with the Best," *J. Stat. Planning and Inference*, 29 (1991).

Clark, G. and W. Yang, "A Bonferroni Selection Procedure When Using Common Random Numbers with Unknown Variances," *Proc.* 1986 *Winter Simulation Conf.*, J. Wilson, J. Henriksen and S. Roberts (Eds.), Washington, DC, 1986, 313–315.

Hochberg, Y. and A. Tamhane, *Multiple Comparison Procedures*, Wiley, New York, 1987.

Hsu, J. and B. Nelson, "Optimization Over a Finite Number of System Designs with One-stage Sampling and Multiple Comparisons with the Best," *Proc.* 1988 *Winter Simulation Conf.*, M. Abrams, P. Haigh and J. Comfort (Eds.), San Diego, CA, 1988, 451–457.

Iglehart, D., "Simulating Stable Stochastic Systems, VII: Selecting the Best System," In *TIMS Studies in the Management Sciences*, 1977, 37–49.

Kleijnen, J., "Analyzing Simulation Experiments with Common Random Numbers," *Management Sci.*, 34 (1988), 65–74.

——, "Regression Metamodels for Simulation with Common Random Numbers: Comparison of Validation Tests and Confidence Intervals," *Management Sci.*, 38 (1992), 1164–1185.

Koenig, L. and A. Law, "A Procedure for Selecting a Subset of Size m Containing the l Best of k Independent Normal Populations, with Applications to Simulation," *Comm. Statistics*, B14 (1985), 719–734.

Law, A. and W. Kelton, *Simulation Modeling and Analysis*, Second Ed., McGraw-Hill, New York, 1991.

Nelson, B., "Some Properties of Simulation Interval Estimators Under Dependence Induction," *Oper. Res. Letters*, 6 (1987), 169–176.

——, "Control-Variate Remedies," *Oper. Res.*, 38 (1990), 974–992.

——, "Statistical Analysis of Simulation Results," *Handbook of Industrial Engineering*, G. Salvendy (Ed.), Wiley, New York, 1992a, 2567–2593.

——, "Common Random Numbers and Multiple Comparisons in Simulation Analysis," *Proc.* 1*st Industrial Engineering Res. Conf.*, G. Klutke, D. Mitta, B. Nnaji and L. Sieford (Eds.), Chicago, IL, 1992b, 463–466.

—— and J. Hsu, "Control-Variate Models of Common Random Numbers for Multiple Comparisons with the Best," Working Pa-

per Series No. 1990-011, Department of Industrial and Systems Engineering, Ohio State University, Columbus, OH.

Nozari, A., S. Arnold and C. Pegden, "Control Variates for Multi-population Simulation Experiments," *IIE Trans.*, 16 (1984), 159–169.

——, —— and ——, "Statistical Analysis for Use with the Schruben and Margolin Correlation Induction Strategy," *Oper. Res.*, 35 (1987), 127–139.

Schruben, L. and B. Margolin, "Pseudorandom Number Assignment in Statistically Designed Simulation and Distribution Sampling Experiments," *J. Amer. Statistical Assoc.*, 73 (1978), 504–525.

Tew, J. and J. Wilson, "Validation of Simulation Analysis Methods for the Schruben-Margolin Correlation-Induction Strategy," *Oper. Res.*, 40 (1992a), 87–103.

—— and ——, "Estimating Simulation Metamodels Using Integrated Variance Reduction Techniques," *IIE Trans.* (1992b).

Tong, Y., *Probability Inequalities in Multivariate Distributions*, Academic Press, New York, 1980.

Wilson, J. and A. Pritsker, "Variance Reduction in Queueing Simulation Using Generalized Concomitant Variables," *J. Statistical Comp. and Simulation*, 19 (1984), 129–153.

Yang, W. and B. Nelson, "Using Common Random Numbers and Control Variates in Multiple-Comparison Procedures," *Oper. Res.*, 39 (1991), 583–591.

Accepted by James R. Wilson, former Departmental Editor; received July 22, 1991. *This paper has been with the authors* 3 *months for* 1 *revision.*