Proceedings of the 1999 Winter Simulation Conference
P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds.

A DECISION-THEORETIC APPROACH TO SCREENING AND SELECTION WITH COMMON RANDOM NUMBERS

Stephen E. Chick
Koichiro Inoue

Department of Industrial and Operations Engineering
The University of Michigan
1205 Beal Avenue
Ann Arbor, MI 48109-2117, U.S.A.

ABSTRACT

This article describes some recently-proposed procedures that identify the best simulated system when common random numbers are used. The procedures are based on a Bayesian average-case analysis, rather than a worst-case indifference zone formulation. The procedures allow decision-makers to focus on reducing either the expected opportunity-cost loss associated with potentially selecting an inferior system, or the probability of incorrect selection. Numerical experiments indicate that the new procedures outperform two existing procedures with respect to several criteria for a well-known selection problem.

1 INTRODUCTION

This paper considers the problem of comparing a small number of systems, say 2 to 20, in terms of the expected value of some given stochastic performance measure. The performance of each system is estimated by a simulation experiment (Law and Kelton 1991; Banks, Carson, and Nelson 1996), and the goal is to efficiently identify the best system, where 'best' is defined as having the maximum expected performance.

The performance of each system must be estimated with a finite number of simulation replications, so it is impossible to guarantee that the best system will be selected. The indifference-zone formulation, the dominant approach to the problem for some time (Bechhofer, Santner, and Goldsman 1995; Goldsman and Nelson 1998), attempts to provide a lower bound on the probability of correct selection, given the assumption that the best system is at least a prespecified amount better than the other systems.

Most relevant research assumes that the simulation output is independent. On the other hand, common random numbers (CRN) is a variance reduction technique that can be used to sharpen comparisons by inducing a positive correlation between the output of each system (Law and Kelton 1991; Banks, Carson, and Nelson 1996). CRN can be implemented, for example, by causing the j-th simulation replication of each system to observe the same demand patterns.

Clark and Yang (1986) and Nelson and Matejcik (1995) are exceptions that present procedures for selecting the best system when CRN is used. The procedure of Clark and Yang (1986), called Procedure $\mathcal{CY}$ here, does not make special assumptions about the covariance induced by CRN, but is based on the statistically conservative least-favorable condition (LFC) that is characteristic of indifference-zone formulations. Nelson and Matejcik (1995) provide an alternate indifference-zone procedure (Procedure $\mathcal{NM}$) that can require fewer replications to achieve the same probability of correct selection, but requires a sphericity assumption for the covariance matrix.

Here, we describe rather different two-stage procedures proposed by Chick and Inoue (1999a) for selecting the best system using CRN. The approach is Bayesian and decision-theoretic, and is a natural extension of earlier work for independent replications (Chick and Inoue 1999b). The number of second-stage replications is determined by an average-case criterion, rather than the LFC of indifference-zone procedures. This allows first-stage information about the unknown means, variances, and covariances to help identify likely contenders for the best. Whereas Procedures $\mathcal{CY}$ and $\mathcal{NM}$ require that all systems be simulated during the a second stage, our procedures might simulate only a subset of systems. This focuses simulation effort on systems that are likely to benefit most from additional replications. Further, our procedures allow for the reduction of either the probability of incorrect selection *or* the expected opportunity cost loss associated with potentially incorrect selections. We do not require a sphericity assumption for the covariance from CRN.

We refer several times to a specific selection problem considered by Koenig and Law (1985) and Nelson and Matejcik (1995). There are five (s, S) inventory policies for controlling the inventory level of a discrete product. If the inventory level drops below s, then an order is placed to bring the level up to S. Different values for s and S lead to different inventory policies. The demand in each period is the only stochastic process in the model. The best system is the policy that has the minimum expected cost per period, evaluated over 30 periods, where cost is measured in thousands of dollars. See Koenig and Law (1985) for further details of the model.

2 BACKGROUND

Let $x_{i,j}$ be the output of the j-th replication of system i, for $i = 1, \ldots, k$, so that $\mathbf{x}_j = (x_{1,j}, \ldots, x_{k,j})$ is the $1 \times k$ vector of outputs across all systems on replication j. We assume throughout that $\mathbf{x}_1, \mathbf{x}_2, \ldots$ are i.i.d. multivariate normal vectors, conditional on the unknown mean vector $\mathbf{w} = (w_1, \ldots, w_k)$ and unknown variance-covariance matrix $\mathbf{\Sigma}$,

$$p_{\mathbf{X}_j|\mathbf{w},\mathbf{\Sigma}}\left(\mathbf{x}_j\right) \sim \mathcal{N}_k\left(\mathbf{w}, \mathbf{\Sigma}\right). \tag{1}$$

For independent replications, $\mathbf{\Sigma}$ is diagonal.

We are interested in comparing the k different systems in terms of their expected performance w_i. In the inventory example, there are $k = 5$ policies, $x_{i,j}$ is the average cost for 30 periods observed on the j-th replication of inventory policy i, and w_i is the expected cost per period of the i-th inventory policy.

Two-stage indifference-zone procedures guarantee a lower bound P^* on the probability of correct selection (PCS), whenever the best system is at least δ^* better than the other systems. Procedure $\mathcal{CY}$ of Clark and Yang (1986) is the first indifference-zone procedure to account for CRN. A first-stage of r_1 replications of each system are observed using CRN across all systems. The number r_2 of second-stage replications depends on δ^*, $t = t_{1-(1-P^*)/(k-1),r_1-1}$, the $1-(1-P^*)/(k-1)$ quantile of the standard t distribution with $r_1 - 1$ degrees of freedom, and the sample variances of the differences in output,

$$\hat{\sigma}^2_{i,j} = \sum_{l=1}^{r_1} \frac{(x_{i,l} - x_{j,l} - (\bar{x}_i - \bar{x}_j))^2}{r_1 - 1}$$

where $\bar{x}_i$ is the first stage sample mean, so that

$$r_2 = \max\left\{0, \left\lceil \max_{j \neq i}\left(\frac{t}{\delta^*}\right)^2 \hat{\sigma}^2_{i,j} - r_1 \right\rceil\right\}.$$

The system with the highest overall sample mean is then selected as best.

Nelson and Matejcik (1995) develop Procedure $\mathcal{NM}$, an alternate indifference-zone procedure. The number of second-stage replications then depends on δ^*, $g = T^{P^*}_{k-1,(k-1)(r_1-1),1/2}$, the P^* quantile of the maximum of a standard multivariate t random variable of dimension $k-1$ with $(k-1)(r_1-1)$ degrees of freedom, and common correlation $1/2$ (see Hochberg and Tamhane 1987), and the sample variance of the difference (given the sphericity assumption),

$$\hat{\sigma}^2 = \frac{2\sum_{i=1}^{k}\sum_{j=1}^{r_1}(x_{i,j} - \bar{x}_i - \bar{x}_j + m)^2}{(k-1)(r_1-1)}$$

where $m = \sum_{i=1}^{k} \bar{x}_i / k$ is the overall sample mean, so that

$$r_2 = \max\left\{0, \left\lceil \left(\frac{g}{\delta^*}\right)^2 \hat{\sigma}^2 - r_1 \right\rceil\right\}.$$

Empirical tests demonstrate that the procedure is somewhat robust to departures from sphericity.

Procedures $\mathcal{CY}$ and $\mathcal{NM}$ both simulate *all* systems with CRN during second-stage replications, regardless of how well the mean for each system is known after the first stage, or whether the first stage sample mean for some systems are clearly inferior. The issue of screening out inferior systems after the first stage is considered by Nelson, Swann, Goldsman, and Song (1998), although the idea is not discussed at length.

3 AN ALTERNATE FORMULATION

The results of Chick and Inoue (1999a) provide new two-stage procedures for identifying the best simulated system that are less conservative than procedures based on the LFC assumption, that incorporate screening during the second stage so that simulation effort is focused on systems that best benefit from additional analysis, and that do not make restrictive assumptions about the covariance induced by CRN. This section summarizes the Bayesian decision-theoretic assumptions behind those procedures.

The first stage of the new procedures described below are structurally similar to Procedures $\mathcal{CY}$ and $\mathcal{NM}$, in that r_1 replications of each system are observed to obtain simulation output $\mathcal{E}_{\mathbf{x}} = (\mathbf{x}_1, \ldots, \mathbf{x}_{r_1})$, and sample statistics are computed,

$$\hat{\boldsymbol{\mu}} = \sum_{j=1}^{r_1} \frac{\mathbf{x}_j}{r_1} \tag{2}$$

$$\hat{\mathbf{\Sigma}} = \bar{\mathbf{S}}/r_1 = \sum_{j=1}^{r_1}(\mathbf{x}_j - \hat{\boldsymbol{\mu}})^t(\mathbf{x}_j - \hat{\boldsymbol{\mu}})/r_1 \tag{3}$$

The selection of the best system after both stages of output are observed is the same as well, in that the system with the maximal estimate of the mean is selected as best.

The second stage of the new procedures, however, differs in four ways.

Screening. *A subset* $\mathcal{C}_2$ of all systems $\mathcal{U} = \{1, \ldots, k\}$ may be simulated r_2 times during the second stage. All systems are eligible for selection, including the systems in $\mathcal{C}_1 = \{1, \ldots, k\} \backslash \mathcal{C}_2$ that are simulated only during the first stage. When all systems are simulated, denote the second-stage vectors of output by $\mathbf{y}_1, \ldots, \mathbf{y}_{r_2}$. The output still has multivariate normal distribution with unknown mean $\mathbf{w}$ and variance $\mathbf{\Sigma}$, but the notation $\mathbf{y}$ distinguishes second-stage from first-stage output. Let $\mathbf{y}_{\mathcal{C}_2, j}$ be the output vector of the j-th second-stage replication when only the systems in $\mathcal{C}_2$ are simulated. This notation is used throughout to denote appropriate subvectors and submatrices for systems in $\mathcal{C}_2$ or $\mathcal{C}_1$, such as $\mathbf{w} = (\mathbf{w}_{\mathcal{C}_2}\, \mathbf{w}_{\mathcal{C}_1})$ for subvectors of the unknown mean, or $\mathbf{\Sigma}_{\mathcal{C}_2\mathcal{C}_2}$ for the variance-covariance matrix of $\mathbf{y}_{\mathcal{C}_2, j}$. Denote all second stage output by

$$\mathcal{E}_{\mathbf{y}} = \left(\mathbf{y}_{\mathcal{C}_2,1}, \ldots, \mathbf{y}_{\mathcal{C}_2,r_2}\right).$$

Loss Function. A second difference from indifference-zone procedures is a choice for the loss criterion. The PCS is an integral part of indifference-zone procedures. Because the 0-1 loss function and the PCS are related, we consider

$$\mathcal{L}_{0-1}(i, \mathbf{w}) = \begin{cases} 0 & \text{when } w_i \geq \max_j w_j \\ 1 & \text{otherwise,} \end{cases} \tag{4}$$

where the loss is 0 if the best system is selected, and 1 if an inferior system is selected.

In many business and engineering applications, the expected opportunity cost loss may be more relevant. We therefore also consider the *opportunity cost loss*

$$\mathcal{L}_{o.c.}(i, \mathbf{w}) = \max_j w_j - w_i \tag{5}$$

associated with selecting i as best when the means are $\mathbf{w}$. The loss is 0 for a correct selection, and is otherwise the difference in means between the best and the selected system.

Average Case Analysis. Third, the new procedures below choose $\mathcal{C}_2$ and r_2 with an average-case analysis rather than the worst-case LFC analysis. The average is taken over the likely values of the mean and variance, as suggested by the first-stage output. To do so, we adopt a Bayesian approach to infer the values of the unknown mean and variance from simulation output and Bayes' rule. Unknown quantities are treated as random variables and are written in upper case, such as $\mathbf{W} = (W_1, \ldots, W_k)$, and specific outcomes are given in lower case, such as $\mathbf{w}$. We therefore refer to the output precision $\mathbf{\lambda} = \mathbf{\Sigma}^{-1}$, and consider its unknown value through the random variable $\mathbf{\Lambda}$, because the symbols $\mathbf{\Sigma}$ and σ already have strong connotations for variance-covariance and summation.

The average-case analysis for selecting the 'best' second stage experiment ($\mathcal{C}_2$ and r_2), is done with respect to the distributions of the unknown mean, given the first stage output, $\mathbf{W} \mid_{\mathcal{E}_{\mathbf{x}}}$. The general setup is therefore similar to the development of Chick and Inoue (1999b) for independent replications.

Although the assessment of probability distributions for the unknown mean and precision of the simulation output is not a feature of indifference zone formulations, there is a rich history for doing so in the Bayesian tradition (de Groot 1970; Bernardo and Smith 1994). Chick and Inoue (1999a) suggest a prior distribution for $\mathbf{W}, \mathbf{\Lambda}$ that leads to a multivariate t distribution after conditioning on the first stage output that has $r_1 - 1$ degrees of freedom, the same degrees of freedom for related frequentist analysis.

$$p_{\mathbf{\Lambda}|\mathcal{E}_{\mathbf{x}}}(\mathbf{\lambda}) \sim \mathcal{W}_k\left(r_1 + k - 2, \bar{\mathbf{S}}\right) \tag{6}$$

$$p_{\mathbf{W}|\mathbf{\lambda},\mathcal{E}_{\mathbf{x}}}(\mathbf{w}) \sim \mathcal{N}_k\left(\hat{\boldsymbol{\mu}}, r_1\mathbf{\lambda}^{-1}\right)$$

$$p_{\mathbf{W}|\mathcal{E}_{\mathbf{x}}}(\mathbf{w}) \sim \mathbf{St}_k\left(\hat{\boldsymbol{\mu}}, r_1(r_1-1)\bar{\mathbf{S}}^{-1}, r_1 - 1\right). \tag{7}$$

where $\mathcal{W}_k(\alpha, \boldsymbol{\beta})$ is a Wishart distribution, and $\mathbf{St}_k(\mu, \kappa, \nu)$ is a multivariate t distribution with mean μ and variance $\kappa^{-1}\nu/(\nu-2)$ when the degrees of freedom $\nu > 2$. See also Inoue and Chick (1998).

Eq. 7 only make sense when $r_1 \geq k$, because $\bar{\mathbf{S}}$ must be invertible. *This differs from the indifference-zone procedures, which do not require a specific relation between* r_1 *and* k.

The Bayesian probability of correct selection when system i is selected as best, is then

$$1 - E_{\mathbf{W}}[\mathcal{L}_{0-1}(i, \mathbf{w})],$$

where the expectation is taken with respect to whatever information is available (e.g., after the first stage, the distribution in Eq. 7 is appropriate). This is an average-case PCS, not a PCS for the worst-case LFC.

Budget Constraint. A fourth difference is that we propose to reduce the expected loss for a specified amount of second-stage simulation effort. Indifference-zone procedures, on the other hand, seek to reduce the number of replications required to achieve a specified PCS. There is practical motivation for this difference (e.g., declare which system is best by tomorrow at 9am). We therefore assume that after the first stage, the decision-maker selects a second-stage budget b, so that $r_2 \cdot |\mathcal{C}_2| = b$, where $|\mathcal{C}_2|$ is the number of systems in $\mathcal{C}_2$.

3.1 Design Of A 'Good' Second Stage

The problem is to select a non-empty subset $\mathcal{C}_2 \subseteq \mathcal{U} = \{1, 2, \ldots, k\}$ of systems for simulation during the second stage, based on information from the first-stage output $\mathcal{E}_\mathbf{x}$, that minimizes the predicted value of the expected loss after running $r_2 = b/|\mathcal{C}_2|$ replications of systems in $\mathcal{C}_2$ with CRN.

Suppose that for a given $\mathcal{C}_2$ and r_2, the second-stage output $\mathcal{E}_\mathbf{y}$ is observed. Denote the expected value of the unknown mean, conditional on all output, by

$$\mathbf{z} = E[\mathbf{W} \mid \mathcal{E}_\mathbf{x}, \mathcal{E}_\mathbf{y}].$$

Denote the system with the posterior expectation for the unknown mean by

$$d_\mathbf{y}^N = \arg \max_{i:i=1,\ldots,k} z_i. \tag{8}$$

Then the expected loss is

$$E_{\mathbf{W}|\mathcal{E}_\mathbf{x},\mathcal{E}_\mathbf{y}}[\mathcal{L}\left(d_\mathbf{y}^N, \mathbf{W}\right) \mid \mathcal{E}_\mathbf{x}, \mathcal{E}_\mathbf{y}],$$

where $\mathcal{L}\left(d_\mathbf{y}^N, \mathbf{W}\right)$ is the relevant loss function from Eq. 4 or Eq. 5. Although $d_\mathbf{y}^N$ is sub-optimal for the 0-1 loss function (Berger 1988), it is used implicitly by Procedures $\mathcal{CY}$ and $\mathcal{NM}$, is intuitive, and allows for a relatively straightforward analysis.

After the first stage, but prior to the second, the as-yet unseen second-stage output

$$\mathcal{E}_\mathbf{Y} = \left(\mathbf{Y}_{\mathcal{C}_2,1}, \ldots, \mathbf{Y}_{\mathcal{C}_2,r_2}\right)$$

is random. Because the second-stage output is random before it is observed, the posterior expectation of the unknown mean is also random,

$$\mathbf{Z} = E[\mathbf{W} \mid \mathcal{E}_\mathbf{x}, \mathcal{E}_\mathbf{Y}].$$

The distribution of $\mathbf{Z}$ determines the probability that a given system will be selected as best, once the second-stage output is observed. Its specific distribution is described in Sec. 4.

The prediction for expected loss, given that $\mathcal{E}_\mathbf{Y}$ will be, but has not yet been observed, is then obtained by averaging over the random second-stage output,

$$\rho^*(r_2, \mathcal{C}_2) = E_{\mathbf{Y}|\mathcal{E}_\mathbf{x}}\left[E_{\mathbf{W}|\mathcal{E}_\mathbf{x},\mathcal{E}_\mathbf{Y}}[\mathcal{L}\left(d_\mathbf{Y}^N, \mathbf{W}\right) \mid \mathcal{E}_\mathbf{x}, \mathcal{E}_\mathbf{Y}]\right].$$

This leads to the optimization problem

$$\begin{aligned} \min_{\mathcal{C}_2} \quad & \rho^*(r_2, \mathcal{C}_2) \\ \text{s.t.} \quad & r_2 = b/|\mathcal{C}_2|. \end{aligned} \tag{9}$$

We require $\mathcal{C}_2$ to be nonempty, although we comment below on situations where a second-stage might be foregone. Our mathematical analysis considers $r_2 = b/|\mathcal{C}_2|$ to be real-valued, but the numerical experiments of Sec. 6 round r_2 down to an integer.

4 FORMULAS FOR PROCEDURES

Several formulas are important for describing the new selection procedures with CRN. See Chick and Inoue (1999a) for a thorough mathematical treatment. Here, let $[i]$ be a permutation that orders the first-stage sample means, $\bar{x}_{[1]} \leq \ldots \leq \bar{x}_{[k]}$. Let $\phi_\nu(s)$ and $\Phi_\nu(s)$ denote the pdf and cdf, respectively, of the standard t-distributed random variable with ν degrees of freedom, and define $\Psi_\nu(s) = \int_s^\infty (x-s)\phi_\nu(x)dx = \frac{\nu+s^2}{\nu-1}\phi_\nu(s) - s(1-\Phi_\nu(s))$.

Anderson (1957) shows that if the systems $\mathcal{C}_2$ are simulated r_2 times in the second stage, the maximum likelihood estimator (MLE) for the unknown mean is

$$\begin{aligned} \hat{\hat{\boldsymbol{\mu}}}_{\mathcal{C}_2} &= (\sum_{i=1}^{r_1} \mathbf{x}_{\mathcal{C}_2,i} + \sum_{j=1}^{r_2} \mathbf{y}_{\mathcal{C}_2,j})/(r_1+r_2) \\ \hat{\hat{\boldsymbol{\mu}}}_{\mathcal{C}_1} &= \hat{\boldsymbol{\mu}}_{\mathcal{C}_1} + \left(\hat{\hat{\boldsymbol{\mu}}}_{\mathcal{C}_2} - \hat{\boldsymbol{\mu}}_{\mathcal{C}_2}\right)\bar{\mathbf{S}}_{\mathcal{C}_2\mathcal{C}_2}^{-1}\bar{\mathbf{S}}_{\mathcal{C}_2\mathcal{C}_1}. \end{aligned} \tag{10}$$

The number of hats, $\hat{}$ or $\hat{\hat{}}$, indicates the number of stages of data used for the estimate.

Further, the posterior expectation for the unknown mean after the second stage is $\mathbf{z} = \hat{\hat{\boldsymbol{\mu}}}$, and the predictive distribution $p_{\mathbf{Z}|\mathcal{E}_\mathbf{x},\mathcal{E}_\mathbf{Y}}(\mathbf{z})$ is multivariate t,

$$\mathbf{St}_k\left(\hat{\boldsymbol{\mu}}, \frac{(r_1+r_2)}{r_2} r_1(r_1-1)\boldsymbol{\beta}_\mathbf{Z}{}^{-1}, r_1-1\right), \tag{11}$$

where $\boldsymbol{\beta}_\mathbf{Z}$ is identical to $\bar{\mathbf{S}}$ except for the submatrix corresponding to systems simulated for only one stage, where $\boldsymbol{\beta}_{\mathbf{Z},\mathcal{C}_1\mathcal{C}_1} = \bar{\mathbf{S}}_{\mathcal{C}_1\mathcal{C}_2}\bar{\mathbf{S}}_{\mathcal{C}_2\mathcal{C}_2}^{-1}\bar{\mathbf{S}}_{\mathcal{C}_2\mathcal{C}_1}$.

The probability that one system will be preferred over another after both stages using predictions based on first-stage output alone, is determined by the distributions of $\mathbf{Z}$. When no further replications are done ($r_2 = 0$), the system selected as best has the highest first-stage sample mean ($d_\mathbf{y}^N = [k]$). When $r_2 > 0$, the predictive distribution $p_{Z_i - Z_j|\mathcal{E}_\mathbf{x},\mathcal{E}_\mathbf{Y}}\left(z_i - z_j\right)$ of the univariate difference $Z_i - Z_j$ in the expected mean performance of systems i and j is determined from Eq. 11 to be:

$$\mathbf{St}\left(\bar{x}_i - \bar{x}_j, \frac{(r_1+r_2)}{r_2}\tau_{\mathcal{C}_2,i,j}, r_1-1\right), \tag{12}$$

where

$$\tau_{\mathcal{C}_2,i,j} = r_1(r_1-1)\left[(e_i-e_j)\boldsymbol{\beta}_Z(e_i-e_j)^t\right]^{-1}, \qquad (13)$$

and e_i is the $k-$dimensional vector of all 0's, except for a 1 in the $i-$th coordinate. If all systems are simulated in the second stage, substitute $\bar{\mathbf{S}}$ for $\boldsymbol{\beta}_Z$ into Eq. 13. Define $\xi_{\mathcal{C}_2,[i]} = \left(\tau_{\mathcal{C}_2,[i],[k]}\right)^{1/2}(\bar{x}_{[i]} - \bar{x}_{[k]})$.

Chick and Inoue (1999a) indicate that the expected opportunity cost loss $\rho^*(\mathcal{C}_2, r_2)$ when $k = 2$ is exactly

$$\rho^*(\mathcal{C}_2, r_2) = \tau_{\mathcal{U},[1],[2]}^{-1/2}\Psi_{r_1-1}\left[\xi_{\mathcal{U},[1]}\right] - \left(\tfrac{r_1+r_2}{r_2}\tau_{\mathcal{C}_2,[1],[2]}\right)^{\frac{-1}{2}}\Psi_{r_1-1}\left[\left(\tfrac{r_1+r_2}{r_2}\right)^{\frac{1}{2}}\xi_{\mathcal{C}_2,[1]}\right] \qquad (14)$$

They use this observation to motivate a Bonferroni-like sum of pairwise losses for use as a surrogate objective function,

$$\mathrm{Bonf}_{o.c.}(r_2, \mathcal{C}_2) = \sum_{i=1}^{k-1}\tau_{\mathcal{U},[i],[k]}^{-1/2}\Psi_{r_1-1}\left[\xi_{\mathcal{U},[i]}\right] - \left(\tfrac{r_1+r_2}{r_2}\tau_{\mathcal{C}_2,[i],[k]}\right)^{\frac{-1}{2}}\Psi_{r_1-1}\left[\left(\tfrac{r_1+r_2}{r_2}\right)^{\frac{1}{2}}\xi_{\mathcal{C}_2,[i]}\right] \qquad (15)$$

that will be optimized rather than $\rho^*(\mathcal{C}_2, r_2)$. If all systems are simulated infinitely often ($r_2 \to \infty$, $\mathcal{C}_2 = \mathcal{U}$), then $\mathrm{Bonf}_{o.c.}(r_2, \mathcal{C}_2) \to 0$, as desired with the value of perfect information. As $r_2 \to 0$, then $\mathrm{Bonf}_{o.c.}(r_2, \mathcal{C}_2)$ approaches the Bonferroni-like bound for expected loss when $[k]$ is selected with no further replications, $\mathrm{Bonf}_{o.c.}(r_2, \mathcal{C}_2) \to \sum_{i=1}^{k-1}\tau_{\mathcal{U},[i],[k]}^{-1/2}\Psi_{r_1-1}\left[\xi_{\mathcal{U},[i]}\right]$.

A similar Bonferroni-like surrogate objective is obtained for the 0-1 loss function. Chick and Inoue (1999a) indicate that with an extra approximation, the following Bonferroni-like surrogate objective function for 0-1 loss can be obtained.

$$\mathrm{Bonf}_{0-1}(r_2, \mathcal{C}_2) = \sum_{i=1}^{k-1}\Phi_{r_1-1}[\xi_{\mathcal{U},[i]}] - \Phi_{r_1-1}\left[\left(\tfrac{r_1+r_2}{r_2}\right)^{1/2}\xi_{\mathcal{C}_2,[i]}\right] \qquad (16)$$

The difference in the summand is a reasonable approximation to the expected loss in a pairwise comparison between systems $[i]$ and $[k]$. When all systems are simulated ($\mathcal{C}_2 = \mathcal{U}$) and the number of replications grows without bound, then $\lim_{r_2\to\infty}\mathrm{Bonf}_{0-1}(r_2, \mathcal{C}_2) = 0$, so the loss is 0 with perfect information. With no additional replications, $\lim_{r_2\to 0}\mathrm{Bonf}_{0-1}(r_2, \mathcal{C}_2) = \sum_{i=1}^{k-1}\Phi_{r_1-1}[\xi_{\mathcal{U},[i]}]$ is exactly the Bonferroni bound on the probability of *incorrect* selection after the first stage.

5 NEW SELECTION PROCEDURES

There are four new procedures presented by Chick and Inoue (1999a) for reducing the risk of incorrect selections by designing the second-stage to minimize the Bonferroni-like surrogate objective functions. There is an exhaustive and heuristic procedure for each of loss function, the expected opportunity cost and 0-1 loss in Eq. 15 and Eq. 16, respectively.

Procedure $\mathcal{OC}_{\mathrm{crn}}$ below is an exhaustive procedure that checks all $2^k - 1$ non-empty subsets of $\mathcal{U}$ to find the subset $\mathcal{C}_2$ that minimizes the surrogate objective $\mathrm{Bonf}_{o.c.}(r_2, \mathcal{C}_2)$. Procedure 0-1$_{\mathrm{crn}}$, the 0-1 loss analog, is obtained by replacing appropriate terms (e.g. $\mathrm{Bonf}_{0-1}(r_2, \mathcal{C}_2)$ for $\mathrm{Bonf}_{o.c.}(r_2, \mathcal{C}_2)$). After observing the first-stage output, the decision-maker selects a second-stage budget b. The selection of b may be guided by average CPU times for replications during the first stage, in order to control the time required to complete the procedure. If the Bonferroni-like surrogate objective is satisfactorily low after the first stage (e.g., $\sum_{i=1}^{k-1}\Phi_{r_1-1}[\xi_{\mathcal{U},[i]}] \le 1 - P^*$), then the second-stage replications may be foregone.

5.1 Procedure $\mathcal{OC}_{\mathbf{crn}}$

1. Pick a first-stage sample size $r_1 \ge k$. Take i.i.d. sample $\mathbf{x}_1, \ldots, \mathbf{x}_{r_1}$, from each of the k systems *using CRN across systems*.
2. Determine the sample means $\bar{\mathbf{x}}$, the permutation $[i]$ so that $\bar{x}_{[1]} \le \ldots \le \bar{x}_{[k]}$, and set $\bar{\mathbf{S}} = \sum_{j=1}^{r_1}(\mathbf{x}_j - \bar{\mathbf{x}})^t(\mathbf{x}_j - \bar{\mathbf{x}})$
3. Select a budget b for second-stage replications.
4. Choose the subset of systems $\mathcal{C}_2$ of $\mathcal{U} = \{1, \ldots, k\}$ that minimizes the Bonferroni-like surrogate function $\mathrm{Bonf}_{o.c.}(r_2, \mathcal{C}_2)$ from Eq. 15, where $r_2 = b/|\mathcal{C}_2|$.
5. Take $r_2 = \lfloor b/|\mathcal{C}_2|\rfloor$ additional i.i.d. observations $\mathbf{y}_{\mathcal{C}_2,1}, \ldots, \mathbf{y}_{\mathcal{C}_2,r_2}$ from each system in $\mathcal{C}_2$, using CRN across systems.
6. Compute the second-stage sample mean for systems in $\mathcal{C}_2$ and the 'regression estimate' for systems in $\mathcal{C}_1 = \mathcal{U}\backslash\mathcal{C}_2$.

$$\mathbf{z}_{\mathcal{C}_2} = (\sum_{i=1}^{r_1}\mathbf{x}_{\mathcal{C}_2,i} + \sum_{j=1}^{r_2}\mathbf{y}_{\mathcal{C}_2,j})/(r_1+r_2)$$

$$\mathbf{z}_{\mathcal{C}_1} = \bar{\mathbf{x}}_{\mathcal{C}_1} + \left(\mathbf{z}_{\mathcal{C}_2} - \bar{\mathbf{x}}_{\mathcal{C}_2}\right)\bar{\mathbf{S}}_{\mathcal{C}_2\mathcal{C}_2}^{-1}\bar{\mathbf{S}}_{\mathcal{C}_2\mathcal{C}_1}$$

7. Select the system with the largest z_i as best.

Because the runtime of the exhaustive procedures are not polynomial in the number of systems, suboptimal heuristics are proposed that check at most $2k$ subsets. Procedure $\mathcal{OC}_{\mathrm{crn:h}}$, the heuristic for opportunity cost, is presented below. The idea is to first evaluate the surrogate objective

when all systems are simulated, $\mathcal{C}_2 = \mathcal{U}$. The surrogate objective is then evaluated again, this time dropping out either the system that contributes least to reducing $\text{Bonf}_{o.c}(r_2, \mathcal{C}_2)$. If screening that system out, or screening out system $[k]$ improves the surrogate objective, then the process is continued until screening out additional systems no longer decreases the surrogate objective. The heuristic requires $O(k^4)$ time, as there are $O(k)$ iterations, and $O(k^3)$ work per iteration for matrix algebra. Procedure 0-$1_{\text{crn:h}}$, the opportunity cost analog of Procedure $\mathcal{OC}_{\text{crn:h}}$, is obtained by substituting terms from $\text{Bonf}_{o.c.}(r_2, \mathcal{C}_2)$ with terms from $\text{Bonf}_{0-1}(r_2, \mathcal{C}_2)$.

Numerical experiments with the heuristic procedures indicate that the screening step (Step 5.2 of Procedure $\mathcal{OC}_{\text{crn:h}}$) should not round r_2 down when evaluating the surrogate objective. If $b/(|\mathcal{C}_2| - 1)$ is rounded down, then too few systems may be screened.

5.2 Procedure $\mathcal{OC}_{\text{crn:h}}$

1. Do steps 5.1 to 5.1 of Procedure $\mathcal{OC}_{\text{crn}}$.
2. Initialize the set of systems $\mathcal{C}_2$ considered for second-stage analysis to be $\mathcal{U} = \{1, \ldots, k\}$.
3. Compute $r_2 = b/|\mathcal{C}_2|$ and evaluate the Bonferroni-like surrogate for expected opportunity cost loss, $\text{Bonf}_{o.c.}(r_2, \mathcal{C}_2)$.
4. If there is only one system under consideration ($|\mathcal{C}_2| = 1$), then go to Step 5.1.
5. Determine the set $[i] \in \mathcal{C}_2 \backslash \{[k]\}$ that contributes least to reducing $\text{Bonf}_{o.c.}(r_2, \mathcal{C}_2)$, e.g., minimizes

$$\left(\frac{r_1 + r_2}{r_2}\right)^{\frac{-1}{2}} \Psi_{r_1 - 1}\left[\left(\frac{r_1 + r_2}{r_2}\right)^{\frac{1}{2}} \xi_{\mathcal{C}_2,[i]}\right],$$

   and set $r_2' = b/(|\mathcal{C}_2| - 1)$. If eliminating i from $\mathcal{C}_2$ reduces the surrogate bound ($\text{Bonf}_{o.c.}(r_2', \mathcal{C}_2 \backslash \{i\}) < \text{Bonf}_{o.c.}(r_2, \mathcal{C}_2)$), then screen i from the set of considered systems $\mathcal{C}_2$ (assign $\mathcal{C}_2 \quad \mathcal{C}_2 \backslash \{i\}$ and $r_2 \quad r_2'$) & to Step 5.2.
6. If $[k] \in \mathcal{C}_2$ and screening out $[k]$ improves the surrogate objective, then do so & go to Step 5.2.
7. Do steps 5.1 to 5.1 of Procedure $\mathcal{OC}_{\text{crn}}$.

6 EMPIRICAL EVALUATION

Here we evaluate the effectiveness of the indifference-zone Procedures $\mathcal{CY}$ and $\mathcal{NM}$; the new exhaustive and heuristic procedures for reducing the expected 0-1 loss (Procedures 0-1_{crn} and 0-$1_{\text{crn:h}}$); and the opportunity cost analogs (Procedures $\mathcal{OC}_{\text{crn}}$ and $\mathcal{OC}_{\text{crn:h}}$). Each procedure is evaluated in terms of two measures of effectiveness, when applied to determine the best of the five (s,S) inventory policies from Koenig and Law (1985). Each measure of effectiveness is estimated by (i) running 3000 independent first stages with $r_1 = 10$, (ii) running independent second stages for each procedure, then (iii) averaging the 3000 outcomes of each procedure.

Each measure is evaluated as a function of the second-stage budget b. Since Procedures $\mathcal{CY}$ and $\mathcal{NM}$ both require all systems to be simulated, they give the same allocation (although they might give a different PCS guarantee for the LFC), and therefore perform identically with respect to our measures. We choose b to be a multiple of 5, and round down r_2 for our new procedures. This gives a slight advantage to the indifference-zone procedures.

The first measure of effectiveness is the widely-used empirical probability of correct selection (EmpPCS), the fraction of times that a selection procedure selects a known best system. In this case, inventory policy 2 is presumed best, based on tens of thousands of simulation runs. For the inventory example, EmpPCS is plotted in Fig. 1. All four new procedures clearly outperform the indifference-zone procedures for this experiment. Procedure 0-1_{crn} performs somewhat less well than the other three new procedures, perhaps due to approximations made in its derivation. The other procedures all performed rather similarly.

A second measure is the Bonferroni-like surrogate for the predicted expected opportunity cost loss (PredBOCL) of Eq. 15. This measure is computed by averaging over many first-stage procedures, rather than considering both stages of output.

Chick and Inoue (1999a) present empirical evidence that the new procedures also outperform Procedures $\mathcal{CY}$ and $\mathcal{NM}$ with respect to the expected P-value after both stages, the expected opportunity cost loss after both stages, and the average predictive estimate of the probablity of correct selection $\text{Bonf}_{0-1}(r_2, \mathcal{C}_2)$ after one stage. Procedure $\mathcal{OC}_{\text{crn:h}}$ seems particularly effective across all measures.

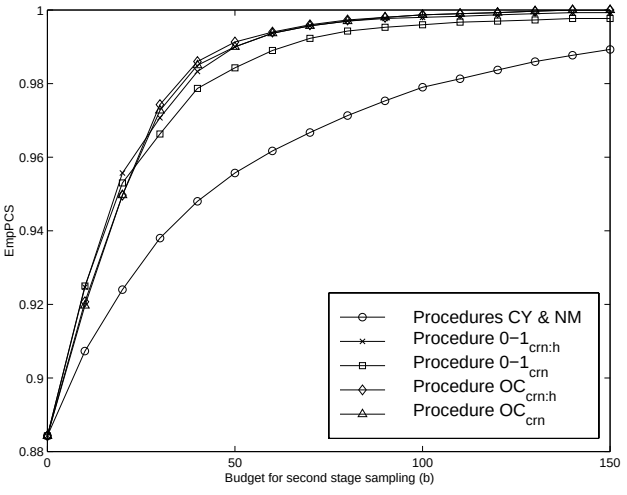

Figure 1: Empirical Fraction of Correct Selections.

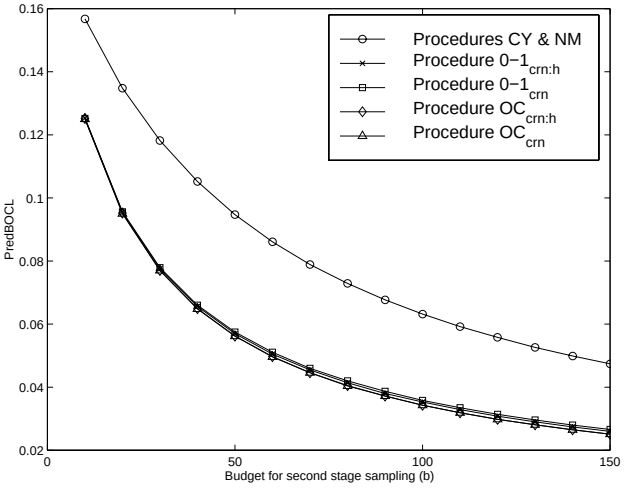

Figure 2: Predicted Expected Opportunity Cost Loss.

7 DISCUSSION

All four new procedures perform better than both indifference zone procedures with respect to all five measures of efficiency in the experiments of Sec. 6. This is not surprising, since two of the systems in the numerical study had performances that could be identified as unlikely contenders for the best after the first stage. These two systems were typically not simulated by our new procedures during the second stage, although the indifference-zone procedures were obligated to run additional replications for them.

It is somewhat surprising that Procedure 0-$1_{\text{crn:h}}$ performs somewhat better than Procedure 0-1_{crn} with respect to EmpPCS. The difference, however, is within sampling error, and might also be associated with slack in the Bonferroni-like approximation for $\text{Bonf}_{0-1}(r_2, \mathcal{C}_2)$. The opportunity cost loss procedures perform strongly as well, perhaps because there are fewer approximations in with the surrogate objective for the opportunity cost loss than for the 0-1 loss function.

A potential criticism of the new procedures is that a bound on the expected loss is not provided. On the other hand, for $k = 2$ the predictive pairwise opportunity cost loss is exact, so the budget b can be increased to insure a given expected opportunity cost loss. The surrogate $\text{Bonf}_{0-1}(r_2, \mathcal{C}_2)$ is only an approximation, not a bound, and therefore cannot provide an average-case PCS guarantee by increasing b. On the other hand, the new procedures provide a higher EmpPCS than the indifference zone procedures as a function of b, at least for the inventory example considered above.

The new procedures require more computation than the indifference-zone procedures. Whereas Procedures $\mathcal{CY}$ and $\mathcal{NM}$ are $O(k^2)$ in the number of systems, the heuristic Procedures 0-$1_{\text{crn:h}}$ and $\mathcal{OC}_{\text{crn:h}}$ are both $O(k^4)$. For the experiments of Sec. 6 the heuristic procedures require about 0.7 CPU seconds to compute the allocations. For large k or small simulation runtimes, CPU time might be better spent running replications rather than calculating allocations. For smaller k, or larger b or simulation runtimes, the benefit of screening is more likely to compensate for the cost of computing $\mathcal{C}_2$.

8 CONCLUSIONS

Indifference-zone formulations have focused on designing procedures that guarantee a minimum probability of correct selection, given the worst-case LFC. The approach here considers the expected probability of correct selection in a Bayesian sense, so that information from first-stage replications can be used in lieu of the LFC. We also consider that business decisions sometimes require an accounting of expected opportunity cost loss, rather than the probability of correct selection.

Although the closed-form solution of our general formulation is not presently known, we present theoretically justifiable procedures for reducing the risk of incorrect selection. The idea is to reduce a Bonferroni-like approximation of the sum of pairwise losses.

For a well-known inventory selection problem, all four new procedures perform favorably relative to two indifference-zone procedures. Procedure $\mathcal{OC}_{\text{crn:h}}$ seems particularly effective.

REFERENCES

Anderson, T. W. 1957. Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association 52*, 200–203.

Banks, J., J. S. Carson, and B. L. Nelson. 1996. *Discrete-Event System Simulation* (2nd ed.). Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

Bechhofer, R. E., T. J. Santner, and D. M. Goldsman. 1995. *Design and Analysis for Statistical Selection, Screening, and Multiple Comparisons*. New York: John Wiley & Sons, Inc.

Berger, J. O. 1988. A Bayesian approach to ranking and selection of related means with alternatives to analysis-of-variance methodology. *Journal of the American Statistical Association 83*(402), 364–373.

Bernardo, J. M. and A. F. M. Smith. 1994. *Bayesian Theory*. Chichester, UK: Wiley.

Chick, S. E. and K. Inoue. 1999a. New procedures for identifying the best simulated system using common random numbers. *in submission*.

Chick, S. E. and K. Inoue. 1999b. New two-stage and sequential procedures for selecting the best simulated system. *in submission*.

Clark, G. M. and W.-N. Yang. 1986. A Bonferroni selection procedure when using common random numbers with unknown variances. In J. Wilson, J. Hendriksen, and S. Roberts (Eds.), *Proceedings of the Winter Simulation Conference*, pp. 313–315. Institute of Electrical and Electronics Engineers, Inc.

de Groot, M. H. 1970. *Optimal Statistical Decisions*. New York: McGraw-Hill, Inc.

Goldsman, D. and B. L. Nelson. 1998. Statistical screening, selection, and multiple comparisons in computer simulation. In D. J. Madeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan (Eds.), *Proceedings of the Winter Simulation Conference*, pp. 159–166. Institute of Electrical and Electronics Engineers, Inc.

Hochberg, Y. and A. C. Tamhane. 1987. *Multiple Comparison Procedures*. New York: John Wiley & Sons.

Inoue, K. and S. E. Chick. 1998. Comparison of Bayesian and frequentist assessments of uncertainty for selecting the best system. In D. J. Medeiros, E. J. Watson, M. Manivannan, and J. Carson (Eds.), *Proceedings of the Winter Simulation Conference*, pp. 727–734. Institute of Electrical and Electronics Engineers, Inc.

Koenig, L. W. and A. M. Law. 1985. A procedure for selecting a subset of size m containing the ℓ best of k independent normal populations, with applications to simulation. *Commun. Statist.-Simulation and Computation 14*(3), 719–734.

Law, A. M. and W. D. Kelton. 1991. *Simulation Modeling & Analysis* (2nd ed.). New York: McGraw-Hill, Inc.

Nelson, B. L. and F. J. Matejcik. 1995. Using common random numbers for indifference-zone selection and multiple comparisons in simulation. *Management Science 41*, 1935–1945.

Nelson, B. L., J. Swann, D. Goldsman, and W. Song. 1998. Simple procedures for selecting the best simulated system when the number of alternatives is large. *Northwestern University, IEMS Technical Report*.

AUTHOR BIOGRAPHY

STEPHEN E. CHICK is an assistant professor of Industrial and Operations Engineering at the University of Michigan, Ann Arbor. In addition to stochastic simulation, his research interests include Bayesian statistics, reliability, decision analysis, and computational methods in statistics. His research is motivated by projects in manufacturing and health care. His work experience includes several years of material handling system design for the automotive industry using simulation analysis.

KOICHIRO INOUE is finishing his Ph.D. thesis at the Department of Industrial and Operations Engineering at the University of Michigan, Ann Arbor. His dissertation work focuses on the application of Bayesian statistics to simulation output analysis, particularly for multiple comparisons problems. He is also interested in multi-attribute decision analysis.