Proceedings of the 2000 Winter Simulation Conference
J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

NEW RESULTS ON PROCEDURES THAT SELECT THE BEST SYSTEM USING CRN

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

One use of simulation is to inform decision makers that seek to select the best of several alternative systems. The system with the highest (or lowest) mean value for simulation output is often selected as best, and simulation output is used to infer the value of the unknown mean of each system. Statistical procedures that help to identify the best system by suggesting an appropriate number of replications for each system are therefore useful tools in simulation. This article explores the performance of representative procedures from two approaches to develop statistical procedures, with the goal of understanding tradeoffs involving the ease of use, computational requirements, and the range of applicability. The focus is primarily on procedures that use common random numbers to sharpen comparisons between systems.

1 INTRODUCTION

One use of simulation is to inform decision makers that want to select the best of several alternative manufacturing or service systems (Law and Kelton 1991; Banks, Carson, and Nelson 1996), where best is defined in terms of the mean performance of each system. This motivates the need for procedures to help simulation experiments efficiently identify the best system.

Four distinct formulations for selecting the best system have emerged. Indifference-zone procedures selects one system that is guaranteed to be best with a prespecified probability P^*, as long as the mean of the best system is better than a smallest practically significant difference δ^* than the others. Matejcik and Nelson (1995) indicate that many indifference-zone procedures can also select a system within δ^* of the best with probability P^*, regardless of the configuration of the means. The subset selection formulation, on the other hand returns a subset of simulated systems that contains the best system with a prespecified probability. Nelson, Swann, Goldsman, and Song (1999) unify these approaches with a two-stage combined procedure (Procedure $\mathcal{C}$) that screens inferior systems with subset selection, then allocates second-stage replications to provide an indifference-zone guarantee. They also show that Procedure $\mathcal{C}$ performs quite well in a battery of empirical tests. For a review of the indifference-zone and subset selection formulations, see Bechhofer, Santner, and Goldsman (1995) or Goldsman and Nelson (1998).

Chen (1996) proposes the optimal computer budget allocation ($\mathcal{OCBA}$), a third formulation for selection procedures. The idea is to sequentially allocate replications in order to improve the evidence for correct selection, based on a thought experiment involving the Bayesian posterior probability of correct selection. Empirical experiments demonstrate that this procedure can result in significant efficiency improvements, as measured by the expected number replications required to obtain a given empirical fraction of correct selections, when the means of each system differ substantially (Chen, Chen, Lin, and Yücesan 1999).

The fourth formulation is the Bayesian decision-theoretic framework developed by Chick and Inoue (2000b). Both two-stage and sequential procedures are available. The idea is to improve the expected value of information of additional replications. One innovation is that replications can be allocated to either improve the probability of correct selection (Procedure 0-1($\mathcal{B}$)), or to reduce the expected opportunity cost of a potentially incorrect decision (Procedure $\mathcal{LL}(\mathcal{B})$).

Both the indifference-zone and subset selection formulations guarantee a bound on the probability of correct selection statements using the smallest number of replications. The Bayesian procedures ($\mathcal{OCBA}$ and work of Chick and Inoue 2000b) attempt to maximally improve the evidence for correct selection using a constrained number of replications. The Bayesian approaches therefore avoid the statistical conservativeness of the indifference-zone approach, but require several approximations to allow for a quickly computable allocation of replications, and do not yet have a provable probability of correct selection guarantee. Given these tradeoffs, empirical evaluations can shed light on which procedure may be best suited for a given application.

A number of papers compare the performance of these various approaches when simulation replications use independent random numbers. Nelson, Swann, Goldsman, and Song (1999) have a large empirical study that describes a variety settings where Procedure $\mathcal{C}$ provides significant efficiencies relative to the well-known two-stage indifference-zone procedure of Rinott (1978). Inoue, Chick, and Chen (1999) indicate that the sequential procedures of Chen (1996) and Chick and Inoue (2000b) outperform Rinott's procedure over a range of empirical experiments, with respect to several measures of effectiveness. Chick and Inoue (2000b) indicate that the two-stage versions of Procedures 0-1($\mathcal{B}$) and $\mathcal{LL}(\mathcal{B})$ seem to select the unique best system somewhat better than Procedure $\mathcal{C}$ in a small empirical study, yet Procedure $\mathcal{C}$ often selects a good system (within δ^* of best) when the unique best system is not selected. Each of these procedures has roughly the same computational burden to allocate the replications for a given stage.

The relative performance of procedures that use common random numbers (CRN) to sharpen comparisons, however, is less well understood. This paper compares the empirical performance of selection procedures that allow the use of CRN. A subset of the numerical results below, supply data that are alluded to, but not presented, can be found in a submitted work (Chick and Inoue 2000a).

2 SELECTION PROCEDURES WITH CRN

Common random numbers (CRN) can be used to sharpen comparisons between systems (Law and Kelton 1991; Banks, Carson, and Nelson 1996). There are a handful of selection procedures that allow for CRN, and each is run as a two-stage procedure. In the first stage, a small number of replications is run for each system. The sample means, variances and covariances from the first stage are then used to determine the number of second stage replications for each system. The second stage of each procedure depends on assumptions about the covariance structure due to CRN, and which of the four selection procedure approaches is taken.

Clark and Yang (1986) introduced the first indifference-zone selection procedure to use CRN (Procedure $\mathcal{CY}$). They do not make assumptions about the correlation structure induced by CRN. Nelson and Matejcik (1995) present an alternative two-stage procedure (Procedure $\mathcal{NM}$) that can reduce the number of replications, but presume a sphericity condition (if $X_{i,j}$ is the output of the j-th replication of system i, then $\mathrm{Var}[X_{i,j}] = 2\phi_i + \tau^2$, and $\mathrm{Cov}[X_{i,j}, X_{i',j}] = \phi_i + \phi_{i'}$ for systems $i \neq i'$). They show empirically that the procedure is robust to deviations from sphericity. Both Procedures $\mathcal{CY}$ and $\mathcal{NM}$ simulate all systems the same number of times during the second stage.

Bayesian decision-theoretic alternatives have also been developed. Chick and Inoue (1999) and Chick and Inoue (2000a) describe two-stage procedures that allocate replications to either improve the probability of correct selection (Procedure $0\text{-}1_{\mathrm{crn}}$) or reduce the expected opportunity cost of a potentially incorrect selection (Procedure $\mathcal{OC}_{\mathrm{crn}}$).

Both Procedures $0\text{-}1_{\mathrm{crn}}$ and $\mathcal{OC}_{\mathrm{crn}}$ allow for screening, in the sense that a subset of systems might be simulated during the second stage, in order to improve the expected value of information from the additional replications. The use of CRN implies that information gained about systems simulated during the second stage also gives information about systems that were screened from the second stage, and the procedures use missing data analysis to account for that information. Unfortunately, Procedures $0\text{-}1_{\mathrm{crn}}$ and $\mathcal{OC}_{\mathrm{crn}}$ impose a large computational burden to screen the systems (if there are k systems, they check all $2^k - 1$ subsets that might be screened out). Chick and Inoue (2000a) therefore present heuristics that check at most $2k$ subsets for screening, Procedures $0\text{-}1_{\mathrm{crn:h}}$ and $\mathcal{OC}_{\mathrm{crn:h}}$.

These procedures and their major assumptions are summarized in Table 1.

Table 1: Characteristics of Two-Stage Selection Procedures that use Common Random Numbers (CRN)

	Assumption for CRN	
	None	Sphericity
Indifference-Zone (PCS $\geq P^*$, no screening)	$\mathcal{CY}$	$\mathcal{NM}$
Bayesian (Value of Info, screening ok)	$0\text{-}1_{\mathrm{crn}}$,$0\text{-}1_{\mathrm{crn:h}}$, $\mathcal{OC}_{\mathrm{crn}}$,$\mathcal{OC}_{\mathrm{crn:h}}$	(none yet)

Inoue (2000) derives the probability models required for a two-stage Bayesian procedure, given a known correlation structure. Since the correlation is typically unknown in practice, that procedure is not evaluated here.

3 SELECTION PROBLEM SETUP

The procedures are applied to both a simple but realistic simulation problem, as well as a stylized problem. The first is an inventory policy selection problem considered initially by Koenig and Law (1985) and analyzed later by Nelson and Matejcik (1995). There are five (s, S) inventory policies for controlling the inventory level of a discrete product with stochastic demand. 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 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.

The stylized problem is in the monotone decreasing means (MDM) configuration, so the means of each system are evenly spaced, $w_i = w_1 - \delta^* \cdot (i-1)/\tau$, for $i = 2, \ldots k$. Here we set $\tau = 2$, $\delta^* = 1/\sqrt{r_1}$, and presume the sphericity condition, testing four combinations of covariance settings when CRN is used (variance $\sigma_i^2 = 1$ and 4; correlation $\rho = 0.5$ and 0.75). Each procedure is evaluated by (i) running a common first stage with $r_1 = 10$ replications of each system, using CRN across systems, (ii) determining the number of replications for each system during the second stage, (iii) then running a separate second stage for each procedure, using CRN across systems. The performance of each system is measured with the figures of merit in Section 4, based on 3000 applications of each procedure.

4 FIGURES OF MERIT

The first figure of merit is the empirical probability of correct selection (PCS), the fraction of times that a procedure correctly selects the best system.

The second figure of merit is the expectation of the Bonferroni bound, BPCS, for the posterior probability of correct selection after output from both stages is observed. Inoue and Chick (1998) indicate that BPCS also has an interpretation as the Bonferroni bound for frequentist evidence for correct selection based on P-values. To compute that bound, the estimate of the mean performance $\hat{\hat{\mu}}_i$ and correlation $\hat{\hat{\sigma}}_{i,j}$ after both stages are required. Let $[k]$ denote the system that is selected as best. We use the MLE to estimate BPCS after observing output from both stages,

$$\text{BPCS} = 1 - \sum_{i:i \neq [k]} (1 - P_i \qquad (1)$$

$$P_i = \Phi_\nu \left[\left(\hat{\hat{\mu}}_{[k]} - \hat{\hat{\mu}}_i \right) \lambda_i^{-1/2} \right] \qquad (2)$$

$$\lambda_i = (e_i - e_{[k]}) \text{Cov} \left[\hat{\hat{\boldsymbol{\mu}}}, \hat{\hat{\boldsymbol{\mu}}} \right] (e_i - e_{[k]})^t$$

where Φ_ν is the cumulative distribution function (cdf) for the standard t-distributed variable with ν degrees of freedom, P_i is the P-value for the hypothesis that the mean for system $[k]$ exceeds the mean for system i, e_i is the unit vector in the i-th coordinate direction, and $\text{Cov}\left[\hat{\hat{\boldsymbol{\mu}}}, \hat{\hat{\boldsymbol{\mu}}}\right]$ is estimated with formulas of Anderson (1957) (see the appendix).

The formulas require a missing data analysis when screening is used. Let $\mathcal{C}_2$ be the set of systems simulated in both stages, and $\mathcal{C}_1$ be the set of systems only simulated during the first stage. Let r_2 be the number of replications per system simulated during the second stage. We use $\nu = r_1 + r_2 - 1$ degrees of freedom when $i, [k] \in \mathcal{C}_2$, and use $\nu = r_1 - 1$ when $i, [k] \in \mathcal{C}_1$. When $i \in \mathcal{C}_2$ and $[k] \in \mathcal{C}_1$ or vice versa, we use a missing-data approximation of Little (1976) for ν (see the appendix). Sample averages from multiple applications of a selection procedure are used to estimate $E[\text{BPCS}]$.

A third figure of merit is PredBPCS, is the predicted Bonferroni-like approximation for the probability of correct selection, given that the first stage has been completed but the second stage has not. See Chick and Inoue (2000a) for a fuller discussion.

Some results also present the average total number of second stage replications (including zeros), denoted 'ANR'.

5 RESULTS

There are different ways to compare the procedures, corresponding to different ways of selecting the second-stage budget b.

5.1 Screening Compared to IZ Procedures

Screening might improve performance by running more replications per system, but simulating fewer systems so that the total number of replications is the same. This section evaluates whether screening can improve the ability to correctly select the best system by assuming that the number of replications for the Bayesian procedures is set to equal the total number of replications suggested by an indifference-zone procedure. After a common first-stage is run for each procedure, the second-stage allocation computed for Procedure $\mathcal{NM}$ is used as the budget $b = k \cdot r_2$ for the number of second stage of replications of the four Bayesian procedures, 0-$1_{\text{crn:h}}$, 0-1_{crn}, $\mathcal{OC}_{\text{crn:h}}$, and $\mathcal{OC}_{\text{crn}}$. 'ANR' is therefore the same for each procedure. A similar process compares Procedure $\mathcal{CY}$ with Procedures 0-$1_{\text{crn:h}}$, 0-1_{crn}, $\mathcal{OC}_{\text{crn:h}}$, and $\mathcal{OC}_{\text{crn}}$.

Table 2 summarizes the figures of merit for the MDM experiment ($k = 5$ systems; common $\sigma_i^2 = 1$ or 4; common correlation $\rho = 0.75$; $r_1 = 10$; $\delta^* = 1/\sqrt{10}$). Some PCS are below $P^* = 0.95$ because the indifference-zone constraint is not satisfied. The benefits of screening and average-case analysis apparently outweigh the deleterious effects of the approximations for this specific experiment, at least for Procedures 0-$1_{\text{crn:h}}$, $\mathcal{OC}_{\text{crn:h}}$, and $\mathcal{OC}_{\text{crn}}$, as these Bayesian procedures outperform both indifference-zone procedures on each figure of merit. This is particularly true when the variance is smaller for each system σ_i^2.

When the variance is larger, however, the benefit is not particularly sizable, as the proximity of the means reduces the ability of the procedure to screen effectively. A larger first stage sampling size may serve to help the screening of the Bayesian procedures by providing more certainty about the differences in means between each system. Procedure 0-1_{crn} performs less well than the other procedures, as it makes an extra approximation in its derivation, and it is therefore not recommended for general use. Further experiments indicate

Table 2: Results for the MDM Selection Problem ($k = 5$, $\rho = 0.75$) in a Test to Evaluate the Benefit Screening of Bayesian Procedures when the Budget is Determined by an Indifference-zone Procedure with $P^* = 0.95$ (see Section 5.1)

Variance	Procedure	ANR	PCS	E[BPCS]
$\sigma_i^2 = 1$	$\mathcal{NM}$	73	0.854	0.769
	0-$1_{\text{crn:h}}$		0.897	0.813
	0-1_{crn}		0.874	0.801
	$\mathcal{OC}_{\text{crn:h}}$		0.896	0.824
	$\mathcal{OC}_{\text{crn}}$		0.892	0.824
	$\mathcal{CY}$	325	0.971	0.917
	0-$1_{\text{crn:h}}$		0.987	0.937
	0-1_{crn}		0.965	0.927
	$\mathcal{OC}_{\text{crn:h}}$		0.985	0.939
	$\mathcal{OC}_{\text{crn}}$		0.984	0.941
$\sigma_i^2 = 4$	$\mathcal{NM}$	449	0.851	0.774
	0-$1_{\text{crn:h}}$		0.863	0.785
	0-1_{crn}		0.835	0.785
	$\mathcal{OC}_{\text{crn:h}}$		0.860	0.797
	$\mathcal{OC}_{\text{crn}}$		0.868	0.793
	$\mathcal{CY}$	1178	0.949	0.892
	0-$1_{\text{crn:h}}$		0.957	0.897
	0-1_{crn}		0.938	0.893
	$\mathcal{OC}_{\text{crn:h}}$		0.957	0.902
	$\mathcal{OC}_{\text{crn}}$		0.955	0.900

that Procedure 0-$1_{\text{crn:h}}$ beats Procedure 0-1_{crn} because it avoids looking at some negative correlation estimates, as discussed below.

Procedure $\mathcal{NM}$ performs better than the Bayesian procedures when $\sigma_i^2 = 4$ and $\rho = 0.5$, but the Bayesian procedures (except 0-1_{crn}) perform better when either $\sigma_i^2 = 1$ and $\rho = 0.5$, or when $\rho = 0.75$. Further experimentation may therefore warranted to evaluate whether Procedure $\mathcal{NM}$ dominates the Bayesian procedures over a broader range of experiments where the variance is larger and the covariance is weak.

When the number of systems is increased from $k = 5$ to 10 (and with r_1 increased to 20), the relative performance of the heuristic Bayesian procedures improves. Table 3 presents the results for $\sigma^2 = 4$, $\rho = 0.75$. Similar results hold when $\sigma^2 = 1$.

All four Bayesian procedures outperform both indifference-zone procedures on each measure of effectiveness for the inventory selection problem when $P^* = 0.95$ and $\delta^* = 2$ (all differences are statistically significant with 95% confidence, except for the comparisons with Procedure $\mathcal{CY}$ with respect to PCS). When $P^* = 0.95$ and $\delta^* = 1$, the PCS is essentially 1, since a very large number of replications are taken, and the Bayesian procedures have a slightly higher E[BPCS] than Procedures $\mathcal{CY}$ and $\mathcal{NM}$, except that Procedure $\mathcal{CY}$ slightly outperforms Procedure 0-1_{crn}.

Table 3: Results for the MDM Selection Problem, Now with $k = 10$ Systems ($\rho = 0.75$) in a Test to Evaluate the Benefit Screening of Bayesian Procedures when the Budget is Determined by an Indifference-zone Procedure (see Section 5.1)

Variance	Procedure	ANR	PCS	E[BPCS]
$\sigma_i^2 = 4$	$\mathcal{NM}$	2170	0.882	0.804
	0-$1_{\text{crn:h}}$		0.928	0.848
	$\mathcal{OC}_{\text{crn:h}}$		0.926	0.862
	$\mathcal{CY}$	5391	0.968	0.912
	0-$1_{\text{crn:h}}$		0.977	0.932
	$\mathcal{OC}_{\text{crn:h}}$		0.982	0.940

5.2 Benefit of Screening with CRN

We compare the performance of the four Bayesian selection procedures with the procedure that does no screening, as a function of the second stage budget b. The procedure that does no screening provides no PCS guarantee, but otherwise resembles Procedures $\mathcal{CY}$ and $\mathcal{NM}$ in that it deterministically allocates the same number of replications (b/k) to each of the k systems during the second stage.

The value of CRN for Bayesian selection procedures is evaluated by comparing the four Bayesian procedures that allow correlation from CRN with two analogous procedures (Procedure 0-1($\mathcal{B}$) to improve the probability of correct selection, and Procedure $\mathcal{LL}$($\mathcal{B}$) to improve the expected opportunity cost) that presume independent replications (Chick and Inoue 2000b). While CRN can sharpen comparisons by inducing a positive correlation, the requirement that some subset $\mathcal{C}_2$ each be simulated the same number of times is somewhat restrictive, and may be a source of inefficiency.

Table 4 summarizes E[BPCS] as a function of the second stage budget b for the inventory selection problem. The rightmost 5 columns of the table support the assertion that screening is of benefit for this problem, as the four Bayesian procedures significantly outperform the procedure that does not screen. The relative performance of the procedures with respect to PCS is very similar. Table 4 also indicates that CRN improves the ability to correctly select the best system. When there are no second stage replications, the use of CRN improves the E[BPCS] from 0.676 to 0.835.

Essentially the same conclusions are obtained for the MDM experiments ($k = 5$, four variations for the covariance matrix). Figure 1 illustrates the PCS improvement with a common variance $\sigma_i^2 = 1$ and correlation $\rho = 0.75$. The output for the procedures that require independent replications has a simulated correlation of 0 in this experiment.

5.3 Predictive Value of Second-Stage

Information about the unknown mean and variance is obtained after observing the first stage output. This gives information about what the output might be, before it is

Table 4: The Expected Bonferroni Bound on the Posterior Probability of Correct Selection (E[BPCS]) on the Inventory Selection Problem, for Two Bayesian Procedures that Require Independent Replications (0-1($\mathcal{B}$) and $\mathcal{LL}(\mathcal{B})$); the Four Bayesian Procedures that Allow Correlation from CRN; and the Procedure that does not Screen (see Section 5.2)

Second-Stage	Procedure						
Budget, b	0-1($\mathcal{B}$)	$\mathcal{LL}(\mathcal{B})$	0-$1_{\text{crn:h}}$	0-1_{crn}	$\mathcal{OC}_{\text{crn:h}}$	$\mathcal{OC}_{\text{crn}}$	No Screen
0	0.676	0.676	0.835	0.835	0.835	0.835	0.835
50	0.777	0.781	0.949	0.948	0.949	0.949	0.894
100	0.818	0.828	0.980	0.979	0.983	0.983	0.928
150	0.846	0.857	0.991	0.990	0.993	0.993	0.954

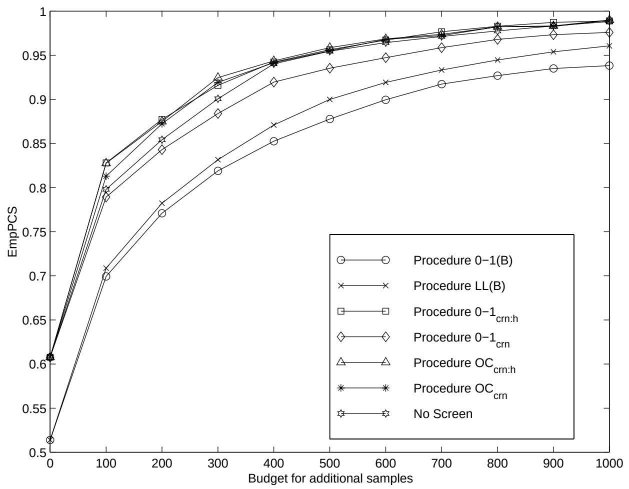

Figure 1: Empirical Fraction of Correct Selections (PCS) as a Function of the Total Budget b for Additional Samples for the MDM Example ($\sigma_i^2 = 1$; $\rho = 0.5$) in Section 5.2

Table 5: Approximation for the Predicted Probability of Correct Selection (PredBPCS) for Two Indifference-zone Procedures (with a Fixed δ^*, P^*) in the Experiment of Section 5.3

Procedure	δ^*	P^*	ANR	PredBPCS
$\mathcal{CY}$	0.1	0.80	42,890	0.9996
		0.90	65,345	0.9998
		0.95	92,080	0.9998
	0.5	0.80	1,665	0.9904
		0.90	2,565	0.9937
		0.95	3,635	0.9955
$\mathcal{NM}$	0.1	0.80	16,345	0.9990
		0.90	26,845	0.9994
		0.95	37,745	0.9996
	0.5	0.80	605	0.9745
		0.90	1,025	0.9846
		0.95	1,460	0.9891

observed, and therefore gives information about how likely it is that the best system will actually be selected as best. In this section, we examine the predictive probability that the correct system will be selected as best for the indifference zone and heuristic Bayesian procedures.

Specifically, the measure of evidence is PredBPCS, a Bonferroni-like approximation to the predictive probability that the best system will actually be selected as best. That is, based on a noninformative prior distribution and the first stage output, what is the probability (approximately) that the best system will actually be chosen as best, given that a certain number of replications will be run.

In this experiment, we the second stage budget for both indifference zone procedures for a specific batch of first stage output from the inventory selection problem for a variety of settings (δ^* = 0.01, 0.05, 0.1, 0.5; P^* = 0.8, 0.9, 0.95, 0.99; simulated 60 rather than 30 months). Some of the results are presented in Table 5. Settings of δ^* and P^* that led to a PredBPCS of roughly 0.99 were then noted for each procedure. For instance, Procedure $\mathcal{CY}$ requires 513 replications per system, or $b = 2565$ second stage replications total, when $\delta^* = 0.5$, $P^* = 0.9$ to achieve PredBPCS = 0.9937.

Procedure 0-$1_{\text{crn:h}}$ was then checked with a variety of second stage budgets b to attempt to achieve the same PredBPCS. Table 6 presents some of the results. It requires $b = 310$ total replications (Procedure $\mathcal{NM}$ required 1025) to achieve PredBPCS $= 0.9846$; and required $b = 780$ to achieve PredBPCS $= 0.9937$ (Procedure $\mathcal{CY}$ required 2565).

Table 6: Approximation for the Predicted Probability of Correct Aelection (PredBPCS) for Procedure 0-$1_{\text{crn:h}}$ in the Experiment of Section 5.3

Procedure	ANR	PredBPCS
0-$1_{\text{crn:h}}$	310	0.9846
	500	0.9904
	780	0.9937
	1,090	0.9955

6 DISCUSSION

The Bayesian procedures with CRN perform as well as or better than both the analogous Bayesian procedures that require independent replications, as well as both indifference-zone procedures with CRN, for most of the experiments in Section 5. Bayesian procedures with CRN can therefore be of practical value. This is not surprising, as CRN is known to sharpen comparisons (Law and Kelton 1991), and two of the systems in the inventory example could typically be screened due to their poorer performance. We believe that the relative value of screening can improve if there are more vastly inferior systems. Procedure 0-1_{crn} is somewhat inferior, empirically, to the other Bayesian procedures, due to approximations in its derivation, so the others are preferred.

There is an exception to the assessment that the Bayesian procedures outperform the indifference-zone procedures. Procedure $\mathcal{NM}$ outperforms the Bayesian procedures for the MDM experiment with a larger variance and smaller correlation. The strong performance of $\mathcal{NM}$ is not surprising, since it assumes sphericity, an assumption that is satisfied in our MDM experiments. We conject that further efficiencies can be obtained by developing a Bayesian procedure that also uses the sphericity assumption.

Even though the heuristic procedures check at most $2k$ of the $2^k - 1 = 31$ subsets $\mathcal{C}_2$ considered by the exhaustive Bayesian procedures, the heuristics perform slightly better. Further experiments indicate that the performance of the exhaustive procedures can be degraded if at least one covariance estimate is negative. If a negative covariance estimate is observed when positive correlation is justified by the structure of the problem, a somewhat larger first stage can be run in order to improve the correlation estimate, and therefore improve the performance of the procedures. Bayesian procedures that presume a sphericity condition might also reduce the chance of obtaining a negative correlation estimate, assuming that CRN induces a positive correlation.

A criticism of the Bayesian procedures is that screening incurs a CPU time cost that is not incurred by Procedures $\mathcal{CY}$ and $\mathcal{NM}$. In Section 5, Procedures 0-1_{crn} and $\mathcal{OC}_{\mathrm{crn}}$ took 5-6 times as long as the heuristics 0-$1_{\mathrm{crn:h}}$ and $\mathcal{OC}_{\mathrm{crn:h}}$ to pick a screening subset $\mathcal{C}_2$, which required an average of about 0.7 CPU seconds (implemented in Matlab, run on a PC). More generally, the heuristics run in $O(k^4)$ time, but Procedures $\mathcal{CY}$ and $\mathcal{NM}$ run in $O(k^2)$ time. For large k or small simulation runtimes, CPU time might be better spent running replications rather than selecting $\mathcal{C}_2$. For small k, or large b or simulation runtimes, the relative benefit of screening seems to improve. The development of more efficient heuristics is an area for further research.

7 CONCLUSIONS

Existing indifference-zone procedures that use CRN guarantee a minimum PCS, given the worst-case least-favorable configuration, but do not allow screening during the second stage. The Bayesian approach does not provide a PCS guarantee, but overcomes statistical conservativeness by allocating replications to improve the expected value of information gained from the second stage by screening. The use of screening entails a missing data analysis, but the difficulty of that analysis can be hidden from an end user with a well-designed software package. An added benefit of the Bayesian approach is that simulation practitioners can consider either the expected opportunity cost of a potentially incorrect selection, rather than the probability of correct selection.

Experiments indicate that the screening of the Bayesian procedures can provide a significant improvement in the ability to identify the best system. Procedure $\mathcal{OC}_{\mathrm{crn:h}}$ seems particularly effective. The Bayesian procedures require more CPU time than the indifference-zone procedures, however, so the indifference-zone approaches may be preferable if the number of systems is rather large, or the runtime of the replications is small. The sphericity assumption behind Procedure $\mathcal{NM}$ provides an improvement over Procedure $\mathcal{CY}$. These observations suggest a potential benefit to developing a new procedure that draws upon the advantages of Bayesian screening and the sphericity assumption.

8 APPENDIX

Let $\mathbf{x}_j$ be the vector of output of the j-th first-stage replication, so the first-stage sample statistics are:

$$\hat{\boldsymbol{\mu}} = \sum_{j=1}^{r_1} \mathbf{x}_j / r_1 \tag{3}$$

$$\hat{\boldsymbol{\Sigma}} = \bar{\mathbf{S}}/r_1 = \sum_{j=1}^{r_1} (\mathbf{x}_j - \hat{\boldsymbol{\mu}})^t (\mathbf{x}_j - \hat{\boldsymbol{\mu}})/r_1. \tag{4}$$

Let $\mathbf{y}_j$ be the vector of output of the j-th second-stage replication. Suppose that a subset $\mathcal{C}_2$ of systems is simulated r_2 times during the second stage, and that $\mathcal{C}_1 = \{1, \ldots, k\} \backslash \mathcal{C}_2$ is the subset of systems simulated during the first stage alone. We use subscripts here to denote appropriate subvectors and submatrices for systems in $\mathcal{C}_2$ or $\mathcal{C}_1$, such as $\hat{\boldsymbol{\mu}} = (\hat{\boldsymbol{\mu}}_{\mathcal{C}_2}\, \hat{\boldsymbol{\mu}}_{\mathcal{C}_1})$ for subvectors of the estimate of the mean, or $\bar{\mathbf{S}}_{\mathcal{C}_2\mathcal{C}_2}$ for the sample variance of systems in $\mathcal{C}_2$. Two hats or bars indicate an estimate based on two stages of output, such as

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

$\mathbf{Cov}\left[\hat{\hat{\boldsymbol{\mu}}}, \hat{\hat{\boldsymbol{\mu}}}\right]$ in Equation 2. Anderson (1957) shows that the MLE for Σ given both stages of output is

$$\begin{aligned}\hat{\hat{\boldsymbol{\Sigma}}}_{\mathcal{C}_2\mathcal{C}_2} &= \bar{\bar{\mathbf{S}}}_{\mathcal{C}_2\mathcal{C}_2}/(r_1+r_2) \\ \hat{\hat{\boldsymbol{\Sigma}}}_{\mathcal{C}_1\mathcal{C}_2} &= \left[\hat{\hat{\boldsymbol{\Sigma}}}_{\mathcal{C}_2\mathcal{C}_1}\right]^t = \hat{\boldsymbol{\Sigma}}_{\mathcal{C}_1\mathcal{C}_2}\hat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{C}_2\mathcal{C}_2}\hat{\hat{\boldsymbol{\Sigma}}}_{\mathcal{C}_2\mathcal{C}_2} \\ \hat{\hat{\boldsymbol{\Sigma}}}_{\mathcal{C}_1\mathcal{C}_1} &= \hat{\boldsymbol{\Sigma}}_{\mathcal{C}_1\mathcal{C}_1} - \hat{\boldsymbol{\Sigma}}_{\mathcal{C}_1\mathcal{C}_2}\hat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{C}_2\mathcal{C}_2}\hat{\boldsymbol{\Sigma}}_{\mathcal{C}_2\mathcal{C}_1} \\ &+ \hat{\boldsymbol{\Sigma}}_{\mathcal{C}_1\mathcal{C}_2}\hat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{C}_2\mathcal{C}_2}\hat{\hat{\boldsymbol{\Sigma}}}_{\mathcal{C}_2\mathcal{C}_2}\hat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{C}_2\mathcal{C}_2}\hat{\boldsymbol{\Sigma}}_{\mathcal{C}_2\mathcal{C}_1}\end{aligned} \tag{5}$$

and that

$$\begin{aligned}\mathrm{Cov}\left[\hat{\hat{\boldsymbol{\mu}}}_{\mathcal{C}_2}, \hat{\hat{\boldsymbol{\mu}}}_{\mathcal{C}_2}\right] &= \hat{\hat{\boldsymbol{\Sigma}}}_{\mathcal{C}_2\mathcal{C}_2}/(r_1+r_2) \\ \mathrm{Cov}\left[\hat{\hat{\boldsymbol{\mu}}}_{\mathcal{C}_1}, \hat{\hat{\boldsymbol{\mu}}}_{\mathcal{C}_2}\right] &= \mathrm{Cov}\left[\hat{\hat{\boldsymbol{\mu}}}_{\mathcal{C}_2}, \hat{\hat{\boldsymbol{\mu}}}_{\mathcal{C}_1}\right]^t = \hat{\hat{\boldsymbol{\Sigma}}}_{\mathcal{C}_1\mathcal{C}_2}/(r_1+r_2) \\ \mathrm{Cov}\left[\hat{\hat{\boldsymbol{\mu}}}_{\mathcal{C}_1}, \hat{\hat{\boldsymbol{\mu}}}_{\mathcal{C}_1}\right] &= \left[\hat{\boldsymbol{\Sigma}}_{\mathcal{C}_1\mathcal{C}_1} - \hat{\boldsymbol{\Sigma}}_{\mathcal{C}_1\mathcal{C}_2}\hat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{C}_2\mathcal{C}_2}\hat{\boldsymbol{\Sigma}}_{\mathcal{C}_2\mathcal{C}_1}\right]/r_1 \\ &+ \bar{\mathbf{S}}_{\mathcal{C}_1\mathcal{C}_2}\bar{\mathbf{S}}^{-1}_{\mathcal{C}_2\mathcal{C}_2}\mathrm{Cov}\left[\hat{\hat{\boldsymbol{\mu}}}_{\mathcal{C}_2}, \hat{\hat{\boldsymbol{\mu}}}_{\mathcal{C}_2}\right]\bar{\mathbf{S}}^{-1}_{\mathcal{C}_2\mathcal{C}_2}\bar{\mathbf{S}}_{\mathcal{C}_2\mathcal{C}_1}\end{aligned} \tag{6}$$

Degrees of freedom in Equation 2. Suppose $i \in \mathcal{C}_2$ and $j \in \mathcal{C}_1$. Let $a^t = (1/(r_1 + r_2), -2/(r_1+r_2), 1/r_2)$, estimate the correlation as $\hat{\hat{\rho}} = \hat{\sigma}_{i,j}(\hat{\sigma}_{i,i}\hat{\sigma}_{j,j})^{-1/2}(\hat{\hat{\sigma}}_{i,i}\hat{\sigma}^{-1}_{i,i})^{1/2}(\hat{\sigma}_{j,j}\hat{\hat{\sigma}}^{-1}_{j,j})^{1/2}$, and let

$$\delta^2(a) = \frac{a_2^2\hat{\hat{\sigma}}_{i,i}\hat{\hat{\sigma}}_{j,j} + 4a_2a_3\hat{\hat{\sigma}}_{i,j}\hat{\hat{\sigma}}_{j,j} + 2a_3^2\hat{\hat{\sigma}}_{j,j}(1-\hat{\hat{\rho}})^2}{2(a_1\hat{\hat{\sigma}}_{i,i} + a_2\hat{\hat{\sigma}}_{i,j} + a_3\hat{\hat{\sigma}}_{j,j})^2}.$$

Little (1976) suggests the missing-data approximation

$$\nu = \frac{r_1+r_2-1}{1 + (\frac{1}{r_1} - 1/(r_1+r_2))(1-\hat{\hat{\rho}}^2)\delta^2(a)}.$$

for the degrees of freedom of the difference in means of systems i and j, given both stages of output.

ACKNOWLEDGMENTS

The authors are pleased to acknowledge the financial support of the Engineering Research Center for Reconfigurable Machining Systems (NSF grant # EEC-9529125), and the assistance from the members of the ERC (both industry and academia).

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.

Chen, C.-H. 1996. A lower bound for the correct subset-selection probability and its application to discrete event simulations. *IEEE Transactions on Automatic Control* 41(8): 1227–1231.

Chen, H.-C., C.-H. Chen, J. Lin, and E. Yücesan. 1999. An asymptotic allocation for simultaneous simulation experiments. In *Proceedings of the Winter Simulation Conference*, ed. P. A. Farrington, H. B. Nembhard, D. Sturrock, and G. Evans, 359–366. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.

Chick, S. E., and K. Inoue. 1999. A decision-theoretic approach to screening and selection with common random numbers. In *Proceedings of the Winter Simulation Conference*, ed. P. A. Farrington, H. B. Nembhard, D. Sturrock, and G. Evans, 603–610. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.

Chick, S. E., and K. Inoue. 2000a. New procedures for identifying the best simulated system using common random numbers. *in resubmission*.

Chick, S. E., and K. Inoue. 2000b. New two-stage and sequential procedures for selecting the best simulated system. *Operations Research*. to appear.

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

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

Inoue, K. 2000. *Decision-theoretic comparison of alternative system configurations using stochastic simulation*. Ph.

D. thesis, The University of Michigan, Ann Arbor, MI. Dept. of Industrial and Operations Engineering.

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

Inoue, K., S. E. Chick, and C.-H. Chen. 1999. An empirical evaluation of several methods to select the best system. *ACM Transactions on Modeling and Computer Simulation* 9(4): *in press*.

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.

Little, R. J. A. 1976. Inference about means from incomplete multivariate data. *Biometrika* 63(3): 593–604.

Matejcik, F. J., and B. L. Nelson. 1995. Two-stage multiple comparisons with the best for computer simulation. *Operations Research* 43: 633–640.

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. 1999. Simple procedures for selecting the best simulated system when the number of alternatives is large. Technical report, Northwestern University, Department of Industrial Engineering and Management Science, Evanston, IL.

Rinott, Y. 1978. On two-stage selection procedures and related probability-inequalities. *Communications in Statistics* A7: 799–811.

AUTHOR BIOGRAPHIES

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, epidemiology models, decision analysis, reliability, and computational methods in statistics. His research is motivated by projects in manufacturing and health care. His work experience includes several years of materials handling system design for the automotive industry using simulation analysis. His email and web addresses are `<sechick@engin.umich.edu>` and `<http://www.engin.umich.edu/~sechick/>`.

KOICHIRO INOUE completed his Ph.D. in Industrial and Operations Engineering at the University of Michigan, Ann Arbor in August 2000. He received a BS degree in Industrial and Systems Engineering from Aoyama Gakuin University, Tokyo in 1992, he received an MS degree in Industrial and Operations Engineering at the University of Michigan, Ann Arbor in 1995. His email address is `<koichiro@engin.umich.edu>`.