Proceedings of the 2005 Winter Simulation Conference
M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, eds.

PERFORMANCE OF VARIANCE UPDATING RANKING AND SELECTION PROCEDURES

Gwendolyn J. Malone

DRS Technologies
Palm Bay, Florida 32905

Seong-Hee Kim
David Goldsman
Demet Batur

School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0205, U.S.A.

ABSTRACT

Kim and Nelson (2005) developed two indifference-zone procedures for steady-state simulation where the goal is to find the system with the largest or smallest expected steady-state performance measure. One of the procedures, called $\mathcal{KN}$++, updates a variance estimate as more observations become available and is proven to be asymptotically valid when there is no dependence across systems (for example, there is no use of common random numbers). Their procedure exhibits significant improvement over other existing procedures for use in steady-state simulation. In this paper, we first present a modification of $\mathcal{KN}$++ that is asymptotically valid with the use of common random numbers. Then, we study how well $\mathcal{KN}$++ works when data within a system are independent and identically distributed, but data between systems may be positively correlated. Specific applications include the finding-the-best problem when (i) the data are normal, and (ii) the data are Bernoulli.

1 INTRODUCTION

In simulation, there are at least four types of "ranking-and-selection" comparison problems that are typically encountered: finding the best system from a set of competitors, comparison with a standard, Bernoulli selection, and multinomial selection (Goldsman and Nelson 1998).

For the finding-the-best problem, we wish to find the scenario with the largest or smallest expected performance measure among a number of simulated alternatives. This problem has received significant attention from the simulation community, and procedures developed for this goal are sometimes divided into two categories depending on whether a terminating or steady-state simulation is applicable given the nature of the process being simulated.

In terminating simulations, we usually have well-defined initial conditions and stopping times for each replication. An example of such a system is a branch office of a bank that opens every morning at 9:00 a.m. completely empty and empties out as soon as the last customer who entered before the closing time of 4:00 p.m. has finished being served. With terminating simulations, we often take within-replication averages as basic observations and these within-replication averages are typically independent and identically distributed (i.i.d.) normal random variables. A number of statistically valid procedures have been developed for the goal of finding the scenario with the largest or smallest expected performance measure among a number of simulated systems, assuming that basic observations are i.i.d. normal. Recent literature includes Nelson et al. (2001), Boesel et al. (2003), and Kim and Nelson (2001). In particular, procedure $\mathcal{KN}$ (Kim and Nelson 2001) has been shown to be highly efficient compared to other procedures in terms of the number of observations required to reach a decision.

On the other hand, for steady-state simulations, there are no clear start-up or terminating conditions. An example of such a situation would be a model of a manufacturing facility which operates 24 hours a day and we are interested in the long-run performance of a given statistic. Any procedure developed assuming i.i.d. normality can be used 'as is' for steady-state simulation if one is willing to make multiple replications or use batch means. However, Kim and Nelson (2005) argue that for several reasons, the multiple-replication design approach or simply taking batch means from a single-replication design are not desirable for steady-state simulation; but without the multiple replications or batching, individual raw observations are usually correlated serially (e.g., the wait times of a series of customers) and therefore problematic. Kim and Nelson (2005) present two asymptotically valid procedures — $\mathcal{KN}$+ and $\mathcal{KN}$++ — that take stationary (and dependent) data from the single replication as basic observations. More precisely, suppose the output data from the single replication satisfy a Functional Central Limit

Theorem (FCLT, Billingsley 1968) and the data are independent across systems—e.g., there is no use of common random numbers (CRN). Then the procedures find the best or near-best system with a probability of correct selection (PCS) of at least a pre-specified constant, as the "indifference-zone" (IZ) parameter—the minimum "practical" difference between systems worth detecting—goes to zero. Procedure $\mathcal{KN}$++ shows significant improvement over other existing procedures for steady-state simulation.

In this paper, we first modify $\mathcal{KN}$++ so that it is still asymptotically valid when CRN is employed. Then, we study the small-sample behavior of $\mathcal{KN}$++ for specific i.i.d. data types and compare its performance with existing specialized procedures for the data type with or without CRN. Specifically, we test the performance of $\mathcal{KN}$++ on the finding-the-best problem when data are i.i.d. normal. Of course, i.i.d. normal data are a special case of stationary data, and so they satisfy a FCLT. Therefore, $\mathcal{KN}$++ still guarantees asymptotic validity, but the exact validity for a small sample has not yet been established. We will compare the performance of $\mathcal{KN}$++ with $\mathcal{KN}$, which is statistically valid for small samples.

We also study the performance of $\mathcal{KN}$++ for the Bernoulli selection problem. In the Bernoulli problem, the goal is to find the system with the largest probability of 'success'. This problem can be considered a finding-the-best problem where data are 0-1 since the success probability is the same as the expectation for Bernoulli distributed data. Bernoulli data are also satisfy the FCLT. Therefore, the procedure is still asymptotically valid, but the exact validity (for small samples) is not guaranteed.

The remainder of the paper is organized as follows. In Section 2, we define notation and give some background. Section 3 compares the performance of $\mathcal{KN}$ and $\mathcal{KN}$++ for i.i.d. normal data. Section 4 compares the performance of $\mathcal{KN}$++ on Bernoulli distributed data with that of a number of existing procedures. We present conclusions in Section 5.

2 BACKGROUND

This section provides the full description of $\mathcal{KN}$++ assuming that data are i.i.d.—with and without CRN.

We assume that we want to find the system with the largest expected value. Let X_{ij} be the jth output observation from system i. For steady-state simulation with a single replication, $X_{i1}, X_{i2}, \ldots$ are stationary (and dependent) data collected after a warm-up period. For a terminating simulation or the Bernoulli selection problem, we assume that $X_{i1}, X_{i2}, \ldots$ are i.i.d. data. System i has the expected (possibly, long-run) value $\mu_i = \mathrm{E}[X_{ij}]$. In the Bernoulli selection problem, $\mu_i = p_i \equiv \Pr\{X_{ij} = 1\}$. For steady-state data, we define the asymptotic variance constant $v^2 \equiv \lim_{r\to\infty} r\mathrm{Var}(\bar{X}(r))$, where $\bar{X}(r)$ is a sample average of r observations. In the special case that the data are i.i.d., the variance constant is simply $\sigma_i^2 = \mathrm{Var}[X_{ij}]$; and the usual sample variance estimator $S_i^2(r)$ based on r observations is used to estimate σ_i^2. For general steady-state simulation data, more-sophisticated variance estimators will be necessary.

Without loss of generality, we can assume that $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_k$. Then $\mathcal{KN}$++ guarantees

$$\Pr\{\text{Select system } k\} \geq 1 - \alpha \text{ provided } \mu_k \geq \mu_{k-1} + \delta.$$

$\mathcal{KN}$++ is distinguished from other procedures in the sense that it contains a recalculation of the variance estimate at various update points where a certain "continuation region" for sequential sampling is redefined. First, we need to define a batching sequence (m_r, b_r), where b_r is the number of batches and m_r is the batch size at stage r of sampling. For the asymptotic validity of $\mathcal{KN}$++ to hold, a variance estimator of the differences between systems i and ℓ—denoted by $m_r V_{il}^2(r)$—must have the strong consistency property, i.e., $m_r V_{il}^2(r) \to v_{i\ell}^2$ with probability one as $m_r \to \infty$ and $b_r \to \infty$, where $v_{i\ell}^2$ is the asymptotic variance constant of the differences between systems i and ℓ. For additional discussion on variance estimators with the strong consistency property, see Damerdji (1994, 1995), Damerdji and Goldsman (1995) and Chien, Goldsman, and Melamed (1997). Kim and Nelson (2005) study the performance of $\mathcal{KN}$++ using various variance estimators along with several different batching sequences that ensure the strong consistency of the chosen variance estimators. Although $\mathcal{KN}$++ shows significant improvement over $\mathcal{KN}$+ in many respects, there is, unfortunately, no convenient way to update the variance estimates without saving raw data for general batching sequences and general variance estimators. Therefore, updating variances might require saving all the past raw data—which can be a problem due to limited memory space when the number of systems is large and when the procedure does not stop for a great deal of time.

Since we are only interested in i.i.d. data in the current paper, we will not deal with complicated variance estimators or batching sequences. The usual sample variance S^2 is strongly consistent without batching (i.e., if the batch size is set to one) as long as the number of observations r goes to infinity. Also, one can easily update $S^2(r)$ *on the fly* without saving data (see Damerdji, Handerson, and Glynn 1997). The upshot is that for i.i.d. data, procedure $\mathcal{KN}$++ can be given as follows:

$\mathcal{KN}$++ Procedure

Setup: Specify the desired confidence level $1/k < 1-\alpha < 1$, IZ parameter $\delta > 0$, and first-stage sample size $n_0 \geq 2$. Set $d = n_0 - 1$. Calculate the constants η and c as described below.

Initialization: Let $I = \{1, 2, \ldots, k\}$ be the set of systems still in contention, and let $h^2 = 2c\eta(n_0 - 1)$.

Obtain n_0 observations X_{ij}, $j = 1, 2, \ldots, n_0$, from each system $i = 1, 2, \ldots, k$.

For all $i \neq \ell$, compute the estimator $S^2_{i\ell}(n_0)$, the sample variance of the difference between systems i and ℓ, based on n_0 observations.

Update: If we have reached the next update point, then for all $i \neq \ell$ $(i, \ell \in I)$, compute η with $d = r - 1$, the constant $h^2 = 2c\eta(r-1)$, and the estimator $S^2_{i\ell}(r)$, the sample variance of the difference between systems i and ℓ based on r observations.

Go to **Screening**.

Screening: Set $I^{\text{old}} = I$. Let

$$I = \left\{ i : i \in I^{\text{old}} \text{ and } \bar{X}_i(r) \geq \bar{X}_\ell(r) - W_{i\ell}(r), \forall \ell \in I^{\text{old}}, \ell \neq i \right\},$$

where

$$W_{i\ell}(r) = \max\left\{0, \frac{\delta}{2cr}\left(\frac{h^2 S^2_{i\ell}(r)}{\delta^2} - r\right)\right\}.$$

Stopping Rule: If $|I| = 1$, then stop and select the system whose index is in I as the best.

Otherwise, take one additional observation $X_{i,r+1}$ from each system $i \in I$, set $r = r + 1$, and go to **Update**.

Constants: The constant c may be any nonnegative integer. The constant η is the solution to the equation

$$\sum_{\ell=1}^{c} (-1)^{\ell+1} \left(1 - \frac{1}{2}\mathcal{I}(\ell = c)\right)\left(1 + \frac{2\eta(2c-\ell)\ell}{c}\right)^{-\frac{d}{2}} = \beta,$$

where $\mathcal{I}$ is the indicator function, $\beta = 1-(1-\alpha)^{1/(k-1)}$ if the systems are simulated independently, and $\beta = \alpha/(k-1)$ if CRN is employed. The constant $c = 1$ is suggested for use if the experimenter has no prior knowledge about the experiment.

The proof of the asymptotic validity of procedure $\mathcal{KN}$++ for the stationary case without CRN is given in Kim and Nelson (2005); the analogous proof for CRN is given in Malone (2005).

3 NORMAL DATA

We now apply $\mathcal{KN}$++ to i.i.d. normal data and compare its performance with that of $\mathcal{KN}$.

3.1 Experimental Setup for I.I.D. Normal Data

We chose the first-stage sample size to be $n_0 = 10$. The number of systems in each experiment varies over $k = 2, 5, 10, 25, 100$.

The IZ parameter, δ, was set to $\delta = \sigma_1/\sqrt{n_0}$, where σ_1^2 is the variance of an observation from the best system. Thus, δ is the standard deviation of the first-stage sample mean of the best system. A sample variance $S^2(r)$ will be updated every 10 observations.

Two configurations of the true means were used: The slippage configuration (SC) and monotonically decreasing means (MDM) configuration. In the SC, μ_1 was set to δ, while $\mu_2 = \mu_3 = \cdots = \mu_k = 0$. This is a difficult configuration for procedures that try to eliminate systems because all of the inferior systems are close to the best. To investigate the effectiveness of the procedures in eliminating non-competitive systems, we also carried out experimentation with the MDM, where the means of all systems were spaced evenly apart: $\mu_i = \mu_1 - \delta(i-1)$, for $i = 2, 3, \ldots, k$.

For each configuration of the means, we examined the effects of equal and unequal variances. In the equal-variance configuration, σ_i was set to 1. In the unequal-variance configuration, the variance of the best system was set either higher or lower than the variances of the other systems. In the MDM, experiments were run with the variances either directly proportional or inversely proportional to the mean of each system. Specifically, we took $\sigma_i^2 = |\mu_i - \delta| + 1$ to examine the effects of increasing variance (INC) as the mean increases, and $\sigma_i^2 = 1/(|\mu_i - \delta| + 1)$ to examine the effects of decreasing variances (DEC) as the mean increases. In addition, some experiments were run with means in the SC, but with the variances of all systems either monotonically decreasing or monotonically increasing (similar to the MDM configuration).

When CRN was employed we assumed that the correlation between all pairs of systems was ρ, and values of $\rho = 0.02, 0.25, 0.5, 0.75$ were tested.

Thus, we tested a total of six configurations: SC with equal variances, MDM with equal variances, MDM with increasing variances, MDM with decreasing variances, SC with increasing variances and SC with decreasing variances. For each configuration, 1000 macroreplications (complete repetitions) of the entire experiment were performed. In all experiments, the nominal probability of correct selection was set at $1 - \alpha = 0.95$. To compare the performance of the different procedures

Table 1: Monte Carlo $\widehat{\text{PCS}}$ results for $k = 5$ in the SC when data are i.i.d. normal.

	INC Var		DEC Var	
ρ	$\mathcal{KN}$	$\mathcal{KN}$++	$\mathcal{KN}$	$\mathcal{KN}$++
0	0.96	0.949	0.968	0.958
0.02	0.959	0.947	0.959	0.953
0.25	0.958	0.946	0.953	0.961
0.50	0.960	0.949	0.964	0.963
0.75	0.965	0.952	0.975	0.967

we recorded the total number of basic observations required by each procedure on each replication and reported the sample averages over all replications along with each procedure's attained PCS.

3.2 Results with I.I.D. Normal Data

Illustrative results are given in Tables 1–4. Each table displays $\widehat{\text{PCS}}$, the Monte Carlo estimate of the probability of a correct selection, or $\widehat{E(N)}$, the Monte Carlo estimate of the expected total number of observations.

Table 2: Monte Carlo $\widehat{\text{PCS}}$ results for $k = 25$ in the SC when data are i.i.d. normal.

	INC Var		DEC Var	
ρ	$\mathcal{KN}$	$\mathcal{KN}$++	$\mathcal{KN}$	$\mathcal{KN}$++
0	0.956	0.952	0.981	0.979
0.02	0.966	0.938	0.982	0.980
0.25	0.965	0.946	0.982	0.980
0.50	0.965	0.948	0.991	0.982
0.75	0.965	0.945	0.984	0.982

Tables 1 and 2 show the estimated PCS in the SC configuration when the numbers of systems are 5 and 25, respectively. The values of the PCS shown in Tables 1 and 2 are often lower for $\mathcal{KN}$++ than for $\mathcal{KN}$, particularly in the SC increasing variance cases. Also, there are some cases for which $\mathcal{KN}$++ does not quite appear to satisfy the PCS requirement—possibly because $\mathcal{KN}$++ is only asymptotically valid. In such cases, the discrepancies are small, e.g., the smallest estimated PCS we observed was 0.938. In addition, it turns out that for larger k with the MDM configuration, $\mathcal{KN}$++ easily satisfies the PCS requirement.

Tables 3 and 4 show the average total number of observations for $k = 25$ when SC and MDM configurations are, respectively, employed. We see that if there is no correlation between the simulated systems, then procedure $\mathcal{KN}$++ always requires significantly fewer observations than $\mathcal{KN}$ while still attaining and often exceeding the desired PCS. For example, when $k = 25$ in the SC with increasing variance, the expected number of observations needed in order to reach a solution is 11,014 if $\mathcal{KN}$ is employed, but only 5,906 when $\mathcal{KN}$++ is used. The results also show that the savings in observations realized by using $\mathcal{KN}$++, while always significant, are even more pronounced when the selection problem is "difficult."

Table 3: Monte Carlo $\widehat{E(N)}$ results for $k = 25$ in the SC when data are i.i.d. normal.

	INC Var		DEC Var	
ρ	$\mathcal{KN}$	$\mathcal{KN}$++	$\mathcal{KN}$	$\mathcal{KN}$++
0	11014.1	5905.5	1140.8	738.4
0.02	10895.4	5802.0	1125.6	726.4
0.25	8853.8	4499.1	896.7	598.2
0.50	6634.9	3034.7	647.4	458.3
0.75	4158.2	1574.0	414.3	328.9

Table 4: Monte Carlo $\widehat{E(N)}$ results for $k = 25$ in the MDM configuration when data are i.i.d. normal.

	INC Var		DEC Var	
ρ	$\mathcal{KN}$	$\mathcal{KN}$++	$\mathcal{KN}$	$\mathcal{KN}$++
0	1443.5	912.2	582.6	436.9
0.02	1443.5	896.0	595.3	433.4
0.25	1151.5	719.9	495.6	382.9
0.50	838.5	535.8	399.9	331.0
0.75	517.4	355.0	307.7	282.3

We see that when CRN is used, both procedures are able to reach a solution much sooner than if the systems are simulated independently. Further, $\mathcal{KN}$++ uniformly requires fewer observations than does $\mathcal{KN}$ over all values of ρ tested.

Thus, whenever the nature of the experiment permits, CRN should be considered. Moreover, we find that $\mathcal{KN}$++ is superior to $\mathcal{KN}$ in terms of minimizing the total number of observations over all scenarios tested. So although $\mathcal{KN}$++ for i.i.d. normal data does not have exact statistical validity, our results show the procedure is capable of efficiently selecting a winner at the cost of only a small loss in the PCS.

4 BERNOULLI DATA

We have seen that the idea of updating variance estimates has the potential to improve the efficiency of existing procedures for i.i.d. normal data. We now move to the selection problem where basic observations are Bernoulli distributed, and where we wish to select that Bernoulli alternative having the largest success parameter. Some classical IZ selection procedures designed specifically for Bernoulli data include Sobel and Huyett (1957), Bechhofer and Kulkarni (1982), Bechhofer, Kiefer, and Sobel (1968), and Paulson (1993). Can we do better with our asymptotic procedure, originally designed for normal data?

The procedure due to Paulson (1993), denoted by $\mathcal{B}_{\rm P}$, is similar to our $\mathcal{KN}$++ in the sense that it is fully sequential and eliminates clearly inferior systems at each stage of sampling. However, it uses another type of IZ known as the odds ratio formulation, i.e., the ratio of the odds of success for the best system to the odds of success for the second-best system. Thus, we wish to select the best system whenever $\frac{p_k/(1-p_k)}{p_{k-1}/(1-p_{k-1})} \geq \theta$, where p_k and p_{k-1} are the probabilities of success for the best and second-best systems respectively, and θ is the user-specified odds ratio.

Since $\mathcal{B}_{\rm P}$ usually requires the fewest observations to satisfy a common PCS constraint among these IZ procedures under study, it will be the procedure used for performance comparisons.

$\mathcal{B}_{\rm P}$ Procedure

Setup: For the given k, specify the odds ratio θ and desired confidence level $1-\alpha$.

Sampling: At the mth stage of experimentation ($m \geq 1$), take a random Bernoulli observation, X_{im}, from each of the systems $i \in R_m$, where R_m is the set of all systems that have not yet been eliminated from contention.

Stopping rule: Let $Y_{im} = \sum_{j=1}^{m} X_{ij}$ $(i = 1, 2, \ldots, k)$ and let e_i $(i = 1, 2, \ldots, k$, with e_i initialized to $\infty)$ denote the stage at which system i was eliminated. Define $n_{im} = \min(m, e_i)$ as the number of observations taken from system i through the mth stage of experimentation. For each system still in contention, let

$$g_i(m) = \sum_{j=1}^{k} \theta^{Y_{j,n_{jm}} - Y_{i,n_{jm}}},$$

where $j \neq i$, $j = 1, 2, \ldots, k$, and $i \in R_m$. After the mth stage of experimentation, eliminate from further consideration and sampling any system still in contention ($i \in R_m$) for which

$$g_i(m) > \frac{k-1}{\alpha},$$

and set $e_i = m$ for such i.

Stop at the first value of m for which only one system remains in contention.

Terminal decision rule: Select the one remaining system as the best.

In order to use $\mathcal{KN}$++ for Bernoulli selection, a minor modification is required. Due to the nature of the underlying Bernoulli random variables in this problem, it is possible that the first n_0 observations will all be either zero or one; and then we will obtain $S^2 = 0$, which causes the procedure to stop immediately and possibly leads to an incorrect decision. To avoid this problem, we need to modify the **Initialization** step of $\mathcal{KN}$++ as follows:

Initialization: Let $I = \{1, 2, \ldots, k\}$ be the set of systems still in contention, and let $h^2 = 2c\eta(n_0 - 1)$.

Obtain n_0 observations X_{ij}, $j = 1, 2, \ldots, n_0$, from each system $i = 1, 2, \ldots, k$.

For all $i \neq \ell$, compute the estimator $S^2_{i\ell}(n_0)$, the sample variance of the difference between systems i and ℓ, based on n_0 observations. *If all variance estimates are > 0, set the observation counter $r = n_0$ and continue to* **Screening**. *Otherwise, take 10 more observations and recalculate the variance estimates. Continue to do so until nonzero estimates are obtained.*

We now describe the experimental setup.

4.1 Experimental Setup for Bernoulli Data

The goal here is to test the performance of $\mathcal{KN}$++ with CRN. Of course, in practice, the experimenter will not be able to induce a specified amount of correlation, but for Monte Carlo testing purposes we consider "toy" problems for which we can induce the same level of correlation between all the systems and then test the performance implications. In any case, we generate correlated multivariate Bernoulli data via the NORTA (NORmal To Anything) method proposed by Cario and Nelson (1997).

We test the procedure under two primary configurations of the underlying success probabilities. The first is an "unfavorable" slippage configuration (SC) of the competing systems' p_i-values. For purposes of procedure evaluation we assume that k is the best system and that under the SC, $p_k = p$ and $p_{k-1} = p_{k-2} = \cdots =$

$p_1 = p - \delta$. The second configuration is a "more favorable" configuration (MFC) of the p_i-values in which $p_k = p$, $p_{k-1} = p - \delta$, and $p_{k-2} = \cdots = p_1 = p - 2\delta$. In the MFC, all systems are still fairly competitive with the best system, but the main competition is between the best and second-best systems. This configuration will represent a situation where the experimenter feels that all systems are viable options for the best (highly competitive with each other) but is potentially more likely to occur in practice than the very restrictive SC.

To compare our results to those of Paulson's procedure, his odds ratio IZ is easily converted into a difference (given p_k) using

$$\theta = \frac{p_k/(1-p_k)}{(p_k-\delta)/(1-p_k-\delta)} \ .$$

In order to determine which value of n_0 should be used, we tested performance for a selection of n_0 values and examined the resulting PCS estimates as well as how our estimates of $E(N)$ compared to those of Paulson's procedure $\mathcal{B}_{\rm P}$. We ran comparisons for various probability configurations, $k = 2, 3, 4, 5$, and $\rho = 0.05$, 0.1, 0.15, 0.2, 0.25, 0.5. Table 5 gives an example of the results for $k = 3$, $\delta = 0.06$, and $1 - \alpha = 0.95$, under a SC in which the best $p_k = 0.85$.

Table 5: Monte Carlo evaluation of n_0 values for $k = 3$, $\delta = 0.06$, and $1 - \alpha = 0.95$, for i.i.d. Bernoulli data in the SC with $p_k = 0.85$.

n_0	ρ	$\widehat{\rm PCS}$	$\widehat{E(N)}$	$\mathcal{B}_{\rm P}$ $\widehat{E(N)}$
10	0.05	0.949	475	501
10	0.10	0.948	451	
10	0.15	0.944	426	
10	0.20	0.943	398	
10	0.25	0.938	374	
10	0.50	0.913	245	
20	0.05	0.948	476	
20	0.10	0.947	450	
20	0.15	0.943	425	
20	0.20	0.941	399	
20	0.25	0.939	373	
20	0.50	0.914	245	
30	0.05	0.950	478	
30	0.10	0.948	451	
30	0.15	0.947	427	
30	0.20	0.944	400	
30	0.25	0.942	374	
30	0.50	0.916	248	

These results indicate that a good choice of n_0 is 10. This value almost always results in the lowest number of total observations. However, some of the PCS values are lower than the desired 0.95, particularly for larger values of ρ. This undercoverage is probably related to the fact that our procedure is only asymptotically guaranteed to meet the PCS constraint; and furthermore, we are dealing with Bernoulli data here instead of normal data. In addition, as ρ increases, the X_{ij} values become more and more alike (particularly when the true differences between the systems are small), making it harder to obtain good variance estimates.

The PCS values increase as n_0 increases since we have more observations upon which to base our initial variance estimates. However, the gain in PCS with increased n_0 is not very significant since the procedure updates the variance estimate anyway every 10 observations (in fact, our PCS values never drop below 0.90 and are often very close to the nominal 0.95). Thus, the choice $n_0 = 10$ is reasonable from a PCS perspective and typically yields low $E(N)$ values. If the experimenter wants a more certain PCS guarantee, e.g., when δ is small, a larger value of n_0 (30, for instance) can always be used.

Each experiment was run for 100,000 replications and results were compared to Paulson's under the same k, p_k, and IZ for $1 - \alpha = 0.95$. We test values of $\delta = 0.03$, 0.06, and 0.09, under the SC and MFC for all values of $k \leq 5$ and $\rho = 0$ (no CRN), 0.05, 0.1, 0.15, 0.2, 0.25, and 0.5. We consider two values of $p_k = 0.35$ and 0.85. All results are for $n_0 = 10$.

4.2 Experimental Results for Bernoulli Data

A small portion of the results that illustrate some of the key findings is shown in Tables 6 and 7. As before, $\widehat{\rm PCS}$ and $\widehat{E(N)}$ denote the Monte Carlo estimates of the PCS and expected value of total observations, respectively.

When $k = 2$, procedure $\mathcal{KN}$++ beats Paulson's procedure over all configurations as long as $\rho \geq 0.10$. The savings when more correlation can be induced can be significant. For instance, with $k = 2$, $p_k = 0.85$ and $\delta = 0.06$, even a ρ value of 0.15 produces considerable savings, with an estimated number of observations of 201 versus 250 for Paulson's. In the case when $k = 3$, $\rho \geq 0.15$ is sufficient to produce $E(N)$ values that are less than those of Paulson's for all configurations. When $k = 4$, a ρ value of 0.20 provides results better than Paulson's for all configurations, though a value of $\rho \geq 0.15$ suffices when the selection problem is difficult (say, in the SC). For $k = 5$, we require a $\rho \geq 0.25$ to beat Paulson's in the MFC case, though $\rho = 0.15$ is all that is necessary under the SC.

We also performed a few tests for a "more-difficult" problem, when $\delta = 0.01$. Table 8 indicates that a value of $\rho \geq 0.1$ is sufficient to provide savings in $E(N)$ when $k = 2$ or $k = 3$, at least in the SC.

Table 6: Monte Carlo results for $k = 2$ and $1-\alpha = 0.95$, for i.i.d. Bernoulli data in the SC with $p_k = 0.85$.

δ	ρ	$\widehat{\text{PCS}}$	$\widehat{E(N)}$	$\mathcal{B}_\text{P}$ $\widehat{\text{PCS}}$	$\mathcal{B}_\text{P}$ $\widehat{E(N)}$
0.03	0.00	0.951	893	0.954	850
0.03	0.05	0.951	856		
0.03	0.10	0.951	804		
0.03	0.15	0.950	761		
0.03	0.20	0.949	713		
0.03	0.25	0.947	669		
0.03	0.50	0.937	430		
0.06	0.00	0.946	238	0.963	250
0.06	0.05	0.944	225		
0.06	0.10	0.943	213		
0.06	0.15	0.942	201		
0.06	0.20	0.941	187		
0.06	0.25	0.937	176		
0.06	0.50	0.923	116		

Table 7: Monte Carlo results for $k = 4$ and $1 - \alpha = 0.95$, for i.i.d. Bernoulli data in the SC and MFC with $p_k = 0.85$.

δ	ρ	Config.	$\widehat{\text{PCS}}$	$\widehat{E(N)}$	$\mathcal{B}_\text{P}$ $\widehat{\text{PCS}}$	$\mathcal{B}_\text{P}$ $\widehat{E(N)}$
0.03	0.00	SC	0.957	2871	0.956	2559
0.03	0.05	SC	0.958	2736		
0.03	0.10	SC	0.955	2584		
0.03	0.15	SC	0.957	2438		
0.03	0.20	SC	0.954	2297		
0.03	0.25	SC	0.953	2155		
0.03	0.50	SC	0.946	1408		
0.03	0.00	MFC	0.983	2277	0.983	1823
0.03	0.05	MFC	0.983	2169		
0.03	0.10	MFC	0.983	2047		
0.03	0.15	MFC	0.984	1934		
0.03	0.20	MFC	0.983	1821		
0.03	0.25	MFC	0.982	1711		
0.03	0.50	MFC	0.979	1127		
0.06	0.00	SC	0.953	767	0.959	731
0.06	0.05	SC	0.949	727		
0.06	0.10	SC	0.948	690		
0.06	0.15	SC	0.947	651		
0.06	0.20	SC	0.943	613		
0.06	0.25	SC	0.939	571		
0.06	0.50	SC	0.907	373		
0.06	0.00	MFC	0.982	623	0.986	518
0.06	0.05	MFC	0.981	593		
0.06	0.10	MFC	0.980	561		
0.06	0.15	MFC	0.979	529		
0.06	0.20	MFC	0.978	498		
0.06	0.25	MFC	0.976	465		
0.06	0.50	MFC	0.961	308		

All of our results have been compared to $\mathcal{B}_\text{P}$ run without the use of CRN, because Paulson's procedure is designed to be used with i.i.d. data. Nevertheless, if we were to arbitrarily apply CRN to Paulson, that procedure actually becomes *less* efficient in terms of $E(N)$ as more correlation is induced (see Table 9). This inefficiency renders as questionable the use of Paulson in the presence of correlation between systems.

Table 8: Monte Carlo results for $\delta = 0.01$ and $1 - \alpha = 0.95$, for i.i.d. Bernoulli data in the SC with $p_k = 0.85$.

k	ρ	$\widehat{\text{PCS}}$	$\widehat{E(N)}$	$\mathcal{B}_\text{P}$ $\widehat{\text{PCS}}$	$\mathcal{B}_\text{P}$ $\widehat{E(N)}$
2	0.0	0.952	7628	0.955	7289
2	0.1	0.951	6842		
3	0.0	0.955	16029	0.954	14621
3	0.1	0.955	14426		

Table 9: Performance of $\mathcal{B}_\text{P}$ with use of CRN, for $1-\alpha = 0.95$.

k	ρ	p_k	$\mathcal{B}_\text{P}$ $\widehat{\text{PCS}}$	$\mathcal{B}_\text{P}$ $\widehat{E(N)}$
2	0.00	0.35	0.955	334.3
2	0.10	0.35	0.967	341.9
2	0.25	0.35	0.983	353.7
2	0.00	0.85	0.963	248.0
2	0.10	0.85	0.974	252.7
2	0.25	0.85	0.988	261.4
3	0.00	0.35	0.958	700.6
3	0.10	0.35	0.973	708.2
3	0.25	0.35	0.988	724.0
3	0.00	0.85	0.963	496.5
3	0.10	0.85	0.974	503.2
3	0.25	0.85	0.989	511.6

Therefore, the user can be fairly certain to achieve savings in $E(N)$ using our procedure when $\rho \geq 0.15$ for $k \leq 4$ and $\rho \geq 0.25$ for $k = 5$. Of course, the user will not necessarily know how much correlation will be induced in a real-world application; but then one could at least calculate a rough estimate of the correlation based on an initial sample. However, even if we cannot estimate the correlation with any accuracy, our procedure is only slightly less efficient than Paulson's when ρ is very small (0.05 to 0.1), particularly for $k \leq 3$; so there is little risk in using it and potentially large savings.

5 CONCLUSIONS

The modified $\mathcal{KN}$++ procedure exhibits significant savings in observations for i.i.d. normal data compared to $\mathcal{KN}$ and competitive performances for bernoulli data compared to $\mathcal{B}_\text{P}$ in certain circumstances. For i.i.d. normal data, the idea of updating the variance estimate

seems very promising in terms of producing savings in the number of observations. However, $\mathcal{KN}$++ does not guarantee the PCS requirement. Therefore, if the PCS requirement is a hard constraint, then $\mathcal{KN}$ or some other statistically valid procedures would be preferred. Otherwise, $\mathcal{KN}$++ should be a good substitute of those procedures.

For the Bernoulli selection problem, there are cases where $\mathcal{KN}$++ clearly defeats $\mathcal{B}_{\mathrm{P}}$ in terms of total number of observations. Use of the modified $\mathcal{KN}$++ is recommended when it is important to be able to detect small differences between the best and second-best systems ($\delta \leq 0.1$) for the case when k is small (say, $k \leq 5$) or when all of the inferior systems are considered to be highly competitive with the best.

ACKNOWLEDGMENTS

This research was supported in part by National Science Foundation grant DMI-0400260.

REFERENCES

Bechhofer, R. E., J. Kiefer, and M. Sobel. 1968. *Sequential Identification and Ranking Procedures (with Special Reference to Koopman-Darmois Populations)*. Chicago: University of Chicago Press.

Bechhofer, R. E. and R. V. Kulkarni. 1982. Closed adaptive sequential procedures for selecting the best of $k \geq 2$ Bernoulli populations. In *Statistical Decision Theory and Related Topics, III*, eds. S. S. Gupta and J. O. Berger, Vol. 1, 61–108, New York: Academic Press.

Billingsley, P. 1968. *Convergence of Probability Measures*. New York: John Wiley & Sons.

Boesel, J., B. L. Nelson, and S.-H. Kim. 2003. Using ranking and selection to 'clean up' after simulation optimization. *Operations Research*, 51:814–825.

Cario, M. C. and B. L. Nelson. 1997. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.

Chien, C., D. Goldsman, and B. Melamed. 1997. Large-sample results for batch means. *Management Science*, 43:1288–1295.

Damerdji, H. 1994. Strong consistency of the variance estimator in steady-state simulation output anaylysis. *Mathematics of Operations Research*, 19:494–512.

Damerdji, H. 1995. Mean-square consistency of the variance estimator in steady-state simulation output analysis. *Operations Research*, 43:282–291.

Damerdji, H. and D. Goldsman. 1995. Consistency of several variants of the standardized time series area variance estimator. *Naval Research Logistics*, 42:1161–1176.

Damerdji, H., S. G. Henderson, and P. W. Glynn. 1997. Computational efficiency evaluation in output analysis, In *Proceedings of the 1997 Winter Simulation Conference*, eds. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson, 208–215. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Goldsman, D. and B. L. Nelson. 1998. Comparing systems via simulation. In *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*, ed. J. Banks, 273–306. New York: John Wiley & Sons.

Kim, S.-H. and B. L. Nelson. 2001. A fully sequential procedure for indifference-zone selection in simulation. *ACM TOMACS*, 11:251–273.

Kim, S.-H. and B. L. Nelson. 2005. On the asymptotic validity of fully sequential selection procedures for steady-state simulation. Accepted for publication in *Operations Research*.

Malone, G. 2005. *Ranking and Selection Procedures for Bernoulli and Multinomial Data*. Doctoral Dissertation, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia.

Nelson, B. L., J. Swann, D. Goldsman, and W.-M. T. Song. 2001. Simple procedures for selecting the best simulated system when the number of alternatives is large. *Operations Research*, 49:950–963.

Paulson, E. 1993. Sequential procedures for selecting the best one of k Koopman-Darmois populations. Unpublished.

Sobel, M. and M. J. Huyett. 1957. Selecting the best one of several binomial populations. *Bell System Technical Journal*, 36:537–576.

AUTHOR BIOGRAPHIES

GWENDOLYN J. MALONE received her her Ph.D. in Industrial and Systems Engineering at the Georgia Institute of Technology in 2005 and is currently a simulation analyst at DRS Technologies in Palm Bay, Florida. Her research interests include simulation modeling and analysis and ranking and selection procedures. She is a member of IIE and INFORMS. Her e-mail address is `<gwen@gwenmalone.com>`

SEONG-HEE KIM is an Assistant Professor in the School of Industrial and Systems Engineering at the Georgia Institute of Technology. She received her Ph.D. in Industrial Engineering and Man-

agement Sciences from Northwestern University in 2001. Her research interests include ranking and selection, quality control, and simulation output analysis. She is a member of INFORMS and IIE, and serves the editorial board of the Simulation Department of *IIE Transactions*. Her e-mail and web addresses are <skim@isye.gatech.edu> and <www.isye.gatech.edu/~skim/>, respectively.

DAVID GOLDSMAN is a Professor in the School of Industrial and Systems Engineering at the Georgia Institute of Technology. He received his Ph.D. in Operations Research and Industrial Engineering from Cornell University. His research interests include simulation output analysis and ranking and selection. He is an active participant in the Winter Simulation Conference, having been Program Chair in 1995, and having served on the WSC Board of Directors since 2002. His e-mail address is <sman@isye.gatech.edu>, and his web page is <www.isye.gatech.edu/~sman/>.

DEMET BATUR is a Ph.D. candidate in the School of Industrial and Systems Engineering at the Georgia Institute of Technology. She received her B.Sc. degree in Industrial Engineering from Marmara University, Istanbul, Turkey, and her M.S. degree in School of Industrial and Systems Engineering from Georgia Institute of Technology in 2001 and 2004, respectively. Her current research interests include ranking and selection and simulation output analysis. Her e-mail address is <dbatur@isye.gatech.edu>.