IIE Transactions (2005) **37**, 623–634

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/07408170590948486

The tradeoff between sampling and switching: New sequential procedures for indifference-zone selection

L. JEFF HONG[1] and BARRY L. NELSON[2]

[1] *Department of Industrial Engineering & Engineering Management, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong*
E-mail: hongl@ust.hk
[2] *Department of Industrial Engineering & Management Sciences, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208-3119, USA*
E-mail: nelsonb@northwestern.edu

Received April 2003 and accepted December 2004

Statistical Ranking and Selection (R&S) is a collection of experiment design and analysis techniques for selecting the "population" with the largest or smallest mean performance from among a finite set of alternatives. R&S procedures have received considerable research attention in the stochastic simulation community, and they have been incorporated in commercial simulation software. One of the ways that R&S procedures are evaluated and compared is via the expected number of samples (often replications) that must be generated to reach a decision. In this paper we argue that sampling cost alone does not adequately characterize the efficiency of ranking-and-selection procedures, and the cost of switching among the simulations of the alternative systems should also be considered. We introduce two new, adaptive procedures, the minimum switching sequential procedure and the multi-stage sequential procedure with tradeoff, that provide the same statistical guarantees as existing procedures and significantly reduce the expected total computational cost of application, especially when applied to favorable configurations of the competing means.

1. Introduction

Ranking-and-Selection (R&S) procedures based on the indifference-zone formulation have been proposed to select the simulated system with the largest or smallest mean performance from among a finite number of alternative systems (see Bechhofer *et al.* (1995) for a summary). Among these, fully sequential procedures, which approximate the sum of differences between two systems as a Brownian motion process and use a triangular continuation region to determine the stopping time of the selection process, were first proposed by Paulson (1964). Figure 1 illustrates how a triangular continuation works. Hartmann (1988, 1991) improved Paulson's procedure by replacing Boole's inequality with a geometric inequality and replacing a large-deviation bound by a Brownian motion bound. These procedures were intended for the case of normally-distributed data with a known or unknown common variance across systems. Recently, Kim and Nelson (2001, 2003) further extended Hartmann's work to allow for unknown and unequal variances, the use of common random numbers and single-replication experiment designs, yielding procedures that are more applicable in computer simulation experiments. In this paper we also refine sequential selection procedures with a triangular continuation region by considering the computation costs that are incurred in simulation experiments.

One important, and relatively recent, application of R&S procedures is within optimization-via-simulation algorithms (see, for instance, Pichitlamken and Nelson (2003) and Hong and Nelson (2004)). Many optimization algorithms attempt to move from a current good solution to an improved solution on each iteration by selecting the best from among a small set of candidates or neighbors. R&S procedures can be embedded within these algorithms to help recognize improved solutions efficiently, and with a statistical guarantee of correctness. Since some optimization algorithms are sequential and attempt to revisit solutions to ensure convergence, Pichitlamken *et al.* (2004) developed fully-sequential selection procedures that exploit the data already obtained on previous visits, even if the sample sizes are unequal. The procedures we introduce in the current paper also extend easily to unequal initial samples, and they are particularly well suited to the optimization setting.

R&S procedures are typically evaluated in terms of the expected number of samples required to reach termination. Sequential procedures with elimination (such as the procedure of Kim and Nelson (2001, 2003) hereafter designated as KN) and Bayesian procedures (such as Chick and Inoue (2001)) reduce the expected total number of samples

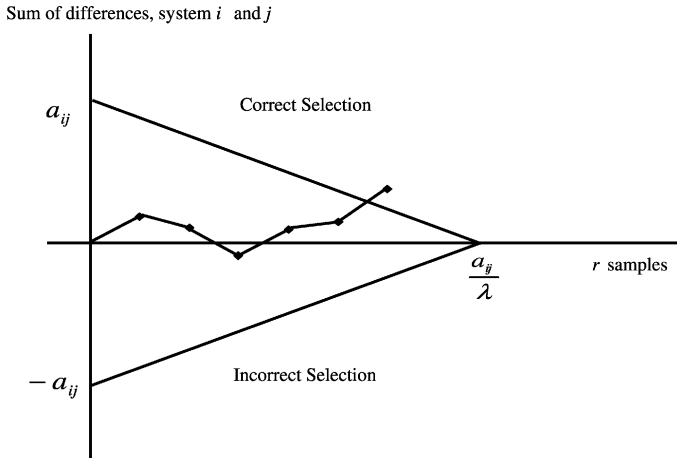

Fig. 1. Triangular continuation region for sequential, indifference-zone procedures. The selection of system i or j as being the best depends on whether the sum of differences exits the region from the top or bottom. In this example, system i has the largest mean so a correct selection exits upwards.

relative to the well-known two-stage indifference-zone selection procedures, such as the procedure of Rinott (1978) hereafter designated as R. Two-stage procedures were designed to provide their guarantees in the face of a "least-favorable configuration", and not to try to adapt to more favorable situations (two-stage procedures with elimination at the first stage have been discussed in, for instance, Nelson *et al.* (2001) and these can be more efficient than procedures such as R).

The focus in R&S research has been on developing procedures that reduce the cost of *sampling*, which means generating data (typically replications) from a simulation model. To achieve this reduction, fully sequential procedures such as KN repeatedly switch among the different simulation models, where a *switch* occurs when changing the model instance from which samples are generated. In fact, the KN procedure requires one switch for each new sample generated after an initial stage of sampling.

Unfortunately, the computational overhead of switching can be significant, sometimes orders of magnitude more than sampling. Furthermore, the cost of switching may be incurred thousands of times if the R&S procedure is embedded within an optimization algorithm that explores a large solution space. The work required to switch from one simulated system to another usually includes storing state information about the current system (including values of the random number seeds); saving all relevant output data; swapping the executable code for the current system out of, and the code for the next system into, active memory; and restoring the state information for the next system to the values it had on the last call. Therefore, focusing solely on the sampling cost is misleading in many applications. One way to solve this problem is to use multiple parallel processors. However, the work required to implement the parallel structure may be substantial and most of the state-of-the-art simulation optimization algorithms are designed for a single processor. Therefore, we focus on the situation of a single processor.

Two-stage indifference-zone selection procedures, such as R, minimize the number of switches; however, they do not adapt to the observed differences among the systems as the sampling progresses. Thus, they are efficient from a switching perspective, but inefficient from a sampling perspective. KN, and similar procedures, are efficient from a sampling perspective, but may not be efficient when switching is considered. In this paper we propose sequential procedures that are adaptive, like the KN procedure, but also balance the cost of sampling and switching to reduce the *total* cost of the experiment. Furthermore, our procedures tend to allocate more samples to the better systems, which is desirable when R&S is also used to add statistical inference after an optimization algorithm has terminated (Boesel *et al.* (2003)).

In Section 2 we provide a new procedure, the Minimum Switching Sequential (MSS) procedure, that consumes the minimal number of switches, just as do the two-stage procedures, while still maintaining sequential sampling. In Section 3 a multi-stage framework is given within which the MSS and KN procedures can be applied adaptively. A procedure for determining the number of stages (switches) and the sample size within each stage for the multi-stage framework is discussed in Section 4. Numerical examples are given in Section 5.

2. A minimum switching procedure

In this section we describe a sequential procedure that guarantees, with a confidence level greater than or equal to $1-\alpha$, that the system ultimately selected has the largest true mean if the true mean of the best system is at least δ better than the second best. When there are inferior systems whose means are within δ of the best, then the procedure guarantees to find one of these *good* systems with the same probability. The parameter δ, which defines the *indifference zone*, is set by the experimenter as the minimum difference in expected performance that it is important to detect. Differences of less than δ are considered practically insignificant.

The MSS procedure, has two stages, similar to the R procedure. A *stage* is a checkpoint at which the maximum number of samples that can be taken from each system until the next checkpoint is determined; it is the "maximum" number of samples because fewer samples may be needed if some systems are eliminated from consideration. All of the procedures considered in this paper, including R and KN, assume unknown output variances, and for that reason they require an initial stage of sampling that is not adaptive and whose size is somewhat arbitrary; we refer to this as the "zeroth stage". The first decision about how to proceed occurs after the zeroth stage.

The MSS procedure works as follows: after obtaining $n_0 \geq 2$ samples from each system in the zeroth stage, it estimates the parameters of the triangular continuation region

and checks if any system can be eliminated immediately. Let I be the set of systems still in play at the end of the zeroth stage, and let B and S *always* denote the systems in I with the best and the second-best first-stage sample means. In stage 1, the MSS procedure takes the maximum number of samples implied by the continuation region from system B, so that no more samples are needed for system B under any circumstances (this contrasts with the R procedure which takes the maximum number of samples from *all* systems, and the KN procedure which obtains one sample at a time from all systems still in play). Then the MSS procedure obtains one sample at a time from system S, comparing a weighted sample mean from system S to a weighted sample mean from system B, with elimination decisions after each sample. If B eliminates S, then the identity of S is updated (since the former system S is no longer in the now-updated set I, and S is always the system in I with the second-largest zeroth-stage sample mean), and the process starts over again. If S eliminates B, then the identities of both B and S are updated, the new system B gets the maximum number of samples, and the process continues. The procedure always examines only two systems at a time, eliminating one, and stopping when there is only one system remaining in I. The MSS procedure is sequential in the way that it obtains samples, but it requires at most k switches after the zeroth stage, where k is the number of systems.

Throughout this paper we use $\mathbf{x}_i$ to denote the ith system and use $X_{i\ell}$ to denote the ℓth independent sample from system i. Think of $\mathbf{x}_i$ as the vector of decision variables that define the ith system. We assume that $X_{i\ell} \sim N(\mu_i, \sigma_i^2)$, with both μ_i and σ_i^2 unknown. The procedure has two stages, denoted by $s = 0, 1$. Let $\bar{X}_i(m) = m^{-1}\sum_{\ell=1}^{m} X_{i\ell}$ denote the sample mean of the first m samples from system i, and let $\bar{X}_i(n; s)$ denote the sample mean of the first n samples from system i *taken in stage* s, where $i = 1, 2, \ldots, k$.

Minimum switching sequential procedure

***Setup*:** Select confidence level $1/k < 1 - \alpha < 1$, indifference-zone parameter $\delta > 0$ and zeroth-stage sample size $n_0 \geq 2$. Select λ such that $0 < \lambda < \delta$ ($\lambda = \delta/4$ is recommended by Paulson (1964)).

***Initialization*:** Take n_0 samples $X_{i\ell}$, $\ell = 1, 2, \ldots, n_0$, from each system $i = 1, 2, \ldots, k$. For all $i \neq j$, calculate:

$$S_{ij}^2 = \frac{1}{n_0 - 1}\sum_{\ell=1}^{n_0}(X_{i\ell} - X_{j\ell} - [\bar{X}_i(n_0; 0) - \bar{X}_j(n_0; 0)])^2, \quad (1)$$

the sample variance of the difference between systems i and j, and let:

$$a_{ij} = \frac{(n_0 - 1)S_{ij}^2}{4(\delta - \lambda)}\left\{\left[1 - (1 - \alpha)^{1/(k-1)}\right]^{-2/(n_0-1)} - 1\right\}. \quad (2)$$

The parameters a_{ij} and λ define the triangular continuation region for systems i and j. Let

$$N_{ij} = \max\left\{0, \left\lceil \frac{a_{ij}}{\lambda} \right\rceil - n_0\right\},$$

where $\lceil c \rceil$ denotes the smallest integer greater than or equal to c, be the maximum additional number of observations required for systems i and j when comparing them, and $N_{ij} \geq 0$. Notice that a_{ij}/λ is the end of the triangular continuation region.

***Initial screening*:** For all $i \neq j$, calculate the first-stage summary statistic:

$$Z_{ij}(n_0) = n_0[\bar{X}_i(n_0; 0) - \bar{X}_j(n_0; 0)].$$

Let

$$I = \{i : Z_{ij}(n_0) \geq \min\{0, -a_{ij} + n_0\lambda\}, \\ i, j = 1, 2, \ldots, k,\ i \neq j\},$$

be the set of systems still in play. If $|I| = 1$, then stop and select the system whose index is in I as the best. Otherwise, sort the elements in I based on the zeroth-stage sample means $\bar{X}_i(n_0; 0)$ and let B and S *always* be the systems in I with the best and the second-best zeroth-stage sample means. Let

$$N_B = \max_{j \in I, j \neq B} N_{Bj},$$

be the maximum additional number of observations required for system B when comparing it against all other systems in I. Take N_B additional samples from system B. Let r be the number of samples taken from system S in stage 1, and set $r = 0$. Let ψ be an elimination indicator. If $\psi = 0$ then system B is still in contention; otherwise, system B has been eliminated. Set $\psi = 0$.

***Screening*:** Take one sample from system S, and let $r = r + 1$. Calculate the combined two-stage summary statistic:

$$Z_{BS}(n_0 + r) = Z_{BS}(n_0) + r[\bar{X}_B(N_B; 1) - \bar{X}_S(r; 1)], \quad (3)$$

and let

$$W_{BS} = \max\{0, a_{BS} - \lambda(n_0 + r)\}.$$

If $Z_{BS}(n_0 + r) \geq W_{BS}$, then let $I = I \setminus \{S\}$ and update S; if $Z_{BS}(n_0 + r) \leq -W_{BS}$, then let $I = I \setminus \{B\}$, update B and S, and let $\psi = 1$; otherwise, go to ***Screening***.

***Stopping rule*:** If $|I| = 1$, then stop and select the system whose index is in I as the best. Otherwise, if $\psi = 1$, then let:

$$N_B = \max_{j \in I, j \neq B} N_{Bj},$$

and take $\max\{0, N_B - r\}$ samples from system B. Always let $r = 0$ and $\psi = 0$, then go to ***Screening***.

Remark 1. $Z_{BS}(n_0 + r)$ is defined by Equation (3) instead of the more-natural cumulative difference:

$$Z_{BS}(n_0 + r) = \sum_{\ell=1}^{n_0+r} (\bar{\bar{X}}_B(n_0 + N_B) - X_{S\ell}). \tag{4}$$

Jennison *et al.* (1982) have shown that when $k > 2$ the sampling rule can only depend on the difference between sample means, not on the individual sample means, if Equation (4) is used. However, the individual sample means are required for sorting in the MSS procedure. Therefore, we use Equation (3) to ensure independence between stages 0 and 1.

To prove the validity of the procedure, we need the following lemmas.

Lemma 1. (Slepian's Inequality, see Tong (1980).) *Let* $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_r)^T$ *be distributed according to a multivariate normal distribution* $N(0, \Sigma)$, *where* Σ *is a correlation matrix. Let* $\mathbf{R} = (\rho_{ij})$ *and* $\mathbf{T} = (\tau_{ij})$ *be two positive semidefinite correlation matrices. If* $\rho_{ij} \geq \tau_{ij}$ *holds for all* i, j, *then*:

$$\Pr_{\Sigma=\mathbf{R}}\left[\bigcap_{i=1}^{r}\{Y_i \geq a_i\}\right] \geq \Pr_{\Sigma=\mathbf{T}}\left[\bigcap_{i=1}^{r}\{Y_i \geq a_i\}\right],$$

holds for all $\mathbf{a} = (a_1, a_2, \ldots, a_r)^T$. *The inequality is strict if* $\mathbf{R}$ *and* $\mathbf{T}$ *are positive definite and if the strict inequality* $\rho_{ij} > \tau_{ij}$ *holds for some* $i \neq j$.

Lemma 2. (After Paulson (1964).) *Let* $Y_1, Y_2, \ldots$ *be i.i.d.* $N(\delta, \sigma^2)$ *random variables with* $\delta > 0$. *Then for any* r_0 *and* r_1, *where* $r_0 < r_1$:

$$\Pr\left\{\bigcap_{r=r_0}^{r_1}\left[\sum_{\ell=1}^{r} Y_\ell \geq \min\{0, -a + \lambda r\}\right]\right\} \geq 1 - e^{-2(\delta-\lambda)a/\sigma^2},$$

where $a > 0$ *and* $0 < \lambda < \delta$.

Remark 2. Suppose there is a triangular continuation region defined by:

$$L(r) = -a + r\lambda,$$
$$U(r) = a - r\lambda.$$

Let $\mathcal{E}$ denote the event that $\sum_{\ell=1}^{r} Y_\ell$ leaves the continuation region for the first time through the upper boundary. Then

$$\Pr\{\mathcal{E}\} \geq \Pr\left\{\bigcap_{r=n_0}^{\lceil a/\lambda \rceil}\left[\sum_{\ell=1}^{r} Y_\ell \geq \min\{0, -a + \lambda r\}\right]\right\},$$
$$\geq 1 - e^{-2(\delta-\lambda)a/\sigma^2}.$$

This is the probability bound used in Paulson (1964) for his sequential selection procedures using a triangular continuation region.

Lemma 3. (After Tamhane (1977).) *Let* $V_1, V_2, \ldots, V_k$ *be independent random variables, and let* $g_j(v_1, v_2, \ldots, v_k)$, $j = 1, 2, \ldots, p$, *be non-negative, real-valued functions, each one nondecreasing in each of its arguments. Then*

$$E\left[\prod_{j=1}^{p} g_j(V_1, V_2, \ldots, V_k)\right] \geq \prod_{j=1}^{p} E[g_j(V_1, V_2, \ldots, V_k)].$$

We are now in a position to prove the main result. Without loss of generality, suppose that the true means of the systems are indexed so that $\mu_k \geq \mu_{k-1} \geq \cdots \geq \mu_1$.

Theorem 1. *Suppose that* $X_{i\ell}$, $\ell = 1, 2, \ldots$, *are i.i.d. normally distributed, and that* X_{ip} *and* X_{jq} *are independent for* $i \neq j$. *Then the MSS procedure selects* $\mathbf{x}_k$ *with a probability of at least* $1 - \alpha$ *whenever* $\mu_k - \mu_i \geq \delta$ *for* $i = 1, 2, \ldots, k - 1$.

Proof. We start with the case of two systems, since only two systems are compared at a time, with one having the maximum number of samples. Let these systems be i and j with system i having N_i samples. For any two systems i and j, $i \neq j$, and for any $0 \leq r \leq N_i$, let,

$$P_{ij}(n_0 + r) = \frac{Z_{ij}(n_0 + r) - (n_0 + r)(\mu_i - \mu_j)}{\sqrt{\mathrm{Var}(Z_{ij}(n_0 + r))}},$$

be the standardized $Z_{ij}(n_0 + r)$ process. For any two systems i and j, $i \neq j$, and for any $0 \leq r \leq N_i$, define:

$$Y_{ij}(n_0 + r) = \sum_{\ell=1}^{n_0+r} (X_{i\ell} - X_{j\ell}).$$

Then we can let

$$Q_{ij}(n_0 + r) = \frac{Y_{ij}(n_0 + r) - (n_0 + r)(\mu_i - \mu_j)}{\sqrt{\mathrm{Var}(Y_{ij}(n_0 + r))}},$$

be the standardized $Y_{ij}(n_0 + r)$ process.

Notice that

$$\mathrm{Var}[Z_{ij}(n_0 + r)] = n_0\left(\sigma_i^2 + \sigma_j^2\right) + r\left(\frac{r}{N_i}\sigma_i^2 + \sigma_j^2\right),$$
$$\mathrm{Var}[Y_{ij}(n_0 + r)] = (n_0 + r)\left(\sigma_i^2 + \sigma_j^2\right).$$

Since $0 \leq r \leq N_i$ we have that:

$$\mathrm{Var}[Z_{ij}(n_0 + r)] \leq \mathrm{Var}[Y_{ij}(n_0 + r)]. \tag{5}$$

For $0 \leq r_1 \leq r_2 \leq N_i$:

$$\mathrm{Cov}[P_{ij}(n_0 + r_1), P_{ij}(n_0 + r_2)] = \frac{n_0\left(\sigma_i^2 + \sigma_j^2\right) + r_1\left((r_2/N_i)\sigma_i^2 + \sigma_j^2\right)}{\sqrt{n_0\left(\sigma_i^2 + \sigma_j^2\right) + r_1\left((r_1/N_i)\sigma_i^2 + \sigma_j^2\right)}\sqrt{n_0\left(\sigma_i^2 + \sigma_j^2\right) + r_2\left((r_2/N_i)\sigma_i^2 + \sigma_j^2\right)}},$$

$$\mathrm{Cov}[Q_{ij}(n_0 + r_1), Q_{ij}(n_0 + r_2)] = \sqrt{\frac{n_0 + r_1}{n_0 + r_2}}.$$

One can easily check that:

$$\begin{aligned}&\mathrm{Cov}[P_{ij}(n_0+r_1),\, P_{ij}(n_0+r_2)]\\ &\quad\geq \mathrm{Cov}[Q_{ij}(n_0+r_1),\, Q_{ij}(n_0+r_2)].\end{aligned}$$

Notice that covariance and correlation are the same here, since P_{ij} and Q_{ij} are standardized. Since $P_{ij}(n_0+r)$ and $Q_{ij}(n_0+r)$, $r=0,1,\ldots,N_i$, are all normally distributed, then by Lemma 1 we have:

$$\Pr\left[\bigcap_{r=0}^{N}\{P_{ij}(n_0+r)\geq a_r\}\right]\geq \Pr\left[\bigcap_{r=0}^{N}\{Q_{ij}(n_0+r)\geq a_r\}\right], \tag{6}$$

for any $0\leq N\leq N_i$.

Now we can find a lower bound on the probability that system k eliminates system i.

$$\begin{aligned}&\Pr\{\mathbf{x}_k \text{ eliminates } \mathbf{x}_i\}\\ &\geq E\Bigg[\Pr\Bigg\{\bigcap_{r=0}^{N_{ki}}[Z_{ki}(n_0+r)\\ &\quad\geq \min\{0,-a_{ki}+(n_0+r)\lambda\}]\Big|S_{ki}^2\Bigg\}\Bigg],\\ &= E\Bigg[\Pr\Bigg\{\bigcap_{r=0}^{N_{ki}}\Bigg[P_{ki}(n_0+r)\\ &\quad\geq \frac{\min\{0,-a_{ki}+(n_0+r)\lambda\}-(n_0+r)(\mu_k-\mu_i)}{\mathrm{Var}[Z_{ki}(n_0+r)]}\Bigg]\Big|S_{ki}^2\Bigg\}\Bigg],\\ &\geq E\Bigg[\Pr\Bigg\{\bigcap_{r=0}^{N_{ki}}\Bigg[Q_{ki}(n_0+r)\\ &\quad\geq \frac{\min\{0,-a_{ki}+(n_0+r)\lambda\}-(n_0+r)(\mu_k-\mu_i)}{\mathrm{Var}\,[Z_{ki}(n_0+r)]}\Bigg]\Big|S_{ki}^2\Bigg\}\Bigg]\\ &\text{(by the inequality of Equation (6), since } N_{ki}\leq N_i),\\ &\geq E\Bigg[\Pr\Bigg\{\bigcap_{r=0}^{N_{ki}}\Bigg[Q_{ki}(n_0+r)\\ &\quad\geq \frac{\min\{0,-a_{ki}+(n_0+r)\lambda\}-(n_0+r)(\mu_k-\mu_i)}{\mathrm{Var}\,[Y_{ki}(n_0+r)]}\Bigg]\Big|S_{ki}^2\Bigg\}\Bigg]\\ &\text{(by the inequality of Equation (5), since } \min\{0,-a_{ki}+(n_0\\ &+r)\lambda\}-(n_0+r)(\mu_k-\mu_i)<0),\\ &= E\Bigg[\Pr\Bigg\{\bigcap_{r=0}^{N_{ki}}\Bigg[\sum_{l=1}^{n_0+r}(X_{kl}-X_{il})\\ &\quad\geq \min\{0,-a_{ki}+(n_0+r)\lambda\}\Bigg]\Big|S_{ki}^2\Bigg\}\Bigg],\\ &\geq E\Bigg[1-\exp\left\{\frac{-2(\mu_k-\mu_i-\lambda)a_{ki}}{\sigma_{ki}^2}\right\}\Big|S_{ki}^2\Bigg]\\ &\text{(by Lemma 2)},\\ &\geq E\Bigg[1-\exp\left\{\frac{-2(\delta-\lambda)a_{ki}}{\sigma_{ki}^2}\right\}\Big|S_{ki}^2\Bigg]\\ &\text{(since } \mu_k-\mu_i\geq\delta),\\ &= E\Bigg[1-\exp\Bigg\{-\frac{1}{2}\{[1-(1-\alpha)^{1/(k-1)}]^{-2/(n_0-1)}-1\}\\ &\quad\times\frac{(n_0-1)S_{ki}^2}{\sigma_{ki}^2}\Bigg\}\Bigg],\\ &= (1-\alpha)^{1/(k-1)}\\ &\big(\text{since } (n_0-1)S_{ki}^2/\sigma_{ki}^2\sim\chi_{n_0-1}^2\\ &\quad\text{and } E\big[e^{t\chi_{n_0-1}^2}\big]=(1-2t)^{-(n_0-1)/2}\big).\end{aligned}$$

Let CS_i denote the event that system k eliminates system i. Then

$$\Pr\{\text{select } \mathbf{x}_k\}\geq \Pr\left\{\bigcap_{i=1}^{k-1}\mathrm{CS}_i\right\},$$

because system i, $i\neq k$, could be eliminated by system j, $j\neq k$. Therefore,

$$\begin{aligned}&\Pr\{\text{select } \mathbf{x}_k\}\geq \Pr\left\{\bigcap_{i=1}^{k-1}\mathrm{CS}_i\right\},\\ &= E\left[\Pr\left\{\bigcap_{i=1}^{k-1}\mathrm{CS}_i\Big|X_{k1},\ldots,X_{k,n_0+N_i},S_{k1}^2,\ldots,S_{k,k-1}^2\right\}\right],\\ &= E\left[\prod_{i=1}^{k-1}\Pr\left\{\mathrm{CS}_i\Big|X_{k1},\ldots,X_{k,n_0+N_i},S_{ki}^2\right\}\right]\\ &\quad\text{(the events are conditionally independent)},\\ &\geq \prod_{i=1}^{k-1}E\big[\Pr\left\{\mathrm{CS}_i\Big|X_{k1},\ldots,X_{k,n_0+N_i},S_{ki}^2\right\}\big]\\ &\quad\big(\text{by Lemma 3, since } \Pr\left\{\mathrm{CS}_i\Big|X_{k1},\ldots,X_{k,n_0+N_i},S_{ki}^2\right\}\\ &\qquad\text{is nondecreasing in } X_{kl} \text{ and } S_{ki}^2\big),\\ &= \prod_{i=1}^{k-1}\Pr\left\{\mathbf{x}_k \text{ eliminates } \mathbf{x}_i\right\},\\ &\geq \prod_{i=1}^{k-1}(1-\alpha)^{1/(k-1)},\\ &= 1-\alpha.\end{aligned}$$

This proves that $\Pr\{\text{select } \mathbf{x}_k\}\geq 1-\alpha$ whenever $\mu_k-\mu_i\geq\delta$ for $i=1,2,\ldots,k$. ■

Corollary 1. *Suppose $\mu_k-\mu_{k-1}<\delta$. Then the MSS procedure selects a solution whose mean is within δ of μ_k with probability at least $1-\alpha$.*

Proof. Suppose $\mu_k\geq\mu_{k-1}\geq\cdots\geq\mu_1$. The statement is trivially true if $\mu_k-\mu_1<\delta$, as we can select any system to satisfy the claim. Let $t>1$ be the first system index within δ of μ_k, i.e., $\mu_k-\mu_t<\delta$, and $\mu_k-\mu_{t-1}\geq\delta$. The correct ("good") selection in this context is that all elements in the set $\mathcal{A}=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{t-1}\}$ get eliminated.

$$\Pr\{\text{eliminate } \mathcal{A}\}\geq \Pr\left\{\bigcap_{i=1}^{t-1}\mathrm{CS}_i\right\},$$

$$\begin{aligned}
&\geq \prod_{i=1}^{t-1} \Pr\{\mathbf{x}_k \text{ eliminates } \mathbf{x}_i\},\\
&\geq \prod_{i=1}^{t-1} (1-\alpha)^{1/(k-1)},\\
&= (1-\alpha)^{(t-1)/(k-1)},\\
&\geq 1-\alpha.
\end{aligned}$$

This proves Corollary 1. ■

Remark 3. We proved the use of a_{ij} defined in Equation (2) by using Paulson's probability bound. The proof based on this bound is concise and provides intuition about why the result is true (greater induced dependence by obtaining more samples on one system). Pichitlamken *et al.* (2002) showed that the probability bound of Fabian (1974) which is the basis of the KN procedure, can also be used, and that the proof carries over directly to our context. Fabian's probability bound is a Brownian motion bound that is tighter than Paulson's probability bound which is a large-deviation bound. If Fabian's probability bound is used, we have:

$$a_{ij} = \frac{(n_0-1)S_{ij}^2}{4(\delta-\lambda)}\left\{\left[2-2(1-\alpha)^{1/(k-1)}\right]^{-2/(n_0-1)}-1\right\}, \quad (7)$$

for $\lambda = \delta/2$. The quantity a_{ij} in Equation (7) is smaller than it is in Equation (2); thus the procedure may terminate sooner.

Remark 4. If some systems have been previously sampled, as discussed in Pichitlamken *et al.* (2004), the MSS procedure can be easily extended to fit the situation. Assume there are n_B and n_S initial samples for systems B and S. Without loss of generality, we assume $n_B > n_0$ and $n_S > n_0$; otherwise, simply consider them as the zeroth-stage samples. Then take $N_B + n_0 - n_B$ samples from system B, calculate $Z_{BS}(n_0 + r)$ using Equation (3) and start screening from $r = n_S - n_0$ instead of $r = 1$.

The MSS procedure has the same number of switches as two-stage procedures, such as R, when both take additional samples for all systems after stage 0. However, the MSS procedure can eliminate clearly inferior systems at the end of stage 0, whereas the R procedure only ceases sampling if n_0 happens to be larger than the total number of samples that it requires. Not surprisingly, our experiments show that the average sample size of the MSS procedure is less than or equal to that of the R procedure in all situations. When the true difference between any system and the best system is larger than the indifference amount δ, which is the usual case in practice, the MSS procedure has a much smaller average sample size than the R procedure. Since two-stage procedures have been implemented in commercial simulation software, such as Automod (Brooks-PRI Automation) and Awesim (Frontstep), the MSS procedure is a good substitute for them.

In fully sequential procedures such as KN, the elimination decisions between the better and the inferior systems occur with each having the same number of samples, since the sampling is synchronized. In the MSS procedure the inferior systems are typically eliminated by one of the (apparently) better systems that has received *all* of the samples that the MSS procedure will ever allocate to it. In some cases this fact will allow the MSS procedure to be more efficient than the KN procedure, even if the switching cost is not significant. Consider, for instance, the case that among $k > 2$ systems the true mean difference between the best and second-best systems is small (less than or equal to δ), and the other systems are clearly inferior to these two. In this case, it is very likely that both the MSS and KN procedures will end up taking nearly the maximum number of samples from the two good systems. However, the MSS procedure will tend to eliminate the clearly inferior systems earlier (fewer samples), because the inferior systems are typically compared to a good system that has received all of its samples (since it is likely that one of the two good systems will have the largest zeroth-stage sample mean, and therefore will immediately be allocated all of its samples).

The disadvantage that the MSS procedure has is that at least one system *always* gets the maximum number of samples, which may be more than needed when true differences between systems are much larger than the indifference-zone parameter δ. In that case the wasted sampling may overcome the savings from not switching. In the next two sections, a multi-stage procedure, and a heuristic for determining the sample size in each stage, are provided to solve this problem.

3. A multi-stage procedure

In this section, we provide a multi-stage sequential framework that yields procedures having the same probability guarantee as the MSS procedure but allowing much more sampling flexibility. The framework/procedure is called the Multi-stage Sequential procedure with Tradeoff (MST). In the MST procedure, a sampling rule for the next stage, including the sampling order and the sample size, can be determined at the end of the previous stage using all information available through that stage. This procedure allows the control of sample size, number of switches and number of stages. Both the KN and MSS procedures fit into this framework: when the number of samples in each stage for each system is fixed to be one, and using Fabian's probability bound, it becomes the KN procedure. When the number of stages is fixed to be two, it becomes the MSS procedure.

As with the MSS procedure, the MST procedure starts by taking $n_0 \geq 2$ samples from each system in the zeroth stage, estimating the parameters of the continuation region and checking if any system can be eliminated. The MST procedure then uses all available information to determine

the maximum number of samples for the next stage, which may (and likely will) be less than the maximum number required to reach the end of the continuation region. Stage 1 is then executed in a manner that is essentially the same as in the MSS procedure (with the additional feature that all surviving systems are used in elimination decisions). Any systems that survive go on to stage 2, whose maximum number of samples is determined before the stage begins using all of the information at hand. The process continues until there is only a single system left in play. In this section we present the procedure in complete generality, and defer proposing a specific sampling rule until Section 4. Let s be the stage count, n_s be the maximum number of samples allocated to a contending system in stage s, and $N_s = \sum_{t=0}^{s} n_t$ be the total (cumulative) sample size at the end of stage s.

Multi-stage sequential procedure with tradeoff

***Setup*:** Select the confidence level $1/k < 1-\alpha < 1$, indifference-zone parameter $\delta > 0$ and first-stage sample size $n_0 \geq 2$. Set $s = 0$ and define $N_{-1} = 0$.

***Initialization*:** Take n_0 samples $X_{i\ell}$, $\ell = 1, 2, \ldots, n_0$ from each system $i = 1, 2, \ldots, k$. For all $i \neq j$, calculate S_{ij}^2, the sample variance of the difference between systems i and j, using Equation (1), and calculate a_{ij} either using Equation (2) and $\lambda = \delta/4$, or using Equation (7) and $\lambda = \delta/2$.

***Initial screening*:** For all $i \neq j$, calculate:

$$Z_{ij}(n_0) = n_0[\bar{X}_i(n_0; 0) - \bar{X}_j(n_0; 0)].$$

Let

$$I = \{i\colon Z_{ij}(n_0) \geq \min\{0, -a_{ij} + n_0\lambda\}, i, j = 1, 2, \ldots, k, i \neq j\}.$$

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

***Determine sampling rule*:** Let $N_s = N_{s-1} + n_s$. Sort the elements in I by $\bar{X}_i(N_s)$ from largest to smallest. Let $[i]$ be the system with the ith largest sample mean. Determine the sample size n_{s+1} for stage $s+1$ using $Z_{ij}(n_0)$, $Z_{ij}(n_0+1), \ldots, Z_{ij}(N_s)$ for all $i \neq j$ (one possible rule is provided in Section 4). Notice that $Z_{ij}(N_s)$, for all $i \neq j$ in I, provide enough information to sort $\bar{X}_i(N_s)$ for all $i \in I$.

Update the stage to $s = s + 1$, and take n_s samples from system [1]. Let J be the set of systems that have completed sampling in stage s and have not been eliminated. Set $J = \{[1]\}$.

***Screening*:** In this step, all surviving systems that are currently not in the set J are compared to the systems in J to see if they can enter J and if some systems in J can be removed from J. Let r be the number of samples taken from system $[t]$ so far in this stage. Let ψ be an elimination indicator: If $\psi = 0$ then system $[t]$ is still in contention; otherwise, system $[t]$ has been eliminated.
Let $t = 2$.

While $t \leq |I|$ do
  Set $r = 0$ and $\psi = 0$.
  While $(r < n_s)$ and $(J \neq \emptyset)$ and $(\psi = 0)$ do
    Take one observation from system $[t]$ and let $r = r + 1$. For every $i \in J$, calculate:

$$Z_{i[t]}(N_{s-1} + r) = Z_{i[t]}(N_{s-1}) + r\big[\bar{X}_i(n_s; s) - \bar{X}_{[t]}(r; s)\big], \quad (8)$$

$$W_{i[t]} = \max\{0, a_{i[t]} - \lambda(N_{s-1} + r)\},$$

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

$$J = \big\{i : i \in J^{\text{old}} \text{ and } Z_{i[t]}(N_{s-1} + r) \geq -W_{i[t]}, \forall i \in J^{\text{old}}\big\},$$

    If $Z_{i[t]}(N_{s-1} + r) \geq W_{i[t]}$ for some $i \in J^{\text{old}}$ then set $\psi = 1$.
  End while
  If $(J = \emptyset)$ or $(\psi = 0)$, then $J = J \cup \{[t]\}$ and take $\max\{n_s - r, 0\}$ samples from system $[t]$.
  $t = t + 1$.
End while
Let $I = J$.

Stopping rule*:** If $|I| = 1$, then stop and select the system whose index is in I as the best. Otherwise, go to ***Determine sampling rule.

Remark 5. Notice that:

$$Z_{ij}(N_s) = \sum_{\ell=1}^{N_s} (X_{i\ell} - X_{j\ell}),$$

Thus, the first s stages can be viewed as a single stage with N_s samples for any two systems in contention at stage $s+1$. Therefore, the validity of the MST procedure can be proved easily using repeated application of the same techniques as those in the proof of the MSS procedure.

The MST procedure provides a general framework to design procedures that consider both sampling cost and switching cost. One can determine the sample size for the next stage, and therefore the number of stages and number of switches, based on the information available at the current stage. In the next section we propose a sampling rule that takes advantage of this flexibility.

4. A heuristic for determining the sampling rule

The sampling rule in the MST procedure must specify the maximum number of samples to take from each system in the next stage. Ideally, we want an optimal allocation

considering all systems and all situations. This turns out to be very difficult since the $Z_{ij}(n)$ processes defined in Equation (8) are not independent of each other and the decision must be made based on sample information.

One approach is to approximate the processes by independent Brownian motion processes assuming that the sample means and variances are true means and variances, and then formulate an optimization problem where the objective includes the probability that a Brownian motion process exits a triangular region before some time t; this involves difficult numerical integrations (Hall, 1997). Solving this optimization problem is also time consuming and the quality of the result depends heavily on how well the sample means and variances approximate the true means and variances.

We have found that a crude approximation still yields significant improvements. To derive the algorithm we need the following result from Kamien and Schwartz (1981).

Lemma 4. (Optimal inspection.) *If a system can fail at any time t, $0 \le t \le T$, let $F(t)$ be the (known) probability of failure between time 0 and t, where F is a nondecreasing differentiable function with $F(0) = 0$ and $F(T) = 1$. Let c_0 be the cost per inspection, τ be the time between the failure and its detection, and $L(\tau)$ be the loss caused by not detecting the system failure for a period of τ time units. If $L(\tau)$ is linear, $L(\tau) = c_1\tau$, then the optimal inspection rate function $r(t)$ (number of inspections per unit time at t) is*:

$$r(t) = \left\{ \frac{c_1 F'(t)}{2c_0[1 - F(t)]} \right\}^{1/2}. \tag{9}$$

Thus, the optimal number of inspections between 0 and t is $\int_0^t r(s)\mathrm{d}s$, and the optimal time for the first inspection can be approximated by solving:

$$\int_0^t r(s)\mathrm{d}s = 1, \tag{10}$$

for t. Notice that $\int_0^t r(s)\mathrm{d}s$ is an increasing function of t, since $r(s)$ is positive. An approximate solution for Equation (10) can therefore be obtained by finding the smallest positive integer h, $h \ge 1$, such that $\sum_{i=1}^{h} r(i\Delta)\Delta \ge 1$, for some $\Delta > 0$, and then approximating the optimal first-inspection time by $\min\{T, h\Delta\}$. The quality of the approximation depends on the size of Δ. The smaller Δ is, the more accurate the solution is, but also the more computational effort needed to compute it.

In the MST procedure, let c be the cost of switching in units of the cost of sampling, so that $c =$ (cost of one switch)/(cost of taking one sample), something that can be estimated during the zeroth stage if it is not known. We assume that the switching cost and sampling cost are the same for all systems. Without loss of generality, we let the sampling cost be one and switching cost be c. Then the total cost is also in units of sampling cost.

Consider two systems, i and j. By analogy to the optimal inspection problem, leaving the triangular continuation region corresponds to a failure; a stage corresponds to an inspection, with cost $c_0 = 2c$ since two additional switches will be required when starting the next stage; and sampling beyond the point at which one of the systems has been eliminated corresponds to the loss, with approximate cost $L(\tau) = c_1\tau = 2\tau$ if elimination occurred τ samples before the end of the stage. To complete the analogy, we assume that the $Z_{ij}(t)$ are continuous processes, and use $Z_{ij}(N_s)/N_s$ and S_{ij}^2 to approximate the true mean difference $\mu_{ij} = \mu_i - \mu_j$ and true variance of the difference $\sigma_{ij}^2 = \sigma_i^2 + \sigma_j^2$ for all $i, j = 1, 2, \ldots, k$ and $i \ne j$.

To approximate the probability that $Z_{ij}(N_s + t)$ leaves the triangular region between $[0, t]$, we use the probability that $Z_{ij}(N_s + t)$ is outside the triangular region at time t. We know that $Z_{ij}(N_s + t) - Z_{ij}(N_s)$ has a normal distribution with mean $\mu_{ij}t$ and variance $\sigma_{ij}^2 t$. Therefore, conditional on $Z_{ij}(N_s) = z_{ij}$ and $S_{ij}^2 = s_{ij}^2$, we treat $Z_{ij}(N_s + t)$ as having approximately a normal distribution with mean $z_{ij} + t z_{ij}/N_s$ and variance $t s_{ij}^2$. Let $\Phi_{ij}(x)$ and $\phi_{ij}(x)$ denote the cumulative distribution function and the density, respectively, of this random variable. Then the probability that process $Z_{ij}(N_s + t)$ is outside the triangular continuation region at time t can be written as:

$$F_{ij}(t) = 1 - \{\Phi_{ij}[a_{ij} - \lambda(N_s + t)] - \Phi_{ij}[-a_{ij} + \lambda(N_s + t)]\}.$$

Clearly, $F_{ij}(t)$ is a nondecreasing differentiable function. Let T_{ij} be the end of the triangular region, $T_{ij} = a_{ij}/\lambda$, so that $F_{ij}(0) = 0$ and $F_{ij}(T_{ij}) = 1$. Thus, $F_{ij}(t)$ satisfies the conditions on $F(t)$ in Lemma 4. Taking the derivative of $F_{ij}(t)$ we get:

$$\begin{aligned} F'_{ij}(t) = \frac{1}{2t} \Bigg\{ & \phi_{ij}[a_{ij} - \lambda(N_s + t)] \\ & \times \left(a_{ij} - \lambda(N_s - t) + \frac{N_s - t}{N_s} z_{ij} \right) \\ & + \phi_{ij}[-a_{ij} + \lambda(N_s + t)] \\ & \times \left(a_{ij} - \lambda(N_s - t) - \frac{N_s - t}{N_s} z_{ij} \right) \Bigg\}. \end{aligned}$$

Now we can easily calculate the optimal inspection rate function, $r_{ij}(t)$, using Equation (9). As described above, we can solve $\int_0^t r_{ij}(s) = 1$ by using $\Delta_{ij} = \max\{T_{ij}/m, 1\}$, where m is the maximum number of evaluations of $r_{ij}(s)$ we want to spend. Let t_{ij}^* denote the approximate solution. Since we need an integer solution and also need to consider all systems that are still in contention at the end of stage s, we let $n_{s+1} = \max_{j \in I, j \ne [1]}\{\lceil t_{[1]j}^* \rceil\}$. In the experiments reported in Section 5 we let $m = 50$.

5. Empirical evaluation

In this section we summarize the results of an extensive empirical evaluation of the MSS procedure and MST procedures relative to the R and KN procedures, which are representatives of two-stage procedures and fully sequential procedures, respectively. The systems are represented by various configurations of k normal distributions and, to assess the impact of non-normality, lognormal distributions whose skewness and kurtosis (standardized third and fourth moments) differed from those of the normal distribution. In all cases system k is the true best (has the largest true mean). We evaluated each procedure on different variations of the systems, examining factors including the number of systems, the configuration of means, the configuration of variances and the relative cost of sampling and switching. The configurations, the experiment design and the results are described below.

5.1. *Configurations and experiment design*

Two configurations of the true means were used: the Slippage Configuration (SC) and the Monotone Increasing Means (MIMs) configuration. In SC, μ_k was set to δ and $\mu_1 = \mu_2 = \cdots = \mu_{k-1} = 0$. This is a difficult configuration in terms of procedure performance since all inferior systems are exactly δ from the best. In MIMs $\mu_i = (i-1)\delta$, $i = 1, 2, \ldots, k$. MIMs is used to investigate the effectiveness of the procedures in more favorable settings.

For each configuration of the means, we examined the effect of variances. There are three configurations of variances: Equal Variances (EVs), Increasing Variances (IVs) and Decreasing Variances (DVs). In EVs, $\sigma_1 = \sigma_2 = \cdots = \sigma_k = 1$; in IVs, $\sigma_i = i$; and in DVs, $\sigma_i = k + 1 - i$.

Let c denote the switching cost in units of sampling cost (the cost of generating a sample). We used $c = 1$, $c = 10$, $c = 100$ and $c = 1000$ to represent different cost structures. Considering the current status of computer simulation software, even $c = 1000$ is not excessive.

We also varied the number of systems in each experiment, with $k = 2$, 5 and 10. In all experiments, the nominal Probability of Correct Selection (PCS) was $1 - \alpha = 0.95$ and the first-stage sample size n_0 was 10. The indifference-zone parameter was set to $\delta = \sigma_k/\sqrt{n_0}$, where σ_k is the standard deviation of the best system. Thus, δ is one standard deviation of the first-stage sample mean of the best system. For each configuration, 1000 macroreplications (complete repetitions) of each procedure were carried out to compare the performance measures, including the observed PCS, Average Number of Samples (ANSas.), Average Number of Switches (ANSws) and Average Total Cost (ATC).

5.2. *Summary of results*

The experiments showed that the MST procedure is superior to the other procedures across all of the configurations we examined in terms of the ATC. The difference between the MST procedure and the best of the KN and MSS procedures is significant when $c = 10$. The MST and KN procedures behave similarly when c is low, whereas the MST and MSS procedures behave similarly when c is high. The MSS procedure is very close to the R procedure in the slippage configuration, but has much better performance than the R procedure in the MIMs configuration, which is more typical in practice. Therefore, the MSS procedure can be a good substitute for the widely used R procedure. We also found that the variance configuration does not significantly change the relative performances of all four procedures.

Table 1. Observed PCS of the MST and MSS procedures using Paulson's and Fabian's bounds

Procedure	$k = 2$	$k = 5$	$k = 10$
MSS Paulson	0.987	0.993	0.995
MSS Fabian	0.969	0.987	0.995
MST Paulson	0.974	0.984	0.993
MST Fabian	0.965	0.973	0.986

5.3. *Some specific results*

We do not attempt to present comprehensive results from such a large simulation study. Instead, we present selected results that emphasize the key conclusions.

5.3.1. *Validity check*

In Sections 2 and 3 we proved the validity of the MSS and MST procedures using Paulson's probability bound. However, we also pointed out that Pichitlamken *et al.* (2004) support the use of Fabian's probability bound which is tighter than Paulson's bound. To evaluate how much improvement is possible, we checked the use of both bounds under the slippage configuration, EVs, and $c = 10$ for different k. The results in Table 1 show that Paulson's probability bound is very conservative and that Fabian's probability bound is closer to the nominal PCS. In the rest of the paper, the MST and MSS procedures refer to the MST and MSS procedures using Fabian's probability bound.

5.3.2. *Effect of mean configuration*

We compared the four procedures using the slippage configuration and a MIMs configuration. We used EVs, $c = 10$

Table 2. Effect of mean configuration

	SC			*MIM*		
Procedure	*ANSas*	*ANSws*	*ATC*	*ANSas*	*ANSws*	*ATC*
R	1845.2	20.0	2045.2	1845.2	20.0	2045.2
KN	977.2	887.2	9848.8	426.6	336.6	3792.4
MSS	1950.2	19.9	2149.3	981.7	18.5	1167.0
MST	1185.7	23.8	1424.1	489.9	20.4	694.2

Table 3. Effect of cost structure

			ATC			
Procedure	*ANSas*	*ANSws*	*c = 1*	*c = 10*	*c = 100*	*c = 1000*
R	1845.2	20.0	1865.2	2045.2	3845.2	21 845
KN	426.6	336.6	763.2	3792.4	34084.0	337 000
MSS	981.7	18.5	1000.2	1167.0	2835.0	19 510
MST($c = 1$)	453.6	24.1	477.7			
MST ($c = 10$)	489.7	20.3		693.1		
MST ($c = 100$)	720.8	18.5			2574.8	
MST ($c = 1000$)	793.3	18.5				19 327.3

and $k = 10$ in both configurations. The results in Table 2 show that the R procedure does not depend on the configuration of means while the other procedures require less work when the mean configuration deviates from the slippage configuration. In both configurations KN has the lowest ANSas, MSS has the lowest ANSws and MST has the lowest ATC. The reason why the MSS procedure has a lower ANSws than the R procedure is because some clearly inferior systems can be eliminated based on zeroth-stage samples in the MSS procedure.

5.3.3. *Effect of the cost structure*

To see the effect of the cost structure in usual situations we used the MIMs and EVs conditions for $k = 10$ systems. Table 3 summarizes the results of experiments. ANSas and ANSws do not change for the R, KN and MSS procedures when cost changes, but the ATC does change linearly as a function of c. We observed a different pattern for the MST procedure. As c goes up, ANSas goes up, ANSws goes down and ATC is constantly below the other three procedures. As we expected, the performance of the MST procedure is closer to the KN procedure when c is low, and is closer to the MSS procedure when c is high.

5.3.4. *Robustness study*

To assess the impact of non-normal data on the procedures, they were applied to lognormally-distributed data with increasing levels of skewness and kurtosis, relative to the normal distribution (which has a skewness of zero and a kurtosis of three). The configurations of the means and variances are the same as the Nelson *et al.* (2001). In all cases $n_0 = 10$, $c = 10$, $\sigma_i = 1$ for all i, $\delta = 1/\sqrt{n_0}$, $\mu_1 = \mu_2 = \cdots = \mu_{k-1} = 0$ and $\mu_k = \delta$. We also used different numbers of systems, $k = 2, 5, 10$. In the low-deviation case the skewness and kurtosis are 1.8 and 9.1; in the medium-deviation case they are 4.0 and 41.0; and in the high-deviation case they are 6.2 and 113.2. Notice that the mean configuration is the slippage configuration which is a difficult configuration with respect to achieving the nominal PCS.

Table 4 shows the estimated PCS for the three lognormal cases for the MSS and MST procedures with the corresponding normal case included for comparison. When the level of deviation from the normal distribution increases, the observed PCS decreases. However, the observed PCS is close to the nominal value until the deviation is quite significant.

Table 5. The five alternative inventory policies

Policy (i)	*s*	*S*	*Expected cost*
1	20	40	114.176
2	20	80	112.742
3	40	60	130.550
4	40	100	130.699
5	60	100	147.382

5.3.5. *An illustrative example*

We provide a system simulation example to compare our procedures with the R and KN procedures. Consider the (s, S) inventory system for the $k = 5$ inventory policies provided in Koenig and Law (1985). The objective of the study is to compare the five policies given in Table 5 on the basis of their corresponding expected average cost per month for the first 30 months of operation and select the policy with the minimum expected cost. The expected costs, which can be analytically computed, are also given in Table 5. We set $\delta = 1$, $n_0 = 10$ and $c = 10$. Table 6 includes the results of the simulation study based on 1000 complete macroreplications. The results are consistent with our observations in the previous experiments. Notice that δ is smaller than the true difference between the best and the second-best systems, which explains the high value of the observed PCS.

Table 4. Observed PCS of the MSS and MST procedures in the robustness study with a nominal level of 0.95

	Normal		*Low deviation*		*Medium deviation*		*High deviation*	
	MSS	*MST*	*MSS*	*MST*	*MSS*	*MST*	*MSS*	*MST*
$k = 2$	0.969	0.965	0.970	0.972	0.960	0.946	0.941	0.926
$k = 5$	0.987	0.973	0.980	0.954	0.953	0.915	0.922	0.901
$k = 10$	0.995	0.986	0.979	0.967	0.950	0.932	0.926	0.888

Table 6. Simulation study of the illustrative example

Procedure	*PCS*	*ANSas*	*ANSws*	*ATC*
R	1.000	1033.1	10.0	1133.0
KN	0.998	235.7	190.7	2142.6
MSS	0.999	635.0	7.56	710.7
MST	0.998	268.5	9.42	362.7

6. Conclusions

In this paper we have presented two sequential indifference-zone selection procedures: MSS and MST. As we discussed in Section 5, the MSS procedure has a similar performance (in both ANSa and ANSws) in the slippage configuration, but a much better performance in more typical configurations, than the widely used R procedure. Therefore, it can be a substitute for the R procedure.

MST is a multi-stage procedure that decides the number of stages and the sample size in each stage based on the tradeoff between sampling cost and switching cost. To minimize the total cost of sampling and switching, we provided a simple heuristic algorithm. Although the algorithm is very crude, it works well based on the numerical examples in Section 5. With the heuristic, the MST procedure is uniformly superior to both the minimum switching procedures (e.g., R and MSS) and the fully sequential procedures (e.g., KN) in terms of the ATC value.

Acknowledgements

Portions of this work were previously published as An indifference-zone selection procedure with minimum switching and sequential sampling, in *Proceedings of the 2003 Winter Simulation Conference*. Chick, S., Sánchez, P. J., Ferrin, D. and Morrice, D. J. (eds.), pp. 474–480. Research for this paper was partially supported by National Science Foundation under grant DMI-0217690. The authors thank Juta Pichitlamken, the Department Editor and the referees for their feedback.

References

Bechhofer, R.E., Santner, T.J. and Goldsman, D.M. (1995) *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons*, Wiley, New York, NY.

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

Chick, S.E. and Inoue, K. (2001) New two-stage and sequential procedures for selecting the best simulated system. *Operations Research*, **49**, 732–743.

Fabian, V. (1974) Note on Anderson's sequential procedures with triangular boundary. *Annals of Statistics*, **2**, 170–176.

Hall, W.J. (1997) The distribution of Brownian motion on linear stopping boundaries. *Sequential Analysis*, **16**, 345–352.

Hartmann, M. (1988) An improvement on Paulson's sequential ranking procedure. *Sequential Analysis*, **7**, 363–372.

Hartmann, M. (1991) An improvement on Paulson's procedure for selecting the population with the largest mean from k normal populations with a common unknown variance. *Sequential Analysis*, **10**, 1–16.

Hong, L.J. and Nelson, B.L. (2005) Discrete optimization via simulation using COMPASS. *Operations Research* (forthcoming).

Jennison, C., Johnston, I.M. and Turnbull, B.W. (1982) Asymptotically optimal procedures for sequential adaptive selection of the best of several normal means, in *Statistical Decision Theory and Related Topics III*. Gupta, S. S. and Berger, J. (eds.), Academic Press, New York, NY, pp. 55–86.

Kamien, M.I. and Schwartz, N.L. (1981) *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management*, North-Holland, New York, NY.

Kim, S.-H. and Nelson, B.L. (2001) A fully sequential procedure for indifference-zone selection in simulation. *ACM Transactions on Modeling and Computer Simulation*, **11**, 251–273.

Kim, S.-H. and Nelson, B.L. (2003) On the asymptotic validity of fully sequential selection procedures for steady-state simulation. Working paper, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, 60208.

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

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

Paulson, E. (1964) A sequential procedure for selecting the population with the largest mean from k normal populations. *Annals of Mathematical Statistics*, **35**, 174–180.

Pichitlamken, J. and Nelson, B.L. (2003) A combined procedure for optimization via simulation. *ACM Transactions on Modeling and Computer Simulation*, **13**, 155–179.

Pichitlamken, J., Nelson, B.L. and Hong, L.J. (2005) A selection-of-the-best procedure for optimization via simulation. *European Journal of Operational Research* (forthcoming).

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

Tamhane, A.C. (1977) Multiple comparisons in model I: one-way ANOVA with unequal variances. *Communications in Statistics*, **A6**, 15–32.

Tong, Y.L. (1980) *Probability Inequalities in Multivariate Distributions*, Academic Press, New York. NY.

Biographies

L. Jeff Hong is an Assistant Professor in the Department of Industrial Engineering & Engineering Management at the Hong Kong University of Science and Technology. He received his B.E. in Industrial Engineering and Automotive Engineering from Tsinghua University, China, his M.S. in Applied Mathematics from the University of Cincinnati, and his Ph.D. in Industrial Engineering from Northwestern University. His

research interests include ranking and selection and optimization via simulation.

Barry L. Nelson is the Krebs Professor of Industrial Engineering and Management Sciences at Northwestern University, and is Director of the Master of Engineering Management Program. His research focuses on the design and analysis of computer simulation experiments for stochastic systems. He has published numerous papers and two books. He has served as the Simulation Area Editor of *Operations Research* and as President of the INFORMS (then TIMS) College on Simulation. He has held many positions for the Winter Simulation Conference, including Program Chair in 1997 and is currently a member of the Board of Directors.

Contributed by the Simulation Department