Fully Sequential Indifference-Zone Selection Procedures with Variance-Dependent Sampling

L. Jeff Hong

Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China

Received 5 August 2005; revised 25 February 2006; accepted 27 February 2006
DOI 10.1002/nav.20155
Published online 5 April 2006 in Wiley InterScience (www.interscience.wiley.com).

Abstract: Fully sequential indifference-zone selection procedures have been proposed in the simulation literature to select the system with the best mean performance from a group of simulated systems. However, the existing sequential indifference-zone procedures allocate an equal number of samples to the two systems in comparison even if their variances are drastically different. In this paper we propose new fully sequential indifference-zone procedures that allocate samples according to the variances. We show that the procedures work better than several existing sequential indifference-zone procedures when variances of the systems are different. © 2006 Wiley Periodicals, Inc. Naval Research Logistics 53: 464–476, 2006

Keywords: ranking and selection; indifference-zone selection; sequential analysis; Brownian motion

1. INTRODUCTION

Selecting the best system is a common problem in computer simulation, where the best system is defined as the system with the largest or smallest mean performance. There are two general approaches to solve this problem: a frequentist approach and a Bayesian approach. The frequentist approach (see Bechhofer, Santner, and Goldsman [3] for a summary) guarantees that the selected system is the best system with a pre-specified probability of correct selection (PCS). To achieve this objective Bechhofer [2] suggested an indifference-zone formulation: guarantee to select the single best system whenever the difference between the mean of the best system and the mean of the second-best system is greater than or equal to δ, where $\delta > 0$ is the smallest difference the experimenter feels worth detecting. Indifference-zone selection procedures, e.g., the two-stage procedure of Rinott [24] and the fully sequential procedure of Kim and Nelson [18,19], allocate samples to different systems to guarantee the PCS even for the slippage configuration (SC), where the differences between the best system and all other systems are all δ. They are typically conservative for average cases in which the differences between the best systems and some other systems are greater than δ. Bayesian procedures, e.g., Chen et al. [5] and Chick and Inoue [7,8], attempt to allocate a finite computation budget to maximize the posterior PCS. They typically allocate different numbers of samples to different systems based on the sample means and sample variances, and the allocation can be done in stages or sequentially. They do not provide a guaranteed (frequentist's) PCS, except for Chen and Yücesan [6], who guarantee a posterior PCS, but they work well for average cases. In this paper we take the frequentist viewpoint. The procedures provided in this paper guarantee a pre-specified PCS under the indifference-zone formulation.

Rinott's procedure [24] and the KN procedure of Kim and Nelson [18] are representative procedures for indifference-zone selection. Rinott's procedure has two stages. In the first stage it takes an initial number of samples from all systems, calculates the sample variances, and determines the second-stage sample sizes. In the second stage it takes the required number of samples from all systems and calculates the overall sample means. Suppose that we want to select the system with the largest mean. Then the system with the largest overall sample mean is selected as the best system. KN procedure is fully sequential. It approximates the sum of differences between two systems as a Brownian motion process with drift and uses a triangular continuation region to determine the stopping time of the selection process. Figure 1 illustrates how a triangular continuation region works. KN procedure also needs a first stage to estimate variances and to determine the corresponding triangular region for each pair of systems. It then keeps allocating one sample to each surviving system and checking whether any system(s) can be eliminated.

Correspondence to: L. J. Hong (hongl@ust.hk)

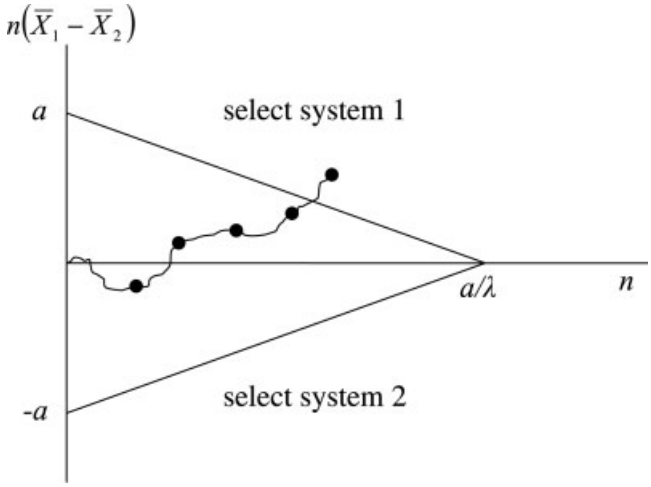

Figure 1. Triangular continuation region for sequential indifference-zone procedures. Selection of system 1 or 2 as best depends on whether the sum of differences exits the region from the top or bottom.

It stops when there is only one system left, and that system is selected as the best.

While both procedures provide a guaranteed PCS, KN often requires a smaller sample size to select the best system than Rinott's procedure when the systems are not in the SC. This is because KN adapts to the difference between the two systems in comparison. When the difference between the two systems is larger than δ, the Brownian motion process constructed in KN has a drift larger than δ, and it tends to exit the triangular continuation region earlier. Then the procedure requires fewer samples than Rinott's procedure to eliminate the inferior system. Rinott's procedure, however, does not adapt to the difference between systems. It requires the same number of samples even if the systems are not in the SC.

KN procedure also has disadvantages when compared to Rinott's procedure. For instance, KN procedure does not adapt to variances of individual systems. Instead, it adjusts to the variance of the difference between each pair of systems. When variances of the two systems in comparison are significantly different, Rinott's procedure allocates more samples to the system with larger variance and fewer samples to the system with smaller variance, but KN allocates an equal number of samples to both systems. In an extreme case where one system has a zero variance and the other has a large variance, Rinott's procedure allocates no second-stage samples to the system with zero variance, but KN may allocate a large number of samples to it after the first stage.

It is common in simulation study to have systems with different variances. Simulation is often used to study queueing systems, e.g., production facilities, communication networks, or service systems, where mean and variance often have a functional relationship. For example, the steady-state mean and variance of the number of customers in an M/M/1 queueing system are $\rho/(1-\rho)$ and $\rho/(1-\rho)^2$, respectively, where ρ is the traffic intensity (Kulkarni [20]). Therefore, a larger mean also implies a larger variance.

In this paper we first propose a new approach, in Section 2, to construct a Brownian motion process with drift that allows different sample sizes and different variances for different systems. We show that the Brownian motion process constructed in KN procedure is a special case of our construction. In Section 3 we introduce a fully sequential procedure assuming all variances are known in advance and use the procedure to illustrate how to control the sampling using variance information. In Section 4 we introduce a fully sequential procedure that allows for unknown variances. Results of a comprehensive empirical study are summarized and reported in Section 5, followed by the conclusions in Section 6.

2. CONSTRUCTION OF BROWNIAN MOTION

Suppose that there are two systems, systems 1 and 2, each normally distributed with mean μ_i and variance σ_i^2, $i = 1, 2$. Let $X_{i\ell}$ denote the ℓth sample from system i and we require that $X_{i\ell}, i = 1, 2$ and $\ell = 1, 2, \ldots$, are mutually independent. We also let $\overline{X}_i(n) = \sum_{\ell=1}^{n} X_{i\ell}/n$ for $i = 1, 2$ and $n = 1, 2, \ldots$.

Let $\mathcal{B}_\Delta(t)$ denote the Brownian motion process with drift Δ. When comparing the means of two systems, there are two approaches to approximate the Brownian motion process in the sequential-analysis literature: using equal sample sizes but allowing unequal variances for both systems (as in Kim and Nelson [18]) or assuming unit variance but allow unequal sample sizes for both systems (as in Robbins and Siegmund [25]). The first approach uses the fact that, for any non-decreasing sequence n of positive integers, the random sequences $n[\overline{X}_1(n) - \overline{X}_2(n)]/\sqrt{\sigma_1^2 + \sigma_2^2}$ and $\mathcal{B}_{(\mu_1-\mu_2)/\sqrt{\sigma_1^2+\sigma_2^2}}(n)$ have the same joint distribution. The second approach shows that, for any sequence of pairs (m, n) of positive integers that is non-decreasing in each coordinate, the random sequences $mn/(m+n)[\overline{X}_1(m) - \overline{X}_2(n)]$ and $\mathcal{B}_{\mu_1-\mu_2}(mn/(m+n))$ have the same joint distribution if $\sigma_1^2 = \sigma_2^2 = 1$.

In this section we generalize these two approaches to allow both unequal sample sizes and unequal variances. Let

$$Z(m,n) = \left[\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right]^{-1} \left[\overline{X}_1(m) - \overline{X}_2(n)\right].$$

The next theorem shows that the sequences $Z(m,n)$ and $\mathcal{B}_{\mu_1-\mu_2}([\sigma_1^2/m + \sigma_2^2/n]^{-1})$ have the same joint distribution. Note that the approaches of Kim and Nelson [18] and Robbins and Siegmund [25] are both special cases of the theorem.

THEOREM 1: For any sequence of pairs (m,n) of positive integers that is non-decreasing in each coordinate, the random sequences $Z(m,n)$ and $\mathcal{B}_{\mu_1-\mu_2}([\sigma_1^2/m+\sigma_2^2/n]^{-1})$ have the same joint distribution.

PROOF: For any finite sequence of (m,n), since $Z(m,n)$'s are linear functions of jointly normal variables, they are jointly normally distributed. Moreover,

$$\mathrm{E}[Z(m,n)] = [\sigma_1^2/m+\sigma_2^2/n]^{-1}(\mu_1-\mu_2),$$
$$\mathrm{Var}[Z(m,n)] = [\sigma_1^2/m+\sigma_2^2/n]^{-1},$$

and for any non-negative integers p and q,

$$\begin{aligned}\mathrm{Cov}[Z(m+p,n+q),Z(m,n)] &= [\sigma_1^2/(m+p)+\sigma_2^2/(n+q)]^{-1}[\sigma_1^2/m+\sigma_2^2/n]^{-1}\mathrm{Cov}[\overline{X}_1(m+p),\overline{X}_1(m)]\\ &\quad+[\sigma_1^2/(m+p)+\sigma_2^2/(n+q)]^{-1}[\sigma_1^2/m+\sigma_2^2/n]^{-1}\mathrm{Cov}[\overline{X}_2(n+q),\overline{X}_2(n)]\\ &= [\sigma_1^2/(m+p)+\sigma_2^2/(n+q)]^{-1}[\sigma_1^2/m+\sigma_2^2/n]^{-1}[\sigma_1^2/(m+p)+\sigma_2^2/(n+q)]\\ &= [\sigma_1^2/m+\sigma_2^2/n]^{-1}.\end{aligned}$$

Therefore, for any finite sequence of (m,n), $Z(m,n)$ and $\mathcal{B}_{\mu_1-\mu_2}([\sigma_1^2/m+\sigma_2^2/n]^{-1})$ have the same joint distribution. By the Kolmogorov's Extension Theorem (Durrett [9]), the conclusion also holds for any infinite sequence of (m,n). □

Theorem 1 gives a general framework that allows unequal sample sizes and unequal variances for both systems. We can combine it with the triangular continuation region to design indifference-zone selection procedures that allow variance-dependent sampling.

3. KNOWN-VARIANCES PROCEDURE

Suppose there are k systems each normally distributed with unknown mean μ_i and known variance σ_i^2, $i=1,2,\ldots,k$. In indifference-zone selection we design an experiment to select a system and to guarantee that the selected system has the largest mean among all k systems with a probability at least $1-\alpha$ if the difference between means of the best and the second best systems is greater than or equal to the indifference-zone parameter δ. In this section we assume that σ_i^2, $i=1,2,\ldots,k$, are known, but can be unequal. We assume known variances to illustrate the essence and benefit of variance-dependent sampling in fully sequential selection procedures. The case of unknown variances is discussed in Section 4.

3.1. The Procedure

In this subsection we introduce a fully sequential procedure for the known-variances case. The procedure works as follows: We first design a sampling rule and then samples are taken one at a time according to the rule. After taking each sample, the procedure checks whether any of the remaining systems can be eliminated. It stops when there is only one system left and the system is selected as the best system. The sampling rule defines clearly how to allocate each sample to the k systems, and it may depend *only* on the variances of the systems. The selection of sampling rule is discussed in Section 3.3.

Let $X_{i\ell}$ denote the ℓth sample from system i. We require that $X_{i\ell}$, $i=1,2,\ldots,k$ and $\ell=1,2,\ldots,$ are mutually independent. Independence of $X_{i\ell}$, $\ell=1,2,\ldots$ is a direct result of making replications. When samples are obtained within a single replication of a steady-state simulation then techniques such as batch means allow the independence to hold approximately (see, for instance, Law and Kelton [21]). Independence between $X_{i\ell}$ and of $X_{j\ell}$ implies that we do not use common random numbers. Although we expect our procedures to work, in the sense of delivering at least the desired PCS, in the presence of common random numbers, they do not exploit them.

Known-Variances Procedure (KVP)

Setup. Select confidence level $1/k<1-\alpha<1$ and indifference-zone parameter $\delta>0$. Let $\lambda=\delta/2$ and

$$a=-\frac{1}{\delta}\ln\left[2-2(1-\alpha)^{\frac{1}{k-1}}\right]. \tag{1}$$

Note that λ and a define the triangular continuation region (see Figure 1).

Initialization. Let $I=\{1,2,\ldots,k\}$ be the set of systems still in contention. Let r be the counter of the total number of samples allocated to all k systems, and let $n_i(r)$ be the number of samples allocated to system i, $i=1,2,\ldots,k$. Note that $\sum_{i=1}^k n_i(r)=r$. Set $r=0$ and $n_1(r)=n_2(r)=\cdots=n_k(r)=0$.

Determining sampling rule. Select a sampling rule that may depend *only* on $\sigma_1^2,\sigma_2^2,\ldots,\sigma_k^2$. A suggested sampling rule, Sampling Rule 1, can be found in Section 3.3.

Screening. Let $r = r + 1$. Take the rth sample according to the sampling rule, and update $n_i(r)$ for all $i \in I$. For all $i, j \in I$, and $i \neq j$, if $n_i(r) \geq 1$ and $n_j(r) \geq 1$, let

$$t_{ij}(r) = \left[\frac{\sigma_i^2}{n_i(r)} + \frac{\sigma_j^2}{n_j(r)}\right]^{-1} \text{ and} \quad (2)$$

$$Z_{ij}(t_{ij}(r)) = t_{ij}(r)\left[\overline{X}_i(n_i(r)) - \overline{X}_j(n_j(r))\right]; \quad (3)$$

otherwise, let $t_{ij}(r) = 0$ and $Z_{ij}(t_{ij}(r)) = 0$. Let $I^{\text{old}} = I$ and let

$$I = I^{\text{old}} \backslash \left\{i \in I^{\text{old}} : Z_{ij}(t_{ij}(r)) < \min[0, -a + \lambda t_{ij}(r)] \text{ for some } j \in I^{\text{old}} \text{ and } j \neq i\right\}, \quad (4)$$

where $A \backslash B = \{x : x \in A \text{ and } x \notin B\}$.

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

REMARKS:

1. Any sampling rule that depends only on variances of the systems can be used in KVP. However, the sampling rule may not depend on sample means of the systems. Therefore, Bayesian sampling rules that exploits the information on sample means, e.g., the OCBA rule of Chen et al. [5] and the LL rule of Chick and Inoue [7], may not be used in KVP.
2. In the *Screening* step, we use Eq. (4) to construct I. However, if the rth sample is allocated to system i, only the comparisons involving system i need to be checked to see whether any system(s) can be eliminated. The comparisons not involving system i will not eliminate any systems, since $t_{pq}(r) = t_{pq}(r-1)$ and $Z_{pq}(r) = Z_{pq}(r-1)$ for all $p, q \in I$ and $p, q \neq i$.
3. The procedure checks whether any system can be eliminated after every sample is taken. If the computational cost of checking is relatively high compared to the cost of taking samples, the procedure can also check after a batch of samples is taken. In practical problems, however, checking cost is often negligible compared to sampling cost.

The idea of forming a sequential test statistic that is equivalent to examining a Brownian motion process at a variable time index $t_{ij}(r)$ is not new. Robbins and Siegmund [25] use it to study the comparison between two systems. Jennison, Johnston, and Turnbull [15] and Jennison and Turnbull [16] use it to study the selection among k systems. However, they all assume equal variances.

3.2. Statistical Validity

Let Λ denote the triangular continuation region formed by $a - \lambda t$ and $-a + \lambda t$, where $a > 0$, $\lambda > 0$ and $t \geq 0$ (see Figure 1). Let $T = \inf\{t > 0 : \mathcal{B}_\Delta(t) \notin \Lambda\}$, T is the random time that $\mathcal{B}_\Delta(t)$ first exits Λ. Then Lemma 1 gives the probability of $\mathcal{B}_\Delta(t)$ exiting Λ from the side that corresponds to an incorrect selection.

LEMMA 1 (Fabian [10]): For a fixed triangular continuation region Λ, if $\lambda = \Delta/2$ and $\Delta > 0$, then

$$\Pr\{\mathcal{B}_\Delta(T) < 0\} = \frac{1}{2}e^{-a\Delta}.$$

If we can only observe the Brownian motion process at a set of discrete times $t_1, t_2, \ldots$, then Lemma 1 cannot be applied directly. The next lemma shows that, under general conditions, the probability of incorrect selection (PICS) of the continuous-time process is an upper bound for the PICS of the discrete-time process.

Let Π be a symmetric continuation region formed by $g(t) \geq 0$ and $-g(t)$. For instance, if we let

$$g(t) = \begin{cases} a - \lambda t & : \text{ if } 0 \leq t \leq a/\lambda \\ 0 & : \text{ if } t > a/\lambda \end{cases},$$

then Π becomes Λ. Let $T_c = \inf\{t : \mathcal{B}_\Delta(t) \notin \Pi\}$ and $T_d = \inf\{t_i : \mathcal{B}_\Delta(t_i) \notin \Pi\}$. T_c is the time that $\mathcal{B}_\Delta(t)$ first exits Π and T_d is the time that $\mathcal{B}_\Delta(t)$ first exits Π at an observed time t_i, $i = 1, 2, \ldots$. Then we have the following lemma.

LEMMA 2 (Jennison, Johnstone, and Turnbull [14]): A discrete-time process is obtained by observing $\mathcal{B}_\Delta(t)$ at an increasing sequence of times $\{t_i; i = 1, 2, \ldots\}$ taking values in a given countable set. The value of t_i depends on $\mathcal{B}_\Delta(t)$ only through its values in the period $[0, t_{i-1}]$. If $T_d < \infty$ almost surely and the conditional distribution of $\{t_i\}$ given $\mathcal{B}_\Delta(t) = b$ is the same as that given $\mathcal{B}_\Delta(t) = -b$, then

$$\Pr\{\mathcal{B}_\Delta(T_d) < 0\} \leq \Pr\{\mathcal{B}_\Delta(T_c) < 0\}.$$

By Theorem 1 we know that the sequence $Z_{ij}(t_{ij}(r))$, $r = 1, 2, \ldots$, behaves like a drifted Brownian motion process $\mathcal{B}_{\mu_i - \mu_j}(t)$ observed at a discrete sequence of times $t_{ij}(r), r = 1, 2, \ldots$. Since $t_{ij}(r)$ does not depend on $\mathcal{B}_{\mu_i - \mu_j}(t)$ and T_d is finite, the conditions of Lemma 2 are satisfied. Then the lemma shows that the PICS caused by $Z_{ij}(t_{ij}(r))$ exiting the incorrect side of Λ is bounded above by the probability of $\mathcal{B}_{\mu_i - \mu_j}(t)$ exiting the incorrect side of Λ.

We also need the following lemma to allocate the PICS to all pairs of comparisons when $k > 2$.

LEMMA 3 (Tamhane [27]): 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

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

Now we are in a position to state and prove the statistical validity of KVP. 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 2: 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 $i \neq j$ and any positive integers p and q. Then the Known-Variances Procedure selects system k with probability at least $1 - \alpha$ whenever $\mu_k - \mu_{k-1} \geq \delta$.

PROOF: Let Λ denote the triangular continuation region formed by a and λ. We define $T_{ij}^{(1)}$, T_{ij} and T be the first times that $Z_{ij}(t_{ij}(r))$, $\mathcal{B}_{\mu_i - \mu_j}(t)$ and $\mathcal{B}_{\delta}(t)$ exit Λ, respectively. Then for any $i = 1, 2, \ldots, k-1$,

$$\begin{aligned}
\Pr\{\text{system } i \text{ eliminates system } k\} &= \Pr\left\{Z_{ki}\left(T_{ki}^{(1)}\right) < 0\right\} \\
&\leq \Pr\left\{\mathcal{B}_{\mu_k - \mu_i}(T_{ki}) < 0\right\} \qquad \text{by Theorem 1, Lemma 2, and } \mu_k - \mu_i > 0 \\
&\leq \Pr\{\mathcal{B}_{\delta}(T) < 0\} \quad \text{since } \mu_k - \mu_i \geq \delta = \frac{1}{2}e^{-a\delta} \quad \text{by Lemma 1} = 1 - (1-\alpha)^{\frac{1}{k-1}}.
\end{aligned}$$

Then

$$\begin{aligned}
&\Pr\{\text{selecting system } k\} \\
&\quad = \Pr\left\{\bigcap_{i=1}^{k-1}\{\text{system } k \text{ eliminates system } i\}\right\} \\
&\quad = \mathrm{E}\left[\Pr\left\{\bigcap_{i=1}^{k-1}\{\text{system } k \text{ eliminates system } i\} \,\middle|\, X_{k1}, X_{k2}, \ldots\right\}\right]. \qquad (5)
\end{aligned}$$

Conditioned on $X_{k1}, X_{k2}, \ldots$, the events that system k eliminates system i are mutually independent for all $i = 1, 2, \ldots, k-1$, and the probability that system k eliminates system i is nondecreasing in $X_{k1}, X_{k2}, \ldots$. Therefore,

$$\begin{aligned}
\text{Eq. (5)} &= \mathrm{E}\left[\prod_{i=1}^{k-1} \Pr\{\text{system } k \text{ eliminates system } i | X_{k1}, X_{k2}, \ldots\}\right] \qquad \text{by independence of the events} \\
&\geq \prod_{i=1}^{k-1} \mathrm{E}\left[\Pr\{\text{system } k \text{ eliminates system } i | X_{k1}, X_{k2}, \ldots\}\right] \qquad \text{by Lemma 3} \\
&= \prod_{i=1}^{k-1} \Pr\{\text{system } k \text{ eliminates system } i\} \geq \prod_{i=1}^{k-1}\left[1 - \left(1 - (1-\alpha)^{\frac{1}{k-1}}\right)\right] = 1 - \alpha.
\end{aligned}$$

This concludes the proof. □

When there are systems whose means are within δ to the mean of the best system, KVP guarantees that one of the good systems, either the best system or a system that is within δ to the best system, will be selected with a probability at least $1 - \alpha$. This result follows directly from Corollary 1 of Kim and Nelson [18] and Theorem 2.

3.3. Selection of Sampling Rule

Suppose there are only two systems 1 and 2 and we want to select the system with the larger mean. Then the sampling process stops if

$$t_{12}(r) = T_{12}^{(1)}, \tag{6}$$

where $T_{12}^{(1)}$ is the first time $Z_{12}(t_{12}(r)), r = 1, 2, \ldots$, exits the triangular region. To reduce the total sample size required to select the better system from systems 1 and 2 by the sampling rule, we want to make Eq. (6) easier to satisfy. Therefore, for a given $r > 0$, either increasing the left-hand side or decreasing the right-hand side of Eq. (6) may do it.

The right-hand side of Eq. (6) can be approximated by T_{12}, the first time $\mathcal{B}_{\mu_1-\mu_2}(t), t > 0$, exits the triangular region. Since the distribution of T_{12} is not affected by the sampling rule, we cannot reduce the right-hand side of Eq. (6). However, increasing the left-hand side of Eq. (6) can be achieved by maximizing $t_{12}(r)$ for any $r = 1, 2, \ldots$. By the definition of $t_{12}(r)$ in Eq. (2), we can formulate the problem as

$$\begin{aligned} \max \quad & \left[\frac{\sigma_1^2}{n_1(r)} + \frac{\sigma_2^2}{n_2(r)}\right]^{-1} \\ \text{s.t.} \quad & n_1(r) + n_2(r) = r, \\ & n_1(r) \text{ and } n_2(r) \text{ are non-negative integers.} \end{aligned} \tag{7}$$

If we relax the integrality constraints on $n_1(r)$ and $n_2(r)$, it is easy to show that the optimal solution to Problem (7) satisfies

$$\frac{n_1(r)}{\sigma_1} = \frac{n_2(r)}{\sigma_2}, \tag{8}$$

which suggests that the optimal sampling rule allocates samples proportionally to the standard deviations of the systems in comparison. Note that the OCBA rules, e.g., Chen et al. [5] and Fu et al. [11], also satisfy Eq. (8) when there are only two systems.

Equation (8) is an interesting result. It minimizes the variance of $\overline{X}_1(n_1(r)) - \overline{X}_2(n_2(r))$ given a fixed r and maximizes $\Pr\{\overline{X}_1(n_1(r)) \geq \overline{X}_2(n_2(r))\}$ given a fixed r if $\mu_1 > \mu_2$. However, it is different from Rinott's procedure, which allocates samples proportionally to the (estimated) variances instead of the standard deviations.

When there are $k > 2$ systems in comparison, the optimal sampling rule depends on the means of the systems (see, for instance, Chen et al. [5]). Since the means of the systems are not known in advance and the sample means cannot be used, we suggest using the sampling rules that satisfy

$$\frac{n_1(r)}{\sigma_1} = \frac{n_2(r)}{\sigma_2} = \cdots = \frac{n_k(r)}{\sigma_k}. \tag{9}$$

By Eq. (8), we know that Eq. (9) guarantees optimality for the comparisons between any pair of systems.

When samples are allocated one at a time, Eq. (9) cannot be satisfied by all r, $r = 1, 2, \ldots$. We use the following sampling rule.

SAMPLING RULE 1: Allocate the first sample to the system (or any system if there are more than one) with the lowest σ_i. Then allocate the next sample, the $(r+1)$st sample, $r = 1, 2, \ldots$, to the surviving system with the lowest $n_i(r)/\sigma_i$. If there are systems with the same $n_i(r)/\sigma_i$, allocate to the system (or any system if there are more than one) with the lowest σ_i among them.

Sampling Rule 1 approximates Eq. (9) to let all systems have similar $n_i(r)/\sigma_i$. When $\sigma_1 = \sigma_2 = \cdots = \sigma_k$, Sampling Rule 1 becomes the sampling rule used in KN. When the variances of the k systems are significantly different, however, Sampling Rule 1 can be significantly better than the KN rule. For instance, if there are two systems with $\sigma_1 = 1$ and $\sigma_2 = 10$ and we let $r = 22$, then KN rule gives $t_{12}(r) = 0.109$, which is significantly smaller than $t_{12}(r) = 0.198$ of Sampling Rule 1.

4. UNKNOWN-VARIANCES PROCEDURE

In most simulation studies the variances of the systems are not known in advance. Therefore, indifference-zone selection procedures used in simulation typically have two or more stages, with the first stage providing variance estimates that help determine the sampling rule in the subsequent stages. If the sampling rule determined at the end of the first stage depends only on the first-stage sample variances, Stein [26] showed that the overall sample means are independent of the first-stage sample variances. In this section we also exploit this property and design a fully sequential procedure that allows unknown variances and variance-dependent sampling and also guarantees the PCS.

4.1. The Procedure

In this subsection we introduce a fully sequential procedure, called the Unknown Variance Procedure (UVP). UVP is similar to KVP except that it requires a first stage to estimate the variances of the systems, since the variances are not known in advance. Once the variances are estimated, the procedure determines a sampling rule and samples according to the rule. The elimination decisions of UVP are made similar to those of KVP. The sampling rule of UVP may depend on the first-stage sample variances. However, it may not depend on the first-stage sample means. Otherwise, the overall sample means are no longer independent of the sampling rule.

Unknown-Variances Procedure (UVP)

Setup. Select confidence level $1/k < 1-\alpha < 1$, indifference-zone parameter $\delta > 0$, and first-stage sample size $n_0 \geq 2$. Let $\lambda = \delta/2$ and a be the solution of the equation

$$\mathrm{E}\left[\frac{1}{2}\exp\left(-\frac{a\delta}{n_0-1}\Psi\right)\right] = 1-(1-\alpha)^{\frac{1}{k-1}}, \quad (10)$$

where Ψ is a random variable whose density function is

$$f_\Psi(x) = 2\left[1-F_{\chi^2_{n_0-1}}(x)\right]f_{\chi^2_{n_0-1}}(x),$$

and $F_{\chi^2_{n_0-1}}(x)$ and $f_{\chi^2_{n_0-1}}(x)$ are the distribution function and density function of a χ^2 distribution with n_0-1 degrees of freedom.

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

$$S_i^2 = \frac{1}{n_0-1}\sum_{\ell=1}^{n_0}\left(X_{i\ell}-\overline{X}_i(n_0)\right)^2, \quad (11)$$

the first-stage sample variance of systems i. Let $I = \{1,2,\ldots,k\}$ be the set of systems still in contention. Let r be the counter of the total number of samples allocated to all systems and $n_i(r)$ be the number of samples allocated to system i, $i = 1,2,\ldots,k$. Set $r = kn_0$, and $n_1(r) = n_2(r) = \cdots = n_k(r) = n_0$.

Determining sampling rule. Select a sampling rule that may depend *only* on $S_1^2, S_2^2, \ldots, S_k^2$. A suggested sampling rule, Sampling Rule 2, is given at the end of this subsection.

Screening. For all $i, j \in I$ and $i \neq j$, let

$$\tau_{ij}(r) = \left[\frac{S_i^2}{n_i(r)}+\frac{S_j^2}{n_j(r)}\right]^{-1} \text{ and } \quad (12)$$

$$Y_{ij}(\tau_{ij}(r)) = \tau_{ij}(r)\left[\overline{X}_i(n_i(r))-\overline{X}_j(n_j(r))\right].$$

Let $I^{\mathrm{old}} = I$ and let

$$I = I^{\mathrm{old}}\Big\backslash\left\{i \in I^{\mathrm{old}} : Y_{ij}(\tau_{ij}(r)) < \min[0, -a+\lambda\tau_{ij}(r)] \right.$$
$$\left. \text{for some } j \in I^{\mathrm{old}} \text{ and } j \neq i\right\}. \quad (13)$$

Stopping rule. If $|I| = 1$, then stop and select the system whose index is in I as the best. Otherwise, let $r = r+1$. Take the rth sample according to the sampling rule, and update $n_i(r)$ for all $i \in I$. Then go to *Screening*.

Comparing the definitions of $t_{ij}(r)$ in Eq. (2) and $\tau_{ij}(r)$ in Eq. (12), we suggest using the following sampling rule.

SAMPLING RULE 2: After the first stage, allocate the next sample, the $(r+1)$st sample, $r = kn_0, kn_0+1, \ldots$, to the surviving system with the lowest $n_i(r)/S_i$. If there are systems with the same $n_i(r)/S_i$, allocate to the system (or any system if there are more that one) with the lowest S_i among them.

4.2. Statistical Validity

Let Π_1 and Π_2 be two symmetric continuation regions. Π_1 is formed by $g_1(t) \geq 0$ and $-g_1(t)$, Π_2 is formed by $g_2(t) \geq 0$ and $-g_2(t)$, and $g_2(t) \geq g_1(t)$ for all $t \geq 0$ (see Figure 2 for an example). Let $T_1 = \inf\{t : \mathcal{B}_\Delta(t) \notin \Pi_1\}$ and $T_2 = \inf\{t : \mathcal{B}_\Delta(t) \notin \Pi_2\}$. Note that to exit Π_2, $\mathcal{B}_\Delta(t)$ must first exit Π_1. Therefore, for each realization of $\mathcal{B}_\Delta(t)$, $T_1 \leq T_2$. Then we have the following lemma.

LEMMA 4: If $\Delta > 0$, then $\Pr\{\mathcal{B}_\Delta(T_1) < 0\} \geq \Pr\{\mathcal{B}_\Delta(T_2) < 0\}$.

PROOF: Let $P_\Delta(\omega)$ denote the probability distribution on the space with elements $\omega = \{b(t); 0 \leq t \leq T_2\}$, where $b(t)$ is the realization of a Brownian motion $\mathcal{B}_\Delta(t)$. Then by Jennison, Johnston, and Turnbull [14],

$$\frac{dP_\Delta}{dP_{-\Delta}}(\omega) = e^{2\Delta b(T_2)},$$

where the derivative is taken relative to the σ-field $\sigma\{b(t); 0 \leq t \leq T_2\}$. Then

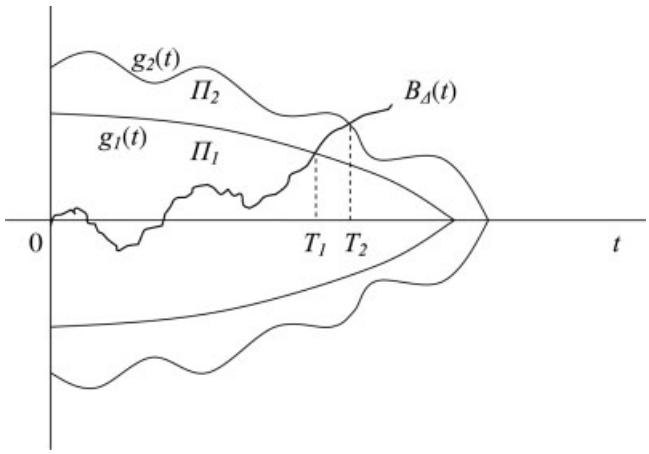

Figure 2. Example of two symmetric continuation regions where $g_1(t) \leq g_2(t)$ for all $t \geq 0$.

$$\begin{aligned}\Pr\{\mathcal{B}_\Delta(T_1) < 0\} - \Pr\{\mathcal{B}_\Delta(T_2) < 0\} &= \Pr\{\omega : b(T_1) < 0\} - \Pr\{\omega : b(T_2) < 0\} \\ &= \left[\Pr\{\omega : b(T_1) < 0, b(T_2) < 0\} + \Pr\{\omega : b(T_1) < 0, b(T_2) > 0\}\right] \\ &\quad - \left[\Pr\{\omega : b(T_1) < 0, b(T_2) < 0\} + \Pr\{\omega : b(T_1) > 0, b(T_2) < 0\}\right] \\ &= \Pr\{\omega : b(T_1) < 0, b(T_2) > 0\} - \Pr\{\omega : b(T_1) > 0, b(T_2) < 0\} = \int_{\Omega_1} dP_\Delta(\omega) - \int_{\Omega_2} dP_\Delta(\omega),\end{aligned}$$

where we define $\Omega_1 = \{\omega : b(T_1) < 0, b(T_2) > 0\}$ and $\Omega_2 = \{\omega : b(T_1) > 0, b(T_2) < 0\}$. Note that Ω_2 can be obtained from Ω_1 by replacing $b(t), t \geq 0$ by $-b(t), t \geq 0$. By symmetry of Π_1 and Π_2,

$$\begin{aligned}\Pr\{\mathcal{B}_\Delta(T_1) < 0\} - \Pr\{\mathcal{B}_\Delta(T_2) < 0\} &= \int_{\Omega_1} dP_\Delta(\omega) - \int_{\Omega_1} dP_{-\Delta}(\omega) \\ &= \int_{\Omega_1} \left[\frac{dP_\Delta}{dP_{-\Delta}}(\omega) - 1\right] dP_{-\Delta}(\omega) \\ &= \int_{\Omega_1} \left[e^{2\Delta b(T_2)} - 1\right] dP_{-\Delta}(\omega) \geq 0.\end{aligned}$$

The last equation follows from the fact that $b(T_2) > 0$ in Ω_1. □

Now we are in a position to state and prove the statistical validity of UVP. 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 3: 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 $i \neq j$ and any positive integers p and q. Then the Unknown-Variances Procedure selects system k with probability at least $1 - \alpha$ whenever $\mu_k - \mu_{k-1} \geq \delta$.

PROOF: In UVP we compare $Y_{ij}(\tau_{ij}(r))$ to $a - \lambda\tau_{ij}(r)$ and $-a + \lambda\tau_{ij}(r)$, which is equivalent to compare $Z_{ij}(t_{ij}(r))$ to $a'_{ij} - \lambda t_{ij}(r)$ and $-a'_{ij} + \lambda t$, where $t_{ij}(r)$ and $Z_{ij}(t_{ij}(r))$ are defined in Eqs. (2) and (3), and

$$a'_{ij} = \frac{S_i^2/n_i(r) + S_j^2/n_j(r)}{\sigma_i^2/n_i(r) + \sigma_j^2/n_j(r)}\, a.$$

Let

$$\gamma(r) = \frac{\sigma_i^2/n_i(r)}{\sigma_i^2/n_i(r) + \sigma_j^2/n_j(r)}.$$

Then $\gamma(r) \in (0, 1)$ and

$$a'_{ij} = \left[\gamma(r)\frac{S_i^2}{\sigma_i^2} + (1 - \gamma(r))\frac{S_j^2}{\sigma_j^2}\right] a. \tag{14}$$

Conditioned on S_i^2, $i = 1, 2, \ldots, k$, a'_{ij} is independent of $Z_{ij}(t_{ij}(r))$. However, a'_{ij} is not a constant since it changes with respect to r. If we let a'_{ij} stay constant between r and $r + 1$, then the boundary of the continuation region is piecewise linear with the same slope λ as illustrated in Figure 3. Lemma 1 cannot be applied directly on the continuation region.

For all $i, j = 1, 2, \ldots, k$ and $i \neq j$, let

$$a''_{ij} = \min\left\{\frac{S_i^2}{\sigma_i^2}, \frac{S_j^2}{\sigma_j^2}\right\} a;$$

a''_{ij} is a constant given S_i^2 and S_j^2. From Eq. (14), it is clear that $a''_{ij} \leq a'_{ij}$ for any r such that $n_i(r) > 0$ and $n_j(r) > 0$. Let $\Pi_{ij}^{(1)}$ be the continuation region formed by a''_{ij} and λ and $\Pi_{ij}^{(2)}$ be the continuation region formed by a'_{ij} and λ. It is clear that $\Pi_{ij}^{(1)}$ is completely inside of $\Pi_{ij}^{(2)}$ (Figure 3). Let $T_{ij}^{(1)}$ and $T_{ij}^{(2)}$ be the first times that $\mathcal{B}_{\mu_i - \mu_j}(t)$ exits $\Pi_{ij}^{(1)}$ and $\Pi_{ij}^{(2)}$, respectively, $T_{ij}^{(3)}$ be the first time that $Z_{ij}(t_{ij}(r))$ exits $\Pi_{ij}^{(2)}$, and T_{ij} be the first time that $\mathcal{B}_\delta(t)$ exits $\Pi_{ij}^{(1)}$. Then,

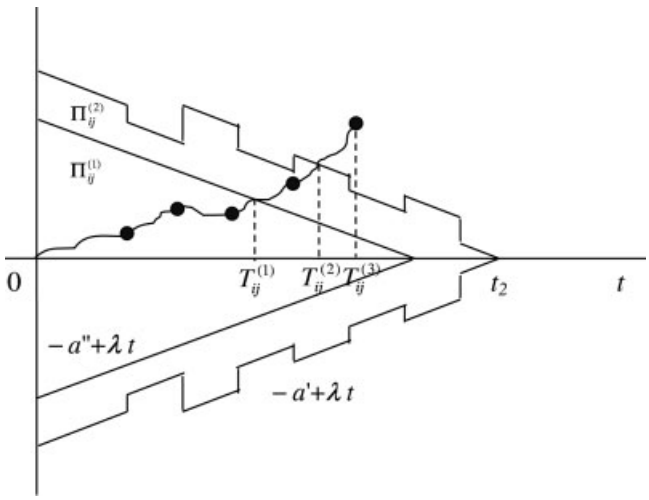

Figure 3. $\Pi_{ij}^{(1)}$ and $\Pi_{ij}^{(2)}$.

$$\begin{aligned}
\Pr\{\text{system } i \text{ eliminates system } k\} &= \mathrm{E}\left[\Pr\left\{\text{system } i \text{ eliminates system } k | S_k^2, S_i^2\right\}\right] = \mathrm{E}\left[\Pr\left\{Z_{ki}(T_{ki}^{(3)}) < 0 | S_k^2, S_i^2\right\}\right] \\
&\le \mathrm{E}\left[\Pr\left\{\mathcal{B}_{\mu_k-\mu_i}(T_{ki}^{(2)}) < 0 | S_k^2, S_i^2\right\}\right] \qquad \text{by Theorem 1, Lemma 2, and } \mu_k - \mu_i > 0 \\
&\le \mathrm{E}\left[\Pr\left\{\mathcal{B}_{\mu_k-\mu_i}(T_{ki}^{(1)}) < 0 | S_k^2, S_i^2\right\}\right] \qquad \text{by Lemma 4} \\
&\le \mathrm{E}\left[\Pr\left\{\mathcal{B}_{\delta}(T_{ki}) < 0 | S_k^2, S_i^2\right\}\right] \quad \text{since } \mu_k - \mu_i \ge \delta \\
&\le \mathrm{E}\left[\frac{1}{2} e^{-a''_{ki}\delta}\right] \qquad \text{by Lemma 1} \\
&= \mathrm{E}\left[\frac{1}{2}\exp\left(-\frac{a\delta}{n_0-1}\min\left\{\frac{(n_0-1)S_k^2}{\sigma_k^2}, \frac{(n_0-1)S_i^2}{\sigma_i^2}\right\}\right)\right]. \qquad (15)
\end{aligned}$$

Let

$$\Psi = \min\left\{\frac{(n_0-1)S_k^2}{\sigma_k^2}, \frac{(n_0-1)S_i^2}{\sigma_i^2}\right\}.$$

Since both $(n_0-1)S_k^2/\sigma_k^2$ and $(n_0-1)S_i^2/\sigma_i^2$ are independent $\chi^2_{n_0-1}$ random variables, where χ^2_f denotes a chi-square random variable with f degrees of freedom, then Ψ has the density function

$$f_\Psi(x) = 2\left[1 - F_{\chi^2_{n_0-1}}(x)\right] f_{\chi^2_{n_0-1}}(x),$$

where $F_{\chi^2_{n_0-1}}(x)$ and $f_{\chi^2_{n_0-1}}(x)$ are the distribution function and density function of a χ^2 distribution with n_0-1 degrees of freedom. By Eq. (10),

$$\text{Eq. (15)} = 1 - (1-\alpha)^{\frac{1}{k-1}}.$$

The rest of the proof follows from the proof of Theorem 2. □

Similarly, when there are systems whose means are within δ to the mean of the best system, UVP guarantees that one of the good systems, either the best system or a system that is within δ to the best system, will be selected with a probability at least $1-\alpha$. This result follows directly from Corollary 1 of Kim and Nelson [18] and Theorem 3.

4.3. The Choice of *a*

To find a using Eq. (10) we must numerically solve an equation that involves integration. Since the left-hand side of Eq. (10) is monotone, the equation is not difficult to solve. Moreover, all pairwise comparisons use the same a, we only need to solve the equation once. In the rest of this subsection we give an upper bound and a lower bound of a and show the connections of the two bounds to some well-known results in ranking and selection. Having an upper bound and a lower bound also helps solve Eq. (10).

Note that

$$\begin{aligned}
\text{Eq. (15)} &= \int_0^\infty \frac{1}{2}\exp\left(-\frac{a\delta}{n_0-1}x\right) f_\Psi(x)dx \\
&= \int_0^\infty \frac{1}{2}\exp\left(-\frac{a\delta}{n_0-1}x\right) 2\left[1 - F_{\chi^2_{n_0-1}}(x)\right] f_{\chi^2_{n_0-1}}(x)dx \\
&\le \int_0^\infty \exp\left(-\frac{a\delta}{n_0-1}x\right) f_{\chi^2_{n_0-1}}(x)dx \\
&= \mathrm{E}\left[\exp\left(-\frac{a\delta}{n_0-1}\chi^2_{n_0-1}\right)\right] \\
&= \left[1 + \frac{2a\delta}{n_0-1}\right]^{-(n_0-1)/2}. \qquad (16)
\end{aligned}$$

If we set Eq. (16) equal to $1-(1-\alpha)^{1/(k-1)}$, we obtain that

$$a_u = \frac{n_0-1}{2\delta}\left\{\left[1-(1-\alpha)^{1/(k-1)}\right]^{-2/(n_0-1)} - 1\right\}.$$

It is easy to show that a_u is greater than or equal to the a obtained by solving Eq. (10). If we substitute a by a_u in UVP, the statistical validity of UVP still holds.

Note that

$$\begin{aligned}
\text{Eq. (15)} &\ge \mathrm{E}\left[\frac{1}{2}\exp\left(-\frac{a\delta}{n_0-1}\frac{(n_0-1)S_k^2}{\sigma_k^2}\right)\right] \\
&= \mathrm{E}\left[\frac{1}{2}\exp\left(-\frac{a\delta}{n_0-1}\chi^2_{n_0-1}\right)\right] \\
&= \frac{1}{2}\left[1 + \frac{2a\delta}{n_0-1}\right]^{-(n_0-1)/2}. \qquad (17)
\end{aligned}$$

If we set Eq. (17) equal to $1-(1-\alpha)^{1/(k-1)}$, we obtain that

$$a_\ell = \frac{n_0-1}{2\delta}\left\{\left[2-2(1-\alpha)^{1/(k-1)}\right]^{-2/(n_0-1)} - 1\right\}.$$

It is easy to show that a_ℓ is less than or equal to the a obtained by solving Eq. (10).

Both a_u and a_ℓ correspond to some well-known results in ranking and selection. a_u is often used in sequential procedures that apply the Paulson probability bound [22] to construct the triangular continuation region, e.g., Hong and Nelson [12], and a_ℓ is often used in sequential procedures that apply the Fabian probability bound of Lemma 1, e.g., Kim and Nelson [18] and Pichitlamken, Nelson, and Hong [23]. The Paulson probability bound is generally easier to apply but less tight than the Fabian probability bound.

Although a_ℓ is a lower bound of a, we conjecture that the use of a_ℓ in UVP can also deliver the desired statistical validity. Suppose that we substitute $\gamma(r)$ and a by any constant $\gamma \in (0,1)$ and a_ℓ in Eq. (14) and denote the resulting a'_{ij} as $\tilde{a}_{ij}$. Then

$$\tilde{a}_{ij} = \left[\gamma \frac{S_i^2}{\sigma_i^2} + (1-\gamma)\frac{S_j^2}{\sigma_j^2}\right] a_\ell.$$

Note that $\tilde{a}_{ij}$ is a constant conditioned on S_i^2 and S_j^2. Let T_{ij} be the first time that $\mathcal{B}_\delta(t)$ exits the triangular continuation region formed by $\tilde{a}_{ij}$ and λ. Then,

$$\Pr\{\mathcal{B}_\delta(T_{ki}) < 0\} = \mathrm{E}\left[\frac{1}{2}e^{-\tilde{a}_{ki}\delta}\right] \quad \text{condition on } S_k^2 \text{ and } S_i^2 \text{ and then apply Lemma 1}$$

$$= \frac{1}{2}\,\mathrm{E}\left[\exp\left(-\frac{\gamma a_\ell \delta}{n_0-1}\frac{(n_0-1)S_k^2}{\sigma_k^2}\right)\right]\mathrm{E}\left[\exp\left(-\frac{(1-\gamma)a_\ell\delta}{n_0-1}\frac{(n_0-1)S_i^2}{\sigma_i^2}\right)\right] \quad \text{by the independence of } S_k^2 \text{ and } S_i^2$$

$$= \frac{1}{2}\left\{\left[1+\frac{2\gamma a_\ell\delta}{n_0-1}\right]\left[1+\frac{2(1-\gamma)a_\ell\delta}{n_0-1}\right]\right\}^{-(n_0-1)/2} \le \frac{1}{2}\left[1+\frac{2a_\ell\delta}{n_0-1}\right]^{-(n_0-1)/2} = 1-(1-\alpha)^{1/k-1}.$$

Although $\gamma(r)$ is not a constant, it is independent of $\mathcal{B}_{\mu_k-\mu_i}(t)$, since it is determined by the sampling rule. Hence, the variation of $\gamma(r)$ does not particularly cause $\mathcal{B}_{\mu_k-\mu_i}(t)$ to exit from the lower side of the triangular region. Therefore, the above analysis gives us reason to conjecture that using a_ℓ instead of a in UVP can also deliver the desired statistical validity. The numerical tests conducted in Section 5 also support our conjecture. Therefore, we recommend using a_ℓ instead of a in UVP.

5. NUMERICAL EXPERIMENTS

In this section we summarize the results of an extensive empirical evaluation of KVP and UVP relative to KN procedure. The original KN procedure is for unknown-variances case. To study the empirical performance of KVP, we also designed a KN-like procedure for known-variances cases. We call the procedure $\mathrm{KN}_{\mathrm{known}}$ procedure.

Except for the job-shop example in Section 5.3, the systems are represented by various configurations of k normal distributions with system k being the best (having the largest mean). Two configurations of means were used: the slippage configuration (SC) and the monotone-increasing-means configuration (MIM). In SC, μ_k was set to δ and $\mu_1 = \mu_2 = \ldots = \mu_{k-1} = 0$. This is a difficult configuration in terms of statistical validity since all inferior systems are exactly δ from the best. In MIM $\mu_i = i\delta, i = 1,2,\ldots,k$. MIM is used to investigate the effectiveness of the procedures in more favorable settings. We also examined the effect of variances. There are three configurations of variances: equal-variances configuration (EV), increasing-variances configuration (IV) and decreasing-variances configuration (DV). In EV, $\sigma_1 = \sigma_2 = \ldots = \sigma_k = 10$; in IV, $\sigma_i = 1 + 9(i-1)/(k-1)$; and in DV, $\sigma_i = 10 - 9(i-1)/(k-1)$. For each configuration of mean and variance, 1000 macroreplications (complete repetitions) of each procedure were carried out to compare the performance measures, including the observed PCS and the average sample size. In all cases we set $\alpha = 0.05$ and $\delta = 1$. If variances are unknown, we set $n_0 = 10$.

5.1. Validity Check

In Sections 3 and 4 we proved the statistical validity of KVP and UVP with a and a_u. We also conjectured that UVP with a_ℓ can also deliver the desired PCS. In this subsection we test the statistical validity of the procedures by checking the observed PCS. We use the SC of means with all three configurations of variances and set $k = 2$. Note that the SC with $k = 2$ is the most difficult case to deliver the desired PCS. The results in Table 1 show that all procedures have the

Table 1. Observed PCS of KVP and UVP.

Var. config.	KVP	UVP with a	UVP with a_u	UVP with a_ℓ
EV	0.953	0.973	0.978	0.961
IV	0.954	0.967	0.982	0.960
DV	0.953	0.978	0.990	0.956

desired PCS. They also support our conjecture of using a_ℓ in UVP. Therefore, we recommend using a_ℓ in UVP. In the rest of this section, we use UVP with a_ℓ to represent UVP.

We also run KVP and UVP for the MIM configuration of means with the three variance configurations and $k = 10$. The observed PCS in all cases are at least 0.99 although the required PCS is only 0.95. Therefore, the procedures generally require more samples than necessary to select the best system, which is a major drawback of almost all of the existing indifference-zone selection procedures (see the empirical study of Branke, Chick, and Schmidt [4]).

5.2. Effectiveness of Variance-Dependent Sampling

We compare the efficiency of KVP and UVP to KN procedures based on the average sample sizes used to select the best system. Results in Tables 2 and 3 show the average sample sizes in different configurations when $k = 2$ and $k = 10$, respectively. Note that the SC and MIM are identical when $k = 2$. For each comparison, either between KVP and KN_{known} or between UVP and KN, we conducted a t test to check whether the average sample sizes of two procedures are equal. When the sample sizes of the two procedures in comparison are different at 95% confidence level, we calculated the percentages of reduction of KVP(or UVP) to KN_{known}(or KN) and reported them in Tables 2 and 3.

The results in Tables 2 and 3 show that the procedures with variance-dependent sampling reduce the total sample size required to select the best system when the variances of the systems are unequal. They are more effective when the number of systems is small. From Table 3 we see that KVP and UVP are less effective in the MIM+DV configuration. In this configuration, both KVP and UVP tend to allocate more samples to systems that are eliminated soon and fewer samples to systems that will survive longer. But ideally, we want systems that survive longer to have more samples since we need more samples to eliminate them. To further reduce the sample size of the fully sequential procedures, information on sample means may be used in the design of sampling rules. It is a topic for future research.

Note that KN allows the use of common random numbers (CRN), which often reduce the variances of pairwise differences and make eliminations easier, and our procedures do not. Therefore, for problems in which CRN are very effective, the modest gains of our procedures may disappear. However, CRN is not always effective and it may not be easy to implement. In Section 5.3 we compare UVP to KN with CRN using a queueing example. For this example, UVP works better than KN with CRN.

5.3. An Illustrative Example

In this section we study a job-shop improvement problem (see Law and Kelton [21] for the simulation model). The job shop consists of five work stations, and at present stations $1, 2, \ldots, 5$ consist of, respectively, $3, 2, 4, 3$, and 1 parallel machines. Assume that jobs arrive to the system with interarrival times that are i.i.d. exponential random variables with mean 0.25 h. There are three types of jobs, and arriving jobs are of type 1, 2, and 3 with respective probabilities 0.3, 0.5, and 0.2. Job types 1, 2, and 3 require 4, 3, and 5 tasks to be done, respectively, and each task must be done at a specified station and in a prescribed order. The routings for the different job types are

Job type	Work station routing
1	3, 1, 2, 5
2	4, 1, 3
3	2, 5, 1, 4, 3

If a job arrives at a particular station and finds all machines in that station already busy, the job joins a single FIFO queue at the station. The time to perform a task at a particular machine is an independent Erlang-2 random variable whose mean depends on the job type and the station to which the

Table 2. Average sample size of KVP and UVP when $k = 2$.

	Known variances			Unknown variances		
Var. config.	KN_{known}	KVP	Reduction	KN	UVP	Reduction
SC + EV	589.96	602.17	—	788.35	753.88	—
SC + IV	305.67	179.66	41.2%	412.56	253.91	38.5%
SC + DV	304.87	180.18	40.1%	386.17	236.15	38.8%

Table 3. Average sample size of KVP and UVP when $k = 10$.

	Known variances			Unknown variances		
Configuration	KN_{known}	KVP	Reduction	KN	UVP	Reduction
SC + EV	6098.4	6057.2	—	10232	10094	—
SC + IV	2841.7	2527.2	11.1%	4909.5	4296.0	12.5%
SC + DV	1561.6	1384.9	11.3%	2804.8	2378.3	15.2%
MIM + EV	2600.7	2614.9	—	4365.5	4314.8	—
MIM + IV	1918.1	1692.8	11.7%	3169.7	2818.1	11.1%
MIM + DV	334.47	325.72	2.7%	595.75	540.94	9.2%

machine belongs. The mean service times for each job type and each task are

Job type	Mean service time for successive tasks, in h
1	0.50, 0.60, 0.85, 0.50
2	1.10, 0.80, 0.75
3	1.20, 0.25, 0.70, 0.90, 1.00

Suppose that the owner of the job shop is concerned that the product waiting times are too long, and he or she is considering improving the performance of the job shop. All machines cost approximately \$300,000 and the owner can at most afford three new machines. He or she also estimates that 1 h of waiting time is equivalent to \$100,000 of loss of goodwill or potential sales and wants to improve the performance of the job shop to minimize the total expected cost, including both machine cost and waiting-time cost (where waiting time is taken to be the weighted average total waiting time, weighted by the fraction of jobs of each type). The engineer of the job shop suggests considering work stations 1, 2, and 4, since the queues of these three stations are often long. The owner decides to compare eight alternatives: keeping the current configuration, adding one machine to one of stations 1, 2, and 4 (totally three alternatives), adding two machines to two of stations 1, 2, and 4 (totally three alternatives), or adding one machine to each of stations 1, 2, and 4. He or she wants to select the alternative with the lowest total expected cost.

We use ranking and selection to solve this problem. Suppose that the owner decides that $\delta = \$50,000$ and $\alpha = 0.05$ are sufficient for making correct decision. Then we can use both KN and UVP to solve this problem since the variances of the alternatives are not known in advance. We set $n_0 = 10$ for both procedures. We run 1000 macroreplications to compare the two procedures. KN procedure selects the last alternative, adding three machines, as the best design 994 times of 1000 times and UVP procedure selects the same alternative 993 times of 1000 times. The average sample size of KN procedure is 601.8 with a standard error 5.9; the average sample size of UVP is 468.9 with a standard error 4.4. UVP is clearly better than KN, with a 22.9% savings in average sample size.

We also implement CRN for this problem where we synchronize the interarrival times, job types, and service times. The (accurately estimated) correlation coefficients of all pairs of systems range from 0.162 to 0.644. Therefore, the CRN is effective for this problem. Then we run the KN procedure that exploits CRN for this problem. KN selects the best design 994 times of 1000 macroreplications and the average sample size is 568.1 with a standard error 8.8. It requires fewer samples compared to the KN procedure that does not exploit CRN, but it requires more samples compared to UVP. The reason that CRN does not bring significant benefit to KN in this problem is because the systems in comparison have significantly different variances. For example, the variance of the original design is more than 3000 times higher than the variance of the best design. Therefore, the variance of the difference of the two design is completely dominated by the variance of the original system no matter whether CRN is used. For problems in which the variances of the systems are not drastically different, CRN may bring more benefit to the KN procedure.

6. CONCLUSIONS

In this paper we present a general approach to construct a Brownian motion process with drift to compare the means of two different systems. This approach allows unequal variances and unequal sample sizes for both systems. We then combine the construction and the triangular continuation region and design fully sequential indifference-zone procedures that allow variance-dependent sampling. We show that the procedures deliver the desired PCS and require fewer samples than the popular KN procedure.

There are several ways in which we might be able to extend the approach developed in this paper. First, it appears possible to combine our construction of Brownian motion process with other types of continuation regions, e.g., the parabolic continuation region by Batur and Kim [1], to design new indifference-zone selection sequential procedures. Second, it seems possible to extend our approach to comparison with a standard using the formulation of Kim [17]. Third, it may be

possible to apply the methodology of Kim and Nelson [19] to UVP to allow autocorrelated samples and deliver the desired asymptotic validity when $\delta \to 0$. Another possible extension is to combine our construction of Brownian motion process with sequential indifference-zone selection procedures used in optimization via simulation [13, 23], where systems are generated sequentially based on previous selection decisions and they often have different numbers of samples.

ACKNOWLEDGMENTS

The author thanks Barry L. Nelson of Northwestern University for providing the simulation code of the job shop example used in Section 5.3. This research was partially supported by Hong Kong Research Grants Council Grant DAG04/05.EG04 and Grant CERG HKUST 613305.

REFERENCES

[1] D. Batur and S.-H. Kim, Fully sequential selection procedures with parabolic boundary, IIE Trans, forthcoming.
[2] R.E. Bechhofer, A single-sample multiple decision procedure for ranking means of normal populations with known variances, Ann Math Statist 25 (1954), 16–39.
[3] R.E. Bechhofer, T. Santner, and D.M. Goldsman, Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons, Wiley, New York, 1995.
[4] J. Branke, S.E. Chick, and C. Schmidt, New developments in ranking and selection: An empirical comparison of the three main approaches, Proceedings of the 2005 Winter Simulation Conference, pp. 708–717.
[5] C.H. Chen, J. Lin, E. Yücesan, and S.E. Chick, Simulation budget allocation for further enhancing the efficiency of ordinal optimization, J Discr Event Dynam Syst 10 (2000), 251–270.
[6] C.H. Chen and E. Yücesan, An alternative simulation budget allocation scheme for efficient simulation, Int J Simul Process Model 1 (2005), 49–57.
[7] S.E. Chick and K. Inoue, New two-stage and sequential procedures for selecting the best simulated system, Oper Res 49 (2001), 732–743.
[8] S.E. Chick and K. Inoue, New procedures to select the best simulated system using common random numbers, Manage Sci 47 (2001), 1133–1149.
[9] R. Durrett, Probability: Theory and Examples, 2nd ed., Duxbury Press, 1995.
[10] V. Fabian, Note on Anderson's sequential procedures with triangular boundary, Ann Statist 2 (1974), 170–176.
[11] M.C. Fu, J.Q. Hu, C.H. Chen, and X. Xiong, Optimal computing budget allocation under correlated sampling, Proceedings of the 2004 Winter Simulation Conference, pp. 595–603.
[12] L.J. Hong and B.L. Nelson, The trade-off between sampling and switching: New sequential procedures for indifference-zone selection, IIE Trans 37 (2005), 623–634.
[13] L.J. Hong and B.L. Nelson, Selecting the best system when systems are revealed sequentially, Working Paper, Department of Industrial Engineering and Logistics Management, The Hong Kong University of Science and Technology, 2005.
[14] C. Jennison, I.M. Johnston, and B.W. Turnbull, Asymptotically optimal procedures for sequential adaptive selection of the best of several normal means, Technical Report, Department of Operations Research and Industrial Engineering, Cornell University, 1980.
[15] C. Jennison, I.M. Johnston, and B.W. Turnbull, Asymptotically optimal procedures for sequential adaptive selection of the best of several normal means, Statist Decision Theor Rel Topics III 2 (1982), 55–86.
[16] C. Jennison and B.W. Turbull, Group Sequential Methods with Applications to Clinical Trials, Chapman and Hall, 2000.
[17] S.-H. Kim, Comparison with a standard via fully sequential procedure, ACM Trans Model Comput Sim 15 (2005), 1–20.
[18] S.-H. Kim and B.L. Nelson, A fully sequential procedure for indifference-zone selection in simulation, ACM Trans Model Comput Simul 11 (2001), 251–273.
[19] S.-H. Kim and B.L. Nelson, On the asymptotic validity of fully sequential selection procedures for steady-state simulation, Oper Res, forthcoming.
[20] V.G. Kulkarni, Modeling and Analysis of Stochastic Systems, Chapman and Hall, 1995.
[21] A.M. Law and W.D. Kelton, Simulation Modeling and Analysis, 3rd ed., McGraw–Hill, 2000.
[22] E. Paulson, A sequential procedure for selecting the population with the largest mean from k normal populations, Ann Math Statist 35 (1964), 174–180.
[23] J. Pichitlamken, B.L. Nelson, and L.J. Hong, A sequential procedure for neighborhood selection-of-the-best in optimization via simulation, Eur J Oper Res, forthcoming.
[24] Y. Rinott, On two-stage selection procedures and related probability-inequalities, Commun Statist A7 (1978), 799–811.
[25] H. Robbins and D. Siegmund, Sequential tests involving two populations, J Am Statist Assoc 69 (1974), 132–139.
[26] C. Stein, A two-sample test for a linear hypothesis whose power is independent of the variance, Ann Math Statist 16 (1945), 243–258.
[27] A.C. Tamhane, Multiple comparisons in model I: One-way ANOVA with unequal variances, Commun Statist A6 (1977), 15–32.