

# Fully Sequential Selection Procedures with Parabolic Boundary

Demet Batur

Seong-Hee Kim

School of Industrial and Systems Engineering

Georgia Institute of Technology

June 8, 2005

## Abstract

We present two fully sequential indifference-zone procedures to select the best system from a number of competing simulated systems when best is defined by maximum or minimum expected performance. These two procedures have parabola shaped continuation regions rather than triangular continuation regions employed in several papers. The procedures we present accommodate unequal and unknown variances across systems and the use of common random numbers. However, we assume that basic observations are independent and identically normally distributed. We compare the performance of our procedures with those of other fully sequential procedures available in the literature.

## 1 Introduction

Among comparison problems we meet in simulation, the problem of finding-the-best is probably the most popular. For the details of different types of comparison problems in simulation, see Goldsman and Nelson (1998). The goal of finding-the-best is to choose the one with the largest or smallest expected performance measure among a finite number of simulated systems. To solve this problem, a number of different approaches have been proposed. The indifference-zone approach tries to choose a system whose performance measure is at least a user-specified constant—called the indifference-zone parameter—better than all the other alternative systems with the probability of correct selection (PCS) no less than a pre-specified amount. The indifference-zone parameter—denoted by  $\delta$ —is set by the experimenter as the

minimum practical difference worth detecting. If there exist systems whose means are within  $\delta$  of the best, then the experimenter is *indifferent* to which of these is selected. Recent references include Nelson et al. (2001), Boesel et al. (2003), and Kim and Nelson (2001, 2005). On the other hand, Chick (1997) and Chick and Inoue (2001ab) propose completely different procedures from a decision-theoretic point of view and Chen et al. (1997, 2000) propose heuristic procedures that maximize PCS under a budget constraint. Ma and Willemain (2004) present a different approach for finding the best on the basis of expected steady-state performance in steady-state simulation.

In this paper, we restrict our interest to terminating simulation where basic observations are often within-replication averages; therefore, independent and identically distributed (i.i.d.) normality assumption makes sense. One of the most recent statistically valid indifference-zone procedures for i.i.d. normal data is a fully sequential procedure,  $\mathcal{KN}$ , proposed by Kim and Nelson (2001). Kim and Nelson (2001) define a fully sequential procedure as the one that takes a single basic observation from each alternative that is still in play at the current stage of sampling and eliminates inferior systems immediately when there is an evidence that they are. Fully sequential procedures have a boundary called a continuation region. Figure 1 shows two different types of continuation regions,  $\mathcal{T}$  and  $\mathcal{P}$  that have triangular and parabolic shapes, respectively. The horizontal axis in Figure 1 represents the stage number and the vertical axis represents the value of a monitoring statistic at each stage. As long as the monitoring statistic stays within the continuation region, sampling continues. When the monitoring statistic exits the continuation region, one system is eliminated depending on which direction—through the upper boundary or the lower boundary—the exit is made.

$\mathcal{KN}$  is closely related to and derived from a hypothesis test on a drift of a Brownian motion process. Lerche (1986) states that for the known drift of a Brownian motion, a triangular boundary is optimal in a sense that it minimizes a certain type of Bayes risk. However, they show that for the unknown drift of a Brownian motion process, parabolic

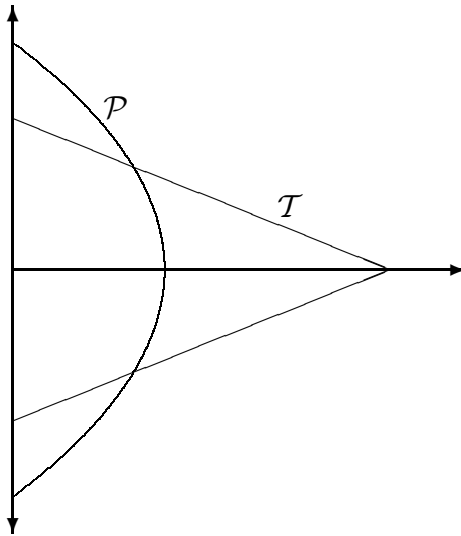


Figure 1: Triangular and parabolic continuation regions.

boundaries are optimal in terms of the expected value of the specific type of Bayes risk.

From this motivation, Zhu et al. (2005) propose a new fully sequential procedure whose continuation region is parabolic for i.i.d. normal data. However, the procedure assumes known variances which is unlikely to be true in practice. In this paper, we propose two fully sequential indifference-zone procedures for finding-the-best problem, one of which is a natural extension of Zhu et al. (2005). Our procedures have parabolic continuation regions and assume unknown and unequal variances across systems. We compare the performance of the proposed procedures with that of  $\mathcal{KN}$ , which is one of the most efficient statistically valid indifference-zone procedures for i.i.d. data in terms of the number of observations required.

There exists another recent ranking and selection (R&S) procedure called  $\mathcal{KN}++$  due to Kim and Nelson (2005). This procedure updates variance estimates as more observations become available and is shown to be very effective in terms of the number of observations required. However, it is specifically designed for steady-state simulation that employs the single-replication design and takes individual observations such as individual wait times—that are, at best, stationary and dependent—as basic observations. Though  $\mathcal{KN}++$  is asymptotically valid for stationary and dependent data and highly efficient, Malone, Kim,

Goldman, and Batur (2005) show that  $\mathcal{KN}++$  does not always satisfy the PCS requirement and is only heuristic for i.i.d. normal data. Since our interests in this paper are not on procedures whose validity is not established for i.i.d. normal data, we do not consider  $\mathcal{KN}++$  when we compare the performance of our procedures to other existing procedures.

The paper is organized as follows: In Section 2, algorithms of the two new procedures are given and we prove that the procedures choose the best or a near-best system with at least a pre-specified PCS. In Section 3, we discuss implementation issues including the determination of the values of parameters for the new procedures. Section 4 provides analytical comparisons among  $\mathcal{KN}$  and the two proposed procedures. Section 5 compares the performance of the proposed procedures with that of  $\mathcal{KN}$  by empirical studies based on i.i.d. normal data, followed by a conclusion in Section 6.

## 2 Procedures

We design two fully sequential indifference-zone procedures. Both have the parabolic continuation regions but with different parameters. We assume that there are  $k$  systems. The best system is defined as the one with the largest expected performance when the difference between the expected performances of the best and the second best is at least  $\delta$ . Our procedures guarantee to select the best or a near-best system with a PCS greater than or equal to  $1 - \alpha$ . If there are inferior systems whose means are within  $\delta$  of the true best system, then those systems are called “good” systems and the procedures select one of these systems with at least  $1 - \alpha$  PCS.

In the procedures,  $X_{ij}$ ,  $i = 1, 2, \dots, k$  and  $j = 1, 2, \dots$ , denotes the  $j$ th observation from the  $i$ th system and  $X_{ij}$ ,  $j = 1, 2, \dots$ , are assumed to be i.i.d. normally distributed with mean  $\mu_i$  and variance  $\sigma_i^2$ . The variances are unknown to the experimenter and may be unequal.

## 2.1 First Procedure

We present the first procedure that we call  $\mathcal{P}_1$ .

### Procedure $\mathcal{P}_1$

**Setup:** Choose nominal PCS  $1 - \alpha$ , indifference zone  $\delta$ , and first stage sample size  $n_0 \geq 2$ .

Then determine  $\lambda$  and  $\xi$  referring to **Parameters**.

**Initialization:** Let  $I = \{1, 2, \dots, k\}$  be the initial set of systems.

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

For all  $i \neq \ell$  compute the sample variance of the difference between systems  $i$  and  $\ell$  which is

$$S_{i\ell}^2 = \frac{1}{n_0 - 1} \sum_{j=1}^{n_0} \left( X_{ij} - X_{\ell j} - [\bar{X}_i(n_0) - \bar{X}_\ell(n_0)] \right)^2,$$

where  $\bar{X}_i(n_0)$  is the sample average of the first  $n_0$  observations from system  $i$ . Let

$$N_{i\ell} = \left\lfloor \frac{(n_0 - 1)^2 S_{i\ell}^4 \xi}{\delta^2} \right\rfloor,$$

where  $\lfloor \cdot \rfloor$  indicates truncation of any fractional part. Then, let

$$N_i = \max_{\ell \neq i} N_{i\ell}.$$

Here,  $N_i + 1$  is the maximum number of observations that can be taken from system  $i$ .

Set the number of observations,  $r$ , equal to  $n_0$ , and go to the next section.

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

$$I = \left\{ i : i \in I^{\text{old}} \text{ and } \sum_{j=1}^r X_{ij} > \sum_{j=1}^r X_{\ell j} - R_{i\ell}(r), \forall \ell \in I^{\text{old}}, \ell \neq i \right\},$$

where

$$R_{i\ell}(r) = \lambda \sqrt{\max \left\{ \frac{(n_0 - 1)^2 S_{i\ell}^4 \xi}{\delta^2} - r, 0 \right\}}.$$

**Stopping Rule:** If  $|I| = 1$ , then stop and select the system  $i \in I$  as the best. Otherwise, take one additional observation  $X_{i,r+1}$  from each system  $i \in I$ , and set  $r = r + 1$ , and go to **Screening**.

(If the objective is to select a subset of size  $m$  containing the best system, then the stopping rule should be  $|I| = m > 1$ .)

**Parameters:** The parameter  $\lambda$  is any positive real number. For  $1 - \alpha = 0.95$ , we recommend taking  $\lambda = 0.9268$  as given in Ferebee (1982). Section 3.1 discusses the choice of  $\lambda$  for other values of  $\alpha$ . For a given  $\lambda$  value,  $\xi$  is calculated as the solution to the equation

$$g(\xi; \lambda, n_0) = \beta, \quad (1)$$

where

$$g(\xi; \lambda, n_0) \equiv \mathbb{E} \left[ \frac{1}{2} - \frac{1}{2\sqrt{2\pi}} \exp \left\{ -\frac{\xi}{2} (\chi_{n_0-1}^2)^2 \right\} \sum_{n=0}^{\infty} \left( \frac{k_{2n+1}(\lambda)}{\ell_{2n+1}(\lambda)} \frac{2^n n!}{(2n+1)!} (\chi_{n_0-1}^2 \sqrt{\xi})^{2n+1} \right) \right].$$

In this function,  $k_n(\cdot)$  and  $l_n(\cdot)$  are certain Hermite functions which are defined in Lemma 1 and  $\chi_{n_0-1}^2$  is a chi-squared random variable with  $n_0 - 1$  degrees of freedom. If common random numbers (CRN) are employed, we choose  $\beta = \alpha/(k - 1)$  and if systems are simulated independently, use  $\beta = 1 - (1 - \alpha)^{1/(k-1)}$ . However, this equation does not have a closed form solution. We make a detailed discussion about how to determine the  $\xi$  values in Section 3.2.

To prove that the procedure satisfies the PCS requirement, we need the following three lemmas:

**Lemma 1 (Ferebee 1982)** *Let  $\mathcal{W}(t, \Delta)$  be a Brownian motion process on  $[0, +\infty)$  with a drift  $\Delta > 0$ . Consider the parabola  $y = a\sqrt{s-t}$ ,  $0 \leq t \leq s$ ,  $a > 0$ . Let  $R$  be  $(-y, y)$  and  $T$  the first time that  $\mathcal{W}(t, \Delta) \notin R$ . That is,*

$$T = \inf\{t : 0 \leq t \leq s \text{ and } |\mathcal{W}(t, \Delta)| = a\sqrt{s-t}\}.$$

Then,

$$\Pr\{\mathcal{W}(T, \Delta) < 0\} = \frac{1}{2} - \frac{1}{2\sqrt{2\pi}} \exp\{-s\Delta^2/2\} \sum_{n=0}^{\infty} \left( \frac{k_{2n+1}(a)}{l_{2n+1}(a)} \frac{2^n n!}{(2n+1)!} (\Delta\sqrt{s})^{2n+1} \right),$$

where  $k_n(\cdot)$  and  $l_n(\cdot)$  are certain Hermite functions which are defined as

$$k_n(x) = \mathbb{E}[(Z+x)^n] \text{ and } l_n(x) = \mathbb{E}[(Z+x)^n \frac{1}{2} \text{sgn}(Z+x)]$$

for  $n = 0, 1, \dots$ , where  $Z$  is a standard  $N(0, 1)$  random variable.

Let  $\phi(x)$  and  $\Phi(x)$  be the pdf and cdf of  $N(0, 1)$ . Then, one can get the recurrence relations

$$\begin{aligned} k_{n+1}(x) &= xk_n(x) + nk_{n-1}(x), \\ l_{n+1}(x) &= xl_n(x) + nl_{n-1}(x), \quad n \geq 1, \end{aligned}$$

where

$$k_0(x) = 1, l_0(x) = \Phi(x) - \frac{1}{2}, k_1(x) = x, \text{ and } l_1(x) = x \left( \Phi(x) - \frac{1}{2} \right) + \phi(x).$$

**Lemma 2 (Jennison et al. 1980)** Suppose that a continuation region  $R$  is  $(-g(t), g(t))$  given by a non-negative function  $g(t), t \geq 0$ . Consider two processes: a continuous process  $\mathcal{W}(t, \Delta), t \geq 0$ , with  $\Delta > 0$  and a discrete process obtained by observing  $\mathcal{W}(t, \Delta)$  at a random, increasing sequence of times  $\{t_i : i = 1, 2, \dots\}$  taking values in a given countable set. Let  $\tau_C = \inf\{t > 0 : \mathcal{W}(t, \Delta) \notin R\}$  and  $\tau_D = \inf\{t_i : \mathcal{W}(t_i, \Delta) \notin R\}$  and assume that  $\tau_D < \infty$  almost surely. Note that  $\tau_D \geq \tau_C$ . The error probabilities are

$$\Pr\{\mathcal{E}_C\} \equiv \Pr\{\mathcal{W}(\tau_C, \Delta) \leq -g(\tau_C)\} = \Pr\{\mathcal{W}(\tau_C, \Delta) < 0\},$$

$$\Pr\{\mathcal{E}_D\} \equiv \Pr\{\mathcal{W}(\tau_D, \Delta) \leq -g(\tau_D)\} = \Pr\{\mathcal{W}(\tau_D, \Delta) < 0\}.$$

Consider an outcome  $\{(b(t); t \geq 0), \{t_i\}\}$  where  $b(t)$  is the path of a Brownian motion. The conditional distribution of  $\{t_i\}$  given  $\mathcal{W}(t, \Delta) = b(t), t \geq 0$ , is the same as that given  $\mathcal{W}(t, \Delta) = -b(t), t \geq 0$ . Under these conditions,

$$\Pr\{\mathcal{E}_D\} \leq \Pr\{\mathcal{E}_C\}.$$

Lemma 1 gives the probability of incorrect selection for a continuous Brownian motion process with a drift. However, in our procedure we only observe the process at integer times. When each observation is i.i.d. normally distributed, the partial sums of the differences behave like Brownian motion process with drift at each integer point but Brownian motion with drift is still only an approximation for our discrete process. However, Lemma 2 states that under very general conditions, the probability of incorrect selection does not increase when the Brownian motion process is observed at discrete times compared to the case where the process is observed continuously; thus, procedures designed for continuous Brownian motion process with a drift provide an upper bound on the probability of incorrect selection for a discrete process.

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

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

Without loss of generality, we assume that  $\mu_1 \geq \mu_2 \geq \dots \geq \mu_k$ . Now, we present the main result.

**Theorem 1** *Let  $\mathbf{X}_j = (X_{1j}, X_{2j}, \dots, X_{kj})'$  be a vector of observations across all  $k$  systems. Suppose that  $\mathbf{X}_1, \mathbf{X}_2, \dots$  are distributed i.i.d. multivariate normal with mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ , where  $\boldsymbol{\mu}$  is an unknown vector with the property  $\mu_1 \geq \mu_2 + \delta$  and  $\boldsymbol{\Sigma}$  is an unknown, positive definite matrix. Then,  $\mathcal{P}_1$  selects system 1 with probability greater than or equal to  $1 - \alpha$ .*

**Proof:** Consider two systems, 1 and  $i$ , such that  $\mu_1 \geq \mu_i + \delta$ . Select a value of  $\xi$  such that  $g(\xi; \lambda, n_0) = \beta$  for some  $0 < \beta < 1/2$ . Let

$$T = \min\left\{r : r \geq n_0 \text{ and } -R_{1i}(r) < \sum_{j=1}^r (X_{1j} - X_{ij}) < R_{1i}(r) \text{ is violated}\right\}.$$



Notice that  $T$  is the stage at which the procedure terminates. Let  $\text{ICS}_i$  be the event of incorrect selection when only two systems, 1 and  $i$ , are considered. Then,

$$\begin{aligned}
\Pr\{\text{ICS}_i\} &= \Pr\left\{\sum_{j=1}^T (X_{1j} - X_{ij}) < -\lambda \sqrt{\max\left\{\frac{(n_0 - 1)^2 S_{1i}^4 \xi}{\delta^2} - T, 0\right\}}\right\} \\
&= \Pr\left\{\sum_{j=1}^T \left(\frac{X_{1j} - X_{ij}}{\sigma_{1i}}\right) < -\lambda \sqrt{\max\left\{\frac{(n_0 - 1)^2 S_{1i}^4 \xi}{\delta^2 \sigma_{1i}^2} - \frac{T}{\sigma_{1i}^2}, 0\right\}}\right\} \\
&\leq \Pr_{\text{SC}}\left\{\sum_{j=1}^T \left(\frac{X_{1j} - X_{ij}}{\sigma_{1i}}\right) < -\lambda \sqrt{\max\left\{\frac{(n_0 - 1)^2 S_{1i}^4 \xi}{\delta^2 \sigma_{1i}^2} - \frac{T}{\sigma_{1i}^2}, 0\right\}}\right\} \\
&= \mathbb{E}\left[\Pr_{\text{SC}}\left\{\sum_{j=1}^T \left(\frac{X_{1j} - X_{ij}}{\sigma_{1i}}\right) < -\lambda \sqrt{\max\left\{\frac{(n_0 - 1)^2 S_{1i}^4 \xi}{\delta^2 \sigma_{1i}^2} - \frac{T}{\sigma_{1i}^2}, 0\right\}} \middle| S_{1i}\right\}\right] \\
&\leq \mathbb{E}\left[\Pr_{\text{SC}}\left\{\sum_{j=1}^T \left(\frac{X_{1j} - X_{ij}}{\sigma_{1i}}\right) < 0 \middle| S_{1i}\right\}\right],
\end{aligned}$$

where  $\sigma_{1i}^2 = \text{Var}(X_{1j} - X_{ij})$  and ‘‘SC’’ denotes the slippage configuration  $\mu_1 = \mu_i + \delta$ .

Since  $(X_{1j} - X_{ij})/\sigma_{1i}, j = 1, 2, \dots$ , are i.i.d.  $N(\delta/\sigma_{1i}, 1)$  random variables under the SC,  $\sum_{j=1}^T (X_{1j} - X_{ij})/\sigma_{1i}$  behaves like a Brownian motion process with drift  $\delta/\sigma_{1i}$  at integer points. Therefore, if we let

$$a = \lambda, \quad s = \frac{(n_0 - 1)^2 S_{1i}^4 \xi}{\delta^2 \sigma_{1i}^2} \quad \text{and} \quad t = \frac{r}{\sigma_{1i}^2}, \quad r = n_0, \dots, N_{1i} + 1,$$

then by Lemma 1 and 2 we get

$$\begin{aligned}
\mathbb{E}\left[\Pr_{\text{SC}}\left\{\sum_{j=1}^T \left(\frac{X_{1j} - X_{ij}}{\sigma_{1i}}\right) < 0 \middle| S_{1i}\right\}\right] &\leq \\
\mathbb{E}\left[\frac{1}{2} - \frac{1}{2\sqrt{2\pi}} \exp\left\{-\frac{(n_0 - 1)^2 S_{1i}^4 \xi}{\sigma_{1i}^4} \frac{\xi}{2}\right\} \sum_{n=0}^{\infty} \left(\frac{k_{2n+1}(\lambda)}{\ell_{2n+1}(\lambda)} \frac{2^n n!}{(2n + 1)!} \left(\frac{(n_0 - 1) S_{1i}^2}{\sigma_{1i}^2} \sqrt{\xi}\right)^{2n+1}\right)\right].
\end{aligned}$$

Since  $(n_0 - 1) S_{1i}^2 / \sigma_{1i}^2$  is chi-square distributed with  $n_0 - 1$  degrees of freedom, the expectation above is equivalent to

$$\mathbb{E}\left[\frac{1}{2} - \frac{1}{2\sqrt{2\pi}} \exp\left\{-\frac{\xi}{2} (\chi_{n_0-1}^2)^2\right\} \sum_{n=0}^{\infty} \left(\frac{k_{2n+1}(\lambda)}{\ell_{2n+1}(\lambda)} \frac{2^n n!}{(2n + 1)!} \left(\chi_{n_0-1}^2 \sqrt{\xi}\right)^{2n+1}\right)\right]. \quad (2)$$

Notice that (2) is the  $g(\xi; \lambda, n_0)$  function, and this expectation is equal to  $\beta$  by the way we choose  $\xi$ . Therefore,

$$\Pr\{\text{ICS}_i\} \leq \beta. \quad (3)$$

Now, assume that we have  $k \geq 2$  systems, and let ICS be the event that an incorrect selection is made by the procedure. If we set  $\beta = \alpha/(k-1)$ , then by the Bonferroni inequality we get

$$\Pr\{\text{ICS}\} \leq \sum_{i=2}^k \Pr\{\text{ICS}_i\} \leq \sum_{i=2}^k \frac{\alpha}{k-1} = \alpha.$$

The above inequality holds whether CRN is employed or systems are simulated independently. However, if systems are simulated independently ( $\Sigma$  is a diagonal matrix in that case), a tighter boundary can be achieved, and its proof goes as follows:

$$\Pr\{\text{CS}\} \geq \Pr\left\{\bigcap_{i=2}^k \text{CS}_i\right\}$$

because the intersection event requires system 1 to eliminate each inferior system  $i$  individually, whereas in reality some system  $\ell \neq 1, i$  could eliminate  $i$ . Thus,

$$\begin{aligned} \Pr\{\text{CS}\} &\geq \Pr\left\{\bigcap_{i=2}^k \text{CS}_i\right\} \\ &= \mathbb{E}\left[\Pr\left\{\bigcap_{i=2}^k \text{CS}_i \mid X_{11}, \dots, X_{1, N_1+1}\right\}\right] \\ &= \mathbb{E}\left[\prod_{i=2}^k \Pr\{\text{CS}_i \mid X_{11}, \dots, X_{1, N_1+1}\}\right], \end{aligned} \quad (4)$$

where the last equality follows because the events are conditionally independent. Clearly, (4) does not increase if we assume the slippage configuration, so we do so from here on.

Now, notice that  $\Pr\{\text{CS}_i \mid X_{11}, \dots, X_{1, N_1+1}\}$  is nondecreasing in  $X_{1j}, j = 1, \dots, N_1 + 1$ . Therefore, by Lemma 3,

$$\begin{aligned} (4) &\geq \prod_{i=2}^k \mathbb{E}[\Pr_{\text{SC}}\{\text{CS}_i \mid X_{11}, \dots, X_{1, N_1+1}\}] \\ &= \prod_{i=2}^k \mathbb{E}[1 - \Pr_{\text{SC}}\{\text{ICS}_i\}] \end{aligned}$$

$$\begin{aligned}
&\geq \prod_{i=2}^k (1 - \beta) \\
&= \left\{ 1 - \left( 1 - (1 - \alpha)^{1/(k-1)} \right) \right\}^{k-1} = 1 - \alpha,
\end{aligned}$$

where the last inequality comes from (3).  $\square$

**Corollary 1** *If  $\mu_1 < \mu_2 + \delta$ , then with probability  $\geq 1 - \alpha$  the proposed procedure selects one of the systems whose means are within  $\delta$  of  $\mu_1$ .*

**Proof:** The proof is similar to that of Corollary 1 of Kim and Nelson (2001).  $\square$

Corollary 1 guarantees that the proposed procedure will choose one of the “good” systems with at least  $1 - \alpha$  guarantee when an alternative is less than  $\delta$  better than the other alternatives.

## 2.2 Second Procedure

In this section, rather than presenting the second procedure that we call  $\mathcal{P}_2$ , we only describe the differences between  $\mathcal{P}_1$  and  $\mathcal{P}_2$ , and give the full details in the Appendix.

Basically,  $\mathcal{P}_2$  is a natural extension of Zhu et al. (2005) to unknown variance case. It is very similar to  $\mathcal{P}_1$  but  $\mathcal{P}_2$  takes a different continuation region  $R_{i\ell}(r)$  and parameter  $\xi$ . More specifically, in **Screening** the continuation region  $R_{i\ell}(r)$  for  $\mathcal{P}_2$  is defined as

$$R_{i\ell}(r) = \lambda S_{i\ell} \sqrt{n_0 - 1} \sqrt{\max \left\{ \frac{(n_0 - 1) S_{i\ell}^2 \xi}{\delta^2} - r, 0 \right\}},$$

and **Parameters** needs to be revised as follows:  $\lambda = 0.3$  is recommended for  $1 - \alpha = 0.95$ —the detail of the choice of  $\lambda$  is discussed in Section 3.1—and the constant  $\xi$  is the solution to the equation

$$h(\xi; \lambda, n_0) = \beta, \tag{5}$$

where

$$h(\xi; \lambda, n_0) \equiv \mathbb{E} \left[ \frac{1}{2} - \frac{1}{2\sqrt{2\pi}} \exp \left\{ -\chi_{n_0-1}^2 \frac{\xi}{2} \right\} \sum_{n=0}^{\infty} \left( \frac{k_{2n+1}(\lambda \sqrt{\chi_{n_0-1}^2})}{\ell_{2n+1}(\lambda \sqrt{\chi_{n_0-1}^2})} \frac{2^n n!}{(2n+1)!} \left( \sqrt{\chi_{n_0-1}^2} \xi \right)^{2n+1} \right) \right].$$

**Theorem 2** *Under the same assumptions as in Theorem 1,  $\mathcal{P}_2$  selects system 1 with probability greater than or equal to  $1 - \alpha$ .*

The theorem is proven in the Appendix.

### 3 Design of the Procedures

In this section, we discuss the implementation issues of  $\mathcal{P}_1$  and  $\mathcal{P}_2$ . The main issues involve the determination of  $\lambda$  and  $\xi$ . Besides the user-specified parameters  $\delta, \alpha$ , and  $n_0$  that are required by all the indifference-zone procedures,  $\mathcal{P}_1$  and  $\mathcal{P}_2$  require two additional parameters  $(\lambda, \xi)$ . The parameter  $\lambda$  is another user-specified parameter but  $\xi$  is obtained from Equation (1) or (5). Procedure  $\mathcal{KN}$  also requires two parameters, say,  $c$  and  $\eta$ . The parameters  $c$  and  $\eta$  of  $\mathcal{KN}$  are similar to  $\lambda$  and  $\xi$ , respectively. The constant  $c$  in  $\mathcal{KN}$  is restricted to any positive integer, and the unique solution of  $\eta$  exists in a closed form with the choice of  $c = 1$  (Kim and Nelson 2001). However, in our procedures,  $\lambda$  can be any positive real number and the solution  $\xi$  to Equation (1) or (5) does not exist in a closed form. Therefore, in this section, we discuss which values of  $\lambda$  to choose and how to determine  $\xi$  quickly such that it satisfies Equation (1) or (5).

#### 3.1 Choice of $\lambda$

Our work is based on Ferebee (1982) that presents a statistical procedure that determines whether the drift of a Brownian motion process is positive or negative when variance is known. They derive the expected first exit time of the process through a parabolic boundary, and recommend that one chooses the value of  $\lambda$  that minimizes the expected first exit time when the drift is zero. Zhu et al. (2005) also determined the values of  $\lambda$  in this manner.

In our case, due to unknown variances, we doubt if there exists a closed form of the expected first exit time. Therefore, we find a value  $\lambda$  that minimizes the area of the continuation region, which is proportional to the product of the vertical and horizontal intercepts.

Kim and Nelson (2001) also used the area of their continuation region to find the best choice of  $c$ . For example, in  $\mathcal{P}_1$ , since the product of the horizontal and vertical intercepts is equal to  $\lambda\xi^{3/2}(n_0 - 1)^3 S_{i\ell}^6 / \delta^3$ , we only need to consider how  $\lambda\xi^{3/2}$  changes with respect to  $\lambda$  since  $n_0$ ,  $\delta$ , and  $S_{i\ell}^2$  are not affected by the choice of  $\lambda$ . In  $\mathcal{P}_2$ , we also get the area of the continuation region proportional to  $\lambda\xi^{3/2}$ . Recall that  $\lambda$  can be arbitrarily chosen as any real positive number while  $\xi$  is the solution to  $g(\xi; \lambda, n_0) = \beta$  or  $h(\xi; \lambda, n_0) = \beta$  where  $\beta$  is either  $\alpha/(k - 1)$  or  $1 - (1 - \alpha)^{1/(k-1)}$  depending on the use of CRN.

Figure 2 shows the graph of  $\lambda\xi^{3/2}$  versus  $\lambda$  when  $n_0 = 10$  and  $\beta = 0.05$  in  $\mathcal{P}_1$ . As  $n_0$  increases, the graph becomes even flatter around its minimum, especially, for  $0.9 < \lambda < 1.6$ . When  $\beta$  decreases due to a decrease in  $\alpha$  or increase in  $k$ , the graph tends to shift to the right. The range of  $\alpha$  of general interest is  $0.01 \leq \alpha \leq 0.1$  and this leads to the range of  $\beta$  as  $0 < \beta \leq 0.1$ . For this range of  $\beta$ , the minimum of  $\lambda\xi^{3/2}$  is still achieved around  $\lambda = 1.0$ . Since we usually take  $n_0$  larger than 10—in this paper we recommend  $n_0 = 24$ , which will be discussed in next subsection—and the graph does not dramatically shift to the right when  $\beta$  changes, we recommend taking  $\lambda = 0.9268$  for all values of  $\alpha$ . This is also the choice of Ferebee (1982) when  $\beta = 0.05$ .

Figure 3 shows the graph of  $\lambda\xi^{3/2}$  versus  $\lambda$  when  $n_0 = 24$  and  $\beta = 0.05$  in  $\mathcal{P}_2$ . The minimum is achieved around  $\lambda = 0.3$ . Unlike  $\mathcal{P}_1$ , as  $n_0$  decreases, the graph of  $\mathcal{P}_2$  becomes flatter around its minimum—we conjecture that this reverse tendency is due to the difference in Equations (1) and (5). For  $0 < \beta \leq 0.01$  the minimum is still achieved around  $\lambda = 0.3$ ; therefore, we recommend taking  $\lambda = 0.3$  for all values of  $\alpha$ .

### 3.2 Choice of $\xi$

Once we pick a value for  $\lambda$ , we can determine the value of  $\xi$  from Equation (1) or (5) for given  $n_0$  and  $\beta$  in  $\mathcal{P}_1$  or  $\mathcal{P}_2$ , respectively. Since the expectations are not analytically solvable and quite complicated to be solved by numerical integration, we find them by simulation. More specifically, for given  $n_0$  and  $\lambda$ , we generated four million chi-squared random variables

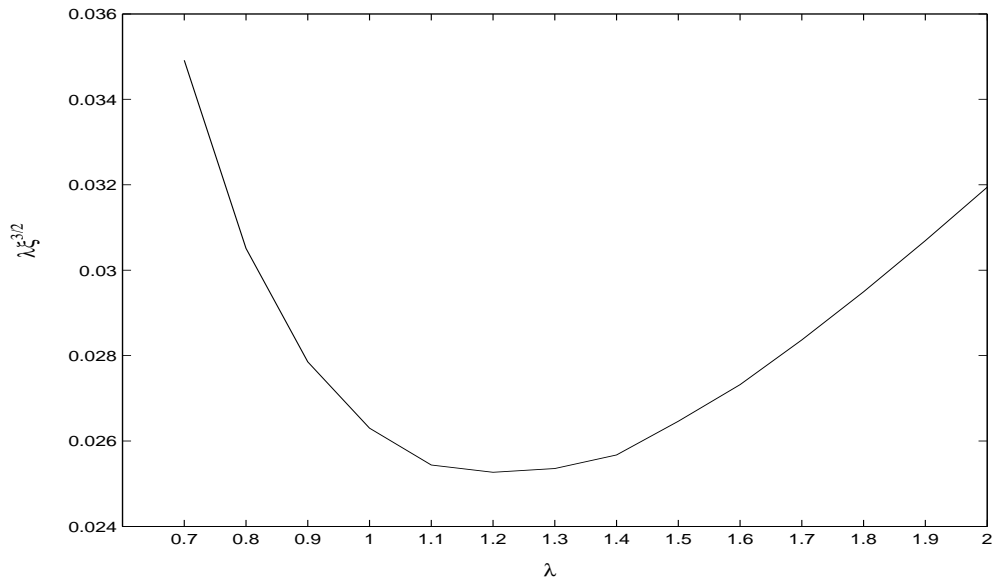


Figure 2:  $\lambda \xi^{3/2}$  vs  $\lambda$  when  $n_0 = 10$  and  $\beta = 0.05$  in  $\mathcal{P}_1$ .

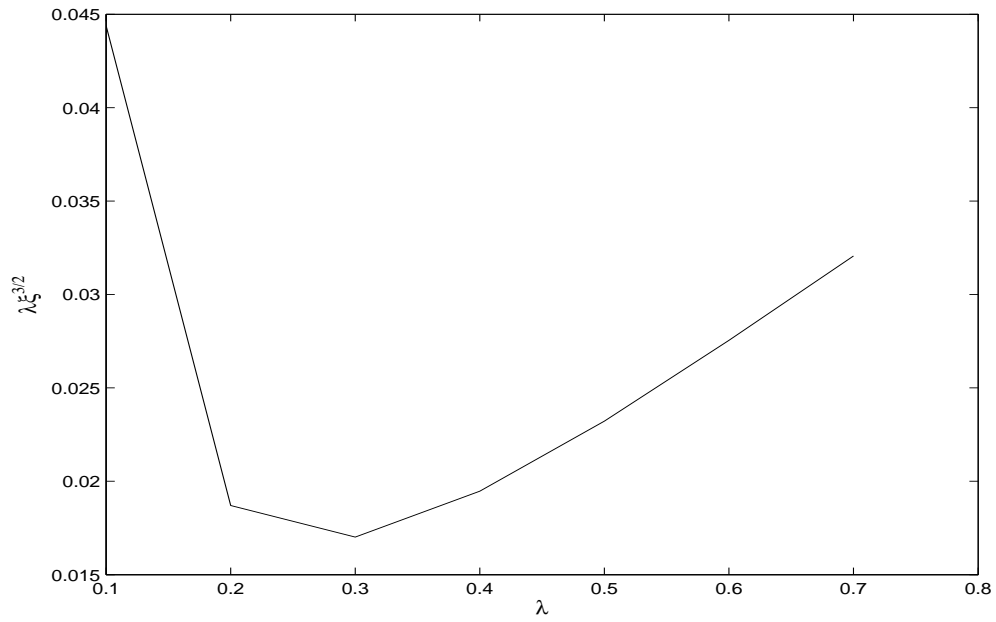


Figure 3:  $\lambda \xi^{3/2}$  vs  $\lambda$  when  $n_0 = 24$  and  $\beta = 0.05$  in  $\mathcal{P}_2$ .

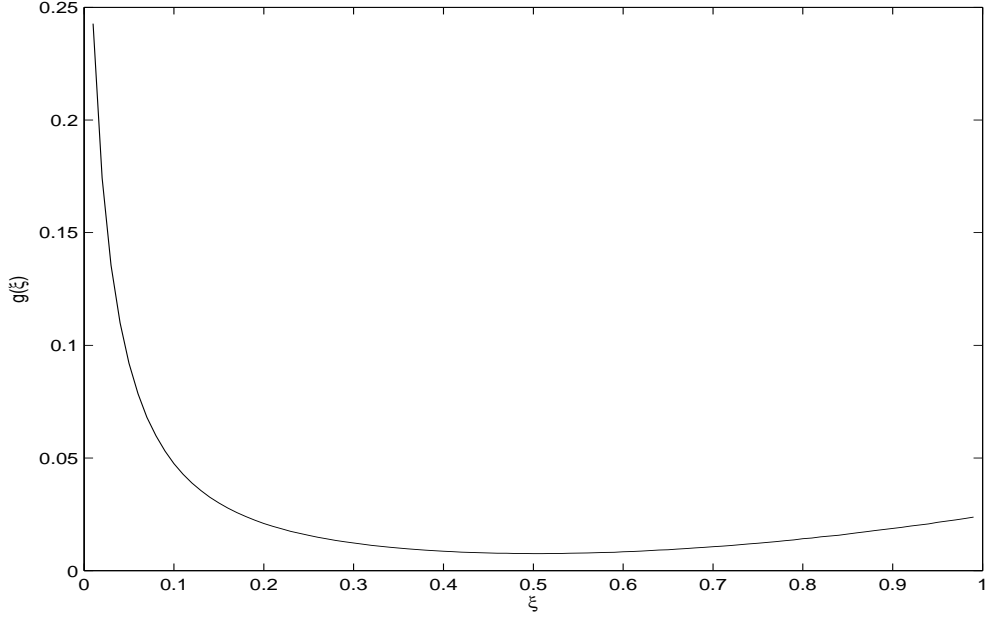


Figure 4:  $g(\xi; \lambda, n_0)$  versus  $\xi$  graph when  $n_0 = 10$  and  $\lambda = 0.9268$  in  $\mathcal{P}_1$ .

to estimate the expectation in Equation (1) or (5) at different values of  $\xi$  and created tables of  $g(\xi; \lambda, n_0)$  and  $h(\xi; \lambda, n_0)$  with respect to  $\xi$  from 0.01 to 1 for  $\mathcal{P}_1$  and from 0.01 to 7 for  $\mathcal{P}_2$  with an increment of 0.01. Then, one can find  $\xi$  at which the estimated expectation is equal to  $\beta$  by doing interpolation.

The difficulty in determination of  $\xi$  is that the existence of  $\xi$  is not guaranteed for all possible values of  $\beta$ , especially, for small  $\beta$ . The parameter  $\beta$ , determined by  $\alpha$  and  $k$ , gets smaller and approaches to zero as  $k$  increases. Therefore, we also need  $g(\xi; \lambda, n_0)$  or  $h(\xi; \lambda, n_0)$  to decrease as  $\xi$  increases in order to guarantee that  $\xi$  exists such that  $g(\xi; \lambda, n_0)$  or  $h(\xi; \lambda, n_0)$  is equal to  $\beta$  for all possible values of  $\beta$ . However,  $g(\xi; \lambda, n_0)$  and  $h(\xi; \lambda, n_0)$  for some  $n_0$  values do not approach to zero as  $\xi$  increases. Instead, they have a non-zero positive minimum which we denote as  $\beta_{min}$ . This implies that if  $\beta$  is smaller than  $\beta_{min}$ , then  $\xi$  does not exist such that  $g(\xi; \lambda, n_0) = \beta$  or  $h(\xi; \lambda, n_0) = \beta$ . Since  $\xi$  can be computed up to a quite small  $\beta$  for  $n_0 > 15$  in  $\mathcal{P}_2$ , this problem is more severe in  $\mathcal{P}_1$  compared to  $\mathcal{P}_2$ . For example, Figure 4 shows that when  $n_0 = 10$  and  $\lambda = 0.9268$ ,  $\beta_{min}$  is 0.0076 for  $\mathcal{P}_1$ . This implies that if  $\alpha = 0.05$ , then  $\xi$  for  $\mathcal{P}_1$  is determined only up to  $k = 7$  in which case  $\beta = \alpha/(k-1) \simeq 0.0083$ .

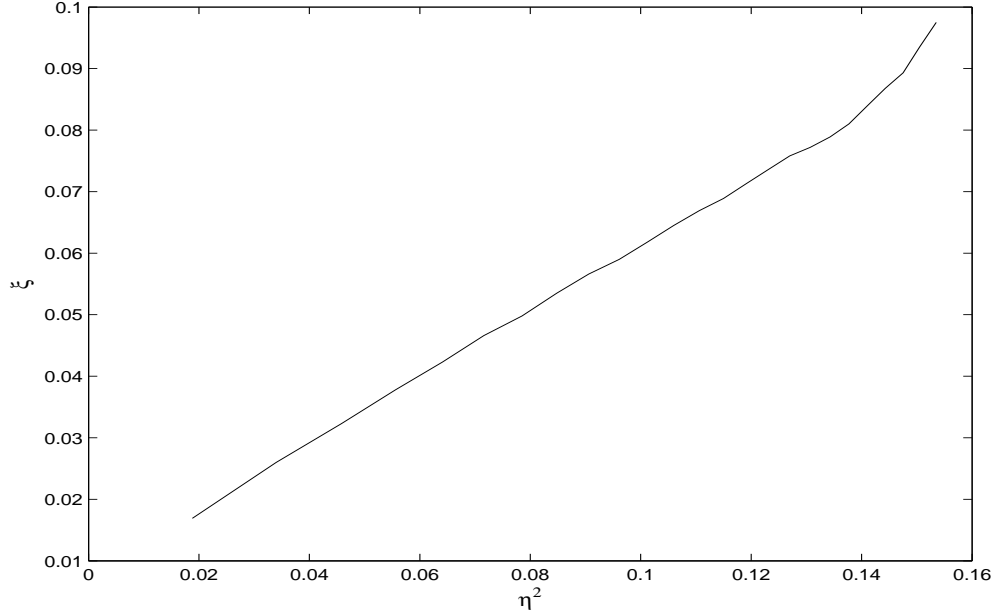


Figure 5:  $\eta^2$  versus  $\xi$  graph when  $n_0 = 20$ ,  $\lambda = 0.9268$ , and  $0.002050 < \beta < 0.05$  in  $\mathcal{P}_1$ .

Kim and Nelson (2001) show that the parameter  $\eta$  when  $c = 1$  is determined by

$$\eta = \frac{1}{2} \left[ (2\beta)^{-2/(n_0-1)} - 1 \right], \quad (6)$$

and we observed that  $\eta^2$  and  $\xi$  seem to have an approximate linear relationship for the range of  $\beta$  where we are able to get  $\xi$  from  $g(\xi; \lambda, n_0) = \beta$  or  $h(\xi; \lambda, n_0) = \beta$ , that is, for the range of  $\beta$  greater than or equal to  $\beta_{min}$ . Figure 5 shows an example of the linear relationship between  $\eta^2$  and  $\xi$  when  $n_0 = 20$ ,  $\lambda = 0.9268$ , and  $0.002050 < \beta < 0.05$  in  $\mathcal{P}_1$ . Similarly, Figure 6 shows the linear relationship between  $\eta^2$  and  $\xi$  when  $n_0 = 10$ ,  $\lambda = 0.3$ , and  $0.00244 < \beta < 0.05$  in  $\mathcal{P}_2$ . We observed a similar linear relationship for all other  $n_0$  values. We will use this linear relationship to get  $\xi$  for  $\beta < \beta_{min}$ .

We will add one more subscript to  $\xi$  and  $\eta$  and use  $\xi_\beta$  and  $\eta_\beta$  to represent values of  $\xi$  and  $\eta$  at specific  $\beta$ , respectively. As shown in Figure 7, we use the points  $(\eta_\alpha^2, \xi_\alpha)$  and  $(\eta_{\beta_{min}}^2, \xi_{\beta_{min}})$  to determine the line

$$Y - \xi_\alpha = (x - \eta_\alpha^2) \frac{\xi_{\beta_{min}} - \xi_\alpha}{\eta_{\beta_{min}}^2 - \eta_\alpha^2}.$$



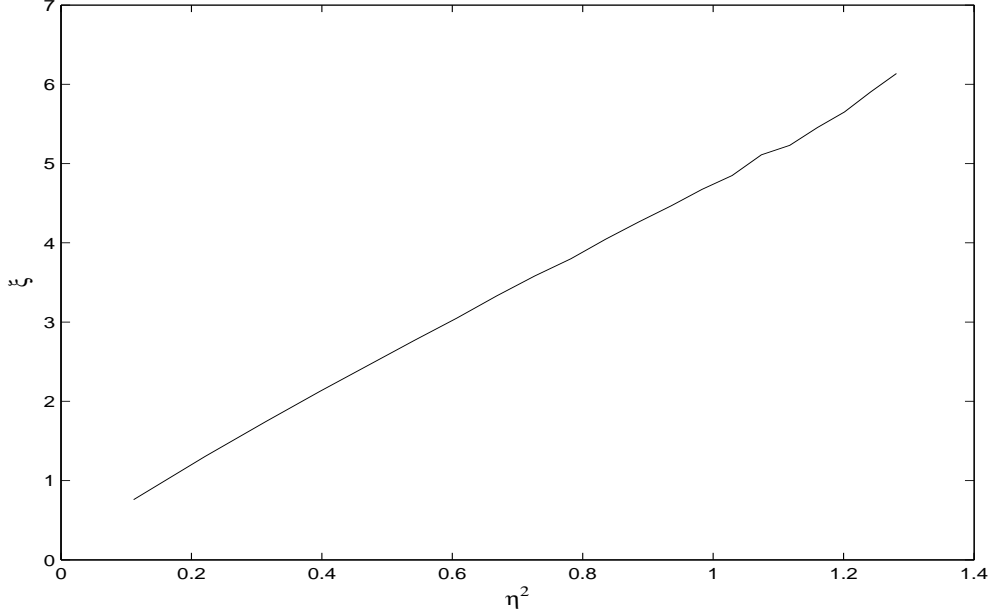


Figure 6:  $\eta^2$  versus  $\xi$  graph when  $n_0 = 10$ ,  $\lambda = 0.3$ , and  $0.00244 < \beta < 0.05$  in  $\mathcal{P}_2$ .

If the linear tendency continues,  $\xi_\beta$  values for  $\beta < \beta_{min}$  can be approximated by

$$Y = (\eta_\beta^2 - \eta_\alpha^2) \frac{\xi_{\beta_{min}} - \xi_\alpha}{\eta_{\beta_{min}}^2 - \eta_\alpha^2} + \xi_\alpha.$$

Therefore, for the approximation of  $\xi$  for  $\beta < \beta_{min}$ , we only need the information about the coordinates of two points:  $(\eta_\alpha^2, \xi_\alpha)$  and  $(\eta_{\beta_{min}}^2, \xi_{\beta_{min}})$ . The values of  $\eta$  can be easily computed from Equation (6) and the values of  $\xi_\alpha$  and  $\xi_{\beta_{min}}$  for the popular choices of  $n_0$  and  $\alpha$  when  $n_0 = 10, 15, 20, 24$  and  $\alpha = 0.01, 0.025, 0.05, 0.10$  can be found in Table 1 for  $\mathcal{P}_1$  and in Table 2 for  $\mathcal{P}_2$ .

Table 3 shows the actual  $\xi_\beta$  found from  $g(\xi; n_0, \lambda) = \beta$  and approximate  $\xi_\beta$  from the linear relationship when  $n_0 = 20$  and  $\lambda = 0.9268$  in  $\mathcal{P}_1$ . The approximation errors are quite small for the range of  $\beta_{min} < \beta \leq \alpha$ . Though there is an approximate linear tendency between  $\eta^2$  and  $\xi$  for all  $n_0$  and  $\beta_{min} < \beta \leq 0.05$ , it does not necessarily guarantee the linear tendency for  $\beta < \beta_{min}$  and a deviation from the linearity assumption is possible as  $\beta$  gets smaller than  $\beta_{min}$ . When  $n_0 = 24$ ,  $\xi$  is determined up to a quite small  $\beta$  by Equations (1) and (5). Therefore, we recommend using 24 as the initial sample size in order to make sure that the linear relationship holds up to a quite small  $\beta$ .

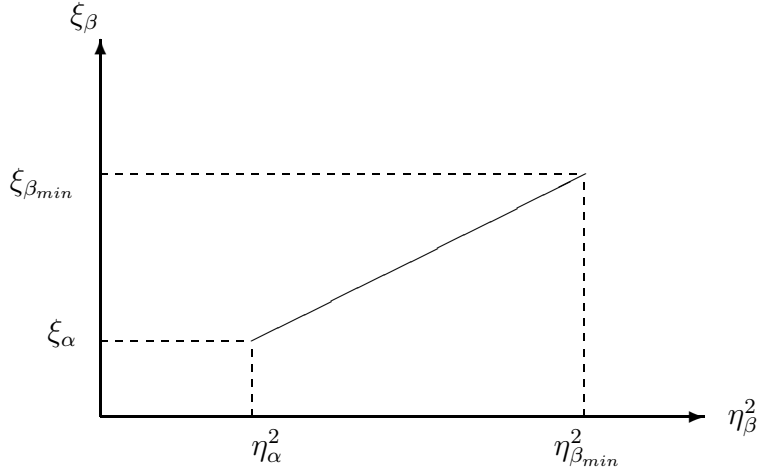


Figure 7: Linearity relationship between  $\eta_\beta^2$  and  $\xi_\beta$ .

Table 1:  $\xi_{\beta_{min}}$  and  $\xi_\alpha$  values when  $\lambda = 0.9268$ ,  $n_0 = 10, 15, 20, 24$ , and  $\alpha = 0.01, 0.025, 0.05, 0.10$  in  $\mathcal{P}_1$ .

$n_0$	$1 - \beta_{min}$	$\xi_{\beta_{min}}$	$\xi_{0.01}$	$\xi_{0.025}$	$\xi_{0.05}$	$\xi_{0.10}$
10	0.9924	0.49	0.3500	0.1744	0.09554	0.04556
15	0.9956	0.17	0.09474	0.05366	0.03194	0.01720
20	0.9980	0.10	0.04282	0.02618	0.01686	0.008229
24	0.99954	0.10	0.02742	0.01755	0.01	0.005457

Table 2:  $\xi_{\beta_{min}}$  and  $\xi_\alpha$  values when  $\lambda = 0.3$ ,  $n_0 = 10, 15, 20, 24$ , and  $\alpha = 0.01, 0.025, 0.05, 0.10$  in  $\mathcal{P}_2$ .

$n_0$	$1 - \beta_{min}$	$\xi_{\beta_{min}}$	$\xi_{0.01}$	$\xi_{0.025}$	$\xi_{0.05}$	$\xi_{0.10}$
10	0.9976	6.14	2.49	1.3167	0.7543	0.3809
15	0.99973	4.11	0.84	0.5062	0.3196	0.1763
20	0.99996	2.96	0.46	0.2956	0.1947	0.1116
24	0.99999	2.30	0.3329	0.2196	0.1476	0.08608

Table 3: Actual and approximate  $\xi_\beta$  values when  $n_0 = 20$  and  $\lambda = 0.9268$  in  $\mathcal{P}_1$ .

$\beta$	Act. $\xi_\beta$	Appr. $\xi_\beta$
0.05	0.0169	0.0169
0.025321	0.0260	0.0260
0.016952	0.0323	0.0330
0.012741	0.0378	0.0389
0.010206	0.0423	0.0440
0.008512	0.0466	0.0485
0.007301	0.0498	0.0526
0.006391	0.0535	0.0564
0.005683	0.0566	0.0598
0.005116	0.0590	0.0631
0.004652	0.0618	0.0662
0.004265	0.0645	0.0691
0.003938	0.0669	0.0718
0.003657	0.0689	0.0744
0.003414	0.0713	0.0769
0.003201	0.0736	0.0793
0.003013	0.0758	0.0816
0.002846	0.0772	0.0838
0.002696	0.0789	0.0860
0.002561	0.0810	0.0881
0.00244	0.0840	0.0901
0.002329	0.0868	0.0920
0.002228	0.0893	0.0939
0.002135	0.0935	0.0957
0.002050	0.0975	0.0975

## 4 Analytical Comparisons of the Continuation Regions

In this section, we perform analytical comparisons among the parabolic regions of  $\mathcal{P}_1$  and  $\mathcal{P}_2$  and the triangular region of  $\mathcal{KN}$ . We will use  $R(t)$  to denote the generic continuation region, and  $H$  and  $V$  to denote intercepts of  $R(t)$  with the horizontal and vertical axes, respectively.

Kim and Nelson (2001) state that if the experimenter has no idea whether there are a few dominant systems or a number of close competitors, the choice of  $c = 1$  appears to be a good compromise solution. Hence, we will take  $c = 1$  throughout this paper. The triangular

continuation region for  $\mathcal{KN}$  when  $c = 1$  is defined as

$$R(t)_{\mathcal{KN}} = \left( -\frac{\eta(n_0 - 1)S_{il}^2}{\delta} + \frac{\delta}{2}t, \frac{\eta(n_0 - 1)S_{il}^2}{\delta} - \frac{\delta}{2}t \right), \quad 0 \leq t \leq \frac{2\eta(n_0 - 1)S_{il}^2}{\delta^2}.$$

Then, the intercepts of  $\mathcal{KN}$ ,  $\mathcal{P}_1$ , and  $\mathcal{P}_2$  are

$$\begin{aligned} V_{\mathcal{KN}} &= \frac{\eta(n_0 - 1)S_{il}^2}{\delta}, & H_{\mathcal{KN}} &= \frac{2\eta(n_0 - 1)S_{il}^2}{\delta^2}, \\ V_{\mathcal{P}_1} &= \frac{\lambda_1\sqrt{\xi_1}(n_0 - 1)S_{il}^2}{\delta}, & H_{\mathcal{P}_1} &= \frac{\xi_1(n_0 - 1)^2S_{il}^4}{\delta^2}, \\ V_{\mathcal{P}_2} &= \frac{\lambda_2\sqrt{\xi_2}(n_0 - 1)S_{il}^2}{\delta}, & H_{\mathcal{P}_2} &= \frac{\xi_2(n_0 - 1)S_{il}^2}{\delta^2}. \end{aligned}$$

The analysis in this section is based on  $\xi_1$ ,  $\xi_2$ , and  $\eta$  values assuming that systems are simulated independently and  $1 - \alpha = 0.95$ . The results are slightly affected if CRN is employed.

Given that all three procedures are statistically valid, we prefer the one with smaller  $V$  and smaller  $H$  since it will provide a tighter screening step and guarantee that the procedure ends earlier if sampling continues to the end stage. More specifically, a smaller  $V$  means tighter screening at the beginning; therefore, a procedure with smaller  $V$  will work efficiently if there exist a few dominant systems. However, if all systems are close to the best system (e.g., the SC), then each system is likely to take a number of observations close to  $H$  and a procedure with a smaller  $H$  will be preferred. Therefore, if the continuation region of a procedure is strictly inside of that of another procedure, then it is clear that the former will dominate the latter. For the comparison, we will focus on the values of  $H$  and  $V$  and the shape of the continuation regions.

**$\mathcal{P}_1$  vs.  $\mathcal{KN}$ :** One can easily compute the ratio of vertical intercepts of  $\mathcal{KN}$  and  $\mathcal{P}_1$ . The ratio  $V_{\mathcal{KN}}/V_{\mathcal{P}_1}$  is equal to  $\eta/(\lambda_1\sqrt{\xi_1})$ . Table 4 shows the ratios of  $V_{\mathcal{KN}}$  and  $V_{\mathcal{P}_1}$  for  $n_0 = 10, 24$  and  $k = 2, 5, 10, 25, 100, 200, 400$ . From these results, we conjecture that  $\mathcal{P}_1$  has a tighter screening at the beginning than  $\mathcal{KN}$ . This tightness tends to be stronger as the number of systems  $k$  increases, but not always.

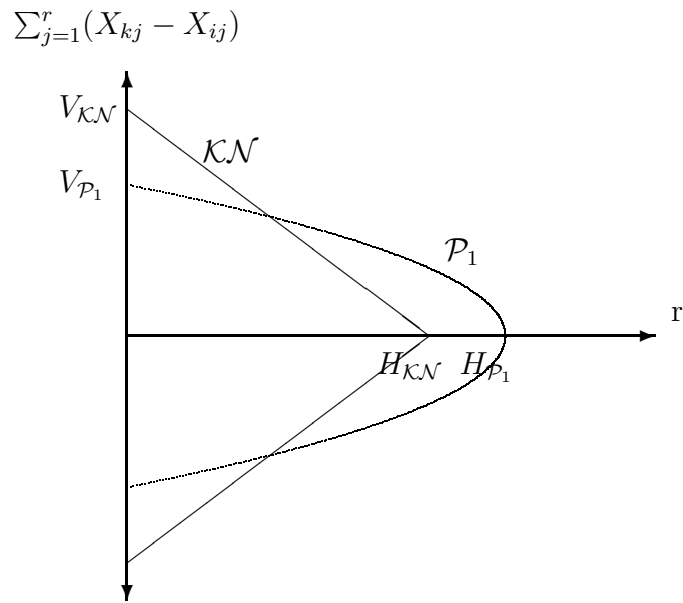


Figure 8: Continuation regions of  $\mathcal{P}_1$  and  $\mathcal{KN}$ .

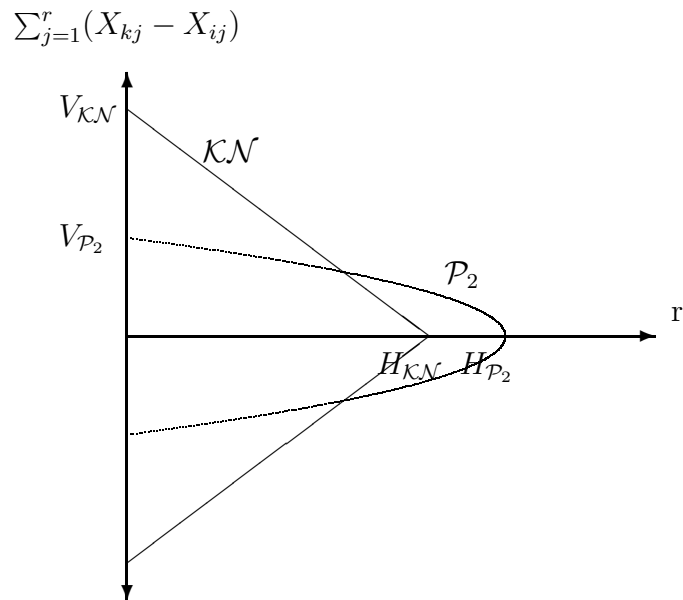


Figure 9: Continuation regions of  $\mathcal{P}_2$  and  $\mathcal{KN}$  for large  $k$ .

Table 4:  $V_{\mathcal{KN}}/V_{\mathcal{P}_1}$  when  $n_0 = 10, 24$  and  $k = 2, 5, 10, 25, 100, 200, 400$ .

	$n_0 = 10$	$n_0 = 24$
$k = 2$	1.166	1.196
$k = 5$	1.260	1.295
$k = 10$	1.218	1.366
$k = 25$	1.264	1.434
$k = 100$	1.269	1.476
$k = 200$	1.270	1.433
$k = 400$	1.271	1.437

Now, consider the horizontal intercepts of the two procedures. The ratio  $H_{\mathcal{KN}}/H_{\mathcal{P}_1}$  is equal to  $2\eta/((n_0 - 1)S_{i\ell}^2\xi_1)$ . Since  $S_{i\ell}^2$  is random, it is hard to evaluate this ratio. However, in general, when  $S_{i\ell}^2 > 1$ , this ratio is less than one and  $H_{\mathcal{P}_1}$  is larger than  $H_{\mathcal{KN}}$ .

From this analysis, one can conjecture that  $\mathcal{P}_1$  starts with a tighter screening than  $\mathcal{KN}$  but after some point, the screening of  $\mathcal{KN}$  becomes tighter as one can see in Figure 8. It implies that when  $k$  is large and there exist a number of dominant systems,  $\mathcal{P}_1$  is expected to be more efficient due to its tighter screening at the beginning. However, under a configuration close to the SC where the procedure is likely to continue to the end stage, we expect  $\mathcal{KN}$  to be more efficient than  $\mathcal{P}_1$ .

$\mathcal{P}_2$  vs.  $\mathcal{KN}$ : The ratio  $V_{\mathcal{KN}}/V_{\mathcal{P}_2}$  is equal to  $\eta/(\lambda_2\sqrt{\xi_2})$ . Table 5 shows  $V_{\mathcal{KN}}/V_{\mathcal{P}_2}$  and  $H_{\mathcal{KN}}/H_{\mathcal{P}_2}$  ratios for  $n_0 = 10, 24$  and  $k = 2, 5, 10, 25, 100, 200, 400$ . The numbers in the table imply that  $\mathcal{P}_2$  is tighter than  $\mathcal{KN}$  at the beginning of the procedure and it becomes much more tighter as  $k$  increases.

The ratio  $H_{\mathcal{KN}}/H_{\mathcal{P}_2}$  is equal to  $2\eta/\xi_2$ .  $H_{\mathcal{KN}}$  is always smaller than  $H_{\mathcal{P}_2}$  when  $n_0 = 10$ . However, when  $n_0 = 24$ ,  $H_{\mathcal{KN}}$  is larger than  $H_{\mathcal{P}_2}$  for  $k < 50$ . Therefore, when  $n_0 = 24$  and  $k$  is small (say  $k < 50$ ),  $\mathcal{P}_2$  is likely to beat  $\mathcal{KN}$  because both  $V$  and  $H$  of  $\mathcal{P}_2$  are smaller than those of  $\mathcal{KN}$ —implying that it is possible that the continuation region of  $\mathcal{P}_2$  is inside that of  $\mathcal{KN}$ . On the other hand, as one can see in Table 5, for large  $k$  ( $k \geq 50$ ), there does not exist a uniform superiority between  $\mathcal{KN}$  and  $\mathcal{P}_2$  because one has a smaller  $V$  but the other has a

Table 5:  $V_{\mathcal{KN}}/V_{\mathcal{P}_2}$  and  $H_{\mathcal{KN}}/H_{\mathcal{P}_2}$  when  $n_0 = 10, 24$  and  $k = 2, 5, 10, 25, 100, 200, 400$

	$n_0 = 10$		$n_0 = 24$	
	$V_{\mathcal{KN}}/V_{\mathcal{P}_2}$	$H_{\mathcal{KN}}/H_{\mathcal{P}_2}$	$V_{\mathcal{KN}}/V_{\mathcal{P}_2}$	$H_{\mathcal{KN}}/H_{\mathcal{P}_2}$
$k = 2$	1.282	0.886	0.961	1.501
$k = 5$	1.441	0.593	1.143	1.250
$k = 10$	1.501	0.476	1.229	1.142
$k = 25$	1.526	0.355	1.329	1.047
$k = 100$	1.541	0.237	1.445	0.919
$k = 200$	1.506	0.187	1.487	0.855
$k = 400$	1.548	0.164	1.532	0.803

Table 6:  $V_{\mathcal{P}_1}/V_{\mathcal{P}_2}$  when  $n_0 = 10, 24$  and  $k = 2, 5, 10, 25, 100, 200, 400$ .

	$n_0 = 10$	$n_0 = 24$
$k = 2$	1.099	0.804
$k = 5$	1.143	0.882
$k = 10$	1.232	0.899
$k = 25$	1.207	0.926
$k = 100$	1.215	0.979
$k = 200$	1.185	1.038
$k = 400$	1.218	1.066

smaller  $H$ . Figure 9 demonstrates possible continuation regions of these two procedures for large  $k$ . From this figure, we conjecture that  $\mathcal{P}_2$  will show better performance than  $\mathcal{KN}$  in general due to a tighter initial screening. However, when most systems are close to the best system like the SC where the procedure is likely to end around  $H$ , it is possible that  $\mathcal{KN}$  beats  $\mathcal{P}_2$ .

**$\mathcal{P}_1$  Vs.  $\mathcal{P}_2$ :** The ratio  $V_{\mathcal{P}_1}/V_{\mathcal{P}_2}$  is equal to  $(\lambda_1\sqrt{\xi_1})/(\lambda_2\sqrt{\xi_2})$ . Table 6 shows the ratios of  $V_{\mathcal{P}_1}$  and  $V_{\mathcal{P}_2}$  for  $n_0 = 10, 24$  and  $k = 2, 5, 10, 25, 100, 200, 400$ . The numbers in the table imply that  $\mathcal{P}_2$  is a little bit tighter than  $\mathcal{P}_1$  either when  $n_0 = 10$  or when  $n_0 = 24$  and  $k$  is large, but  $V$  of  $\mathcal{P}_1$  is not significantly different from that of  $\mathcal{P}_2$ .

The ratio  $H_{\mathcal{P}_1}/H_{\mathcal{P}_2}$  is equal to  $\xi_1(n_0 - 1)S_{i\ell}^2/\xi_2$ . Although  $S_{i\ell}^2$  is random, it is clear that this ratio will be greater than one in general. Actually, it can be quite large because of

$(n_0 - 1)$  and  $S_{i\ell}^2$  factors, implying that  $H$  of  $\mathcal{P}_1$  is likely to be a lot more larger than that of  $\mathcal{P}_2$ .

As a result, we can say that except the case when  $n_0 = 24$  and  $k$  is small,  $\mathcal{P}_2$  is likely to outperform  $\mathcal{P}_1$  because both  $V$  and  $H$  of  $\mathcal{P}_2$  are smaller than those of  $\mathcal{P}_1$ . When  $n_0 = 24$  and  $k$  is small,  $\mathcal{P}_1$  has a little bit tighter screening than  $\mathcal{P}_2$  at the beginning. However,  $H_{\mathcal{P}_1}$  is usually much larger than  $H_{\mathcal{P}_2}$  given that  $n_0$  is larger than 10 and  $S_{i\ell}^2$  is likely to be larger than one. This can overshadow the advantage of  $\mathcal{P}_1$  having a smaller  $V_{\mathcal{P}_1}$ . Thus, it is not clear which procedure will do better for small  $k$ .

## 5 Experimental Results

In this section, we compare the experimental results of  $\mathcal{KN}$ ,  $\mathcal{P}_1$ , and  $\mathcal{P}_2$  based on i.i.d. normal data. In Kim and Nelson (2001), the performance of  $\mathcal{KN}$  procedure is compared to that of Rinott's (1978) procedure and a two-stage screen-and-select procedure proposed by Nelson, Swann, Goldsman, and Song (2000), which are both statistically valid indifference-zone selection procedures. The performance is evaluated in terms of the sample average of the total number of basic observations required by each procedure when all the procedures achieve the nominal PCS. These results show that  $\mathcal{KN}$  procedure is highly efficient for finding the best system compared to the other two procedures. Therefore, we use  $\mathcal{KN}$  as our benchmark for comparison and test the performance of our procedures on the same scenarios Kim and Nelson (2001) used.

Two configurations of the true means are employed: slippage configuration (SC) and monotonically decreasing mean (MDM) configuration. For each mean configuration, three variance configurations are tested: increasing variance (IV), decreasing variance (DV), and common variance (CV). When we combine these configurations, we get a total of six different configurations. The particular mean and variance values that we used for experiments are given in Table 7. Notice that system 1 is set to be the best system.

The nominal PCS is set at  $1 - \alpha = 0.95$ . As suggested in Section 3,  $\lambda$  and  $n_0$  are set



Table 7: The configurations of means and variances used in the experiments.

SC-CV	mean variance	$\mu_1 = \delta, \mu_2 = \mu_3 = \dots = 0$ $\sigma_1^2 = \sigma_2^2 = \dots = 1$
MDM-CV	mean variance	$\mu_i = \delta - (i - 1)\delta, i \geq 1$ $\sigma_1^2 = \sigma_2^2 = \dots = 1$
MDM-IV	mean variance	$\mu_i = \delta - (i - 1)\delta, i \geq 1$ $\sigma_i^2 = 1 + (i - 1)\delta, i \geq 1$
MDM-DV	mean variance	$\mu_i = \delta - (i - 1)\delta, i \geq 1$ $\sigma_i^2 = 1 / (1 + (i - 1)\delta), i \geq 1$
SC-IV	mean variance	$\mu_1 = \delta, \mu_2 = \mu_3 = \dots = 0$ $\sigma_i^2 = 1 + (i - 1)\delta, i \geq 1$
SC-DV	mean variance	$\mu_1 = \delta, \mu_2 = \mu_3 = \dots = 0$ $\sigma_i^2 = 1 / (1 + (i - 1)\delta), i \geq 1$

to  $\lambda = 0.9268$  for  $\mathcal{P}_1$  and 0.3 for  $\mathcal{P}_2$ , and  $n_0 = 24$  for all configurations, respectively. Each result is obtained from 500 macro-replications of the entire experiment. In order to check if each procedure is statistically valid, we record the actual PCS for each configuration. If all procedures seem statistically valid, the comparison will be done based on the sample average of the total number of observations required. With 500 macro-replications, first two digits of these sample averages are statistically meaningful.

Our experimental results support that both  $\mathcal{P}_1$  and  $\mathcal{P}_2$  are statistically valid since the actual PCS are all higher than the nominal PCS when systems are simulated independently, as shown in Table 8. For  $\mathcal{P}_2$ , we observed that there were a few configurations where the actual PCS were lower than 0.95 when we made 500 macro-replications. However, when the number of macro-replications was increased to 5000, the actual PCS of  $\mathcal{P}_2$  were all over 0.95. The numbers in Table 8 are from 500 macro-replications except the actual PCS of  $\mathcal{P}_2$  for  $k = 2$  and 5 that are from 5000 macro-replications. The actual PCS increase when CRN is employed though we did not report them in this paper. Therefore,  $\mathcal{KN}$ ,  $\mathcal{P}_1$ , and  $\mathcal{P}_2$  are all statistically valid and from now on, we will focus on the comparison among three procedures based on sample sizes required until we reach a decision.

We categorize our experimental results into two cases: independent and correlated cases.

In the independent case, all systems are simulated independently. In the correlated case, a constant correlation factor,  $\rho$ , is used between all pairs of systems.

## 5.1 Independent Case

We first compare the performance of the two new procedures with  $\mathcal{KN}$  when systems are simulated independently. Overall,  $\mathcal{P}_2$  shows the best performance among three in terms of the total number of observations required. However, the computational time needed to search for the  $\xi_2$  parameter for a given  $\beta$  is very long compared to that of  $\xi_1$  unless a table for  $\xi_2$  already exists. Therefore, if the gain in the simulation time when  $\mathcal{P}_2$  is applied is not significant enough to compensate the time loss in computing the  $\xi_2$  parameter, one may prefer to use  $\mathcal{P}_1$ , which also shows better performance than  $\mathcal{KN}$  under MDM configurations when  $k$  is large.

**$\mathcal{P}_1$  vs.  $\mathcal{KN}$ :** Table 9 shows the sample average of the total number of observations taken in  $\mathcal{KN}$ ,  $\mathcal{P}_1$ , and  $\mathcal{P}_2$  when  $n_0 = 24$ ,  $\delta = 1/\sqrt{n_0}$ , and  $1 - \alpha = 0.95$  as a function of  $k$ . The results of  $\mathcal{KN}$  are taken from Kim and Nelson (2001). Numbers in boldface represent the cases that  $\mathcal{P}_1$  or  $\mathcal{P}_2$  shows better performance—i.e., spends a smaller number of observations—than  $\mathcal{KN}$ . Numbers in a box show which procedure among those three spends the smallest observations for each configuration.

When  $k = 2$  or  $5$ ,  $\mathcal{KN}$  is superior to  $\mathcal{P}_1$ . However, as  $k$  increases, we observe that  $\mathcal{P}_1$  defeats  $\mathcal{KN}$  in all the MDM configurations and SC with decreasing variances. As expected in Section 4, this is due to the fact that  $\mathcal{P}_1$  has a tighter screening at the beginning and this tightness becomes even stronger as  $k$  increases. Therefore, under MDM configurations with large  $k$ ,  $\mathcal{P}_1$  does a better job in detecting inferior systems early. Similarly, under SC, if we have decreasing variances as means become inferior, it is easier to detect inferior systems due to low variances of inferior systems. These low variances help  $\mathcal{P}_1$  to eliminate inferior systems early in the experiment and, thus, to show better performance than  $\mathcal{KN}$  in SC-DV. However, in the SC-CV and SC-IV, most of elimination events are likely to take place at a

Table 8: Estimated PCS in  $\mathcal{KN}$ ,  $\mathcal{P}_1$ , and  $\mathcal{P}_2$  when  $n_0 = 24$ ,  $\delta = 1/\sqrt{n_0}$ , and  $1 - \alpha = 0.95$  as a function of  $k$  when systems are simulated independently.

		SC-CV	MDM-CV	MDM-IV	MDM-DV	SC-IV	SC-DV
$k = 2$	$\mathcal{KN}$	0.958	0.958	0.954	0.960	0.954	0.960
	$\mathcal{P}_1$	0.982	0.982	0.968	0.972	0.968	0.972
	$\mathcal{P}_2$	0.957	0.957	0.952	0.954	0.952	0.954
$k = 5$	$\mathcal{KN}$	0.960	0.988	0.986	0.982	0.958	0.972
	$\mathcal{P}_1$	0.982	0.992	0.992	0.996	0.972	0.986
	$\mathcal{P}_2$	0.962	0.987	0.988	0.989	0.960	0.963
$k = 10$	$\mathcal{KN}$	0.964	0.992	0.994	0.996	0.958	0.968
	$\mathcal{P}_1$	0.970	0.996	0.994	0.998	0.986	0.974
	$\mathcal{P}_2$	0.978	0.988	0.998	0.996	0.966	0.972
$k = 25$	$\mathcal{KN}$	0.964	0.998	1.000	1.000	0.950	0.982
	$\mathcal{P}_1$	0.992	1.000	0.998	1.000	0.984	0.980
	$\mathcal{P}_2$	0.962	0.996	1.000	0.998	0.964	0.984
$k = 100$	$\mathcal{KN}$	0.956	0.998	1.000	1.000	0.960	0.996
	$\mathcal{P}_1$	0.980	1.000	1.000	1.000	0.982	1.000
	$\mathcal{P}_2$	0.948	0.998	1.000	0.998	0.950	0.992

Table 9: Sample average of the total number of observations in  $\mathcal{KN}$ ,  $\mathcal{P}_1$ , and  $\mathcal{P}_2$  when  $n_0 = 24$ ,  $\delta = 1/\sqrt{n_0}$ , and  $1 - \alpha = 0.95$  as a function of  $k$ .

		SC-CV	MDM-CV	MDM-IV	MDM-DV	SC-IV	SC-DV
$k = 2$	$\mathcal{KN}$	158	158	175	147	175	147
	$\mathcal{P}_1$	187	187	207	167	207	167
	$\mathcal{P}_2$	172	172	185	154	185	154
$k = 5$	$\mathcal{KN}$	738	456	542	403	939	594
	$\mathcal{P}_1$	800	485	568	409	1050	643
	$\mathcal{P}_2$	725	448	523	380	940	575
$k = 10$	$\mathcal{KN}$	1727	761	981	630	2868	1149
	$\mathcal{P}_1$	1861	710	910	615	3049	1185
	$\mathcal{P}_2$	1689	695	890	612	2758	1124
$k = 25$	$\mathcal{KN}$	5015	1333	2009	1157	13399	2312
	$\mathcal{P}_1$	5109	1217	1647	1091	14154	2241
	$\mathcal{P}_2$	4705	1185	1639	1072	12336	2252
$k = 100$	$\mathcal{KN}$	23956	3483	6312	3215	189401	5331
	$\mathcal{P}_1$	24634	3303	4682	3101	19276	4990
	$\mathcal{P}_2$	21123	3235	4653	3052	170987	5210

stage close to the end of the procedure. We have seen that the screening of  $\mathcal{P}_1$  is tighter than that of  $\mathcal{KN}$  only at the beginning of experimentation, but it soon becomes loose and its horizontal-intercept is usually larger than that of  $\mathcal{KN}$ . Therefore,  $\mathcal{KN}$  tends to show better performance under these configurations.

The superiority of  $\mathcal{P}_1$  is most noticeable under MDM-IV. Under MDM-DV, it is already easy to detect inferior systems due to their small variances. Thus, having a tighter continuation region does not help that much, and  $\mathcal{KN}$  and  $\mathcal{P}_1$  show similar performance. However, under MDM-IV where it is more difficult to detect inferior systems due to their high variances, the tightness of  $\mathcal{P}_1$  stands out and helps  $\mathcal{P}_1$  to outperform  $\mathcal{KN}$ .

**$\mathcal{P}_2$  vs.  $\mathcal{KN}$ :** When  $k = 2$ ,  $\mathcal{KN}$  is better than  $\mathcal{P}_2$  under all configurations we tested. This is because  $V_{\mathcal{KN}} < V_{\mathcal{P}_2}$ —i.e.  $\mathcal{KN}$  has a tighter screening—although  $H_{\mathcal{KN}} > H_{\mathcal{P}_2}$  when  $k$  is small. However, for  $k \geq 5$ ,  $\mathcal{P}_2$  outperforms  $\mathcal{KN}$  in all configurations as expected in Section 4 because  $V_{\mathcal{KN}}$  becomes greater than  $V_{\mathcal{P}_2}$  as  $k$  increases although  $H_{\mathcal{KN}}$  becomes slightly less than  $H_{\mathcal{P}_2}$ .

In Table 5, we observe that when  $n_0 = 10$ , the ratio  $V_{\mathcal{KN}}/V_{\mathcal{P}_2}$  increases as  $k$  increases. However, due to the sharp decrease in the ratio  $H_{\mathcal{KN}}/H_{\mathcal{P}_2}$ ,  $\mathcal{KN}$  shows better performance than  $\mathcal{P}_2$  under SC for some values of  $k$ , especially when  $k$  is large as expected in Section 4. But, in practice the SC configuration is unrealistic when  $k$  is large.

**$\mathcal{P}_1$  vs.  $\mathcal{P}_2$ :** From Section 4, we know that when  $n_0 = 10$ , the continuation region of  $\mathcal{P}_2$  is inside that of  $\mathcal{P}_1$ . Hence,  $\mathcal{P}_2$  is expected to show better performance compared to  $\mathcal{P}_1$  under all configurations and the experimental results support this. When  $n_0 = 24$  and  $k > 150$ , we observe a similar result. However, when  $n_0 = 24$  and  $k \leq 150$ , since  $V_{\mathcal{P}_1}$  is slightly less than  $V_{\mathcal{P}_2}$ , it is probable that under some configurations  $\mathcal{P}_1$  beats  $\mathcal{P}_2$ . As shown in Table 9 when  $k = 25$  or  $k = 100$  under SC-DV,  $\mathcal{P}_2$  requires slightly more observations than  $\mathcal{P}_1$ .

Table 10: Sample average of the total number of observations in  $\mathcal{KN}$ ,  $\mathcal{P}_1$ , and  $\mathcal{P}_2$  when  $n_0 = 24, \delta = 1/\sqrt{n_0}, 1 - \alpha = 0.95$ , and  $k = 25$  as a function of  $\rho$ .

		SC-CV	MDM-CV	MDM-IV	MDM-DV	SC-IV	SC-DV
$\rho = 0.02$	$\mathcal{KN}$	4926	1313	1965	1143	13241	2304
	$\mathcal{P}_1$	5001	<b>1194</b>	<b>1628</b>	<b>1074</b>	14226	<b>2178</b>
	$\mathcal{P}_2$	<b>4509</b>	<b>1177</b>	<b>1612</b>	<b>1052</b>	<b>12342</b>	<b>2181</b>
$\rho = 0.25$	$\mathcal{KN}$	3804	1093	1554	983	10736	1792
	$\mathcal{P}_1$	3873	<b>1017</b>	<b>1286</b>	<b>949</b>	11021	<b>1634</b>
	$\mathcal{P}_2$	<b>3513</b>	<b>1001</b>	<b>1290</b>	<b>912</b>	<b>10047</b>	<b>1715</b>
$\rho = 0.50$	$\mathcal{KN}$	2564	870	1118	813	7750	1288
	$\mathcal{P}_1$	<b>2426</b>	<b>823</b>	<b>949</b>	<b>786</b>	8001	<b>1119</b>
	$\mathcal{P}_2$	<b>2349</b>	<b>825</b>	<b>961</b>	<b>779</b>	<b>7225</b>	<b>1243</b>
$\rho = 0.75$	$\mathcal{KN}$	1289	682	749	667	4960	822
	$\mathcal{P}_1$	<b>1125</b>	<b>659</b>	<b>692</b>	<b>649</b>	<b>4927</b>	<b>728</b>
	$\mathcal{P}_2$	<b>1218</b>	<b>663</b>	<b>698</b>	<b>655</b>	<b>4479</b>	<b>812</b>

## 5.2 Correlated Case

We tested four different correlation factors ( $\rho$ ): 0.02, 0.25, 0.50, and 0.75. Kim and Nelson (2001) showed that a minimum correlation required for a CRN case to beat an independent case is approximately 0.02 for a triangular continuation region. We found that when  $\rho = 0.02$ , both procedures  $\mathcal{P}_1$  and  $\mathcal{P}_2$  also spend slightly smaller but very close number of observations on average compared to their independent cases in most of the configurations. Thus, the correlation in the amount of 0.02 still seems to be an approximate minimum amount of correlation required for the new procedures with CRN to be as efficient as independent cases. For  $\rho = 0.25, 0.50, \text{ or } 0.75$ , all experiment designs show a lot better performance compared to their independent cases, and the average total number of observations reduces as the correlation factor increases. The simulation results of  $\mathcal{KN}$ ,  $\mathcal{P}_1$ , and  $\mathcal{P}_2$  for different  $\rho$  values when  $n_0 = 24, \delta = 1/\sqrt{n_0}, 1 - \alpha = 0.95$ , and  $k = 25$  are given in Table 10.

Interestingly, when systems are simulated independently,  $\mathcal{P}_1$  worked better than  $\mathcal{KN}$  only under all the MDM configurations and some SC with decreasing variances for  $k \geq 10$ . However, as  $\rho$  increases,  $\mathcal{P}_1$  shows better performance than  $\mathcal{KN}$  under more configurations.

For example, when  $\rho = 0.75$ ,  $\mathcal{P}_1$  outperforms  $\mathcal{KN}$  in all the configurations. In Section 4, we show that the ratio of vertical intercepts of  $\mathcal{KN}$  and  $\mathcal{P}_1$  is  $V_{\mathcal{KN}}/V_{\mathcal{P}_1} = \eta/(\lambda_1\sqrt{\xi_1})$  and does not depend on sample variance of the difference between two systems  $S_{i\ell}^2$ . Therefore, this ratio will not change with the use of CRN. However, the horizontal intercepts of  $\mathcal{KN}$  and  $\mathcal{P}_1$  ( $H_{\mathcal{KN}}$  and  $H_{\mathcal{P}_1}$ ) are a function of  $S_{i\ell}^2$  and  $S_{i\ell}^4$ , respectively. This implies that if there is a decrease in  $S_{i\ell}^2$  due to the use of CRN, then the decrease in  $H_{\mathcal{P}_1}$  will be much larger than that in  $H_{\mathcal{KN}}$ . This explains why  $\mathcal{P}_1$  catches up the performance of  $\mathcal{KN}$  even under SC when the effect of CRN becomes larger.

The performance of  $\mathcal{P}_2$  is better than that of  $\mathcal{KN}$  under all configurations for all values of  $\rho$  we tested.

For small  $\rho$ ,  $\mathcal{P}_2$  is still better than  $\mathcal{P}_1$  like the independent case. However, as  $\rho$  increases,  $\mathcal{P}_1$  starts to show slightly better results than  $\mathcal{P}_2$ . This is again because of the fact that the horizontal intercepts of  $\mathcal{P}_1$  and  $\mathcal{P}_2$  ( $H_{\mathcal{P}_1}$  and  $H_{\mathcal{P}_2}$ ) are a function of  $S_{i\ell}^4$  and  $S_{i\ell}^2$ , respectively.

## 6 Conclusion

In this paper, we propose two fully sequential procedures with parabolic boundaries for i.i.d. normal data when variances are unknown and unequal. Our procedures are appropriate for the use in simulation environments unlike the procedure due to Zhu et al. (2005) that assumes known variances. The procedures allow for the use of CRN. The  $\mathcal{P}_1$  procedure shows competitive performance compared to  $\mathcal{KN}$ , and the superiority of  $\mathcal{P}_1$  is more noticeable when  $k$  is large and/or the effect of CRN is stronger. Therefore, we recommend that one uses  $\mathcal{KN}$  when  $k$  is small but  $\mathcal{P}_1$  when  $k$  is larger than ten and/or CRN is employed. The  $\mathcal{P}_2$  procedure is better than both  $\mathcal{KN}$  and  $\mathcal{P}_1$  when the total number of observations required is considered as the only criterion. However, the disadvantage of  $\mathcal{P}_2$  is that the computational time of the  $\xi$  parameter is very long compared to that of the  $\xi$  parameter in  $\mathcal{P}_1$ . Therefore, if the gain in the simulation time when  $\mathcal{P}_2$  is applied is significant enough to compensate the time loss in computing the  $\xi$  parameter, one might prefer to use  $\mathcal{P}_2$ .

## Acknowledgement

We appreciate very useful and helpful comments from Dr. David Goldsman. This material was supported by the National Science Foundation under Grant Number DMI-0400260.

## Appendix

### Procedure $\mathcal{P}_2$

**Setup:** Choose nominal level of PCS  $1 - \alpha$ , indifference zone  $\delta$ , and first stage sample size  $n_0 \geq 2$ . Then determine  $\lambda$  and  $\xi$  referring to **Parameters**.

**Initialization:** Let  $I = \{1, 2, \dots, k\}$  be the initial set of systems.

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

For all  $i \neq \ell$  compute the sample variance of the difference between systems  $i$  and  $\ell$  which is

$$S_{i\ell}^2 = \frac{1}{n_0 - 1} \sum_{j=1}^{n_0} \left( X_{ij} - X_{\ell j} - [\bar{X}_i(n_0) - \bar{X}_\ell(n_0)] \right)^2.$$

Let

$$N_{i\ell} = \left\lfloor \frac{(n_0 - 1)S_{i\ell}^2 \xi}{\delta^2} \right\rfloor,$$

where  $\lfloor \cdot \rfloor$  indicates truncation of any fractional part. Then, let

$$N_i = \max_{\ell \neq i} N_{i\ell}.$$

Here,  $N_i + 1$  is the maximum number of observations that can be taken from system  $i$ .

Set the number of observations,  $r$ , equal to  $n_0$ , and go to the next section.

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

$$I = \left\{ i : i \in I^{\text{old}} \text{ and } \sum_{j=1}^r X_{ij} > \sum_{j=1}^r X_{\ell j} - R_{i\ell}(r), \forall \ell \in I^{\text{old}}, \ell \neq i \right\},$$

where

$$R_{i\ell}(r) = \lambda S_{i\ell} \sqrt{n_0 - 1} \sqrt{\max \left\{ \frac{(n_0 - 1) S_{i\ell}^2 \xi}{\delta^2} - r, 0 \right\}}.$$

**Stopping Rule:** If  $|I| = 1$ , then stop and select the system  $i \in I$  as the best. Otherwise, take one additional observation  $X_{i,r+1}$  from each system  $i \in I$ , and set  $r = r + 1$ , and go to **Screening**.

(If the objective is to select a subset of size of  $m$  containing the best system, then the stopping rule should be  $|I| = m > 1$ .)

**Parameters:** The parameter  $\lambda$  is any positive real number. For  $1 - \alpha = 0.95$ , we recommend taking  $\lambda = 0.3$ . Section 3.1 discusses the choice of  $\lambda$  for other values of  $\alpha$ . For a given  $\lambda$  value,  $\xi$  is calculated as the solution to the equation

$$h(\xi; \lambda, n_0) = \beta, \tag{7}$$

where

$$h(\xi; \lambda, n_0) \equiv \mathbb{E} \left[ \frac{1}{2} - \frac{1}{2\sqrt{2\pi}} \exp \left\{ -\chi_{n_0-1}^2 \frac{\xi}{2} \right\} \sum_{n=0}^{\infty} \left( \frac{k_{2n+1}(\lambda \sqrt{\chi_{n_0-1}^2})}{\ell_{2n+1}(\lambda \sqrt{\chi_{n_0-1}^2})} \frac{2^n n!}{(2n+1)!} \left( \sqrt{\chi_{n_0-1}^2} \xi \right)^{2n+1} \right) \right].$$

Now we prove that  $\mathcal{P}_2$  also provides the predetermined probability of correct selection.

**Proof of Theorem 2:** Consider two systems 1 and  $i$  such that  $\mu_1 \geq \mu_i + \delta$ . Select a value of  $\xi$  such that  $h(\xi; \lambda, n_0) = \beta$  for some  $0 < \beta < 1/2$ . Let

$$T = \min \left\{ r : r \geq n_0 \text{ and } -R_{1i}(r) < \sum_{j=1}^r (X_{1j} - X_{ij}) < R_{1i}(r) \text{ is violated} \right\}.$$

Notice that  $T$  is the stage at which the procedure terminates. Let  $\text{ICS}_i$  be the event of incorrect selection when only two systems, 1 and  $i$ , are considered. Then,

$$\begin{aligned} \Pr \{ \text{ICS}_i \} &= \Pr \left\{ \sum_{j=1}^T (X_{1j} - X_{ij}) < -\lambda S_{1i} \sqrt{n_0 - 1} \sqrt{\max \left\{ \frac{(n_0 - 1) S_{1i}^2 \xi}{\delta^2} - T, 0 \right\}} \right\} \\ &= \Pr \left\{ \sum_{j=1}^T \left( \frac{X_{1j} - X_{ij}}{\sigma_{1i}} \right) < -\lambda \frac{S_{1i}}{\sigma_{1i}} \sqrt{n_0 - 1} \sqrt{\max \left\{ \frac{(n_0 - 1) S_{1i}^2 \xi}{\delta^2} - T, 0 \right\}} \right\} \end{aligned}$$



$$\begin{aligned}
&\leq \Pr_{\text{RSC}} \left\{ \sum_{j=1}^T \left( \frac{X_{1j} - X_{ij}}{\sigma_{1i}} \right) < -\lambda \frac{S_{1i}}{\sigma_{1i}} \sqrt{n_0 - 1} \sqrt{\max \left\{ \frac{(n_0 - 1) S_{1i}^2 \xi}{\delta^2} - T, 0 \right\}} \right\} \\
&= \mathbb{E} \left[ \Pr_{\text{RSC}} \left\{ \sum_{j=1}^T \left( \frac{X_{1j} - X_{ij}}{\sigma_{1i}} \right) < -\lambda \frac{S_{1i}}{\sigma_{1i}} \sqrt{n_0 - 1} \sqrt{\max \left\{ \frac{(n_0 - 1) S_{1i}^2 \xi}{\delta^2} - T, 0 \right\}} \right\} \middle| S_{1i} \right] \\
&\leq \mathbb{E} \left[ \Pr_{\text{RSC}} \left\{ \sum_{j=1}^T \left( \frac{X_{1j} - X_{ij}}{\sigma_{1i}} \right) < 0 \middle| S_{1i} \right\} \right]. \tag{8}
\end{aligned}$$

If we let

$$a = \lambda \sqrt{n_0 - 1} \frac{S_{1i}}{\sigma_{1i}}, \quad s = \frac{(n_0 - 1) S_{1i}^2 \xi}{\delta^2}, \quad t = n_0, \dots, N_{1i} + 1,$$

then by using Lemma 1 and 2 we get

$$\begin{aligned}
\mathbb{E} \left[ \Pr_{\text{RSC}} \left\{ \sum_{j=1}^T \left( \frac{X_{1j} - X_{ij}}{\sigma_{1i}} \right) < 0 \middle| S_{1i} \right\} \right] &\leq \\
&\mathbb{E} \left[ \frac{1}{2} - \frac{1}{2\sqrt{2\pi}} \exp \left\{ -\frac{(n_0 - 1) S_{1i}^2 \xi}{\sigma_{1i}^2} \frac{\xi}{2} \right\} \sum_{n=0}^{\infty} \left( \frac{k_{2n+1} \left( \lambda \frac{S_{1i} \sqrt{n_0 - 1}}{\sigma_{1i}} \right)}{\ell_{2n+1} \left( \lambda \frac{S_{1i} \sqrt{n_0 - 1}}{\sigma_{1i}} \right)} \frac{2^n n!}{(2n + 1)!} \left( \sqrt{\frac{(n_0 - 1) S_{1i}^2 \xi}{\sigma_{1i}^2}} \right)^{2n+1} \right) \right].
\end{aligned}$$

Since  $(n_0 - 1) S_{1i}^2 / \sigma_{1i}^2$  is chi-square distributed with  $n_0 - 1$  degrees of freedom, the expectation above is equivalent to

$$\mathbb{E} \left[ \frac{1}{2} - \frac{1}{2\sqrt{2\pi}} \exp \left\{ -\chi_{n_0-1}^2 \frac{\xi}{2} \right\} \sum_{n=0}^{\infty} \left( \frac{k_{2n+1} \left( \lambda \sqrt{\chi_{n_0-1}^2} \right)}{\ell_{2n+1} \left( \lambda \sqrt{\chi_{n_0-1}^2} \right)} \frac{2^n n!}{(2n + 1)!} \left( \sqrt{\chi_{n_0-1}^2} \xi \right)^{2n+1} \right) \right]. \tag{9}$$

Notice that (9) is  $h(\xi; \lambda, n_0)$ . So, this expectation is equal to  $\beta$ .

The proof beyond this point is same as that of Theorem 1.  $\square$

## References

- BOESEL, J., B. L. NELSON, AND S.-H. KIM. 2003. Using ranking and selection to clean up after a simulation search. *Operations Research*, 51:814–825.
- CHEN, H.-C., C.-H. CHEN, L. DAI, AND E. YÜCESAN. 1997. New development of optimal computing budget allocation for discrete event simulation, In *Proceedings of the 1997 Winter Simulation Conference*, ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B.L. Nelson, 334–341. Piscataway, New Jersey: IEEE.

- CHEN, H. C., C. H. CHEN, AND E. YÜCESAN. 2000. Computing efforts allocation for ordinal optimization and discrete event simulation. *IEEE Transactions on Automatic Control*, 45: 960–964.
- CHICK, S. E. 1997. Selecting the best system: A decision-theoretic approach. In *Proceedings of the 1997 Winter Simulation Conference*, ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B.L. Nelson, 326-333. Piscataway, New Jersey: IEEE.
- CHICK, S. AND K. INOUE. 2001a. New two-stage and sequential procedures for selecting the best simulated system. *Operations Research*, 49:1609–1624.
- CHICK, S. AND K. INOUE. 2001b. New procedures for identifying the best simulated system using common random numbers. *Management Science*, 47: 1133–1149.
- FEREBEE, B. 1982. Tests with parabolic boundary for the drift of a Wiener process. *Annals of Statistics*, 10:882-894.
- GOLDSMAN, D. AND B. L. NELSON. 1998. Comparing Systems via Simulation. In *Handbook of simulation: Principles, Methodology, Advances, Applications, and Practice*, ed. J. Banks, 273–306. New York: John Wiley & Sons.
- JENNISON, C., I. M. JOHNSTONE, AND B. W. TURNBULL. 1980. Asymptotically optimal procedures for sequential adaptive selection of the best of several normal means. Technical Report, Department of ORIE, Cornell Univ., Ithaca, NY.
- KIM, S.-H. AND B. L. NELSON. 2001. A fully sequential procedure for indifference-zone selection in simulation. *ACM TOMACS*, 11:251–273.
- KIM, S.-H. AND B. L. NELSON. 2005. On the Asymptotic Validity of Fully Sequential Selection Procedures for Steady-State Simulation. Accepted to *Operations Research*.
- LERCHE. 1986. Boundary Crossing of Brownian Motion. In *Lecture Notes in Statistics*, ed. D. Brillinger, S. Fienberg, J. Gani, J. Hartigan, and K. Krickeberg, 98–139. New York: Springer-Verlag.
- MA, H. H. AND T. R. WILLEMAIN. 2004. Better selection of the best, In *Proceedings of the 2004 Winter Simulation Conference*, ed. R. G. Ingalls, M. D. Rossetti, J. S. Smith,

- and B. A. Peters, 639–644. Piscataway, New Jersey: IEEE.
- MALONE, G. J., S.-H. KIM, D. GOLDSMAN, AND D. BATUR. 2005. Performance of Variance Updating Procedures on Various Data. Technical Report. School of Industrial Systems and Engineering, Georgia Institute of Technology, Atlanta, GA 30332.
- NELSON, B. L., J. SWANN, D. GOLDSMAN, AND W. SONG. 2001. Simple procedures for selecting the best simulated system when the number of alternatives is large. *Operations Research*, 49:950–963.
- RINOTT, Y. 1978. On two-stage selection procedures and related probability-inequalities. *Comm. Stat.-Thy. and Meth.*, A7:799-811.
- TAMHANE, A. C. 1977. Multiple comparisons in model I one-way ANOVA with unequal variances. *Communications in Statistics*, A6:15–32.
- ZHU, L., D. GOLDSMAN, S.-H. KIM, AND D. BATUR. 2005. A Sequential Procedure with Parabolic Boundary for Known-Variance Normal Population Selection. Technical Report, School of Industrial Systems and Engineering, Georgia Institute of Technology, Atlanta, GA 30332.