Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

A FULLY SEQUENTIAL PROCEDURE FOR KNOWN AND EQUAL VARIANCES BASED ON MULTIVARIATE BROWNIAN MOTION

A. B. Dieker
Seong-Hee Kim

Georgia Institute of Technology
Atlanta, GA 30332-0205, USA

ABSTRACT

We consider the problem of identifying the system with the largest expected mean among a number of simulated systems. We provide a new fully sequential procedure whose continuation region is developed based on multivariate Brownian motion when the variances of the systems are known and equal. We provide an approximation to determine the procedure parameters and we show experimental results.

1. INTRODUCTION

Ranking and Selection (R&S) aims to find the best system among a number of systems for which noisy performance information is accessible through simulation. The simulation noise gives rise to a fundamental trade-off between quickly selecting a system and correctly identifying the best system.

The finding-the-best problem, one type of R&S problems, is well-studied in the literature, and there exist at least three different approaches: the Bayesian approach, the optimal computing budget allocation (OCBA) approach, and the indifference-zone (IZ) approach. We refer to Chick (2006) for a review of the Bayesian and OCBA approaches. This paper employs the IZ approach. In the IZ approach, the decision maker sets a practically meaningful difference (namely the IZ parameter) worth detecting, and it is the aim to identify a system whose mean is within the IZ parameter of the best system. Fully-sequential IZ procedures, such as the widely used KN procedure by Kim and Nelson (2001), sequentially take one observation from competing systems and eliminate clearly inferior systems as observations become available. A generally accepted assumption is that the discrete-time system observations can be reasonably approximated by continuous-time stochastic processes such as (variants of) Brownian motion. The mean configuration where the true difference between the best and other alternative systems is exactly the IZ parameter is called the *slippage configuration* (SC) and is known to have the worst-case probability of correct selection.

Given a target probability of correct selection, one aims to devise procedures that use as few simulation replications as possible. A key challenge is to choose the parameters of the procedure, for which one often uses bounds on the worst-case probability of correct selection. These bounds tend to become conservative as the number of systems increases, for instance when using a Bonferroni-type bound. This conservativeness is a major source of inefficiencies and we have been exploring a possible way to overcome this fundamental issue, by developing procedures that treat observations as high-dimensional. This paper works under the assumption of known and equal variances, but our framework is in principle more flexible.

Several recent works recognize the importance of tight estimates on the probability of correct selection. Frazier (2014) presents the Bayes-inspired indifference zone (BIZ) procedure whose lower bound on worst-case probability of correct selection is tight in continuous time, and nearly tight in discrete time. Kim and Dieker (2011) propose a fully-sequential procedure for three systems when variances are known and equal based on the behaviors of a bivariate Brownian motion exiting an ellipse. Dieker and Kim (2012) extend the procedure to unequal variances, but the procedure is still based on grouping three systems. These works

are closely related to the present paper, but here we work with multivariate (possibly high-dimensional) Brownian motion.

The paper is organized as follows. Section 2 defines our problem and introduces our notation. Section 3 proposes our new fully-sequential procedure. In Section 4, we give an approximation in order to set the parameter values of our procedure. Experimental results are presented in Section 5, followed by the conclusions in Section 6.

2. PROBLEM AND NOTATION

This section introduces our notation and assumptions, and defines the problem. Which system is best depends on the problem at hand, but in this paper we assume that the system with the largest mean is the best.

We assume there are k systems, among which we have to select the best system ($k \geq 2$). The set of all systems is defined as $\Omega = \{1,\ldots,k\}$. Let X_{ij} represent an observation from replication (or batch) j of system i for $i = 1,\ldots,k$ and $j = 1,2,\ldots$. The mean and variance of the outputs from system i are defined as $\mu_i = \mathbb{E}[X_{ij}]$ and $\sigma^2 = \mathbb{V}\text{ar}[X_{ij}]$, respectively. In addition to equal variances across systems, we also need the following assumptions:

Assumption 1. *Let X_{ij} represent the jth observation from system i. Then*

$$X_{ij} \overset{i.i.d.}{\sim} \boldsymbol{N}(\mu_i,\sigma^2), \quad j = 1,2,\ldots,$$

where $\overset{i.i.d.}{\sim}$ represents 'are independent and identically distributed as' and $\boldsymbol{N}$ denotes the normal distribution with mean μ_i and variance σ^2. Moreover, $(X_{1j},\ldots,X_{kj})$ and $(X_{1j'},\ldots,X_{kj'})$ are independent for any $j \neq j'$.

Assumption 2. *$\mu_1 \leq \mu_2 \leq \ldots \leq \mu_{k-1} \leq \mu_k - \delta$ for $\delta \in \mathbb{R}^+$.*

Assumption 1 implies that the output data from each system is marginally i.i.d. normally distributed and systems are simulated independently (thus no common random numbers). Without loss of generality, Assumption 2 assumes that system k is the best and its mean is at least δ better than any alternative system. The parameter δ is a user-specified parameter known as the IZ parameter, a practically meaningful difference worth detecting.

It is our aim to devise a method that observes systems sequentially, on the basis of which it can decide that some systems are inferior and thus eliminate them from further consideration. The method stops once only one system remains, and this system is declared the best system. Due to random fluctuations, the true best system could have been eliminated and thus there is a trade-off between computationally advantageous quick elimination and the probability of correct selection. Our method takes in a target probability of incorrect selection α and we study its performance in terms of the actual probability of incorrect selection and the number of simulation replications required.

Additional notation is needed for later sections:

$$
\begin{aligned}
n &\equiv \text{the current number of observations or the current stage number;}\\
I &\equiv \text{set of competing systems at the } n\text{th stage;}\\
\bar{X}_i(n) &\equiv \frac{1}{n}\sum_{j=1}^{n} X_{ij}\text{, the sample mean of system } i \text{ based on the first } n \text{ observations;}\\
\boldsymbol{X}(n) &\equiv k\times 1 \text{ vector of } \textstyle\sum_{j=1}^{n} X_{ij} \text{ for } i=1,\ldots,k;\\
\Lambda_s &\equiv (s-1)\times(s-1) \text{ matrix with diagonal elements equal to 1 and non-diagonal elements equal to 1/2}\\
&\quad \text{for any positive integer } s;\\
\delta_k^2 &\equiv \frac{\delta^2}{\sigma^2}\frac{k-1}{k}.
\end{aligned}
$$

3. OUR PROCEDURE

This section presents a new procedure, which we call $\mathcal{N}$. The procedure is based on a continuation region related to a quadratic form $\mathcal{S}_I$. For $x\in\mathbb{R}^k$ and $I\subset\{1,\ldots,k\}$ with $I=\{i_1,\ldots,i_s\}$, it is defined as

$$
\mathcal{S}_I(x)=\frac{1}{2\sigma^2}\begin{bmatrix} x_{i_2}-x_{i_1}\\ \vdots\\ x_{i_s}-x_{i_1}\end{bmatrix}^T \Lambda_{|I|}^{-1}\begin{bmatrix} x_{i_2}-x_{i_1}\\ \vdots\\ x_{i_s}-x_{i_1}\end{bmatrix}.
$$

From this definition it is not immediate why $\mathcal{S}_I$ only depends on the set I, and not on the order in which its elements are listed. This becomes apparent from the next lemma; see also Lemma 1 in Dieker and Kim (2012). The proof is given in the appendix.

Lemma 1. *For $x\in\mathbb{R}^k$, we have*

$$
\mathcal{S}_I(x)=\frac{1}{\sigma^2}\frac{1}{|I|}\sum_{\substack{i<\ell\\ i,\ell\in I}}(x_i-x_\ell)^2.
$$

We can now describe Procedure $\mathcal{N}$.

New Procedure $\mathcal{N}$

Setup: Select the nominal level $1-\alpha$ and the IZ parameter δ. Choose η (which will be discussed in Section 4). Set $I=\{1,2,\ldots,k\}$ and take one observation from each system. Set $n=1$ and go to **Calculation**.

Calculation: Calculate $\mathcal{S}_I(\boldsymbol{X}(n))$.

Screening: If $\mathcal{S}_I(\boldsymbol{X}(n))\geq\left(\frac{\eta}{\delta_k}\right)^2$, then eliminate the system with the smallest $\bar{X}_i(n)$ among $i\in I$. Update I and go back to **Calculation**. Otherwise, go to **Stopping Rule**.

Stopping Rule: If $|I|=1$, stop and declare the surviving system as the best. Otherwise, take one more observation for all $i\in I$, set $n=n+1$, and go to **Calculation**.

Note that our screening condition is

$$
\mathcal{S}_I(x)\geq\frac{\eta^2}{\delta^2}\frac{k}{k-1}\sigma^2. \tag{1}
$$

The next lemma implies that we in fact automatically check the screening condition for *all* sets $J \subseteq I$ if we check it for the *largest* set $J = I$, which explains why we start with $I = \{1, 2, \ldots, k\}$ and shrink it successively. The procedure thus automatically considers (1) for *all* possible subsets. A proof of the lemma is given in the appendix.

Lemma 2. *For any $x \in \mathbb{R}^k$, we have $\mathcal{S}_{I_1}(x) \leq \mathcal{S}_{I_2}(x)$ if $I_1 \subseteq I_2$.*

4. APPROXIMATIONS

This section presents an approximation for the probability of incorrect selection under the new procedure $\mathcal{N}$ presented in Section 3. We use these approximations *in lieu* of possibly conservative bounds in order to choose the parameter η of the procedure $\mathcal{N}$, thus bypassing a main source of inefficiencies.

The event of incorrect selection can be partitioned according to when the best system is eliminated. If the best system is eliminated first, then we say that the level of elimination is 1. Similarly, if the second system to be eliminated is the best system, then we say that the level of elimination is 2. Thus, the possible levels of incorrect elimination are $1, \ldots, k-1$. The key building block for our approximation scheme is an approximation for the probability of incorrect selection at the first elimination level, which we discuss in Section 4.1. Other levels of incorrect elimination are studied in Section 4.2. With this, we devise a procedure for choosing the parameter η in Section 4.3.

4.1 Immediate (Level 1) Elimination of the Best System

Our approximation for the probability of eliminating system k first is based on an asymptotic analysis as the number of systems $k \to \infty$. Our results use the commonly employed idea of replacing the (discrete) Gaussian observation sequence with a (continuous) Brownian motion.

Throughout this section we use the following notation. For a given a vector $x \in \mathbb{R}^k$,

$$\mathbb{E}_k(x) = \frac{1}{k}\sum_{i=1}^{k} x_i, \quad \mathbb{V}\mathrm{ar}_k(x) = \mathbb{E}_k(x^2) - \mathbb{E}_k(x)^2,$$

where x^2 should be understood componentwise.

In the continuous analog of our problem, the discrete observation window is replaced with a continuous one. The analog of the random walk $\boldsymbol{X}(n)$ is $\sigma B(t)$, where $B(t)$ is a standard Brownian motion in $\mathbb{R}^k$ with drift $(0, \ldots, 0, \delta/\sigma)$.

It is relevant to note that $B(t) - \mathbb{E}_k(B(t))$ is a standard Brownian motion with drift $(-1/k, \ldots, -1/k, (1 - 1/k)) \times \delta/\sigma$ on the hyperplane

$$H = \left\{ x \in \mathbb{R}^k : \sum_{i=1}^{k} x_i = 0 \right\}.$$

Setting $r = \eta/\delta_k$, we define a $(k-1)$-dimensional ball in H by

$$C = \left\{ x \in \mathbb{R}^k : \sum_{i=1}^{k} x_i = 0, \|x\| = r \right\}.$$

Elimination of the best system can be formulated as $B(t) - \mathbb{E}_k(B(t))$ hitting C in the region

$$E_k = \{x \in C : x_k = \min(x_1, \ldots, x_k)\}.$$

We now state the main result of this section. (Note that $\delta_k r = \eta$.)

Lemma 3. *Suppose that $Z_1,\ldots,Z_k$ are i.i.d. standard normal. The probability that the process $B(t)-\mathbb{E}_k(B(t))$ first hits C in the part E_k where the best system k gets eliminated equals*

$$\frac{\int_{-r}^{r} e^{\delta_k x} d\mathbb{P}(Z_k = \min(Z_1,\ldots,Z_k), r(Z_k-\mathbb{E}_k(Z)) \le x\sqrt{k\mathbb{V}\mathrm{ar}_k(Z)})}{\left(\frac{\delta_k r}{2}\right)^{-\nu}\Gamma(\nu+1)I_\nu(\delta_k r)}, \tag{2}$$

where $\nu = k/2-1$, Γ stands for the Gamma function, and I_ν for the modified Bessel function of the first kind.

Proof. The norm of the drift of $B(t)-\mathbb{E}_k(B(t))$ is δ_k. Therefore, the hitting place of $B(t)-\mathbb{E}_k(B(t))$ on C has density f with respect to the uniform distribution μ_C on C with (e.g., Rogers and Pitman (1981))

$$f(x) = \frac{1}{\int_C e^{\delta_k x_k}\mu_C(dx)} e^{\delta_k x_k}, \quad x \in C.$$

This distribution is known as the von Mises distribution. It is known that (e.g., Rogers and Pitman (1981))

$$\int_C e^{\delta_k x_k}\mu_C(dx) = \left(\frac{\delta_k r}{2}\right)^{-\nu}\Gamma(\nu+1)I_\nu(\delta_k r).$$

Note that larger values of $B_k(t)-\mathbb{E}_k(B(t))$ are more likely than smaller values when the process hits C, which should be expected because system k is the best one.

The probability of eliminating the best system first equals

$$\begin{aligned}\int_{E_k} f(x)\mu_C(dx) &= \mathbb{E}[1(X_k = \min(X_1,\ldots,X_k))f(X)] \\ &= \frac{\mathbb{E}[1(X_k = \min(X_1,\ldots,X_k))e^{\delta_k X_k}]}{\int_C e^{\delta_k x_k}\mu_C(dx)},\end{aligned}$$

where X has a uniform distribution on C and 1 stands for the indicator function. The random vector

$$X = \frac{r(Z_1-\mathbb{E}_k(Z),\ldots,Z_k-\mathbb{E}_k(Z))}{\sqrt{k\mathbb{V}\mathrm{ar}_k(Z)}}$$

is a sample from the uniform distribution on C by symmetry. Therefore the sought probability equals

$$\frac{\mathbb{E}[1(Z_k = \min(Z_1,\ldots,Z_k))e^{\delta_k r(Z_k-\mathbb{E}_k(Z))/\sqrt{k\mathbb{V}\mathrm{ar}_k(Z)}}]}{\int_C e^{\delta_k x_k}\mu_C(dx)},$$

as claimed. □

To approximate the probability (2), we replace several of its components by asymptotic approximations. For instance, as $k\to\infty$, the random variables $\mathbb{E}_k(Z)$ and $\mathbb{V}\mathrm{ar}_k(Z)$ converge in distribution to 0 and 1, respectively, by the strong law of large numbers. The rate of convergence is relatively fast (order $1/\sqrt{k}$ by the central limit theorem). We therefore approximate those variables by their deterministic asymptotic approximations. The term with the minimum is slightly more complicated. Writing

$$c_k = \sqrt{2\log k} - \frac{\log\log k + \log(4\pi)}{2\sqrt{2\log k}},$$

$\min(Z_1,\ldots,Z_{k-1})+c_{k-1}$ converges in distribution to 0, cf. Example 3.3.29 in (Embrechts, Klüppelberg, and Mikosch 1997). The rate of convergence is relatively slow (order $1/\sqrt{2\log k}$), so we use an approximation based on the fact that

$$\sqrt{2\log k}(\min(Z_1,\ldots,Z_{k-1})+c_{k-1})$$

converges in distribution to a Gumbel distributed random variable G, which is equal in distribution to $-\log(-\log(U))$ where U is standard uniformly distributed. Even when the central limit theorem is used for the sum instead of the law of large numbers, the minimum and sum are asymptotically independent (e.g., Chow and Teugels (1978)).

This motivates the approximation, for $x\in(-r,r)$,

$$\begin{aligned} &d\mathbb{P}(Z_k=\min(Z_1,\ldots,Z_k), r(Z_k-\mathbb{E}_k(Z))\le x\sqrt{k\mathbb{V}\mathrm{ar}_k(Z)})\\ \approx\quad &d\mathbb{P}(Z_k\le G/\sqrt{2\log k}-c_{k-1}, rZ_k\le x\sqrt{k}), \end{aligned}$$

where Z_k and G are independent.

We are now ready to formulate our approximation for (2).

Lemma 4. *For fixed $g\in\mathbb{R}$, we have*

$$\begin{aligned} &\int_{-r}^{r} e^{\delta_k x} d\mathbb{P}(Z_k\le g/\sqrt{2\log k}-c_{k-1}, rZ_k/\sqrt{k}\le x)\\ &= \exp\left(\frac{\delta_k^2 r^2}{2k}\right)\left[\Psi\left(-\sqrt{k}-\frac{\delta_k r}{\sqrt{k}}\right)-\Psi\left(\frac{g}{\sqrt{2\log k}}-c_{k-1}-\frac{\delta_k r}{\sqrt{k}}\right)\right] \end{aligned}$$

where $\Psi(\cdot)$ is the cumulative distribution function of the standard normal random variable.

Proof. We prove the statement $g=0$ without loss of generality since one could shift c_{k-1} appropriately. Letting Y be a centered Gaussian variable with variance r^2/k, we have

$$\begin{aligned} &\int_{-r}^{r} e^{\delta_k x} d\mathbb{P}(Z_k\le -c_{k-1}, rZ_k/\sqrt{k}\le x)\\ &= \int_{-r}^{-rc_{k-1}/\sqrt{k}} e^{\delta_k y} d\mathbb{P}(Y\le y)\\ &= \int_{-r}^{-rc_{k-1}/\sqrt{k}} e^{\delta_k y}\frac{\sqrt{k}}{r\sqrt{2\pi}}\exp\left(-\frac{ky^2}{2r^2}\right)dy\\ &= e^{\frac{\delta_k^2 r^2}{2k}}\frac{\sqrt{k}}{\sqrt{2\pi}r}\int_{-r}^{-rc_{k-1}/\sqrt{k}}\exp\left(-\frac{\left(y-\frac{\delta_k r^2}{k}\right)^2}{2r^2/k}\right)dy, \end{aligned}$$

and the claim readily follows. □

We thus approximate the probability of first eliminating the best system by

$$\frac{\exp\left(\frac{\delta_k^2 r^2}{2k}\right)\left[\Psi\left(-\sqrt{k}-\frac{\delta_k r}{\sqrt{k}}\right)-\mathbb{E}\Psi\left(\frac{G}{\sqrt{2\log k}}-c_{k-1}-\frac{\delta_k r}{\sqrt{k}}\right)\right]}{(\delta_k r/(2))^{-\nu}\Gamma(\nu+1)I_\nu(\delta_k r)}. \tag{3}$$

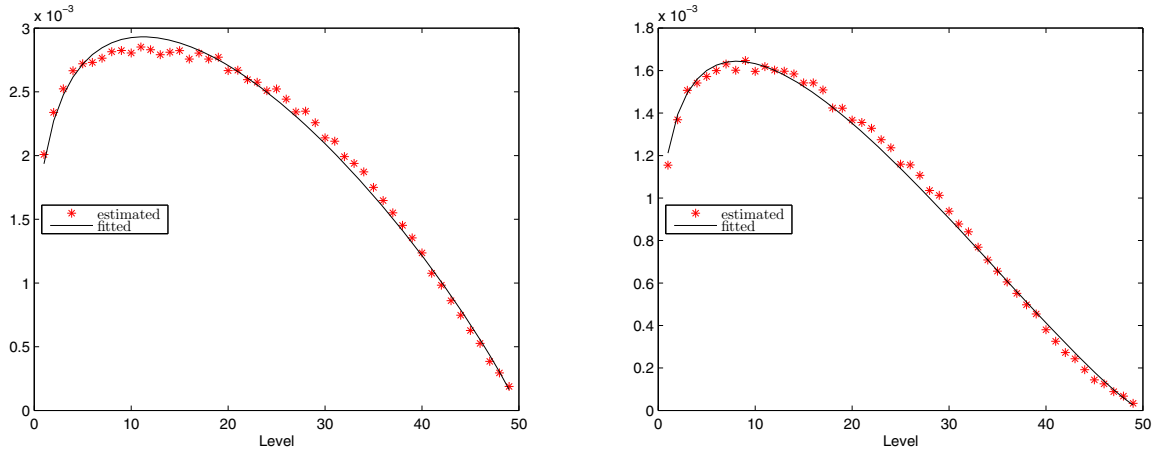
Note that $r=\eta/\delta_k$ (i.e., $\delta_k r=\eta$), so (3) is a function of η and k independent of δ and σ^2. On the computer we estimate the term with the expectation by sampling one million replications of G. We also take logs to avoid numerical overflows and underflows in the denominator, since the Gamma term can be very large and the Bessel term can be very small.

4.2 Other Ways of Incorrect Selection

The preceding subsection has dealt in depth with elimination level 1, and we now study the probabilities of incorrect selection on other levels.

We currently have no mathematical results for the other levels of elimination, so we devise an approximation by 'lifting' some computations for a lower-dimensional problem to a higher-dimensional setting. The lower-dimensional problem requires two inputs: the number of systems $k_0 \ll k$ and a reasonable guess for the radius η_0. The quality of the guess is not so important because the output of the lower-dimensional problem demonstrates its quality and one could then improve it if necessary.

We then run simulations to estimate the incorrect selection probabilities on each level for the lower-dimensional problem. A representative outcome is depicted in Figure 1 when $k_0 = 50$. The probability of incorrect selection is the sum of the probabilities over the different levels. If this is far off from the target then one could revise the choice of η_0.



(a) $1-\alpha = 0.9$ (b) $1-\alpha = 0.95$

Figure 1: The incorrect selection probabilities on each level as a function of level when $k_0 = 50$ and $\mu_1 = \ldots = \mu_{k_0-1} = \mu_{k_0} - \delta = 0$ with $\delta = 0.3$ and $\sigma^2 = 1$.

The lower-dimensional problem thus gives rise to a curve, and a scaled version of this curve is used to approximate the elimination probabilities for the high-dimensional problem. We rely on the idea that these curves resemble the shape of Beta densities. Formally, if $p_0^1, \ldots, p_0^{k_0-1}$ are the estimated probabilities of incorrect selection at levels $1, \ldots, k_0 - 1$, respectively, we fit the points $(1/k_0, p_0^1), \ldots, ((k_0-1)/k_0, p_0^{k_0-1})$ to a function in a form of a Beta density (up to a constant) through an appropriate regression; say

$$C_0 \left(\frac{x}{k_0}\right)^{\alpha_0} \left(1 - \frac{x}{k_0}\right)^{\beta_0} \quad \text{for} \quad x = 1, \ldots, k_0 - 1,$$

where C_0, α_0, β_0 are estimated constants.

Let p_i denote the probability of level i incorrect selection for k number of systems. Then the fitted curve for k_0 is used to approximate p_i as follows:

$$p_i = \frac{k_0}{k} C_0 \left(\frac{i}{k}\right)^{\alpha_0} \left(1 - \frac{i}{k}\right)^{\beta_0} \quad \text{for } i = 1, \ldots, k-1. \tag{4}$$

To get a good fitted graph, k_0 should not be too small but not too large because extremely small probabilities are hard to estimate. We found that $k_0 = 50$ works well. For example, when $1 - \alpha = 0.9$,

$\delta = 0.3$ and $\sigma^2 = 1$ with mean configuration $\mu_1 = \ldots = \mu_{k_0-1} = \mu_{k_0} - \delta = 0$, the fitted curve with $k_0 = 50$ is

$$50 \times \exp^{-5.22} y^{0.258}(1-y)^{0.892} \quad \text{for } y \in (0,1).$$

Then, for any k, we can get the probability of level i incorrect selection p_i as follows:

$$p_i = \frac{50}{k} \exp^{-5.22} \left(\frac{i}{k}\right)^{0.258} \left(1 - \frac{i}{k}\right)^{0.892} \quad \text{for } i = 1, 2, \ldots, k-1.$$

4.3 Choosing η

We now give an algorithm for choosing the parameter η for general k.

Determining η

0. Given the nominal level $1-\alpha$, select $k_0 \ll k$ and find η_0 such that estimated probability of correct selection is approximately $1-\alpha$.

1. Run simulations with k_0 systems to estimate the probabilities of incorrect selections on each level. Calculate C_0, α_0, and β_0 through regression.

2. Determine p_1 by (4) and numerically find η for k systems such that (3) equals p_1.

Remark: In Step **0**, η_0 such that our approximation of the level 1 probability of incorrect selection (3) equals α/k_0 provides a good starting point.

5. EXPERIMENTS

In this section, we compare the performance of the new procedure with a procedure (Procedure $\mathcal{P}$) due to Wang and Kim (2011), which is a version of the KN procedure for known and equal variances. We start with a description of Procedure $\mathcal{P}$.

Procedure $\mathcal{P}$

Setup: Select the nominal level $1-\alpha$ and the IZ parameter δ. Calculate $\eta = -\ln 2\beta$ where $\beta = \alpha/(k-1)$. Set $I = \{1, 2, \ldots, k\}$, take one observation from each system. Set $r = 1$ and go to **Screening**.

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

$$I = \left\{ i : i \in I^{\text{old}} \text{ and } \sum_{j=1}^{r} (X_{ij} - X_{\ell j}) \geq -\frac{\eta(2\sigma^2)}{\delta} + \frac{\delta}{2} r \text{ for } i, \ell \in I^{\text{old}} \text{ and } i \neq \ell \right\}.$$

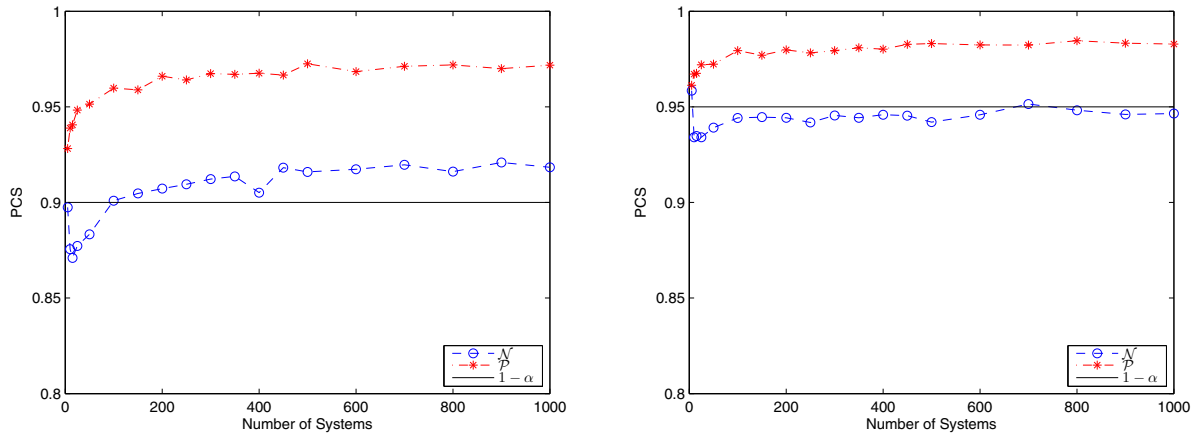
Stopping Rule: If $|I| = 1$, then stop and select the remaining system as the best one. Otherwise, take one additional observation $X_{i,r+1}$ from each active system $i \in I$, set $r = r+1$, and go to **Screening**.

The number of systems k varies over the set $\{5, 25, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000\}$. Only one mean configuration is considered: slippage configuration (SC). Under SC, $\mu_k = \delta = 0.3$ and $\mu_i = 0$ for $i = 2, \ldots, k$, which is considered to be the most difficult mean configuration with the worst-case

probability of incorrect selection in many ranking and selection procedures. Variances across systems are assumed to be equal and set to $\sigma^2 = 1$. The nominal confidence level is set to $1-\alpha = 0.9$ or 0.95. Estimated probability of correct selection (PCS) and average total number of observations until a decision is made (REP) are reported based on 10,000 macro replications.

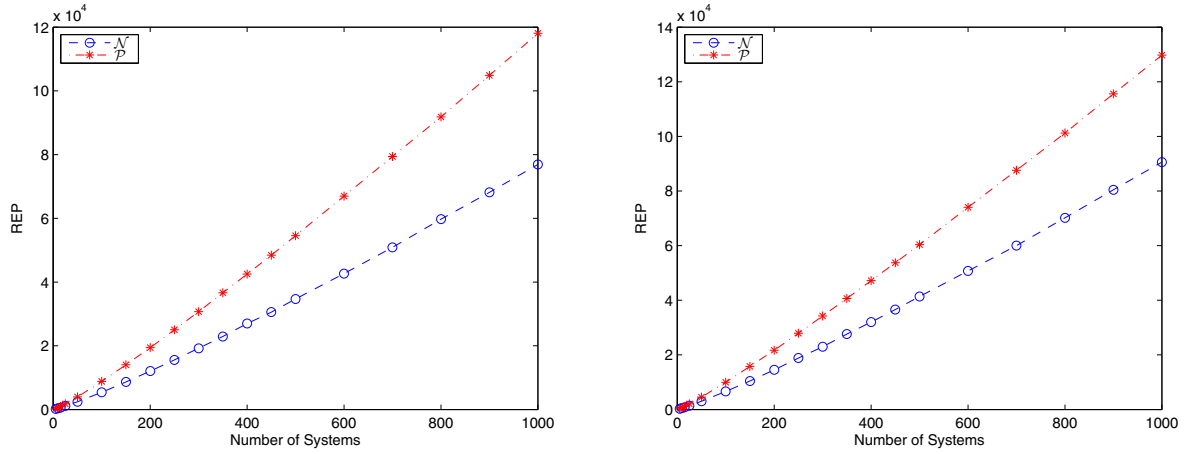
Figure 2 shows the estimated PCS for the procedures $\mathcal{N}$ and $\mathcal{P}$ under the slippage configuration. The $\mathcal{N}$ procedure is tight and produces estimated PCS close to $1-\alpha$ for all values of k tested though some PCS are slightly lower than $1-\alpha$.

Figures 3 reports REP under the SC. Under the SC, the $\mathcal{N}$ procedure spends $20\% \sim 39\%$ fewer observations than the $\mathcal{P}$ procedure for $1-\alpha = 0.9$. The percentage of savings is $12\% \sim 34\%$ for $1-\alpha = 0.95$.



(a) $1-\alpha = 0.9$ (b) $1-\alpha = 0.95$

Figure 2: Estimated PCS when $1-\alpha$ is 90%(left) and 95%(right) under the slippage mean configuration as a function of k.



(a) $1-\alpha = 0.9$ (b) $1-\alpha = 0.95$

Figure 3: Estimated REP when $1-\alpha$ is 90%(left) and 95%(right) under the slippage mean configuration as a function of k.

6. CONCLUSIONS

We present a new fully-sequential procedure whose continuation region is derived based on multivariate Brownian motion. Compared to the existing fully-sequential IZ procedure, which is a version of the KN procedure, the proposed procedure shows a tight worst-case probability of incorrect selection and significant savings in the number of observations needed until a decision is made when the mean configuration follows the slippage configuration. The extension to unequal variances is currently work in progress.

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation under grant CMMI-1131047.

A. Proofs of Lemmas 1 and 2

Proof of Lemma 1. We first derive an explicit expression for $\Lambda_{|I|}^{-1}$:

$$\Lambda_{|I|}^{-1} = \frac{2}{|I|}\begin{bmatrix} |I|-1 & -1 & \cdots & \cdots & -1 \\ -1 & |I|-1 & -1 & \cdots & -1 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & -1 \\ -1 & \cdots & \cdots & -1 & |I|-1 \end{bmatrix}. \tag{5}$$

Assume without loss of generality that $I = \{1,\dots,s\}$. Then we have

$$\begin{aligned} \mathcal{S}_I(x) &= \frac{1}{2\sigma^2}\begin{bmatrix} x_2-x_1 \\ \vdots \\ x_s-x_1 \end{bmatrix}^T \Lambda_s^{-1}\begin{bmatrix} x_2-x_1 \\ \vdots \\ x_s-x_1 \end{bmatrix} \\ &= \frac{1}{\sigma^2}\frac{s-1}{s}\left\{\sum_{i=2}^{s}(x_i-x_1)^2 - \frac{2}{s-1}\sum_{2\le i<\ell\le s}(x_i-x_1)(x_\ell-x_1)\right\} \\ &= \frac{1}{\sigma^2}\frac{1}{|I|}\sum_{\substack{i<\ell \\ i,\ell\in I}}(x_i-x_\ell)^2, \end{aligned}$$

as claimed. □

Proof of Lemma 2. We may set $\sigma = 1$, and it suffices to prove the claim for $|I_2| = |I_1|+1$. By relabeling systems if necessary, we work with $I = I_1 = \{1,\dots,s\}$ and $I_2 = \{1,\dots,s+1\}$. For $i = 1,\dots,k$, we set

$$V_i = \left\{x \in \mathbb{R}^k : \sum_{\ell=1}^{k} x_\ell = 0, x_{i+1} = \cdots = x_k = 0\right\}.$$

It suffices to prove the claim for $x \in V_{s+1}$ since $\mathcal{S}_I(x)$ is invariant under shifting each component of x by the same amount and both $\mathcal{S}_{I_1}(x)$ and $\mathcal{S}_{I_2}(x)$ do not depend on $x_{s+2},\dots,x_k$.

We show that $\mathcal{S}_{I_1}(x) = \mathcal{S}_{I_2}(\Pi_s x)$, where Π_{s+1} is the orthogonal projection matrix on V_s. Since projecting decreases any quadratic form, the claim follows.

By Lemma 1 and the remark preceding it, we have

$$\mathcal{S}_{I_2}(x) = \frac{1}{2}\begin{bmatrix} x_1-x_{s+1} \\ \vdots \\ x_s-x_{s+1} \end{bmatrix}^T \Lambda_{s+1}^{-1}\begin{bmatrix} x_1-x_{s+1} \\ \vdots \\ x_s-x_{s+1} \end{bmatrix}.$$

From (5),

$$\Lambda_{s+1}^{-1} = \frac{2}{s+1}\{(s+1)\,\mathrm{id}_s - \mathbf{1}_s\}$$

where id_s is the $s \times s$ identity matrix and $\mathbf{1}_s$ is the $s \times s$ matrix of ones.

Setting $x_{s+1} = 0$ (i.e., applying the projection matrix Π_s) and noting that Λ_{s+1}^{-1} equals $2\mathrm{id}_s$ when it acts on the space $\{y \in \mathbb{R}^s : \sum_{i=1}^s y_i = 0\}$, we get

$$\mathcal{S}_{I_2}(\Pi_s x) = \sum_{i=1}^{s} x_i^2.$$

We now show that this same expression holds for $\mathcal{S}_{I_1}(x)$. By definition of $\mathcal{S}_{I_1}(x)$, we have

$$\mathcal{S}_{I_1}(x) = \frac{1}{2} x^T V^T \Lambda_s^{-1} V x,$$

where V is the $(s-1) \times s$ matrix given by

$$V = \begin{bmatrix} & & -1 \\ & \mathrm{id}_{s-1} & \vdots \\ & & -1 \end{bmatrix},$$

and we find that $\mathcal{S}_{I_1}(x) = \sum_{i=1}^s x_i^2$ upon noting that $V^T \Lambda_s^{-1} V = 2\mathrm{id}_s$. □

REFERENCES

Chick, S. E. 2006. "Subjective Probability and Bayesian Methodology". In *Handbooks in Operations Research and Management Science: Simulation*, edited by S. G. Henderson and B. L. Nelson. Oxford: Elsevier Science.

Chow, T. L., and J. L. Teugels. 1978. "The Sum and the Maximum of I.I.D. Random Variables". In *Proceedings of the Second Prague Symposium on Asymptotic Statistics*, edited by P. Mandl and M. Huskova, 81–92. New York: North-Holland.

Dieker, A. B., and S.-H. Kim. 2012. "Selecting the Best by Comparing Simulated Systems in a Group of Three When Variances are Known and Unequal". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 1–7. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Embrechts, P., C. Klüppelberg, and T. Mikosch. 1997. *Modelling Extremal Events for Insurance and Finance*. New York: Sringer.

Frazier, P. 2014. "A Fully Sequential Elimination Procedure for Indifference-zone Ranking and Selection with Tight Bounds on Probability of Correct Selection". to appear.

Kim, S.-H., and A. B. Dieker. 2011. "Selecting the Best by Comparing Simulated Systems in a Group of Three". In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, 3992–4002. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kim, S.-H., and B. L. Nelson. 2001. "A Fully Sequential Procedure for Indifference-zone Selection in Simulation". *ACM Transactions on Modeling and Computer Simulation* 11:251–273.

Rogers, L., and J. W. Pitman. 1981. "Markov Functions". *The Annals of Probability* 9:573–582.

Wang, H., and S.-H. Kim. 2011. "Reducing the Conservativeness of Fully Sequential Indifference-Zone Procedures". *IEEE Transactions on Automatic Control* 58:1613–1619.

AUTHOR BIOGRAPHIES

A. B. DIEKER is an Associate Professor in H. Milton Stewart School of Industrial and Systems Engineering at Georgia Institute of Technology. He received his Ph.D. in Mathematics from the University of Amsterdam in 2006. His research interests include service engineering, stochastic processes, and probability theory. He is a member of INFORMS and IMS, and serves as an associate editor for *Operations Research* and *Mathematics of Operations Research*. His email address is ton.dieker@isye.gatech.edu.

SEONG-HEE KIM is an Associate Professor in H. Milton Stewart School of Industrial and Systems Engineering at Georgia Institute of Technology. She received her Ph.D. in Industrial Engineering and Management Sciences from Northwestern University in 2001. She serves as an associate editor for *INFORMS Journal on Computing*. Her email address is skim@isye.gatech.edu.