

SELECTING THE BEST BY COMPARING SIMULATED SYSTEMS IN A GROUP OF THREE

Seong-Hee Kim
A. B. Dieker

Georgia Institute of Technology
765 Ferst Dr NW
Atlanta, GA 30332, USA

ABSTRACT

We present a new fully sequential procedure for selecting the best among a finite number of simulated systems. While many fully sequential selection procedures make a decision based on pairwise comparison, the new procedure compares systems in a group of three and uses some properties of a bivariate Brownian motion process exiting a circle or an ellipse for its derivation.

1. INTRODUCTION

The selection-of-the-best problem is to select the best system among a finite number of simulated systems and has been studied actively in the context of simulation ranking and selection (R&S). There exist at least three different approaches for the selection-of-the-best problem: the indifference-zone (IZ) method, the Bayesian method, and the optimal computing budget allocation (OCBA) method. The indifference-zone (IZ) method finds the best or a near-best with a guarantee on the probability of correct selection (PCS) while the other two methods maximize the PCS under a limited computational budget. In the IZ approach, a decision maker is assumed to be indifferent among systems whose means are within a user-specified constant of the true best system and the user-specified constant is called the indifference zone parameter. Kim and Nelson (2006b) gives a review of R&S procedures with the IZ method. See Chick (2006) for a thorough review of the Bayesian and OCBA methods.

Among selection procedures that take the IZ approach, fully sequential procedures are shown to be highly efficient. Fully sequential procedures with elimination take a single basic observation from each alternative still in play at each stage and eliminate systems from further consideration when there is a clear evidence that they are inferior. Kim and Nelson (2001) present fully sequential procedures that are useful in simulation environments. Kim and Nelson (2006a) and Malone et al. (2005) show that if variance estimates are updated as more observations are available, the performance of fully sequential procedures can be even further improved.

Many fully sequential procedures observe the partial sums of differences between a pair of systems and their statistical validity is derived using some properties of a univariate Brownian motion process. In this paper, we present a new fully sequential selection procedure that compares systems in a group of three rather than in a pair. The new procedure still takes the IZ approach but uses some properties of bivariate Brownian motion processes exiting a circle or an ellipse. The idea is to eliminate a system when it is worse than two systems simultaneously by some considerable amount although the amount is not big enough to eliminate the system in pairwise comparison. We also approximate an upper bound on the probability of correct selection and provide some arguments on why the procedure should work.

The remainder of the paper is organized as follows. In Section 2, we define the problem and provide notation and assumptions. Section 3 presents our new procedure and Section 4 provides some arguments on a guarantee about correct selection of the new procedure. In Section 5 experimental results are discussed, followed by concluding remarks in Section 6.

2. BACKGROUND

In this section, we introduce notation and assumptions. Then we define the selection-of-the-best problem and a brief overview of a popular fully sequential procedure based on pairwise comparisons.

2.1 Notation and Assumptions

We assume that there are k simulated systems ($k \geq 2$). Let X_{ij} be an observation from replication (or batch) j of system i for $i = 1, \dots, k$ and $j = 1, 2, \dots$. The set of all systems is defined as $S = \{1, \dots, k\}$. The mean and variance of the outputs from system i are defined as $\mu_i = E[X_{ij}]$ and $\sigma_i^2 = \text{Var}[X_{ij}]$, respectively.

Assumption 1. Let X_{ij} represent the j th observations from system i . Then

$$X_{ij} \stackrel{\text{IID}}{\sim} N(\mu_i, \sigma_i^2), \quad j = 1, 2, \dots,$$

where $\stackrel{\text{IID}}{\sim}$ represents ‘are independent and identically distributed as’ and N denotes normal distribution with mean μ_i and variance σ_i . Moreover, X_{ij} and $X_{\ell j}$ are independent for any $j \neq j'$.

Assumption 1 implies that the output data from each system is marginally IID normally distributed and systems are simulated independently (thus no common random numbers).

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

Without loss of generality, we assume that system 1 is the best and at least δ better than any alternative system. The user-specified parameter δ is the IZ parameter, a practically meaningful difference worth detecting. Then selection-of-the-best is to guarantee that under Assumptions 1 and 2 the following probability statement holds:

$$\text{PCS} = \Pr\{\text{select system 1}\} \geq 1 - \alpha.$$

Some notation is defined as follows. Let c be a positive integer, $\mathcal{I}(\cdot)$ be the indicator function, an integer $n_0 \geq 2$ be the initial sample size of each system, r be the current number of replications. For $i, \ell = 1, 2, \dots, k$, we define

$$\begin{aligned} \bar{X}_i(r) &\equiv \frac{1}{r} \sum_{j=1}^r X_{ij}, \text{ the sample mean of system } i \text{ based on the first } r \text{ observations;} \\ \sigma_{i\ell}^2 &\equiv \sigma_i^2 + \sigma_\ell^2, \text{ the variance of the difference between systems } i \text{ and } \ell; \\ R(r; a, b, d) &\equiv \max\left\{0, \frac{bd}{a} - \frac{a}{2c}r\right\} \text{ for a positive integer } c \text{ and any } a, b, d \in \mathbb{R}^+; \\ g(\eta) &\equiv \sum_{\ell=1}^c (-1)^{\ell+1} \left(1 - \frac{1}{2}\mathcal{I}(\ell = c)\right) \exp\left\{\frac{-\eta(2c - \ell)\ell}{c}\right\} \text{ for a positive integer } c; \\ W(t) &\equiv \text{a standard Brownian motion process;} \\ W(t, \Delta) &\equiv \text{a standard Brownian motion process with drift } \Delta; \end{aligned}$$

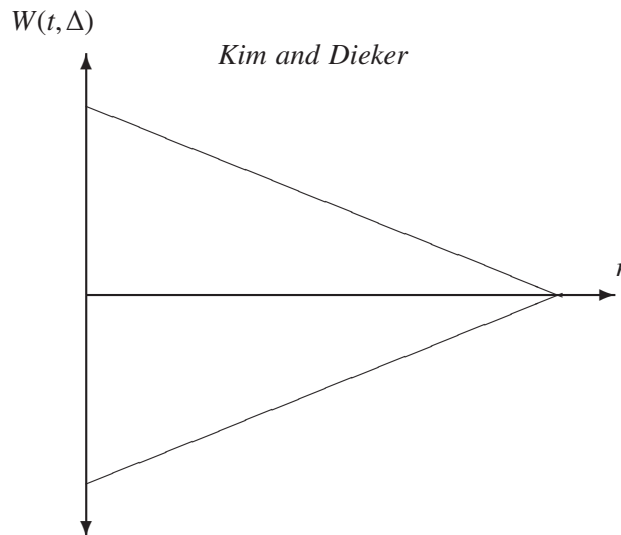


Figure 1: Triangular continuation region and decisions.

To illustrate how a typical fully sequential procedure works, we present a procedure from Wang and Kim (2011) under known variances, which is an extended version of a procedure due to Paulson (1964).

Procedure (\mathcal{P})

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

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

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

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**.

Remark: When $c < \infty$, $R(r; \cdot)$ defines a triangular continuation region as shown in Figure 1 but when $c = \infty$, it becomes a parallel line.

A typical fully sequential procedure observes $\sum_{j=1}^r (X_{ij} - X_{\ell j})$, a partial sum of differences between systems i and ℓ and one system is eliminated whenever the partial sum exits a triangular continuation region defined by $(-R(r; \cdot), R(r; \cdot))$.

3. NEW PROCEDURE

In this section, we propose a new procedure under the assumption that variances are known. The new procedure is as follows:

New Procedure

Setup: Select the nominal level $1 - \alpha$ and the IZ parameter δ . Set $\beta = 2\alpha/(k - 1)$ and calculate $\eta = -\lambda \ln \frac{\beta}{1-\beta}$ for $\lambda \in \mathbb{R}^+$ (the choice of λ will be discussed later in Section 4). Set $I = \{1, 2, \dots, k\}$ and take one observation from each system. Set $r = 1$ and go to **Screening**.

Screening: If $|I| \geq 3$, for each possible group of 3 systems in I , apply [**Screening-a**]. Else if $|I| = 2$, apply [**Screening-b**]:

Screening-a: Sort systems based on sample means $\bar{X}_{[1]}(r) \geq \bar{X}_{[2]}(r) \geq \bar{X}_{[3]}(r)$ where $[i]$ is the identity of a system with the i th largest sample mean in the current group. System [3] is eliminated from I if

$$\left(\frac{\sum_{j=1}^r (X_{[1]j} - X_{[3]j})}{\eta \times \frac{\sigma_{[1][3]}^2}{\delta}} \right)^2 + \left(\frac{\sum_{j=1}^r (X_{[2]j} - X_{[3]j})}{\eta \times \frac{\sigma_{[2][3]}^2}{\delta}} \right)^2 \geq 1. \tag{1}$$

Screening-b: Sort the survived systems $\bar{X}_{[1]}(r) \geq \bar{X}_{[2]}(r)$ and system [2] is eliminated from I if

$$\left(\frac{\sum_{j=1}^r (X_{[1]j} - X_{[2]j})}{\eta \times \frac{\sigma_{[1][2]}^2}{\delta}} \right)^2 \geq 1. \tag{2}$$

Stopping Rule: If $|I| = 1$, return the survived system as the best. Otherwise, set $r = r + 1$, obtain one additional observations for all $i \in I$, and repeat [**Screening**].

Note that (1) defines an ellipse and the inequality implies that system [3] is eliminated when the coordinate of the partial sums of differences between systems [1] and [3] and systems [2] and [3] exits the region defined by the ellipse. When there are only two systems still in play, (2) implies that system [2] is eliminated when the partial sum between systems [1] and [2] is greater than $\eta \times \sigma_{[1][2]}^2 / \delta$, which is the same rule as in Procedure (\mathcal{P}) when $c = \infty$.

4. PROBABILITY OF CORRECT SELECTION

The new procedure is not statistically valid but heuristic. However, we can approximate the lower bound on probability of correct selection (PCS) for $k = 3$ when variances are known and equal and means follow a slippage configuration (SC) in which $\mu_1 - \delta = \mu_2 = \mu_3$. The SC is considered as the most difficult mean configuration in many R&S procedures for the-selection-of-the-best problem (Kim and Nelson 2006b). Then we discuss how to use the bound for $k = 3$ to make the new procedure work for general k at the end of this section.

Under the known and equal variances, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2$ and $\sigma_{it}^2 = 2\sigma^2$. We assume that the three systems satisfy Assumptions 1. There are three ways of making an incorrect selection (ICS) which defines the following three events:

- ICS₁ \equiv the event that system 1 is eliminated first;
- ICS₂ \equiv the event that system 2 is eliminated first and then system 1 is eliminated; and
- ICS₃ \equiv the event that system 3 is eliminated first and then system 1 is eliminated.

Then the probability of incorrect selection is defined as follows:

$$\Pr\{\text{ICS}\} = \Pr\{\text{eliminate system 1}\} = \Pr\{\text{ICS}_1\} + \Pr\{\text{ICS}_2\} + \Pr\{\text{ICS}_3\}.$$

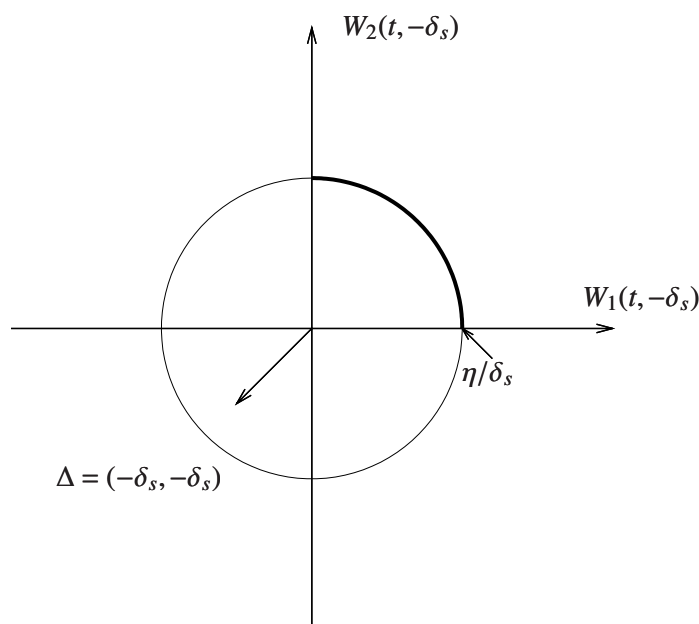


Figure 2: The first quadrant of the circle for the new procedure.

Some additional notation is needed.

$$\begin{aligned}
 C_{i\ell}(r) &\equiv \sum_{j=1}^r \frac{X_{ij} - X_{\ell j}}{\sigma_{i\ell}}; \\
 C_i(r) &\equiv (C_{1i}(r), \dots, C_{i-1,i}(r), C_{i+1,i}(r), \dots, C_{ki}(r)); \text{ and} \\
 O(a) &\equiv \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = a^2\}.
 \end{aligned}$$

4.1 ICS₁

Consider $\Pr\{\text{ICS}_1\}$. The event ICS_1 occurs if $C_1(r) = (C_{21}(r), C_{31}(r))$ exits the inside region defined by the circle $O(\eta\sqrt{2}\sigma/\delta)$ through the first quadrant and $C_2(r)$ and $C_3(r)$ have not crossed the circle through the first quadrant yet. The first quadrant of the circle is shown in Figure 2. Let ICS_b represent the event $C_1(r)$ hits $O(\eta\sqrt{2}\sigma/\delta)$ in the first quadrant. Then $\Pr\{\text{ICS}_1\} \leq \Pr\{\text{ICS}_b\}$. Note that if Assumption 1 holds, then it is clear that a discrete process $C_1(r)$ behaves like a continuous process

$$\mathbf{W}(t, \Delta) = \begin{bmatrix} W_1(t, -\delta_s) \\ W_2(t, -\delta_s) \end{bmatrix} = \Delta t + \Sigma^{1/2} \mathbf{W}(t),$$

where

$$\delta_s = \frac{\delta}{\sqrt{2}\sigma}, \quad \Delta = \begin{bmatrix} -\delta_s \\ -\delta_s \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}, \quad \mathbf{W}(t) = \begin{bmatrix} W_1(t) \\ W_2(t) \end{bmatrix},$$

and $W_1(t)$ and $W_2(t)$ are independent standard Brownian motion processes. Moreover, $\Sigma = Q\Lambda Q^T$ where

$$Q = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 1.5 & 0 \\ 0 & 0.5 \end{bmatrix}, \quad \text{and } Q^T \text{ is a transpose of } Q.$$

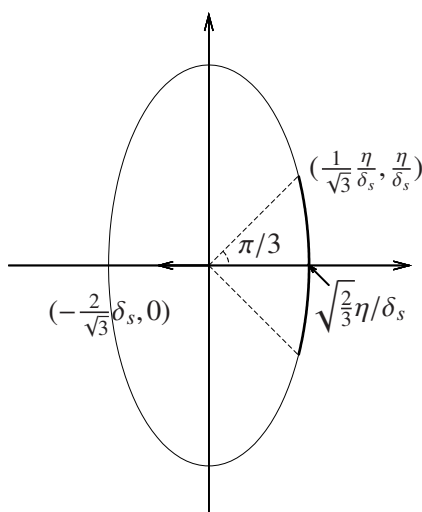


Figure 3: The circle from Figure 2 becomes an ellipse under the linear transformation $\Lambda^{-1/2}Q^T$.

The event ICS_b can be approximated by the event that $\mathbf{W}(t, \Delta)$ first hits $O(\eta/\delta_s)$ through the first quadrant. Whether or not $\mathbf{W}(t, \Delta)$ exits the circle requires studying the Euclidean norm of $\mathbf{W}(t, \Delta)$. The difficulty is that the coordinate processes are not independent. We first transform $\mathbf{W}(t, \Delta)$ into a vector with independent coordinate processes by left-multiplying by $\Lambda^{-1/2}Q^T$. Note that Q^T performs a clockwise rotation by $\pi/4$ and that $\Lambda^{-1/2}$ stretches the rotated circle into an ellipse as shown in Figure 3.

After this linear transformation, the process becomes

$$\Lambda^{-1/2}Q^T \mathbf{W}(t, \Delta) = \begin{bmatrix} -\frac{2}{\sqrt{3}}\delta_s \\ 0 \end{bmatrix} t + \mathbf{W}(t),$$

a vector of independent standard Brownian motion processes with drift. It is clear that the event ICS_b is the event that a Brownian motion with drift $(-\frac{2}{\sqrt{3}}\delta_s, 0)$ first hits the ellipse at a point with $-\pi/3 < \theta < \pi/3$ or equivalently $\frac{1}{\sqrt{3}}\frac{\eta}{\delta_s} \leq x \leq \sqrt{\frac{2}{3}}\frac{\eta}{\delta_s}$. We next discuss how to find a tight upper bound for $\Pr\{ICS_b\}$ using intuition from the theory of large deviations.

Consider two lines $L_U = \{(x, y) \in \mathbb{R}^2 : x = \frac{1}{\sqrt{3}}\frac{\eta}{\delta_s}\}$ and $L_L = \{(x, y) \in \mathbb{R}^2 : x = -\sqrt{\frac{2}{3}}\frac{\eta}{\delta_s}\}$ and a circle $O(\sqrt{\frac{2}{3}}\eta/\delta_s)$. Figure 4 shows L_U , L_L , and the circle. Let C be the curve which consists of part of the line L_U and part of the circle. That is,

$$C = \left\{ (x, y) \in \mathbb{R}^+ \times \mathbb{R}^+ : x^2 + y^2 = \left(\sqrt{\frac{2}{3}}\frac{\eta}{\delta_s} \right)^2 \text{ for } \frac{1}{\sqrt{3}}\frac{\eta}{\delta_s} \leq x \leq \sqrt{\frac{2}{3}}\frac{\eta}{\delta_s} \text{ and } \left(\frac{1}{\sqrt{3}}\frac{\eta}{\delta_s}, y \right), \text{ otherwise} \right\}.$$

Then the following is true:

$$\Pr\{ICS_b\} \leq \Pr\{\text{the Brownian motion with drift } (-2\delta_s/\sqrt{3}, 0) \text{ hits the curve } C \text{ before } L_L\} \quad (3)$$

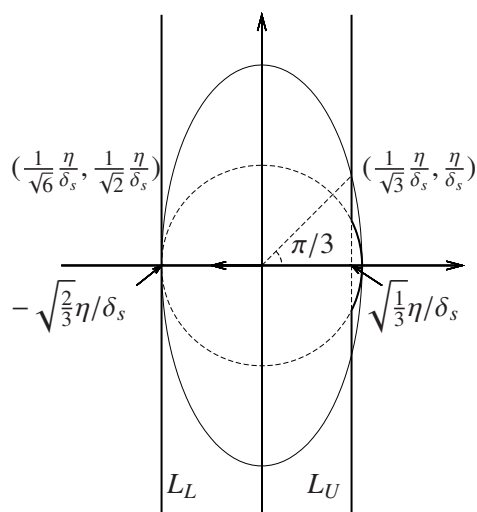


Figure 4: The lines L_U , L_L , and the circle $O(\sqrt{\frac{2}{3}}\eta/\delta_s)$.

The hitting distribution of a multidimensional drifting Brownian motion through fixed lines on either side of the stochastic process is well-known and readily implementable (e.g., Karlin and Taylor 1975). For this reason, instead of (3), it may be tempting to bound $\Pr\{\text{ICS}_b\}$ from above by the probability that the drifting Brownian motion hits L_U before L_L . This probability is small, but we can invoke the theory of large deviations (Dembo and Zeitouni 1998) to deduce that this bound will not be sharp. Indeed, large deviation theory predicts that the most likely way in which this event happens is for the process to hit L_U around the point $(\frac{1}{\sqrt{3}}\frac{\eta}{\delta_s}, 0)$. Since this point is ‘far’ from the thick line in Figure 4, the bound will not be sharp. By allowing a ‘dent’ in L_U in the right place as in Figure 4, if the process hits L_U before L_L , then large deviation theory predicts that each point along the dent is equally likely to be the hitting point. It is therefore a tighter upper bound. As the curve C is in part line and in part circle, the analysis of the probability on the right-hand side of (3) is considerably more delicate.

We condition on the point where the drifting Brownian motion first hits the circle $O(\sqrt{\frac{2}{3}}\eta/\delta_s)$.

$$\begin{aligned}
 \Pr\{\text{ICS}_b\} &\leq \Pr\{\text{hit the curve } C \text{ before hitting the ellipse}\} \\
 &\leq \Pr\{\text{hit the circle in } -\frac{\pi}{3} < \theta < \frac{\pi}{3}\} + \\
 &\quad \Pr\{\text{hit curve } C \text{ before hitting the ellipse} \mid \text{hit the circle in } \frac{\pi}{3} < \theta < \frac{5\pi}{3}\} \times \\
 &\quad \Pr\{\text{hit the circle in } \frac{\pi}{3} < \theta < \frac{5\pi}{3}\} \\
 &\leq \Pr\{\text{hit the circle in } -\frac{\pi}{3} < \theta < \frac{\pi}{3}\} + \\
 &\quad \Pr\{\text{hit line } L_U \text{ before } L_L \mid \text{hit the circle in } \frac{\pi}{3} < \theta < \frac{5\pi}{3}\} \times \Pr\{\text{hit the circle in } \frac{\pi}{3} < \theta < \frac{5\pi}{3}\} \\
 &= \int_{-\pi/3}^{\pi/3} \frac{e^{\{-\frac{2\sqrt{2}}{3}\eta \cos \theta\}}}{2\pi I_0(\frac{2\sqrt{2}}{3}\eta)} d\theta + 2 \int_{\pi/3}^{\pi} \frac{e^{\{\frac{4\sqrt{2}}{3}\eta \cos \theta\}} - e^{\{-\frac{4\sqrt{2}}{3}\eta\}}}{e^{\{\frac{4}{3}\eta\}} - e^{\{-\frac{4\sqrt{2}}{3}\eta\}}} \frac{e^{\{-\frac{2\sqrt{2}}{3}\eta \cos \theta\}}}{2\pi I_0(\frac{2\sqrt{2}}{3}\eta)} d\theta, \tag{4}
 \end{aligned}$$

where $I_\nu(z)$ is a Bessel function defined as $I_\nu(z) = \sum_{q=0}^{\infty} \frac{(z/2)^{2q+\nu}}{q!\Gamma(q+\nu+1)}$ for the gamma function $\Gamma(\cdot)$ and $\Pr\{\text{hit line } L_U \text{ before } L_L \mid \text{hit the circle in } \frac{\pi}{3} < \theta < \frac{5\pi}{3}\}$ is from Theorem 5.2 of Karlin and Taylor (1975). Note that the hitting distribution of a drifting Brownian motion to a circle is known explicitly and called the Von Mises distribution. This distribution is closely related to the theory of Markov intertwining and Bessel processes, see Rogers and Pitman (1981) for details.

4.2 ICS₂ and ICS₃

Now consider $\Pr\{\text{ICS}_2\}$. This implies that $C_2(r) = (C_{12}(r), C_{32}(r))$ exits the inner region defined by the circle through the first quadrant before $C_1(r)$ or $C_3(r)$ does. The corresponding continuous process is

$$\mathbf{W}(t, \Delta) = \begin{bmatrix} W_1(t, \delta_s) \\ W_2(t, 0) \end{bmatrix} = \Delta t + \Sigma^{1/2} \mathbf{W}(t),$$

where

$$\delta_s = \frac{\delta}{\sqrt{2}\sigma}, \quad \Delta = \begin{bmatrix} \delta_s \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}, \quad \mathbf{W}(t) = \begin{bmatrix} W_1(t) \\ W_2(t) \end{bmatrix}.$$

We apply the linear transformation to make the coordinate process independent and obtain an independent coordinate process

$$\Lambda^{-1/2} Q^T \mathbf{W}(t, \Delta) = \begin{bmatrix} \frac{\delta_s}{\sqrt{3}} \\ -\delta_s \end{bmatrix} t + \mathbf{W}(t).$$

For system 2 to be eliminated first, $\Lambda^{-1/2} Q^T \mathbf{W}(t, \Delta)$ should hit the ellipse for $-\frac{\pi}{3} < \theta < \frac{\pi}{3}$. Remember that $\Lambda^{-1/2} Q^T \mathbf{W}(t, \Delta)$ is a corresponding continuous process of $\Lambda^{-1/2} Q^T C_2(r)$ and

$$\Lambda^{-1/2} Q^T C_2(r) = \Lambda^{-1/2} Q^T \begin{bmatrix} \frac{\sum_{j=1}^r (X_{1j} - X_{2j})}{\sqrt{2}\sigma} \\ \frac{\sum_{j=1}^r (X_{3j} - X_{2j})}{\sqrt{2}\sigma} \end{bmatrix} = \begin{bmatrix} \frac{\sum_{j=1}^r (X_{1j} - 2X_{2j} + X_{3j})}{\sqrt{6}\sigma} \\ \frac{\sum_{j=1}^r (X_{3j} - X_{1j})}{\sqrt{2}\sigma} \end{bmatrix}.$$

Thus one can see that the value of the vertical axis of $\Lambda^{-1/2} Q^T \mathbf{W}(t, \Delta)$ when a hitting occurs is approximately the value of $C_{31}(r)$ at that time. Once the hitting to the ellipse occurs, $C_{31}(r)$ starting from the hitting point should hit η/δ_s before $-\eta/\delta_s$ to ensure that system 1 is eliminated (thus, ICS₂).

Now we derive the value of the vertical axis, y , as a function of θ when a hitting occurs on the ellipse for $-\frac{\pi}{3} < \theta < \frac{\pi}{3}$. As the equation of the ellipse is

$$\left(\frac{x}{\sqrt{\frac{2}{3}} \frac{\eta}{\delta_s}} \right)^2 + \left(\frac{y}{\sqrt{2} \frac{\eta}{\delta_s}} \right)^2 = 1,$$

we get the coordinate of the horizontal axis

$$x = \sqrt{\frac{2}{3}} \frac{\eta}{\delta_s} \sqrt{1 - \left(\frac{y}{\sqrt{2} \frac{\eta}{\delta_s}} \right)^2} \quad \text{for} \quad -\frac{\pi}{3} < \theta < \frac{\pi}{3}.$$

Then

$$\tan \theta = \frac{y}{\sqrt{\frac{2}{3}} \frac{\eta}{\delta_s} \sqrt{1 - \left(\frac{y}{\sqrt{2} \frac{\eta}{\delta_s}} \right)^2}} = \frac{\sqrt{3}\gamma}{\sqrt{1-\gamma^2}} \quad \text{where} \quad \gamma = \frac{y}{\sqrt{2} \frac{\eta}{\delta_s}}.$$

After some algebra, we get

$$\gamma = \pm \sqrt{\frac{(\tan \theta)^2}{3 + (\tan \theta)^2}} \quad \text{and} \quad y = \pm \sqrt{2} \frac{\eta}{\delta_s} \sqrt{\frac{(\tan \theta)^2}{3 + (\tan \theta)^2}}.$$

This implies that when $\Lambda^{-1/2} Q^T W(t, \Delta)$ hits the ellipse for $-\frac{\pi}{3} < \theta < 0$,

$$C_{31}(r) \approx \begin{cases} -\sqrt{2} \frac{\eta}{\delta_s} \sqrt{\frac{\tan \theta^2}{3 + \tan \theta^2}}, & -\frac{\pi}{3} < \theta < 0 \\ \sqrt{2} \frac{\eta}{\delta_s} \sqrt{\frac{\tan \theta^2}{3 + \tan \theta^2}}, & 0 < \theta < \frac{\pi}{3}. \end{cases}$$

Thus,

$$\begin{aligned} \Pr\{\text{ICS}_2\} &\leq \Pr\{C_{31}(r) \text{ reaches } \eta/\delta_s \text{ before } -\eta/\delta_s \mid \text{starting from the hitting point}\} \times \\ &\quad \Pr\{C_2(r) \text{ first hits the ellipse for } -\frac{\pi}{3} < \theta < \frac{\pi}{3}\} \\ &\leq \Pr\{C_{31}(r) \text{ reaches } \eta/\delta_s \text{ before } -\eta/\delta_s \mid \text{starting from the hitting point}\} \times \\ &\quad \Pr\{C_2(r) \text{ first hits circle } O(\sqrt{\frac{2}{3}}\eta/\delta_s) \text{ for } -\frac{\pi}{3} < \theta < \frac{\pi}{3}\} \\ &= \int_{-\frac{\pi}{3}}^0 \frac{\exp(-2\sqrt{2}\eta\sqrt{\frac{\tan \theta^2}{3 + \tan \theta^2}}) - \exp(-2\eta)}{\exp(2\eta) - \exp(-2\eta)} \frac{e^{\{\frac{2\sqrt{2}}{3}\eta\cos(\theta + \pi/3)\}}}{2\pi I_0(\frac{2\sqrt{2}}{3}\eta)} d\theta + \\ &\quad \int_0^{\frac{\pi}{3}} \frac{\exp(2\sqrt{2}\eta\sqrt{\frac{\tan \theta^2}{3 + \tan \theta^2}}) - \exp(-2\eta)}{\exp(2\eta) - \exp(-2\eta)} \frac{e^{\{\frac{2\sqrt{2}}{3}\eta\cos(\theta + \pi/3)\}}}{2\pi I_0(\frac{2\sqrt{2}}{3}\eta)} d\theta. \end{aligned} \tag{5}$$

Under the assumption of equal variances and $\mu_2 = \mu_3$, the probability of ICS_3 is equal to that of ICS_2 and thus it will have the same bound as in (5).

We take η in a form of $\eta = -\lambda \ln \frac{\alpha}{1-\alpha}$ for $k = 3$ because the form is in a similar form of a constant in Procedure (\mathcal{P}) or Paulson (1964). More specifically, when $c = \infty$, Procedure (\mathcal{P}) takes the constant equal to $-0.5 \ln(\frac{\beta}{1-\beta})$ where $\beta = \alpha/(k-1)$. The values of (4) and (5) for various λ are given in Table 1 when $\sigma^2 = 1$, $\mu_1 = \mu_2 = 0$, $\mu_3 = 0.3$, $\delta = 0.3$ and $\alpha = 0.05$ with 10,000 macro replications:

Table 1: Approximate values of bounds (4) and (5) under three different values of λ .

λ	(4)	(5)
0.6	0.0685	0.0181
0.65	0.0550	0.0143
0.7	0.0441	0.0113

The number of possible groups of three that contains system 1 and two alternatives that are not members of other groups is $(k-1)/2$. To apply the procedure for a general k , we heuristically takes $\eta = -\lambda \ln \frac{\beta}{1-\beta}$ where $\beta = 2\alpha/(k-1)$, which is same as splitting the overall error α into $(k-1)/2$ groups of three using the Bonferroni bounds. Paulson's procedure splits α into $(k-1)$ pairs.

5. EXPERIMENTS

In this section, we test the performance of the new procedure under various mean configurations and compare it with Procedure (\mathcal{P}). The number of systems k varies over 3, 5, 10, 25, and 100 and the nominal

confidence level is $1 - \alpha = 0.95$. Two mean configurations are considered: SC and monotonic decreasing means configuration (MDM). Under SC, $\mu_1 = \delta$ and $\mu_i = 0$ for $i = 2, \dots, k$. Under MDM, $\mu_i = (2 - i)\delta$. We consider equal variances only and variances are set to $\sigma_i^2 = 1$ for all i .

Tables 2 and 3 show some preliminary results of the new procedure: average total number of observations (REP) and estimated PCS based on 10,000 macro replications for the new procedure for various k under the SC and MDM configurations, respectively. They also shows REP for Paulson's procedure from Section 2 for comparison. At least the first three digits in REP are meaningful and the estimated PCS are meaningful up to the hundredth digit.

The new procedure outperforms Paulson for all configurations for both values of λ except $\lambda = 0.7$ and $k = 3$. We see more savings under MDM configuration and the amount of savings increases as k increases. This implies that the new procedure is more effective in eliminating inferior systems than Procedure (\mathcal{P}). Although Table 1 shows that $\lambda = 0.7$ is needed to guarantee the lower bound on PCS is close to the nominal level, our results show that $\lambda = 0.6$ results in estimated PCS more close to the nominal value while satisfying the PCS requirement.

Table 2: Total number of replications and estimated PCS under the SC configuration.

k	New ($\lambda = 0.6$)		New ($\lambda = 0.7$)		Procedure (\mathcal{P})
	REP	PCS	REP	PCS	REP
3	122	0.955	145	0.973	141
5	238	0.953	290	0.974	292
10	529	0.952	638	0.973	693
25	1423	0.954	1715	0.979	2044
100	6108	0.952	7603	0.982	9978

Table 3: Total number of replications and estimated PCS under the MDM configuration.

k	New ($\lambda = 0.6$)		New ($\lambda = 0.7$)		Procedure (\mathcal{P})
	REP	PCS	REP	PCS	REP
3	96	0.974	114	0.986	117
5	144	0.992	169	0.996	192
10	209	0.999	244	0.999	306
25	296	1.000	348	1.000	479
100	421	1.000	505	1.000	847

6. CONCLUSION

We present a new fully sequential procedure that takes a different paradigm than existing procedures. The new procedure compares systems in a group of three and inferior systems tend to be eliminated faster than the procedures based on pairwise comparisons. The discussion in the paper is based on the assumption that variances are known and equal, which is unrealistic. The relaxation of the assumption of equal variances requires more complicated linear transformation to obtain independent components in the continuous bivariate Brownian motion process. A procedure designed under known variances can be extended to unknown variances by adopting variance updates as in Kim and Nelson (2006a) or Malone et al. (2005). Extending the new procedure to unknown and unequal variances is a topic of our ongoing research.

ACKNOWLEDGMENTS

We thank Peter Frazier and David Goldsman for helpful discussions and Huizhu Crystal Wang for her help in drawing figures.

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, 225–257. Oxford: Elsevier Science.
- Dembo, A., and O. Zeitouni. 1998. *Large Deviations Techniques and Applications*. New York: Springer.
- Karlin, S., and H. M. Taylor. 1975. *A First Course in Stochastic Processes*. New York: Academic Press.
- 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.
- Kim, S.-H., and B. L. Nelson. 2006a. "On the asymptotic validity of fully sequential selection procedures for steady-state simulation". *Operations Research* 54:475–488.
- Kim, S.-H., and B. L. Nelson. 2006b. "Selecting the best system". In *Handbooks in Operations Research and Management Science: Simulation*, edited by S. G. Henderson and B. L. Nelson, 501–534. Oxford: Elsevier Science.
- Malone, G. J., S.-H. Kim, D. Goldsman, and D. Batur. 2005. "Performance of variance updating procedures on various data". In *Proceedings of the 2005 Winter Simulation Conference*, edited by M. E. Kuhl, N. M. Steiger, F. B. Armstrong and J. A. Joines, 825–832. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.
- Paulson, E. 1964. "A sequential procedure for selecting the population with the largest mean from k normal populations". *Annals of Mathematical Statistics* 35:174–180.
- Rogers, L.C.G., and J. W. Pitman. 1981. "Markov functions". *The Annals of Probability* 9:573–582.
- Wang, H., and S.-H. Kim. 2011. "On the conservativeness of fully sequential indifference-zone procedures". Technical report, H. Milton Stewart School of Industrial Systems and Engineering, Georgia Institute of Technology, Atlanta, GA.

AUTHOR BIOGRAPHIES

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. Her research interests include ranking and selection, optimization via simulation, quality control, simulation output analysis, and applications of simulation methods to environmental management. She is a member of INFORMS, and serves as an associate editor in the simulation area of *Operations Research* and the OR/simulation area of *The American Statistician*. She also served on the editorial board of the Simulation Department of *IIE Transactions*. Her e-mail and web addresses are skim@isye.gatech.edu and <http://www.isye.gatech.edu/~skim>, respectively.

ANTONIUS (TON) DIEKER is an Assistant Professor in H. Milton Stewart School of Industrial and Systems Engineering at Georgia Institute of Technology. He received his Ph.D. in Mathematics from University of Amsterdam in 2006. His research interests include service engineering, stochastic processes, and probability theory. He is a member of INFORMS, and serves as an associate editor in Operations Research. His e-mail and web addresses are ton.dieker@isye.gatech.edu and <http://www.isye.gatech.edu/~adieker3>, respectively.