

SELECTING THE BEST BY COMPARING SIMULATED SYSTEMS IN A GROUP OF THREE WHEN VARIANCES ARE KNOWN AND UNEQUAL

A. B. Dieker
Seong-Hee Kim

Georgia Institute of Technology
Atlanta, GA 30332, USA

ABSTRACT

This paper presents a fully sequential procedure for selecting the best system among a finite number of simulated systems with variances that are known but not necessarily equal. Our procedure compares systems in groups of three and is based on exit properties of a bivariate Brownian motion from an ellipse. The procedure is a modification of the one proposed in Kim and Dieker (2011) for the equal-variance case.

1. INTRODUCTION

Ranking and selection in simulation aims to select the best system among a finite number of simulated systems. How one defines ‘best’ depends on the problem at hand; here we assume the best system is the one with the largest expected performance measure. A number of statistical procedures have been developed for ranking and selection problems. In this paper, we consider procedures that take the indifference-zone (IZ) approach first presented by Bechhofer (1954). In the IZ approach, the decision maker is assumed to be indifferent to systems whose true means are within a practically meaningful difference, $\delta > 0$, from the best performance; and tries to find the best or near-best system with a guarantee on the probability of correct selection (PCS). The positive constant δ is called the IZ parameter.

Sequential procedures take one or more basic observations from each active system at each stage and tend to have a small number of stages. For example, Rinott (1978) presents a sequential procedure with two stages. Rinott’s procedure takes initial observations from each system in the first stage, calculates how many additional observations are needed to make a decision with a PCS guarantee, then take additional observations from each system in the second stage. On the other hand, fully sequential procedures take only one basic observation from each active system at each stage and tend to have a large number of stages. For example, see Kim and Nelson (2001, 2006ab). They propose fully sequential procedures with elimination that 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.

Many fully sequential procedures observe the partial sums of differences between a pair of systems and elimination is based on pairwise comparison. The work by Kim and Dieker (2011) departs from this idea by presenting a new fully sequential selection procedure that compares systems in a group of three rather than in a pair. Their procedure takes the IZ approach and uses some properties of bivariate Brownian motion processes exiting a circle. A related line of work is Frazier (2012), who presents a Bayesian-inspired indifference-zone procedure that is quite efficient when the number of systems is huge. Although these new procedures are promising, their developments are restricted to known and equal variances. In this paper, we propose a fully sequential selection procedure based on bivariate Brownian motion processes when variances are known but not necessarily equal. This is the first such contribution to the best of our knowledge.

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. Section 4 summarizes experimental results, followed by concluding remarks in Section 5.

2. BACKGROUND

This section introduces notation and assumptions and defines our problem.

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 observation 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^2 . 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 $\delta \in \mathbb{R}^+$.

This assumption states (without loss of generality) that system 1 is the best and that it is at least δ better than any alternative system. The user-specified parameter δ is the IZ parameter, a practically meaningful difference worth detecting.

We first need some notation. Let r be the current number of replications, and consider a group of three systems labeled [1], [2], and [3]. Moreover, b represents the identity of the system with the largest mean μ_i and i_1 and i_2 represent the remaining two systems in the group. Finally, we define

$$\begin{aligned} \bar{X}_i(r) &\equiv \frac{1}{r} \sum_{j=1}^r X_{ij}, \text{ the sample mean of system } i \text{ based on the first } r \text{ observations;} \\ \sigma_{i\ell}^2 &\equiv \sigma_i^2 + \sigma_\ell^2, \text{ the variance of the difference between systems } i \text{ and } \ell; \\ Y_1(r) &\equiv \sum_{\ell=1}^r \frac{X_{[1]\ell} - X_{[3]\ell}}{\sigma_{[1][3]}}; \\ Y_2(r) &\equiv \sum_{\ell=1}^r \frac{X_{[2]\ell} - X_{[3]\ell}}{\sigma_{[2][3]}}; \\ Y(r) &\equiv (Y_1(r), Y_2(r))^T, \text{ where the superscript } T \text{ stands for transpose;} \\ \Lambda &\equiv \begin{bmatrix} 1 & \frac{\sigma_{[3]}^2}{\sigma_{[1][3]}\sigma_{[2][3]}} \\ \frac{\sigma_{[3]}^2}{\sigma_{[1][3]}\sigma_{[2][3]}} & 1 \end{bmatrix}; \\ \Lambda_b &\equiv \begin{bmatrix} 1 & \frac{\sigma_b^2}{\sigma_{b,i_1}\sigma_{b,i_2}} \\ \frac{\sigma_b^2}{\sigma_{b,i_1}\sigma_{b,i_2}} & 1 \end{bmatrix}; \end{aligned}$$

$$\underline{\delta} \equiv \left(\frac{\delta}{\sigma_{b,i_1}}, \frac{\delta}{\sigma_{b,i_2}} \right)^T;$$

$$\delta_s \equiv \sqrt{\underline{\delta}^T \Lambda_b^{-1} \underline{\delta}}.$$

3. OUR PROCEDURE

In this section, we propose a new procedure under the assumption that variances are known but not necessarily equal. We call it Procedure \mathcal{E} since it is based on ellipses:

Procedure \mathcal{E}

Setup: Given the IZ parameter δ , select some $\eta > 0$; this choice is discussed later in this section. 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: For systems labeled [1], [2], and [3], eliminate the system with the smallest sample mean from I if

$$\mathbf{Y}(r)^T \Lambda^{-1} \mathbf{Y}(r) \geq \left(\frac{\eta}{\delta_s} \right)^2. \tag{1}$$

Screening-b: For systems labeled [1] and [2], eliminate the system with the smaller sample mean 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**].

The procedure differs slightly from a procedure due to Kim and Dieker (2011). The difference is that Kim and Dieker’s procedure uses a circle to determine which system to eliminate, while our Procedure \mathcal{E} uses a special ellipse, namely the one determined by the quadratic form associated with Λ^{-1} . Although this may appear to be a minor difference, there are two main reasons why Procedure \mathcal{E} holds promise compared to the approach in Kim and Dieker (2011). First, using the ellipse determined by Λ^{-1} allows us to perform an explicit error analysis when we replace the Gaussian random walk $\mathbf{Y}(r) = (Y_1(r), Y_2(r))$ by its Brownian counterpart; this is discussed below. In Kim and Dieker (2011), we instead used delicate bounds in our error analysis; for Procedure \mathcal{E} such bounds are not needed and this simplifies calculations greatly. We anticipate that this insight is particularly advantageous in order to develop our methodology in a higher-dimensional setting. A second key advantage of the specific ellipse chosen here is that Procedure \mathcal{E} is invariant under permuting the system labels. Indeed, the following lemma shows that we only need to consider a single quadratic form for each triplet of systems. This also has a computational advantage: unlike in Kim and Dieker (2011), we do not need to sort the system sample means for each triplet to verify whether the process has escaped from a circle or ellipse.

Lemma 1. For any $(x_1, x_2, x_3) \in \mathbb{R}^3$, we have

$$\begin{aligned} (x_2 - x_1, x_3 - x_1) \begin{bmatrix} \sigma_{12}^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_{13}^2 \end{bmatrix}^{-1} \begin{pmatrix} x_2 - x_1 \\ x_3 - x_1 \end{pmatrix} &= (x_1 - x_2, x_3 - x_2) \begin{bmatrix} \sigma_{12}^2 & \sigma_2^2 \\ \sigma_2^2 & \sigma_{23}^2 \end{bmatrix}^{-1} \begin{pmatrix} x_1 - x_2 \\ x_3 - x_2 \end{pmatrix} \\ &= (x_1 - x_3, x_2 - x_3) \begin{bmatrix} \sigma_{13}^2 & \sigma_3^2 \\ \sigma_3^2 & \sigma_{23}^2 \end{bmatrix}^{-1} \begin{pmatrix} x_1 - x_3 \\ x_2 - x_3 \end{pmatrix}. \end{aligned}$$

Proof. We only prove the first equality; the second follows by symmetry. Since

$$\begin{pmatrix} x_1 - x_2 \\ x_3 - x_2 \end{pmatrix} = \begin{bmatrix} -1 & 0 \\ -1 & 1 \end{bmatrix} \begin{pmatrix} x_2 - x_1 \\ x_3 - x_1 \end{pmatrix},$$

we need to prove that

$$\begin{bmatrix} -1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{12}^2 & \sigma_2^2 \\ \sigma_2^2 & \sigma_{23}^2 \end{bmatrix}^{-1} \begin{bmatrix} -1 & 0 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} \sigma_{12}^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_{13}^2 \end{bmatrix}^{-1}.$$

By taking inverses on both sides, we find that this is equivalent with

$$\begin{bmatrix} -1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{12}^2 & \sigma_2^2 \\ \sigma_2^2 & \sigma_{23}^2 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \sigma_{12}^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_{13}^2 \end{bmatrix},$$

and it is straightforward to verify this equality. □

Procedure \mathcal{E} requires a real-valued parameter η , and we next discuss how this parameter can be chosen. We define

$$\rho = \frac{\sigma_b^2}{\sigma_{b,i_1}\sigma_{b,i_2}}, \quad \theta_1 = \arctan\left(\sqrt{\frac{1+\rho}{1-\rho}} \cdot \frac{\sigma_{b,i_1} - \sigma_{b,i_2}}{\sigma_{b,i_1} + \sigma_{b,i_2}}\right), \quad \text{and } \theta_2 = \arctan\left(\sqrt{\frac{1+\rho}{1-\rho}}\right).$$

We let η be a solution to the following equation:

$$\int_{\theta_1 - \theta_2}^{\theta_1 + \theta_2} \frac{\exp(-\eta \cos \theta)}{2\pi I_0(\eta)} d\theta = \frac{2\alpha}{k-1} \tag{3}$$

where $I_\nu(z)$ is a modified Bessel function of the first kind defined as, for $z \geq 0$,

$$I_\nu(z) = \sum_{q=0}^{\infty} \frac{(z/2)^{2q+\nu}}{q! \Gamma(q+\nu+1)}$$

in terms of the gamma function $\Gamma(\cdot)$. We think of the parameter α as a proxy for the probability of incorrect selection, although we do not give a formal performance guarantee here. We intend to address this in future research. A lower α leads to a higher η and therefore a bigger ellipse; more observations are then required for a system to be eliminated.

Since η depends on variances of systems, it changes whenever a different group of three is considered. The left-hand side of Equation (3) is the probability that a continuous Brownian process with drift $\underline{\delta}$ exits the domain $\{(y_1, y_2) : (y_1, y_2) \Lambda^{-1}(y_1, y_2)^T < (\eta/\delta_s)^2\}$ in the first quadrant. When there are only two systems remaining, we use the η determined by Equation (3) for the last group of three.

Our algorithm requires knowledge of δ_s . In reality, for a given triplet of systems, we do not know which system is the best and therefore we cannot calculate δ_s . One way to circumvent this issue is to calculate three possible η/δ_s for each possibility $b = [1], [2],$ or $[3]$, and then take the largest η/δ_s . Using the largest η/δ_s will result in a wider ellipse than the one determined with knowledge of the best system. As a result, the probability of correct selection (PCS) will be higher, but the number of required observations to reach a decision will increase. If the smallest value of η/δ_s is used, then fewer observations will be used but the PCS will decrease. As a compromise, in practice one can choose the second largest (middle) value for η/δ_s . However, we view this paper as a step towards an increased understanding of the use of a high-dimensional Brownian motion in ranking and selection, rather than that it provides a procedure that can be implemented in practice. We leave the topic of designing an algorithm that does not use any system knowledge for further research, and note that such an algorithm should not take the system covariance structure as input. In next section, we thus present our experimental results assuming knowledge of δ_s .

4. EXPERIMENTS

In this section, we test the performance of Procedure \mathcal{E} under two mean and three variance configurations and compare it with a procedure (Procedure \mathcal{P}) due to Wang and Kim (2012), an extended version of a procedure due to Paulson (1964). For known variances, Procedure \mathcal{P} is the most efficient statistically-valid fully sequential IZ procedure.

Procedure \mathcal{P}

Setup: Select the nominal level $1 - \alpha$ and the IZ parameter δ . Calculate $\eta_p = -0.5 \ln \beta / (1 - \beta)$ where $\beta = \alpha / (k - 1)$. 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

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

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**.

We let the number of systems k vary over 3, 5, 10, and 25, and we set $\alpha = 0.05$ and $\delta = 0.3$. 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 test three variance configurations:

- Equal variances (EQUAL): $\sigma_i^2 = 1$ for $i = 1, \dots, k$;
- Increasing variances (INC): $\sigma_i^2 = 1 + (i - 1)\delta$ for $i = 1, \dots, k$; and
- Decreasing variances (DEC): $\sigma_i^2 = \frac{1}{1 + (i - 1)\delta}$ for $i = 1, \dots, k$.

Thus, we have six configurations total: SC-EQUAL, SC-INC, SC-DEC, MDM-EQUAL, MDM-INC, and MDM-DEC. The estimated probability of correct selection (PCS) and the average number of observations until a decision is made (REP) are reported based on 4000 macro replications.

Table 1 shows the estimated PCS of our procedure. Estimated PCS are close or over the nominal level except the SC-INC case with $k = 10$ or 25.

Table 1: Estimated PCS.

k	SC			MDM		
	EQUAL	INC	DEC	EQUAL	INC	DEC
3	0.957	0.962	0.963	0.983	0.982	0.980
5	0.957	0.956	0.961	0.990	0.991	0.990
10	0.952	0.935	0.964	0.997	0.996	0.997
25	0.950	0.874	0.973	0.999	0.999	0.999

Tables 2 and 3 compare the average number of replications of our procedure and Procedure \mathcal{P} . Procedure \mathcal{E} works better than Procedure \mathcal{P} , and shows similar performance as the procedure due to Kim and Dieker (2011) for the equal-variance case. There is some overlap between the numerical results for Procedure \mathcal{E} and the results reported in Kim and Dieker (2011) when the variances are equal. This is because the ellipse from our procedure becomes a circle in the equal-variance setting, and because we use the same experimental setting as in Kim and Dieker (2011), including the same random number generator. Moreover, some of the experiments use the same the critical parameter η .

Table 2: Average total number of replications under the SC configuration.

k	EQUAL		INC		DEC	
	\mathcal{E}	\mathcal{P}	\mathcal{E}	\mathcal{P}	\mathcal{E}	\mathcal{P}
3	135	134	163	166	114	112
5	252	292	341	412	200	222
10	547	693	936	1273	357	426
25	1444	2044	4065	6259	602	790

Table 3: Average total number of replications under the MDM configuration.

k	EQUAL		INC		DEC	
	\mathcal{E}	\mathcal{P}	\mathcal{E}	\mathcal{P}	\mathcal{E}	\mathcal{P}
3	107	106	124	125	92	92
5	153	192	185	243	132	158
10	218	306	278	437	180	236
25	311	479	453	908	234	330

5. CONCLUSIONS

We present a modification of the procedure Kim and Dieker (2011) to cope with known but not necessarily equal variances. The new procedure compares systems in a group of three and uses an ellipse to determine if an inferior system can be eliminated. The procedure works better than Procedure \mathcal{P} but experiences degradation in PCS under the SC configuration with increasing variances and $k \geq 10$. It requires knowledge of δ_s and the system variances, which is unlikely to be available in practice. We are currently investigating how these drawbacks can be overcome.

ACKNOWLEDGMENTS

This work is supported by NSF grant CMMI-1131047.

REFERENCES

- Bechhofer, R. E. 1954. "A single-sample multiple decision procedure for ranking means of normal populations with known variances". *Annals of Mathematical Statistics* 25:16–39.
- 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.
- Frazier, P. 2012. "Indifference-Zone Ranking and Selection for More Than 15,000 Alternatives". Technical Report, Cornell University, Ithaca, NY.
- 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.
- 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.
- 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. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- 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.
- Rinott, Y. 1978. "A two-stage selection procedures and related probability inequalities". *Communications in Statistics* A7:799–811.
- Wang, H., and S.-H. Kim. 2012. "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

ANTONIUS (TON) B. 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 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 e-mail and web addresses are ton.dieker@isye.gatech.edu and <http://www.isye.gatech.edu/~adieker3>, respectively.

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 OR/simulation area of *The American Statistician*. She also served on the editorial board of the Simulation Department of *IIE Transactions* and *Operations Research*. Her e-mail and web addresses are skim@isye.gatech.edu and <http://www.isye.gatech.edu/~skim>, respectively.