Fully Sequential Procedures for Comparing Constrained Systems via Simulation

Sigrún Andradóttir, Seong-Hee Kim

The H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0205

Received 16 October 2007; revised 16 September 2009 and 26 December 2009; accepted 7 February 2010
DOI 10.1002/nav.20408
Published online 9 April 2010 in Wiley InterScience (www.interscience.wiley.com).

Abstract: We consider the problem of finding the system with the best primary performance measure among a finite number of simulated systems in the presence of a stochastic constraint on a single real-valued secondary performance measure. Solving this problem requires the identification and removal from consideration of infeasible systems (Phase I) and of systems whose primary performance measure is dominated by that of other feasible systems (Phase II). We use indifference zones in both phases and consider two approaches, namely, carrying out Phases I and II sequentially and carrying out Phases I and II simultaneously, and we provide specific example procedures of each type. We present theoretical results guaranteeing that our approaches (general and specific, sequential and simultaneous) yield the best system with at least a prespecified probability, and we provide a portion of an extensive numerical study aimed at evaluating and comparing the performance of our approaches. The experimental results show that both new procedures are useful for constrained ranking and selection, with neither procedure showing uniform superiority over the other. Naval Research Logistics 57: 403–421, 2010

Keywords: ranking and selection; constraints; multiple performance measures; fully sequential procedures

1. INTRODUCTION

Our objective in this article is to select the best or near-best system from a set of competing systems, in that we want to maximize or minimize a primary performance measure, while simultaneously ensuring that the selected system satisfies a stochastic constraint on a secondary performance measure. This is a challenging task because of randomness in output data.

Over the last decade, there have been fruitful efforts in developing statistically valid or heuristic ranking and selection (R&S) procedures that are useful for finding a system with the best expected performance among a finite number of simulated systems. For example, Boesel et al. [4], Nelson et al. [23], Kim and Nelson [18, 19], Goldsman et al. [12], Chen [6], Chen et al. [7], Chick [8], and Chick and Inoue [9, 10] all developed procedures that are a great deal more efficient than classical R&S procedures such as that due to Rinott [26].

A limitation of these procedures lies in the fact that they consider only one performance measure. In reality, because of physical or managerial limits placed on a system, we often face constraints on performance measures other than the primary performance measure. When such constraints are stochastic in that the secondary performance measures need to be evaluated using simulation, existing R&S procedures lose their statistical guarantee about a correct choice, defeating the main advantage of using R&S procedures over heuristic procedures.

We now provide two examples where stochastic constraints are present:

EXAMPLE 1: A job shop consists of several machine groups, each with a number of identical machines. The company is interested in determining where it should add capacity (more machines). The primary performance measure of interest is the expected profit, and the decision maker wants to keep the expected cycle time in the system no greater than a certain level.

EXAMPLE 2: An internet-based company wants to identify an inventory policy with the smallest cost among several possible policies. Competition in the on-line market is heavy, and frequent failure in delivering customers' demands on time affects the reputation of the company. Therefore, the company wants to make sure that the percentage of delayed demands is less than a certain level.

There is not a rich literature on solving discrete optimization problems with stochastic constraints using simulation.

Correspondence to: S.-H. Kim (skim@isye.gatech.edu)

Butler et al. [5] combined multiple attribute utility theory from economics with the Rinott [26] procedure to handle multiple performance measures. Their procedure finds the system that gives the highest utility, but the procedure does not handle constraints, and finding an appropriate attribute utility function itself can be difficult. Santner and Tamhane [27] proposed a procedure that is designed to find the system with the largest mean among those whose variances are smaller than a constant. This procedure handles only a special case of our problem and cannot be applied for a general stochastic constraint. Another issue with the procedures of Butler et al. [5] and Santner and Tamhane [27] (and many R&S procedures in general) is that they become inefficient for 20 or more systems (see, e.g., [4]).

The procedures discussed above are two-stage procedures where initial observations are taken from each system in the first stage, the simulation stops to calculate how many additional observations are needed to satisfy the desired probability requirement, and then the additional second-stage observations are collected from each system. By contrast, *fully sequential* procedures take only a single basic observation from each system still in play at each stage, and many sequential procedures also apply a decision strategy at every stage to eliminate apparently inferior systems early in the experimentation process. For this reason, fully sequential procedures are expected to reduce the overall simulation effort required to find the best system. Paulson [24] and Hartmann [13, 14] presented fully sequential procedures for the case of independent and identically distributed (IID) normal data with known or unknown and equal variances, and Kim and Nelson [18] extended their procedures to both unequal and unknown variances across systems and dependence among systems because of the use of common random numbers (CRNs).

In this article, we provide the first formulation of a new problem, namely, constrained R&S. We also propose efficient fully sequential procedures that handle comparisons in the presence of a stochastic constraint on a single real-valued secondary performance measure. This requires the identification and removal from consideration of infeasible systems (Phase I) and of systems whose primary performance measure is dominated by that of other feasible systems (Phase II). Two general approaches are presented, namely carrying out Phases I and II sequentially and carrying out Phases I and II simultaneously. We prove that our procedures guarantee that the best feasible system is selected correctly with a prespecified probability and provide an extensive numerical study aimed at comparing the performance of sequentially running and simultaneously running procedures.

To our knowledge, running two statistical comparison procedures (e.g., R&S procedures) simultaneously has never before been proposed in the simulation literature. Thus, our simultaneously running approach introduces a completely new way of constructing combined R&S techniques. On the other hand, previous work has considered the sequential application of two R&S procedures developed for the common goal of finding the best system (see, e.g., [4, 23]). Our sequentially running approach is different in that we combine two R&S procedures developed for different goals to find the best feasible system. Also, note that the first phases of the procedures in Nelson et al. [23] and Boesel et al. [4] involve taking a predetermined number of samples from each system, while the completion time of the first phase of our sequentially running approach is random.

This article is organized as follows: In Section 2, we formulate the problem of interest. In Section 3, we present a procedure that finds a set of systems that satisfy a stochastic constraint. Then, Sections 4 and 5 cover our sequentially running and simultaneously running procedures, respectively. In both cases, we first provide a framework for solving the constrained R&S problem that a user can patch his/her own algorithms into, along with assumptions that the algorithms should satisfy. Then, we propose example algorithms that fit within the frameworks. Section 6 shows experimental results for the proposed procedures, and some concluding remarks are provided in Section 7. Finally, the Appendix contains the proof of one of the main results in this article. An overview of preliminary work on this article appeared in Andradóttir et al. [1].

2. PROBLEM FORMULATION

In this section, we formulate our problem and define notation for the article. Let X_{ij} be a univariate real-valued observation associated with the primary performance measure from replication (or batch) j of system i, and let Y_{ij} be a univariate real-valued observation associated with the constraint (secondary performance measure) from replication (or batch) j of system i. There are $k \geq 2$ systems. The primary and secondary performance measures are defined as $x_i = \mathrm{E}[X_{ij}]$ and $y_i = \mathrm{E}[Y_{ij}]$ for $i = 1, \ldots, k$, respectively. Our problem is to find

$$\begin{aligned} \arg \max_{i=1,\ldots,k} \quad & x_i \\ \text{s.t.} \quad & y_i \leq Q, \end{aligned}$$

where the constant Q is specified by the user and $x_i, y_i \in \mathbb{R}$. We need the following assumption:

ASSUMPTION 1: For each $i = 1, 2, \ldots, k$,

$$\begin{bmatrix} X_{ij} \\ Y_{ij} \end{bmatrix} \overset{\text{IID}}{\sim} \mathbf{BN}\left(\begin{bmatrix} x_i \\ y_i \end{bmatrix}, \boldsymbol{\Sigma}_i \right), \quad j = 1, 2, \ldots,$$

where $\overset{\text{IID}}{\sim}$ and **BN** represent "are independent and identically distributed as" and "bivariate normal," respectively, and $\boldsymbol{\Sigma}_i$

is the 2×2 positive-definite variance-covariance matrix of a vector (X_{ij}, Y_{ij}). Also, the vectors (X_{ij}, Y_{ij}) and $(X_{\ell\ell'}, Y_{\ell\ell'})$ are independent for $(i, j) \neq (\ell, \ell')$.

Assumption 1 implies that output data from each system are marginally IID normally distributed with positive variances and that all systems are simulated independently (CRNs are not considered in this article). IID normality is plausible, for example, when the basic observations of system performance are either within-replication averages (from a transient or steady-state simulation) or batch means with a large batch size (from a steady-state simulation), see, for example, Chapter 9 of Law and Kelton [21]. For steady-state simulation, we have implicitly assumed that the warm-up period is large enough to remove initialization bias. Also, we allow for dependence between X_{ij} and Y_{ij} because these random variables are likely to be correlated in practice. For instance, the average cycle time for a job shop is usually positively correlated with the average number of jobs in the system.

Although we only consider comparison problems with two performance measures (primary and secondary) in this article, these performance measures could be expectations involving functions of several outputs of the simulation. Also, a decision maker may be interested in performance measures involving probabilities, variances, quantiles, or functions of several expectations. Our problem formulation is general enough to handle such problems because probabilities and variances are expectations of indicator random variables and sample variances, respectively, and quantiles and functions of expectations can be viewed as expectations of quantile estimates and functions of random variables, respectively (ignoring bias in the quantile estimates and functions of random variables when the functions are nonlinear). For a constraint on a probability, basic (0-1) observations Y_{ij} can be obtained from each replication. For a constraint on a variance or quantile of IID observations, batching is needed, and the sample variances or quantile estimates from the batches will be used as basic observations Y_{ij}. For a function of expectations, basic observations could be obtained from the individual replications, but batching is preferred to reduce bias when the function of interest is nonlinear.

The non-normality of basic observations may be a problem. Kim and Nelson [18] and Malone et al. [22] show that fully sequential R&S procedures tend to be robust to non-normality. Also, any non-normality can be lessened by using batches of non-normal data as basic observations (as in [18]). In Section 6.3, we test the robustness of our procedures to IID non-normal data, with and without batching. Handling-correlated observations are more complicated and difficult. We do not discuss it in this article as our interest is in IID observations.

For stochastic systems, it is not always possible to guarantee that we determine correctly whether each system satisfies the stochastic constraint. Instead, we adopt an idea similar to that of the indifference-zone approach of Bechhofer [3] to find a set of feasible or near-feasible systems. A decision maker will be asked to specify a range $[Q^-, Q^+]$ around the constant Q such that $Q^- < Q^+$. Then, three regions are defined:

- $y_i \leq Q^-$: This is the desirable region. Any system in this range is feasible.
- $Q^- < y_i < Q^+$: This is the acceptable region. We assume that one is willing to accept a system in this range as the best if it is declared "feasible," and our procedure chooses it as the best—even if $y_i > Q$. Similarly, we assume that one is willing to decline systems in this region—even if $y_i \leq Q$.
- $y_i \geq Q^+$: This is the unacceptable region. A system in this range is infeasible and should be declined.

Further, the notations S_D, S_A, and S_U denote the sets of all desirable, acceptable, and unacceptable systems, respectively.

With this setup, there are two types of risk for systems in the acceptable region:

- If a feasible system i is in the acceptable region (so that $Q^- < y_i \leq Q$), then it may be declared infeasible and eliminated (Risk I).
- If an infeasible system i is in the acceptable region (so that $Q < y_i < Q^+$), then it may be declared feasible and not eliminated (Risk II).

The Risk I and II concepts are similar to Type I and II errors for a hypothesis test. The values of Q^- and Q^+ should be chosen depending on which risk a decision maker thinks is more important. For example, suppose that a constraint is imposed on the expected number of delayed demands per month for an inventory system, and that it should be no more than 100. If a decision maker defines $(Q^-, Q^+) = (100, 110)$, it means that she thinks Risk I should be minimized and Risk II is relatively less important. On the other hand, if she defines $(Q^-, Q^+) = (90, 100)$, it implies that Risk II should be minimized, and Risk I is not that critical. By choosing $Q^- < Q$ and $Q^+ > Q$, one can achieve a balance between Risks I and II.

For given Q^- and Q^+, we define $q = (Q^- + Q^+)/2$ and $\epsilon = (Q^+ - Q^-)/2$, and henceforth our procedures will be presented in terms of q and ϵ rather than Q, Q^-, and Q^+. Roughly speaking, q is a target value that behaves as a cutoff point between desirable and unacceptable systems, and ϵ is a tolerance level that specifies how much we are willing to be off from q.

In addition to q and ϵ, the decision maker needs to choose the indifference-zone parameter δ for the primary performance measure. This is the smallest absolute difference in the primary measure that the decision maker thinks is important to detect. Thus, only "practical differences" matter, and we are indifferent among desirable and acceptable systems whose primary performance measures are within δ of the true best. Any system whose primary performance measure is at least δ smaller (larger) than system i is called an inferior (superior) system to system i.

To solve the overall constrained optimization problem, we propose two phases: Phase I to identify desirable systems and Phase II to compare desirable systems. Actually, for finding the best feasible system, Phase I only needs to ensure that the best feasible system survives and that all infeasible systems with better primary performance measures than the best feasible system (superior infeasible systems) are eliminated (it does not matter what feasibility decision is made about inferior systems). However, it is not possible to take advantage of this relaxation of the goal of Phase I because it is not known in advance how many systems are feasible or infeasible and which infeasible systems are inferior or superior to the best feasible system. Consequently, we design a procedure for Phase I that guarantees an event that is a bit more strict; namely, finding a set that contains all the desirable systems, possibly with some acceptable systems, but without any unacceptable systems. That is, a correct decision (CD) for Phase I is defined as the selection of a set F such that $S_D \subseteq F \subseteq (S_D \cup S_A)$ with at least a predetermined probability.

Both phases need to be completed correctly with high probability. One can perform Phases I and II sequentially (so that Phase II is applied to the survivors from the completed Phase I). Alternatively, both Phases I and II can be performed simultaneously. A procedure that can be used in Phase I is presented in Section 3. We propose sequentially running procedures in Section 4 and simultaneously running procedures in Section 5.

3. A PROCEDURE FOR FEASIBILITY DETERMINATION

In this section, we provide a procedure for Phase I that attempts to eliminate all the unacceptable systems and return a set consisting of all the desirable systems, possibly including some acceptable systems. First, we define c as a positive integer, $\mathcal{I}$ as the indicator function, and

$n_0 \equiv$ initial sample size for each system ($n_0 \geq 2$);

$r \equiv$ the current stage number ($r \geq n_0$);

$S_i^2 \equiv$ sample variance of $Y_{i1}, \ldots, Y_{in_0}$ for $i = 1, 2, \ldots, k$;

$$R(r; v, w, z) \equiv \max\left\{0, \frac{wz}{2cv} - \frac{v}{2c}r\right\}, \quad \text{for any } v, w, z \in \mathbb{R},\ v \neq 0;$$

$$g(\eta) \equiv \sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1 - \frac{1}{2}\mathcal{I}(\ell = c)\right) \times \left(1 + \frac{2\eta(2c-\ell)\ell}{c}\right)^{-(n_0-1)/2}.$$

Let $|A|$ denote the number of elements in a set A. Now we give an algorithm for Phase I.

Algorithm $\mathcal{F}$: Feasibility Determination Procedure

Setup: Select $n_0 \geq 2$ and nominal probability of correct decision $0 < 1 - \alpha_1 < 1$. For the constraint under consideration, choose ϵ and q. Compute $\eta_1 > 0$, a solution to

$$g(\eta_1) = 1 - (1 - \alpha_1)^{1/k}. \tag{1}$$

Initialization: Let $M = \{1, 2, \ldots, k\}$ and $F = \emptyset$ be the set of systems whose feasibility is not determined yet and the set of feasible systems, respectively. Let $h_1^2 = 2c\eta_1(n_0 - 1)$. Obtain n_0 observations Y_{ij}, $j = 1, 2, \ldots, n_0$, from each system i. Compute S_i^2. Set the stage counter $r = n_0$ and go to Feasibility Check.

Feasibility Check: For each system $i \in M$, if $\sum_{j=1}^{r}(Y_{ij} - q) \leq -R(r; \epsilon, h_1^2, S_i^2)$, then move i from M to F; else if $\sum_{j=1}^{r}(Y_{ij} - q) \geq +R(r; \epsilon, h_1^2, S_i^2)$, then eliminate i from M.

Stopping Rule: If $|M| = 0$, then return F as a set of feasible systems.
Otherwise, take one additional observation $Y_{i,r+1}$ from each system $i \in M$, set $r = r + 1$, and go to Feasibility Check.

REMARK 1: A selection problem usually has $1/k$ as a lower bound for the nominal confidence level because $1/k$ is the probability that a randomly selected system is the best. However, Algorithm $\mathcal{F}$ is not for selection but for feasibility determination, and thus care is needed even if the nominal confidence level is smaller than $1/k$.

REMARK 2: In practice, the choice $c = 1$ is recommended because it guarantees a unique and easy solution of Eq. (1). Also, Kim and Nelson [18] argue that $c = 1$ is a good choice when one does not have any information about the configuration of the systems' means. The proofs in this article are given for a general value of c, but our experimental study will be performed with $c = 1$.

REMARK 3: Algorithm $\mathcal{F}$ can be performed without simulating the systems simultaneously. This avoids the overhead of switching among systems, a typical disadvantage of fully sequential procedures applied to selecting the best among a number of simulated systems.

Algorithm $\mathcal{F}$ is fully sequential and has triangular continuation regions that are defined by $R(r;\cdot)$. Sampling from system i continues until the partial sums $\sum_{j=1}^{r}(Y_{ij}-q)$ depart from the continuation region. If the exit occurs through the upper (lower) boundary, then we conclude that the system is infeasible (feasible). Also, note that Algorithm $\mathcal{F}$ is closely related to the procedures for comparison with a standard presented by Kim [17]. In particular, Kim's procedures [17] can be interpreted as special cases of Algorithm $\mathcal{F}$ with $Q^- = \mu_0 - \delta$, $Q^+ = \mu_0$, $q = \mu_0 - \delta/2$, and $\epsilon = \delta/2$, where μ_0 denotes the expected performance of the standard and δ is the indifference-zone parameter between the standard and alternative systems.

We have the following result whose proof is provided in the Appendix.

THEOREM 1: Under Assumption 1, Algorithm $\mathcal{F}$ guarantees

$$\Pr\{\mathrm{CD}\} = \Pr\{S_{\mathrm{D}} \subseteq F \subseteq (S_{\mathrm{D}} \cup S_{\mathrm{A}})\} \geq 1 - \alpha_1.$$

4. SEQUENTIALLY RUNNING PROCEDURES

This section discusses how procedures for Phases I and II can be applied sequentially. In particular, in Section 4.1, we provide a framework that enables us to combine procedures for Phases I and II, and in Sections 4.2 and 4.3, a number of example procedures are presented.

4.1. A Framework for Sequentially Running Procedures

Let $\mathcal{P}_{\mathrm{I}}$ and $\mathcal{P}_{\mathrm{II}}$ be procedures for Phases I and II, respectively. In this subsection, we present conditions that $\mathcal{P}_{\mathrm{I}}$ and $\mathcal{P}_{\mathrm{II}}$ should satisfy to guarantee the construction of statistically valid sequentially running procedures that will identify the best or near-best desirable or acceptable system with prespecified probability.

Suppose that $\mathcal{P}_{\mathrm{I}}$ satisfies the following assumption.

ASSUMPTION 2: $\mathcal{P}_{\mathrm{I}}$ is a procedure for Phase I that obtains observations Y_{ij} associated with the secondary performance measure from each system i and determines a random subset F of $\{1, 2, \ldots, k\}$ such that $\Pr\{\mathrm{CD}\} \geq 1 - \alpha_1$.

Suppose that $[b]$ represents the identity of the best desirable system and that any system $i \in (S_D \cup S_A)$ such that $x_i > x_{[b]} - \delta$ is called a "good" system. Let $J_\ell, \ell = 1, 2, \ldots, 2^{|S_A|}$, be the distinct subsets of $(S_D \cup S_A)$ such that $S_D \subseteq J_\ell \subseteq (S_D \cup S_A)$. We define a good selection (GS) event as the event that $\mathcal{P}_{\mathrm{II}}$ chooses any system with $x_i > x_{[b]} - \delta$ from the returned set F by $\mathcal{P}_{\mathrm{I}}$. The correct selection (CS) event is defined as the event that a good desirable or acceptable system is chosen by the combined procedure, $\mathcal{P}_{\mathrm{I}} + \mathcal{P}_{\mathrm{II}}$. Note that if the set of systems F returned by $\mathcal{P}_{\mathrm{I}}$ contains an unacceptable system with $x_i > x_{[b]} - \delta$, then a GS in Phase II does not guarantee a CS. Thus, we need both a CD of $\mathcal{P}_{\mathrm{I}}$ and a GS of $\mathcal{P}_{\mathrm{II}}$ to guarantee a CS.

Let E be a set that contains all systems in $(S_D \cup S_A)$ and additional $|S_U|$ number of "imaginary" desirable or acceptable systems with primary performance measures at most $x_{[b]} - \delta$ (so that E contains k systems). The additional $|S_U|$ number of systems in E do not actually exist. Adding these imaginary systems increases the likelihood of mistakenly choosing an inferior system as the best and hence leads to a lower bound on $\Pr\{\mathrm{CS}\}$.

Moreover, let $\mathrm{GS}(J_\ell)$ be the (good selection) event that all systems $i \in J_\ell$ with $x_i \leq x_{[b]} - \delta$ are eliminated when $\mathcal{P}_{\mathrm{II}}$ is applied to J_ℓ, and let $\mathrm{GS}_{\mathrm{pSC}}(J_\ell)$ be the event that system $[b]$ eventually eliminates all systems $i \in J_\ell$ with $x_i \leq x_{[b]} - \delta$ when $\mathcal{P}_{\mathrm{II}}$ is applied to observations X'_{ij} such that

$$X'_{ij} = \begin{cases} X_{ij} - x_{[b]}, & \text{for } i \in J_\ell \text{ and } x_i > x_{[b]} - \delta, \\ X_{ij} - x_i - \delta, & \text{for } i \in J_\ell \text{ and } x_i \leq x_{[b]} - \delta, \end{cases}$$

where $j = 1, 2, \ldots$ (we say that system $[b]$ eventually eliminates system i when system $[b]$ would eliminate system i if systems i and $[b]$ were considered in isolation). Notice that with this setting of observations, the mean of system $[b]$ is zero and all inferior systems have mean $-\delta$. This configuration is similar to the slippage configuration (SC) where all systems are assumed to have mean $-\delta$ except for the best system, which has mean zero. The SC configuration is known as a difficult configuration in the sense that it is hard to detect inferior systems because all of the inferior systems are close to the best. Our configuration with X'_{ij} is not exactly in the SC because there may be multiple good systems $i \in S_A$ with $x_i > x_{[b]} - \delta$, and thus we call this setting a pseudo slippage configuration (pSC). The definition of $\mathrm{GS}_{\mathrm{pSC}}(E)$ is analogous to that of $\mathrm{GS}_{\mathrm{pSC}}(J_\ell)$.

Then, the procedure $\mathcal{P}_{\mathrm{II}}$ needs the following assumption:

ASSUMPTION 3: $\mathcal{P}_{\mathrm{II}}$ satisfies

1. $\mathrm{GS}_{\mathrm{pSC}}(J_\ell) \subseteq \mathrm{GS}(J_\ell)$ for all J_ℓ;
2. $\mathrm{GS}_{\mathrm{pSC}}(E) \subseteq \mathrm{GS}_{\mathrm{pSC}}(J_\ell)$ for all J_ℓ; and
3. $\Pr\{\mathrm{GS}_{\mathrm{pSC}}(E)\} \geq 1 - \alpha_2$.

The first statement means that if system $[b]$ eventually eliminates all inferior desirable and acceptable systems in J_ℓ

with $\mathcal{P}_{\mathrm{II}}$ under the pSC configuration, then $\mathcal{P}_{\mathrm{II}}$ will eliminate all inferior systems in J_ℓ under the original mean configuration (x_i) of the systems in J_ℓ. This statement holds for any procedure whose elimination decision is based on the differences in observations between two systems and whose continuation regions do not depend on the means of the systems, if there are no good systems in $S_D \cup S_A$ other than $[b]$. The second statement implies that if system $[b]$ would eliminate all inferior desirable and acceptable systems when $\mathcal{P}_{\mathrm{II}}$ is applied to E under the pSC configuration, then system $[b]$ would do so if the procedure was applied to any J_ℓ under the pSC configuration. The third statement means that system $[b]$ will eliminate all inferior desirable and acceptable systems with probability at least $1-\alpha_2$ when $\mathcal{P}_{\mathrm{II}}$ is applied to E under the pSC configuration. Any statistically valid fully sequential procedure will satisfy the second and third statements if its statistical validity is established on the basis of correctly eliminating an inferior system in individual pairwise comparisons with system $[b]$ under the pSC configuration on E.

Let $\mathrm{GS} = (\cap_{\ell=1}^{|S_A|} \mathrm{GS}(J_\ell))$ and $\mathrm{GS}_{\mathrm{pSC}} = (\cap_{\ell=1}^{|S_A|} \mathrm{GS}_{\mathrm{pSC}}(J_\ell))$. Now, we present the main lemma.

LEMMA 1: Under Assumptions 2 and 3, the combined procedure $\mathcal{P}_{\mathrm{I}} + \mathcal{P}_{\mathrm{II}}$ guarantees

$$\begin{aligned}\Pr\{\mathrm{CS}\} &= \Pr\{\text{select a system } i \text{ such that } x_i > x_{[b]} - \delta \\ &\qquad \text{and } i \in (S_D \cup S_A)\} \\ &\geq 1 - (\alpha_1 + \alpha_2).\end{aligned}$$

PROOF: Any outcome belonging to the event GS results in choosing a good system, provided that the subset of systems considered by $\mathcal{P}_{\mathrm{II}}$ is one of the J_ℓ. The event CD of $\mathcal{P}_{\mathrm{I}}$ guarantees that this is the case. Consequently, we have that

$$\begin{aligned}\Pr\{\mathrm{CS}\} &\geq \Pr\{\mathrm{CD} \cap \mathrm{GS}\} \\ &= \Pr\{\mathrm{CD}\} + \Pr\{\mathrm{GS}\} - \Pr\{\mathrm{CD} \cup \mathrm{GS}\} \\ &\geq \Pr\{\mathrm{CD}\} + \Pr\{\mathrm{GS}_{\mathrm{pSC}}\} - 1 \\ &\geq \Pr\{\mathrm{CD}\} + \Pr\{\mathrm{GS}_{\mathrm{pSC}}(E)\} - 1 \\ &\geq (1-\alpha_1) + (1-\alpha_2) - 1,\end{aligned}$$

where the second inequality holds because of Statement 1 of Assumption 3, the third equality follows from Statement 2 of Assumption 3, and the fourth inequality follows by Assumption 2 and Statement 3 of Assumption 3. □

A procedure for Phase II of sequentially running procedures, $\mathcal{P}_{\mathrm{II}}$, can either (i) start a completely new simulation of the survivors from the completed Phase I to obtain X_{ij} or (ii) it can save the values of X_{ij} while the Y_{ij} are collected during Phase I and use the saved values of X_{ij} in Phase II later (together with additional observations of X_{ij} collected in Phase II if needed). We present sequential procedures of the first kind in Section 4.2 and sequential procedures of the second kind in Section 4.3.

4.2. Statistically Valid Sequentially Running Procedures

In this section, we propose a specific sequentially running procedure, which we call the Restarting Procedure, and prove that this procedure is statistically valid. In particular, we use Algorithm $\mathcal{F}$ as $\mathcal{P}_{\mathrm{I}}$ in Phase I where we determine the feasibility of each system. Consequently, Assumption 2 holds under Assumption 1, see Theorem 1.

If one is willing to restart from scratch the simulation of each survivor in Phase II, then it is not necessary to save the values of X_{ij} during Phase I. Further, all systems surviving from the completed Phase I start with the same number of (new) observations n_0. Therefore, any statistically valid fully sequential procedure that satisfies Assumption 3 can be used in Phase II. Example procedures include $\mathcal{KN}$ of Kim and Nelson [18], $\mathcal{SSM}$ of Pichitlamken, et al. [25], and those of Hong and Nelson [15] and Batur and Kim [2].

The procedure description will be given for Algorithm $\mathcal{F}$ and $\mathcal{KN}$. Before we present the new procedure, we need some additional notation. In particular, let

$$\begin{aligned}S_{i\ell}^2 &\equiv \text{sample variance of } X_{i1} - X_{\ell 1}, \ldots, X_{in_0} - X_{\ell n_0}, \\ &\qquad \text{the differences between systems } i \text{ and } \ell \text{ for all } i \neq \ell; \\ \bar{X}_i(r) &\equiv \text{sample average of } X_{ij},\ j = 1, 2, \ldots, r.\end{aligned}$$

Restarting Procedure

Setup: Select the overall nominal confidence level $1/k \leq 1-\alpha < 1$ and choose $\alpha_1 > 0$ and $\alpha_2 > 0$ such that $\alpha_1 + \alpha_2 = \alpha$. Choose ϵ and q. Also select the indifference-zone parameter δ and first-stage sample size $n_0 \geq 2$. Calculate η_1 from Eq. (1).

Initialization for Phase I: Same as in Algorithm $\mathcal{F}$.

Feasibility Check: Same as in Algorithm $\mathcal{F}$.

Stopping Rule for Phase I: If $|M| = 0$, then

- if $|F| = 0$, stop and conclude no feasible system;
- if $|F| = 1$, stop and return the system in F as the best; or
- if $|F| > 1$, go to "Initialization for Phase II."

Otherwise, take one additional observation $Y_{i,r+1}$ for each $i \in M$. Set $r = r + 1$. Then, go to "Feasibility Check."

Initialization for Phase II: Let η_2 be a solution to $g(\eta_2) = \alpha_2/(k-1)$ and $h_2^2 = 2c\eta_2(n_0 - 1)$. Let $M = F$ be the set of systems still in contention. For each system $i \in M$, perform an entirely new simulation and obtain

$X_{i1}, \ldots, X_{in_0}$ independent of any Y_{ij} generated in Phase I. Compute $\bar{X}_i(n_0)$ and $S^2_{i\ell}$ for $i, \ell \in M$ and $i \neq \ell$. Set $r = n_0$ and and go to "Comparison."

Comparison: Set $M^{\text{old}} = M$. Let

$$M = \Big\{ i : i \in M^{\text{old}} \text{ and } r\bar{X}_i(r) > r\bar{X}_\ell(r) - R(r; \delta, h_2^2, S^2_{i\ell}), \quad \forall \ell \in M^{\text{old}}, \ell \neq i \Big\}.$$

Stopping Rule for Phase II: If $|M| = 1$, then stop and select the system whose index is in M as the best. Otherwise, for each system $i \in M$, take one additional observation $X_{i,r+1}$ and compute $\bar{X}_i(r+1)$. Then, set $r = r + 1$ and go to "Comparison."

To prove the validity of the Restarting Procedure, we need the following assumption on the mean configuration of the x_i:

ASSUMPTION 4: For any $i \in S_D \cup S_A$ with $i \neq [b]$, $x_i \leq x_{[b]} - \delta$.

Assumption 4 implies that there is only one good system $[b]$ and that the rest of the desirable and acceptable systems are inferior to system $[b]$. This assumption is useful in the R&S literature (see for example [20]).

Now, we present the two main results for the Restarting Procedure.

THEOREM 2: Under Assumptions 1 and 4, the Restarting Procedure guarantees $\Pr\{\text{CS}\} \geq 1 - \alpha$.

PROOF: We have shown earlier that Algorithm $\mathcal{F}$ satisfies Assumption 2 when Assumption 1 holds. Moreover, $\mathcal{KN}$ satisfies the first two statements of Assumption 3 since (i) its elimination decision is based on the differences in observations between two systems, Assumption 4 holds, and the continuation region defined by $R(r; \cdot)$ does not depend on the mean configuration of systems (Statement 1) and (ii) elimination is done via individual pairwise comparison and the continuation region does not depend on the number of surviving systems (Statement 2). Finally, since

$$\Pr\{\text{GS}_{\text{pSC}}(E)\} \geq 1 - \sum_{i \in E \setminus \{[b]\}} \Pr\{\text{system } i \text{ would eventually eliminate system } [b] \text{ in the pSC}\},$$

it follows from Assumptions 1 and 4 and the proof of Theorem 1 of Kim and Nelson [18] that Statement 3 of Assumption 3 holds. As we choose α_1 and α_2 such that $\alpha_1 + \alpha_2 = \alpha$, the result now follows from Lemma 1. □

THEOREM 3: If (i) $\alpha_1 > 0$ and $\alpha_2 > 0$ are chosen such that $(1 - \alpha_1)(1 - \alpha_2) = 1 - \alpha$ in "Setup" and (ii) in "Initialization for Phase II," η_2 is a solution to the equation $g(\eta_2) = \alpha_2/(m-1)$ where $m = |F|$, the number of survivors from Phase I that need to be compared in Phase II, then the Restarting Procedure guarantees $\Pr\{\text{CS}\} \geq 1 - \alpha$ under Assumptions 1 and 4.

PROOF: Assumptions 1 and 4, together with the independence between the data (and hence decisions) in the two phases and Theorem 1 of Kim and Nelson [18], imply that $\Pr\{\text{GS} | F = J_\ell\} \geq 1 - \alpha_2$ for all J_ℓ. Therefore,

$$\begin{aligned}\Pr\{\text{CS}\} &\geq \sum_{\ell=1}^{2^{|S_A|}} \Pr\{F = J_\ell\} \Pr\{\text{GS} \mid F = J_\ell\} \\ &\geq \sum_{\ell=1}^{2^{|S_A|}} \Pr\{F = J_\ell\}(1 - \alpha_2) = \Pr\{\text{CD}\}(1 - \alpha_2) \\ &\geq (1 - \alpha_1)(1 - \alpha_2) = 1 - \alpha,\end{aligned}$$

where the last inequality follows from Theorem 1, and the last equality follows from the way we choose α_1 and α_2. □

We use the term Restarting Procedure II to refer to the variant of the original Restarting Procedure with parameters determined as in Theorem 3. As $(1 - \alpha_1)(1 - \alpha_2)$ is always larger than $1 - (\alpha_1 + \alpha_2)$, Restarting Procedure II can achieve the same overall probability $1 - \alpha$ with slightly larger values of α_1 and α_2. Larger values of α_1 and α_2 result in smaller values of η_1 and η_2 when $c = 1$. This helps increase the efficiency of Phases I and II although the difference $\alpha_1\alpha_2$ between $(1 - \alpha_1)(1 - \alpha_2)$ and $1 - (\alpha_1 + \alpha_2)$ is typically quite small (in the third decimal place or beyond). In addition, as η_2 is determined in terms of m rather than k, Restarting Procedure II is more efficient than the original Restarting Procedure, and thus we consider Restarting Procedure II only in the comparison with other procedures in Section 5.3.

The values of α_1 and α_2 can affect the performance of the two restarting procedures. If Phase I is difficult because of the secondary performance measures of many systems being close to q, then we may want to choose α_1 larger than α_2 to increase the efficiency of Phase I. On the other hand, if Phase I is easy relative to the comparisons between systems in Phase II, then we may want to let α_2 be larger than α_1 to make Phase II more efficient. However, information about the mean configurations of the primary and secondary performance measures is usually not available in advance. Hence, as in Nelson et al. [23], we choose $\alpha_1 = \alpha_2 = \alpha/2$.

4.3. Variations of the Restarting Procedure

The use of the two restarting procedures presented in Section 4.2 means that all the X_{ij} we could have collected

during Phase I will not be used. Given that doing so yields guaranteed statistical validity, restarting may not be a bad idea if Phase I ends early, say with r around n_0. However, it is clear that restarting can be quite wasteful, especially when Phase I is long and many systems survive Phase I. Hence, it is desirable to develop sequential approaches that store the values of X_{ij} generated in Phase I for use in Phase II. However, this approach may require significant memory space if we have to save the values of all individual observations X_{ij} taken during Phase I. Also, it may be the case that different numbers of data points X_{ij} will be available for each survivor at the end of Phase I, which means that we need to start with different initial sample sizes across systems in Phase II. Therefore, it is desirable for $\mathcal{P}_{\text{II}}$ to have the following property:

PROPERTY 1: $\mathcal{P}_{\text{II}}$ makes efficient use of memory in that it only requires saving sample means and sample variances (rather than the raw observations) and is capable of handling different (random) initial sample sizes across systems at the beginning of the simulation.

The sequential selection with memory procedure ($\mathcal{SSM}$) due to Pichitlamken et al. [25] extends the procedure of Kim and Nelson [18] for use within an optimization-via-simulation algorithm when simulation is costly and (partial or complete) information on alternatives previously visited is maintained. Systems are allowed to start with different (deterministic) initial sample sizes. Moreover, $\mathcal{SSM}$ requires saving only sample means and the number of observations taken so far, which removes the memory space issue discussed above.

In this section, we adopt the $\mathcal{SSM}$ procedure as $\mathcal{P}_{\text{II}}$ and present variations of the Restarting Procedure that save the values of X_{ij} during Phase I and use the saved values of X_{ij} in Phase II later. The following algorithm simply applies Algorithm $\mathcal{F}$ first to find a set F of feasible systems (Phase I) and then uses $\mathcal{SSM}$ to complete Phase II and select the best system in F with the X_{ij} collected during Phase I as initial samples from each system i.

Procedure $\mathcal{AK}$

Setup: Same as in the Restarting Procedure.

Initialization for Phase I: Same as in the Restarting Procedure except for the following additional steps:
Obtain n_0 observations X_{ij}, $j = 1, 2, \ldots, n_0$, from each system, $i = 1, 2, \ldots, k$.
For all $i \neq \ell$, compute the estimators $S^2_{i\ell}$.
For each system i, compute $\bar{X}_i(n_0)$ and set $n_i = n_0$, the number of observations taken so far from system i during Phase I.

Feasibility Check: Same as in the Restarting Procedure. Section 3.

Stopping Rule for Phase I: Same as in the Restarting Procedure except that whenever one additional observation $Y_{i,r+1}$ is taken for $i \in M$, we also take one additional observation $X_{i,r+1}$, set $n_i = n_i + 1$, and compute $\bar{X}_i(n_i)$ for each $i \in M$.

Initialization for Phase II: Same as in the Restarting Procedure except that we do not perform any new simulations and η_2 is now a solution to $g(\eta_2) = \alpha_2/(m-1)$, where $m = |F|$.

Comparison: Set $M^{\text{old}} = M$. Let

$$M = \left\{ i : i \in M^{\text{old}} \text{ and } r\bar{X}_i(n_i) > r\bar{X}_\ell(n_\ell) - R\left(r; \delta, h_2^2, S_{i\ell}^2\right), \forall \ell \in M^{\text{old}}, \ell \neq i \right\}.$$

Stopping Rule for Phase II: If $|M| = 1$, then stop and select the system whose index is in M as the best.
Otherwise, for each system $i \in M$ with $n_i \leq r$, take one additional observation $X_{i,r+1}$, set $n_i = n_i + 1$, and compute $\bar{X}_i(n_i)$. Then, set $r = r + 1$ and go to "Comparison."

Although we have not proved the validity of $\mathcal{AK}$, we did not find a single experimental example where the validity is violated. This will be discussed in more detail in Section 6. There are two difficulties with respect to proving validity of $\mathcal{AK}$. The first difficulty is that the number of observations X_{ij} saved in Phase I for surviving system $i \in F$ is random and depends on the observations Y_{ij} collected in Phase I. As X_{ij} and Y_{ij} are likely to be correlated, this dependence may impact the comparisons between systems in Phase II. Similarly, the second difficulty is that continuation region used by $\mathcal{AK}$ in Phase II depends on the number of survivors m of Phase I, another source of dependency. The first difficulty mentioned above can be avoided through the use of the restarting approach. The second difficulty can be avoided by defining η_2 based on k rather than m, as if all k systems were either desirable or acceptable systems and as if they all survived Phase I. Then η_2 in the "Initialization for Phase II" step of $\mathcal{AK}$ would be a solution to the equation

$$g(\eta_2) = \frac{\alpha_2}{k-1}. \tag{2}$$

We call this approach Procedure $\mathcal{C}$ (for "conservative"). When $c = 1$, $g(\eta_2)$ is decreasing in η_2, and thus η_2 in $\mathcal{AK}$ will be smaller than that in Procedure $\mathcal{C}$. This implies that $\mathcal{AK}$ always reaches a decision earlier for a particular sample path of X_{ij} and Y_{ij} and is therefore more efficient than Procedure $\mathcal{C}$ when $c = 1$. $\mathcal{AK}$ is also more efficient than the Restarting Procedure and Restarting Procedure II. The fact that $\mathcal{AK}$ also

appears to yield statistical validity in practice (although this is not guaranteed) suggests that $\mathcal{AK}$ is a desirable choice of a sequentially running approach for comparing constrained systems.

5. SIMULTANEOUSLY RUNNING PROCEDURES

In this section, we introduce an approach that implements Phases I and II simultaneously, rather than sequentially, to solve the constrained optimization problem considered in this article. We provide a lemma that guarantees the construction of statistically valid simultaneously running procedures in Section 5.1, followed by an example procedure in Section 5.2.

5.1. A Framework for Simultaneously Running Procedures

To present a key lemma that guarantees the construction of statistically valid simultaneously running procedures, we adopt notation similar to that in Section 4.1. In particular, let $\mathcal{P}'_{\mathrm{I}}$ and $\mathcal{P}'_{\mathrm{II}}$ be procedures for Phases I and II of a simultaneously running procedure, respectively. Also, let CD_i represent the event of a correct decision on the feasibility of system $i \in \{1, \ldots, k\}$ if $\mathcal{P}'_{\mathrm{I}}$ was applied to system i independently from $\mathcal{P}'_{\mathrm{II}}$. Note that when $i \in S_A$, the decision on the feasibility of system i is correct regardless of whether system i is declared feasible or infeasible. We need the following assumption for $\mathcal{P}'_{\mathrm{I}}$:

ASSUMPTION 5: Procedure $\mathcal{P}'_{\mathrm{I}}$ guarantees $\Pr\{\cap_{i \in S} \mathrm{CD}_i\} \geq (1-\beta)^t$ for any $1 \leq t \leq k$ and any subset S of $\{1, \ldots, k\}$ with cardinality t.

Unlike sequentially running procedures, we do not have a memory space issue for $\mathcal{P}'_{\mathrm{II}}$ because the X_{ij} need not be saved for later use. Also, $\mathcal{P}'_{\mathrm{II}}$ is not required to handle different initial sample sizes across systems. For all $i \in \{1, \ldots, k\} \setminus \{[b]\}$, let GS_i represent the event that system $[b]$ eliminates system i when systems $[b]$ and i are compared in isolation (i.e., $[b]$ would eventually eliminate i if neither system has been eliminated earlier by other systems). Any R&S procedure for the finding-the-best problem can be used for Phase II as long as it satisfies the following assumption:

ASSUMPTION 6: Procedure $\mathcal{P}'_{\mathrm{II}}$ guarantees $\Pr\{\cap_{i \in S} \mathrm{GS}_i\} \geq (1-\beta)^t$ for any $1 \leq t \leq k-1$ and any subset S of $\{i \in \{1, \ldots, k\} : x_i \leq x_{[b]} - \delta\}$ with cardinality t.

The general simultaneously running procedure works as follows:

$\mathcal{P}'_{\mathrm{I}} + \mathcal{P}'_{\mathrm{II}}$ Procedure

1. Determine the values of the required parameters. Put all systems in M, the set of undetermined systems. Let F and SS_i be empty, where F and SS_i denote the set of feasible systems and the set of superior systems to i, respectively, and $i \in \{1, \ldots, k\}$. Take the first-stage sample of size n_0 from all systems.
2. Do Feasibility Check. If i is declared feasible, move i from M to F, and for all ℓ with $i \in SS_\ell$, eliminate ℓ from all sets and delete SS_ℓ. If i is declared infeasible, eliminate i from all sets and delete SS_i.
3. Do Comparison. For all $i, \ell \in (M \cup F), i \neq \ell$, whose superiority is not determined yet (so that $\ell \notin SS_i$ and $i \notin SS_\ell$), compare i and ℓ. If system ℓ is found superior to i and $\ell \in F$, then eliminate i from all sets and delete SS_i. If system ℓ is found superior to i and $\ell \in M$, then add system ℓ to SS_i.
4. If $|M| = 0$ and there is only one or zero element in F, then stop. Otherwise, take additional observations from all survivors (i.e., systems in $M \cup F$) except those whose feasibility and superiority versus all other survivors are already determined and go to Feasibility Check.

For example, suppose that $M = \{2, 3\}$, $F = \{1\}$, $SS_1 = \{2, 3\}$, and $SS_2 = SS_3 = \emptyset$. Then, we already know that system 1 is feasible but inferior to systems 2 and 3. Therefore, there is no need to obtain additional observations for system 1, but we do need more observations for systems 2 and 3 to determine their feasibility and superiority versus each other.

The following lemma holds for the combined procedure $\mathcal{P}'_{\mathrm{I}} + \mathcal{P}'_{\mathrm{II}}$:

LEMMA 2: Under Assumptions 4, 5, and 6, the combined procedure $\mathcal{P}'_{\mathrm{I}} + \mathcal{P}'_{\mathrm{II}}$ guarantees that

$$\Pr\{\mathrm{CS}\} \geq 1 - \left\{\beta + 2\left[1 - (1-\beta)^{(k-1)/2}\right]\right\}.$$

PROOF: Recall that the combined procedure allows the elimination of a system only when it is declared infeasible or when it is found to be inferior to a system declared feasible. Suppose that there are k systems, including j unacceptable systems, where j is an unknown constant. CS is guaranteed when all unacceptable systems and all inferior desirable and acceptable systems are eliminated. We need the following definitions for two events:

$\mathcal{A}^*$: all systems $i \in S_U$ would eventually be eliminated by being declared infeasible; and

$\mathcal{B}^*$: system $[b]$ is declared feasible and all systems $l \in (S_D \cup S_A) \setminus \{[b]\}$ would eventually be eliminated by being declared inferior to system $[b]$.

We have

$$\begin{aligned}\Pr\{\text{CS}\} &= \Pr\{\text{all } i \in S_U \text{ and all } l \in (S_D \cup S_A) \text{ with } x_l \le x_{[b]} - \delta \text{ are eliminated}\} \\ &\ge \Pr\{\mathcal{A}^* \cap \mathcal{B}^*\} \ge \Pr\{\mathcal{A}^*\} + \Pr\{\mathcal{B}^*\} - 1.\end{aligned}$$

Note that

$$\begin{aligned}\Pr\{\mathcal{A}^*\} &= \Pr\{\text{CD}_i \text{ for all } i \in S_U\} \\ &\ge (1-\beta)^j \text{ (by Assumption 5), and} \\ \Pr\{\mathcal{B}^*\} &= \Pr\{\text{CD}_{[b]} \cap (\text{GS}_l \text{ for all } l \in (S_D \cup S_A) \text{ with } l \ne [b])\} \\ &\ge \Pr\{\text{CD}_{[b]}\} + \Pr\{\cap_{l \in (S_D \cup S_A)\setminus\{[b]\}} \text{GS}_l\} - 1 \\ &\ge (1-\beta) + (1-\beta)^{k-j-1} - 1 \\ &\quad \text{(by Assumptions 4, 5, and 6).}\end{aligned}$$

Therefore,

$$\begin{aligned}\Pr\{\text{CS}\} &\ge (1-\beta)^j + (1-\beta) + (1-\beta)^{k-j-1} - 2 \\ &\ge 1 - \beta - 2[1-(1-\beta)^{(k-1)/2}],\end{aligned} \tag{3}$$

as the minimum of the right-hand side of Eq. (3) is achieved at $j = (k-1)/2$. □

REMARK 4: Assumptions 5 and 6 can be generalized in terms of β_1 and β_2, respectively (instead of taking $\beta_1 = \beta_2 = \beta$). In this case, the lower bound on Pr{CS} given in Eq. (3) becomes

$$(1-\beta_1)^j + (1-\beta_1) + (1-\beta_2)^{k-j-1} - 2, \tag{4}$$

and we can identify β_1 and β_2 so that the minimum of (4) is equal to α. However, the issue of choosing the values of (β_1, β_2) is similar to that of choosing the values of (α_1, α_2) for sequentially running procedures as discussed in Section 4.2. Given that information about the mean configurations is usually not available in advance, the choice of $\beta_1 = \beta_2 = \beta$ is reasonable and keeps the determination of the minimum of (4) simple.

5.2. A Statistically Valid Simultaneously Running Procedure

In this subsection, we present a specific simultaneously running procedure that we call $\mathcal{AK}+$. We use Algorithm $\mathcal{F}$ as $\mathcal{P}'_{\text{I}}$ and the $\mathcal{KN}$ procedure due to Kim and Nelson [18] as $\mathcal{P}'_{\text{II}}$. $\mathcal{SSM}$ can still be used as $\mathcal{P}'_{\text{II}}$, but $\mathcal{KN}$ provides tighter screening than $\mathcal{SSM}$ when systems are simulated independently. The $\mathcal{AK}+$ procedure is provided below.

Procedure $\mathcal{AK}+$

Setup: Select the overall confidence level $1/k \le 1-\alpha < 1$. Choose ϵ, q, δ, and $n_0 \ge 2$. Find η, a solution to $g(\eta) = \beta$, where $\beta \in (0,1)$ is the solution to the equation

$$\beta + 2[1-(1-\beta)^{(k-1)/2}] = \alpha. \tag{5}$$

Initialization: Let $M = \{1, 2, \ldots, k\}$ and $F = \emptyset$ be the set of undetermined systems and the set of feasible systems, respectively. Also, let $SS_i = \emptyset$ be the set of superior systems to system i in terms of x_i. Let $h^2 = 2c\eta(n_0-1)$. Obtain n_0 observations X_{ij} and Y_{ij}, $j = 1, 2, \ldots, n_0$, from each system $i = 1, 2, \ldots, k$.
For all i and $\ell \ne i$, compute the estimators S_i^2 and $S_{i\ell}^2$. Set the observation counter $r = n_0$ and go to "Feasibility Check."

Feasibility Check: For $i \in M$, if $\sum_{j=1}^{r}(Y_{ij} - q) \le -R(r;\epsilon,h^2,S_i^2)$, move i from M to F. If $i \in SS_\ell$, eliminate ℓ from M or F and delete SS_ℓ. Else if $\sum_{j=1}^{r}(Y_{ij} - q) \ge +R(r;\epsilon,h^2,S_i^2)$, eliminate i from M and any existing SS_ℓ and delete SS_i.

Comparison: For each i and $\ell \in (M \cup F)$ such that $i \ne \ell$, $\ell \notin SS_i$, $i \notin SS_\ell$, and

$$\sum_{j=1}^{r} X_{ij} \le \sum_{j=1}^{r} X_{\ell j} - R\left(r;\delta,h^2,S_{i\ell}^2\right),$$

if $\ell \in F$, then eliminate i from M or F and delete SS_i; otherwise, if $\ell \notin F$, then add index ℓ to SS_i.

Stopping Rule: If $|M| = 0$ and $|F| = 1$, then stop and select the system whose index is in F as the best. If $|M| = 0$ and $|F| = 0$, then stop and report that there is no feasible system. Otherwise, take one additional observation $X_{i,r+1}$ and $Y_{i,r+1}$ from all $i \in M$ and $\ell \in F$ with $|SS_\ell| < |M|$. Set $r = r+1$ and go to "Feasibility Check."

REMARK 5: Note that there is always a unique solution to Eq. (5) because the left-hand side of Eq. (5) increases monotonically from zero to three with β when $0 < \beta < 1$.

We now prove that $\mathcal{AK}+$ identifies a good system with at least $1-\alpha$ confidence.

THEOREM 4: Under Assumptions 1 and 4, $\mathcal{AK}+$ guarantees $\Pr\{\text{CS}\} \ge 1-\alpha$.

PROOF: Algorithm $\mathcal{F}$ satisfies Assumption 5 because of the independence between the data obtained from different systems; see also the proof of Theorem 1. Similarly, it follows from Assumptions 1 and 4 and the proof of Theorem 2 of Kim and Nelson [18] that $\mathcal{KN}$ also satisfies Assumption 6 (since

the systems are simulated independently). Then, Lemma 2 implies that

$$\Pr\{\text{CS}\} \geq 1 - \left\{\beta + 2\left[1 - (1-\beta)^{(k-1)/2}\right]\right\} = 1 - \alpha,$$

where the last equation follows by how we chose β in "Setup." □

5.3. Sequentially Running vs. Simultaneously Running Procedures

It is a natural question to ask whether sequentially running or simultaneously running procedures work better. One way to compare them is through their continuation regions, see $R(r; v, w, z)$. We consider three sequentially running procedures, namely Restarting Procedure II, $\mathcal{AK}$, and Procedure $\mathcal{C}$. Notice that the continuation regions of the three sequentially running procedures and $\mathcal{AK}+$ differ only in the w part of $R(r; v, w, z)$: the sequentially running procedures take $h_1^2 = 2c\eta_1(n_0 - 1)$ and $h_2^2 = 2c\eta_2(n_0 - 1)$, whereas $\mathcal{AK}+$ takes $h^2 = 2c\eta(n_0 - 1)$ as w in the "Feasibility Check" and "Comparison" steps. As a decision is made when a system exits a continuation region, a smaller continuation region leads to faster decision making, and hence we prefer smaller η. For the sequentially running procedures, it is reasonable to take $\alpha_1 = \alpha_2 = \alpha/2$ (see Section 4.2), and thus the equations $g(\eta_1) = 1-(1-\alpha/2)^{1/k}$ and $g(\eta_2) = \alpha/(2(k-1))$ (or $\alpha/(2(m-1))$) determine the values of η_1 and η_2, respectively. In $\mathcal{AK}+$, the equation $g(\eta) = \beta$ determines η.

With the choice of $c = 1$ (see Section 3), $g(\eta)$ is a decreasing function of η. Consequently, a larger probability between $1-(1-\alpha/2)^{1/k}$ and β and between $\alpha/2(k-1)$ (or $\alpha/2(m-1)$) and β leads to a smaller continuation region for the "Feasibility Check" and "Comparison" steps, respectively. Figure 1 shows that β is always greater than or equal to the values of $1 - (1 - \alpha/2)^{1/k}$ and $\alpha/(2(k-1))$ when $\alpha = 0.05$ for $2 \leq k \leq 500$, and that the three probabilities are quite close for large k. For $\alpha = 0.01$ and 0.1, we also observed the same pattern, although we do not present the graphs here. This implies that $\mathcal{AK}+$ has a tighter screening for the "Feasibility Check" step than sequentially running procedures. The graph also suggests that $\mathcal{AK}+$ has a tighter screening for the "Comparison" step than Procedure $\mathcal{C}$, and thus that $\mathcal{AK}+$ will generally be more efficient than Procedure $\mathcal{C}$. Note, however, that there are circumstances when Procedure $\mathcal{C}$ would be expected to outperform $\mathcal{AK}+$ (e.g., if $k = 2$, $x_1 \gg x_2$, and $y_2 \ll y_1 \leq Q$, in which case the two algorithms would likely collect a similar number of observations for system 1, but Procedure $\mathcal{C}$ would probably collect fewer observations for system 2 than $\mathcal{AK}+$).

$\mathcal{AK}$ determines η_2 based on $\alpha/(2(m-1))$, where m is the number of survivors from Phase I. We first calculate m' such

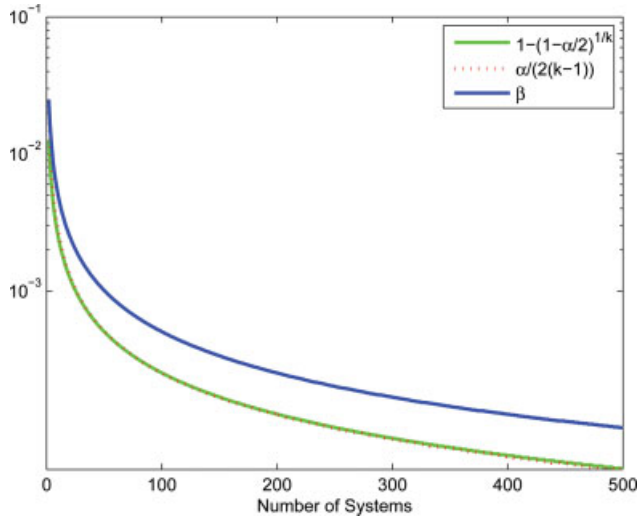

Figure 1. Comparison in a semilog scale among $1-(1-\alpha/2)^{1/k}$, $\frac{\alpha}{2(k-1)}$, and β when $\alpha = 0.05$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

that $\alpha/(2(m'-1)) = \beta$, which means that $\mathcal{AK}$ and $\mathcal{AK}+$ will have the same continuation region for Phase II (the continuation region for Restarting Procedure II will be slightly smaller). From Eq. (5) and $\alpha/(2(m'-1)) = \beta$, one can show that

$$m' = \frac{3}{2} + \frac{1}{\beta} - \frac{(1-\beta)^{\frac{k-1}{2}}}{\beta}$$

(note that the value of m' above need not be integer). If $m < m'$, then $\alpha/(2(m-1))$ is larger than β. Figure 2 shows the ratio of m' and k approaches 50% as k grows when $\alpha = 0.05$ [recall that β depends on k, see Eq. (5)]. Thus, if k is large and fewer than half of the systems survive Phase I, then Restarting Procedure II and $\mathcal{AK}$ provide tighter screening for the "Comparison" step than $\mathcal{AK}+$; the reverse is true when more than half the systems survive Phase I.

Although comparison through continuation regions is useful, it does not tell the entire story. In the next section, we compare the performance of sequentially running and simultaneously running procedures by experiments.

6. EXPERIMENTAL RESULTS

In this section, we illustrate and compare the performance of the new procedures. Sections 6.1 and 6.2 consider the case where (X_{ij}, Y_{ij}) are bivariate normal random variables with correlation ρ. More specifically, Section 6.1 sets up the mean and variance configurations for (X_{ij}, Y_{ij}) that we use, and Section 6.2 provides the experimental results. Finally,

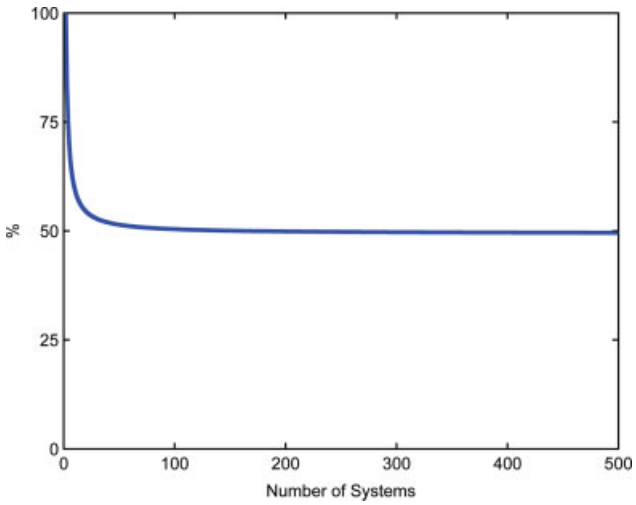

Figure 2. m'/k vs. k when $\alpha = 0.05$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

in Section 6.3, we present numerical results aimed at investigating the robustness of our approaches to the normality assumption.

6.1. Configurations for Normally Distributed Data

We can assume that the target value q for the secondary measure is $q = 0$ without loss of generality. Let b and a be the numbers of desirable and acceptable systems, respectively, and let system b be the best desirable system. We assume that the vector (X_{ij}, Y_{ij}) is taken from replication j of system i and that X_{ij} and Y_{ij} are correlated with correlation ρ. We consider two mean configurations, namely the difficult means (DM) configuration and the monotonically increasing means (MIM) configuration. In both configurations, we assume that all unacceptable systems have larger x_i than desirable systems. This is a difficult testing scenario, because if we fail to eliminate one or more unacceptable systems in Phase I, then an unacceptable system is likely to be chosen as the best because of its large x_i, and this will increase the probability of incorrect selection.

To test the validity of the proposed procedures, we use the DM configuration. In the DM configuration, all desirable systems have y_i exactly ϵ smaller than q, all acceptable systems have y_i equal to q, and all unacceptable systems have y_i exactly ϵ larger than q. Therefore, it is difficult to distinguish between feasible and infeasible systems. In addition, all desirable and acceptable systems have x_i very close to that of the true best desirable system, which makes it difficult to eliminate inferior systems. On the other hand, all unacceptable systems have much larger x_i than that of the true best desirable system. More specifically, the DM configuration has

$$y_i = \begin{cases} -\epsilon, & i = 1, 2, \ldots, b, \\ 0, & i = b+1, \ldots, b+a, \\ \epsilon, & i = b+a+1, \ldots, k, \end{cases}$$

and

$$x_i = \begin{cases} 0, & i = 1, 2, \ldots, b-1, \\ \delta, & i = b, \\ 0, & i = b+1, \ldots, b+a, \\ (i-1)\delta, & i = b+a+1, \ldots, k. \end{cases}$$

To investigate the effectiveness of the procedures in eliminating infeasible or noncompetitive systems, we use the MIM configuration. In the MIM configuration,

$$y_i = \begin{cases} -(b-i+1)\epsilon, & i = 1, 2, \ldots, b, \\ 0, & i = b+1, \ldots, b+a, \\ (i-b)\epsilon, & i = b+a+1, \ldots, k, \end{cases}$$

and

$$x_i = \begin{cases} (i-1)\delta, & i = 1, 2, \ldots, b, \\ (b-2)\delta, & i = b+1, \ldots, b+a, \\ (i-1)\delta, & i = b+a+1, \ldots, k. \end{cases}$$

Let $\sigma^2_{x_i}$ and $\sigma^2_{y_i}$ denote the variances of the primary and secondary measures of system i, respectively. To test the performance of the proposed procedures under various variance configurations, we consider five configurations: CONST $\sigma^2_{x_i}$ and CONST $\sigma^2_{y_i}$, INC $\sigma^2_{x_i}$ and INC $\sigma^2_{y_i}$, INC $\sigma^2_{x_i}$ and DEC $\sigma^2_{y_i}$, DEC $\sigma^2_{x_i}$ and INC $\sigma^2_{y_i}$, and DEC $\sigma^2_{x_i}$ and DEC $\sigma^2_{y_i}$, where CONST stands for constant, INC stands for increasing, and DEC stands for decreasing. In the CONST configuration, all variances are set to one. In the INC configuration, $\sigma^2_{x_i} = 1+(i-1)\delta$ and $\sigma^2_{y_i} = 1+(i-1)\epsilon$. In the DEC configuration, $\sigma^2_{x_i} = 1/(1+(i-1)\delta)$ and $\sigma^2_{y_i} = 1/(1+(i-1)\epsilon)$. Similar variance configurations are considered in other articles, including Kim and Nelson [18].

We test cases with $k \in \{2, 15, 25, 101\}$ systems and correlation $\rho \in \{-0.9, -0.6, 0, 0.3, 0.6, 0.9\}$. The initial sample size n_0 is set to $n_0 = 20$. We set $\delta = \epsilon = 1/\sqrt{n_0}$ to connect the parameters to the variability of the estimates of x_i and y_i from the first stage. Our required overall confidence level is $1 - \alpha = 0.95$. We make 10,000 macroreplications in all cases except for Table 1 where we make 20,000 macroreplications to achieve more precision. With these numbers of macroreplications, the first four digits of the estimated total numbers of replications (REP) in our tables are meaningful, and similarly the estimated probabilities of correct selection (PCS) are meaningful up to the 0.001th digit. Therefore, we

Table 1. The performance of $\mathcal{AK}$ and $\mathcal{AK}+$ for the DM configuration with acceptable systems.

		$b=13$ $a=0$	$b=12$ $a=1$	$b=10$ $a=3$	$b=7$ $a=6$	$b=3$ $a=10$	$b=1$ $a=12$
$\mathcal{AK}$	REP	4063	4109	4184	4319	4502	4604
	PCS	0.973	0.973	0.973	0.974	0.975	0.975
$\mathcal{AK}+$	REP	3763	3749	3726	3686	3615	3581
	PCS	0.960	0.962	0.963	0.963	0.966	0.968

do not report standard errors for REP and PCS to simplify the presentation.

6.2. Main Results for Normally Distributed Data

Our main results are as follows:

Acceptable Systems

Table 1 shows REP and PCS of $\mathcal{AK}$ and $\mathcal{AK}+$ for various numbers of acceptable systems a when $k = 25$, $a + b = 13$, $\rho = 0$, the means are in the DM configuration, and we have constant variances. Table 1 suggests that for given k, the estimated PCS of both procedures either increases or stays about the same when acceptable systems are included, when compared with the case without any acceptable systems. Numerical results omitted from the article to conserve space indicate that if acceptable systems have $x_i \geq \delta$, then the estimated PCS increases even more than the results in Table 1 indicate. These results are reasonable because it does not matter what feasibility decision is made on acceptable systems, which tends to increase the probability of CD in Phase I. Therefore, we do not consider any acceptable systems from now on (thus, $a = 0$) as the existence of acceptable systems does not appear to affect the statistical validity of our procedures.

Effects of Correlation

Table 2 shows the estimated REP and PCS for different values of the correlation ρ between observations of the primary and secondary performance measures when $k = 25$, $b = 13$, the means are in the DM configuration, and the variances are all increasing. It is clear that the estimated REP and PCS do not depend much on ρ. We observed the same tendency for all configurations we tested. This is expected because the derivations of $\mathcal{AK}$ and $\mathcal{AK}+$ are based on a Bonferroni-type inequality, which dissolves the effects of dependence (although $\mathcal{AK}$ has not yet been proven to be statistically valid). Because of this insensitivity of the procedures to ρ, we will only consider the $\rho = 0$ case from now on.

The Validity of $\mathcal{AK}$ and $\mathcal{AK}+$

Tables 3 and 4 show the estimated REP and PCS for $\mathcal{AK}$ and $\mathcal{AK}+$, respectively, under various variance configurations when the DM configuration is used with $b = 1+(k-1)/2, k \in \{5, 15, 25, 101\}$, and $\rho = 0$. The first row of the tables indicates the variance configuration of $\sigma^2_{x_i}, \sigma^2_{y_i}$. For example, "INC, DEC" means increasing $\sigma^2_{x_i}$ and decreasing $\sigma^2_{y_i}$. The estimated PCS values for both procedures are well over 0.95 for all configurations tested.

Experimental results comparing Procedures $\mathcal{C}$ and $\mathcal{AK}$ (which we omit because of space limitations) show that the estimated PCS of both procedures is over $1-\alpha$, but the estimated REP and PCS of $\mathcal{AK}$ are always slightly lower than those of Procedure $\mathcal{C}$. This is expected because Procedures $\mathcal{C}$ and $\mathcal{AK}$ have the same value of η_1, whereas the value of η_2 of $\mathcal{AK}$ is smaller than that of Procedure $\mathcal{C}$. Experimental results for Restarting Procedure II and $\mathcal{AK}$ (which we also omit because of space limitations) show that $\mathcal{AK}$ always provides a lower estimated REP, but the estimated PCS is higher in most

Table 2. Effects of correlation between primary and secondary performance measures for the DM configuration.

	$\rho=-0.9$		$\rho=-0.6$		$\rho=0$		$\rho=0.6$		$\rho=0.9$	
	REP	PCS	REP	PCS	REP	PCS	REP	PCS	REP	PCS
$\mathcal{AK}$	14101	0.976	14143	0.977	14110	0.974	14051	0.975	13982	0.974
$\mathcal{AK}+$	13308	0.964	13355	0.963	13149	0.962	13193	0.960	13225	0.962

Table 3. Performance of $\mathcal{AK}$ for the DM configuration.

$\sigma^2_{x_i}, \sigma^2_{y_i}$	CONST, CONST		INC, INC		INC, DEC		DEC, INC		DEC, DEC	
k	REP	PCS	REP	PCS	REP	PCS	REP	PCS	REP	PCS
5	576	0.969	802	0.967	588	0.965	669	0.972	431	0.972
15	2233	0.971	5433	0.976	2957	0.971	4120	0.983	1083	0.973
25	4063	0.973	14110	0.974	6819	0.978	10892	0.987	1527	0.975
101	20566	0.974	232786	0.979	88043	0.989	191525	0.989	3669	0.986

Table 4. Performance of $\mathcal{AK}+$ for the DM configuration.

$\sigma_{x_i}^2, \sigma_{y_i}^2$	CONST, CONST		INC, INC		INC, DEC		DEC, INC		DEC, DEC	
k	REP	PCS	REP	PCS	REP	PCS	REP	PCS	REP	PCS
5	556	0.960	785	0.957	594	0.961	640	0.956	415	0.959
15	2072	0.963	5108	0.959	2976	0.965	3672	0.958	991	0.962
25	3763	0.963	13149	0.962	6791	0.969	8980	0.960	1390	0.970
101	18737	0.961	211482	0.962	87802	0.986	134857	0.970	3357	0.986

cases than for Restarting Procedure II. Although Restarting Procedure II and $\mathcal{AK}$ use similar values of η_1 and η_2, Restarting Procedure II takes an additional n_0 observations for all surviving systems in Phase II and starts Phase II with less information about the primary performance measures x_i (as $\mathcal{AK}$ uses all the observations X_{ij} collected during Phase I), and this contributes toward lower estimated REP but higher estimated PCS for $\mathcal{AK}$ than Restarting Procedure II.

Finally, for $\mathcal{AK}$ whose validity is not proven, we tested several different configurations, including ones that are not described in Section 6.1, but did not find a single counterexample to the conjecture that $\mathcal{AK}$ is statistically valid. Note that if we choose a difficult mean configuration for y_i to increase the chance of error in Phase I, then the systems are likely to require a large number of replications in Phase I. Therefore, Phase II starts with a very precise estimate of x_i for each survivor, and the chance of making an error in Phase II is small. On the other hand, if the mean configuration of y_i is easy, then the chance of error in Phase I becomes small. This partly explains why the estimated PCS of $\mathcal{AK}$ is mostly over 0.975 and provides more evidence that $\mathcal{AK}$ is statistically valid. Therefore, we will focus on the comparison between $\mathcal{AK}$ and $\mathcal{AK}+$ from now on.

The Effectiveness of $\mathcal{AK}$ and $\mathcal{AK}+$

Tables 5 and 6 show the estimated REP and PCS for $\mathcal{AK}$ and $\mathcal{AK}+$, respectively, under various variance configurations when the MIM configuration is used with $b = 1 + (k-1)/2$, $k \in \{5, 15, 25, 101\}$, and $\rho = 0$. It is clear that the estimated PCS values are much larger than under the DM configuration. This is expected because it is much easier to detect noncompetitive systems when the differences between the true best and other systems are large. Tables 5 and 6 also show a dramatic reduction in the average total number of replications compared with Tables 3 and 4, respectively. For example, for $k = 101$, $b = 51$, and increasing $\sigma_{x_i}^2$ and $\sigma_{y_i}^2$, $\mathcal{AK}$ under the MIM configuration exhibits better than an 85% decrease in REP relative to the DM configuration.

Comparison of $\mathcal{AK}$ and $\mathcal{AK}+$

By comparing the results for $\mathcal{AK}$ and $\mathcal{AK}+$ in Tables 1–3 vs. 4, and 5 vs. 6, we notice that one procedure does not show a uniform superiority over the other procedure. However, under the DM configuration, $\mathcal{AK}+$ usually performs slightly better than $\mathcal{AK}$ in the sense that it spends a smaller

Table 5. Performance of $\mathcal{AK}$ for the MIM configuration.

$\sigma_{x_i}^2, \sigma_{y_i}^2$	CONST, CONST		INC, INC		INC, DEC		DEC, INC		DEC, DEC	
k	REP	PCS	REP	PCS	REP	PCS	REP	PCS	REP	PCS
5	459	0.981	652	0.981	491	0.979	521	0.981	337	0.983
15	1024	0.994	2615	0.994	1695	0.993	1763	0.997	496	0.995
25	1387	0.996	5015	0.996	3141	0.997	3420	0.998	613	0.997
101	3326	0.999	31484	0.999	18624	1.000	22497	1.000	2031	0.999

Table 6. Performance of $\mathcal{AK}+$ for the MIM configuration.

$\sigma_{x_i}^2, \sigma_{y_i}^2$	CONST, CONST		INC, INC		INC, DEC		DEC, INC		DEC, DEC	
k	REP	PCS	REP	PCS	REP	PCS	REP	PCS	REP	PCS
5	466	0.977	661	0.971	509	0.977	532	0.972	343	0.971
15	1070	0.991	2661	0.990	1711	0.993	1724	0.991	515	0.991
25	1488	0.995	5177	0.994	3166	0.995	3099	0.995	629	0.995
101	3590	0.998	34009	0.999	18662	1.000	20704	0.999	2032	1.000

Table 7. Effects of b when $k = 25$ under the DM configuration.

		$b = 1$		$b = 5$		$b = 13$		$b = 20$		$b = 25$	
Variance	Procedure	REP	PCS	REP	PCS	REP	PCS	REP	PCS	REP	PCS
CONST $\sigma_{x_i}^2$ / CONST $\sigma_{y_i}^2$	$\mathcal{AK}$	2965	0.979	3247	0.961	4063	0.973	4889	0.977	5527	0.984
	$\mathcal{AK}+$	(2567)	0.962	(3079)	0.957	(3763)	0.963	(4260)	0.960	(4586)	0.966
INC $\sigma_{x_i}^2$ / INC $\sigma_{y_i}^2$	$\mathcal{AK}$	10818	0.976	11310	0.966	14110	0.974	18678	0.982	23012	0.989
	$\mathcal{AK}+$	(9319)	0.960	(10400)	0.952	(13149)	0.962	(16334)	0.968	(19043)	0.974
DEC $\sigma_{x_i}^2$ / INC $\sigma_{y_i}^2$	$\mathcal{AK}$	10826	0.979	10875	0.971	6819	0.978	10908	0.992	10902	0.999
	$\mathcal{AK}+$	(9279)	0.959	(9526)	0.956	(6791)	0.969	(6985)	0.971	(4932)	0.981
INC $\sigma_{x_i}^2$ / DEC $\sigma_{y_i}^2$	$\mathcal{AK}$	1095	0.983	(1885)	0.971	10892	0.987	14547	0.984	21907	0.984
	$\mathcal{AK}+$	(982)	0.969	2199	0.968	(8980)	0.960	(12881)	0.975	(18361)	0.976
DEC $\sigma_{x_i}^2$ / DEC $\sigma_{y_i}^2$	$\mathcal{AK}$	1096	0.983	(1279)	0.971	1527	0.975	1671	0.981	1746	0.981
	$\mathcal{AK}+$	(972)	0.969	1289	0.966	(1390)	0.970	(1404)	0.965	(1399)	0.966

The numbers in parentheses indicate the smaller REPs for $\mathcal{AK}$ and $\mathcal{AK}+$ under each configuration.

number of replications while satisfying the PCS requirement. On the other hand, under the MIM configuration $\mathcal{AK}$, seems to be slightly better than $\mathcal{AK}+$.

To compare $\mathcal{AK}$ and $\mathcal{AK}+$ further, we tested cases where b varies, considering that the performance of Phase II of $\mathcal{AK}$ depends on the number of feasible systems b (due to the choice of η_2). Tables 7 and 8 show the performance of $\mathcal{AK}$ and $\mathcal{AK}+$ for various values of b when $k = 25$ under the DM and MIM configurations, respectively. For each variance configuration ($\sigma_{x_i}^2$ and $\sigma_{y_i}^2$), the performance of both $\mathcal{AK}$ and $\mathcal{AK}+$ is shown for $b \in \{1, 5, 13, 20, 25\}$. Tables 7 and 8 shows that although both methods satisfy the PCS requirement in all cases, $\mathcal{AK}+$ usually works better under the DM configuration, and $\mathcal{AK}$ mostly works better or is competitive under the MIM configuration. When $b = 1$ (so that the selection of the best feasible system is likely to be done by the "Feasibility Check" step), $\mathcal{AK}+$ works better due to smaller continuation region for the "Feasibility Check" step. Similarly, when $b = 25$, $\mathcal{AK}+$ usually works better because of $\mathcal{AK}$ performing similarly to Procedure $\mathcal{C}$ in this setting. Finally, when $\mathcal{AK}$ performs better, the savings in REP compared with $\mathcal{AK}+$ are generally small, whereas the savings associated with using $\mathcal{AK}+$ rather than $\mathcal{AK}$ can be significant, especially for large b.

The reason why $\mathcal{AK}+$ seems to outperform $\mathcal{AK}$ for the DM configuration may be that $\mathcal{AK}$ needs to wait a long time for Phase I to be completed under the DM configuration before applying Phase II. By contrast, $\mathcal{AK}+$ can eliminate a system when the system is found to be infeasible or inferior to other feasible systems, whichever comes first. Similarly, the reason

Table 8. Effects of b when $k = 25$ under the MIM configuration.

		$b = 1$		$b = 5$		$b = 13$		$b = 20$		$b = 25$	
Variance	Procedure	REP	PCS	REP	PCS	REP	PCS	REP	PCS	REP	PCS
CONST $\sigma_{x_i}^2$ / CONST $\sigma_{y_i}^2$	$\mathcal{AK}$	836	0.999	(1163)	0.993	(1387)	0.996	(1457)	0.998	(1288)	0.999
	$\mathcal{AK}+$	(785)	0.998	1303	0.995	1488	0.995	1495	0.995	1326	0.997
INC $\sigma_{x_i}^2$ / INC $\sigma_{y_i}^2$	$\mathcal{AK}$	1575	0.999	(2624)	0.994	(5015)	0.996	(6953)	0.997	(6895)	0.998
	$\mathcal{AK}+$	(1384)	0.998	2717	0.994	5177	0.994	7039	0.995	7086	0.996
DEC $\sigma_{x_i}^2$ / INC $\sigma_{y_i}^2$	$\mathcal{AK}$	1573	0.999	2269	0.996	(3141)	0.997	3994	0.998	2878	0.999
	$\mathcal{AK}+$	(1379)	0.999	(2160)	0.993	3166	0.995	(3402)	0.995	(2341)	0.997
INC $\sigma_{x_i}^2$ / DEC $\sigma_{y_i}^2$	$\mathcal{AK}$	722	0.998	(1285)	0.993	3420	0.998	5208	0.997	6868	0.998
	$\mathcal{AK}+$	(690)	0.998	1559	0.994	(3099)	0.995	(4734)	0.996	(5876)	0.998
DEC $\sigma_{x_i}^2$ / DEC $\sigma_{y_i}^2$	$\mathcal{AK}$	722	0.998	(779)	0.992	(613)	0.997	560	0.998	541	0.999
	$\mathcal{AK}+$	(688)	0.998	871	0.995	629	0.995	(557)	0.997	(533)	0.998

The numbers in parentheses indicate the smaller REPs for $\mathcal{AK}$ and $\mathcal{AK}+$ under each configuration.

why $\mathcal{AK}$ seems to outperform $\mathcal{AK}+$ in the MIM configuration may be that Phase I of $\mathcal{AK}$ does not take long in this case, and using $m = |F|$ to choose η_2 in Eq. (2) increases the efficiency of $\mathcal{AK}$ further. Thus, an analyst may want to decide on whether to use $\mathcal{AK}$ or $\mathcal{AK}+$ based on how difficult he/she expects Phase I (feasibility determination) to be relative to Phase II (comparison) for the problem at hand.

If one wishes to use a procedure with provable statistical validity, then $\mathcal{AK}+$ is recommended. If one is less concerned about provable statistical validity and more concerned about frequently switching between simulating different systems, then $\mathcal{AK}$ is recommended.

6.3. Robustness to Non-normality

One may be interested in incorporating a constraint on a probability, variance, or quantile. Constraints involving probabilities and variances can be expressed as expectation constraints of the form considered in this article. However, the basic observations X_{ij} and Y_{ij} are in general not normally distributed, and hence Assumption 1 is violated. Kim and Nelson [18] and Malone et al. [22] show that fully sequential procedures for the finding-the-best problem are robust to non-normality in IID data. In this section, we test the robustness of $\mathcal{AK}$ and $\mathcal{AK}+$ to non-normality, with and without batching, based on the extreme case of IID Bernoulli distributed X_{ij} and Y_{ij}.

As actual PCS tends to be greater in the MIM configuration than in the DM configuration, we consider the DM case only and adopt the following configuration:

$$y_i = \begin{cases} q - \epsilon, & i = 1, 2, \ldots, \dfrac{k+1}{2}, \\ q + \epsilon, & i = \dfrac{k+3}{2}, \ldots, k, \end{cases}$$

and

$$x_i = \begin{cases} q - \delta, & i = 1, 2, \ldots, \dfrac{k-1}{2}, \\ q, & i = \dfrac{k+1}{2}, \\ \min\{1, q + (i-1)\delta\}, & i = \dfrac{k+3}{2}, \ldots, k. \end{cases}$$

Table 9. PCS of $\mathcal{AK}$ and $\mathcal{AK}+$ for IID Bernoulli observations with $\epsilon = \delta = 0.05$ and no batching.

	$\mathcal{AK}$			$\mathcal{AK}+$		
k	$q = 0.5$	$q = 0.7$	$q = 0.9$	$q = 0.5$	$q = 0.7$	$q = 0.9$
5	0.979	0.974	0.922	0.971	0.967	0.903
15	0.984	0.981	0.923	0.979	0.973	0.909
25	0.989	0.979	0.924	0.984	0.973	0.907
101	0.990	0.980	0.915	0.987	0.972	0.895

Table 10. PCS of $\mathcal{AK}$ and $\mathcal{AK}+$ for IID Bernoulli observations with $q = 0.95$, $\epsilon = \delta = 0.01$, and batches of size b_s.

	$\mathcal{AK}$				$\mathcal{AK}+$			
k	$b_s = 1$	$b_s = 2$	$b_s = 5$	$b_s = 10$	$b_s = 1$	$b_s = 2$	$b_s = 5$	$b_s = 10$
5	0.691	0.875	0.946	0.955	0.698	0.856	0.928	0.951
15	0.680	0.865	0.949	0.964	0.687	0.852	0.931	0.952
25	0.679	0.870	0.946	0.962	0.692	0.847	0.928	0.955
101	0.672	0.857	0.931	0.965	0.695	0.833	0.904	0.956

This is similar to the DM configuration described in Section 6.1 with $a = 0$ and $b = (k+1)/2$, except we no longer let $q = 0$ (because the y_i are probabilities), and we have added $q - \delta$ to all the x_i with $i \leq b$ and have added q to all x_i with $i > b$ to the extent possible (again because the x_i are probabilities). The nominal confidence level is $1 - \alpha = 0.95$. We set $\epsilon = \delta = 0.05$ and $n_0 = 20$, choose $k \in \{5, 15, 25, 101\}$ and $q \in \{0.5, 0.7, 0.9\}$, and make 10,000 macroreplications.

Table 9 reports the observed PCS without batching. It is clear that the observed PCS decreases as q increases for both $\mathcal{AK}$ and $\mathcal{AK}+$. This is not surprising because estimating a probability close to 0 or 1 (so that either the event of interest or its complement is rare) is known to be difficult. The degradation in PCS is not substantial. Even for $q = 0.9$, the observed PCS is close to 92% for $\mathcal{AK}$ and 90% for $\mathcal{AK}+$.

To test an extreme case, we set $\epsilon = \delta = 0.01$ and $q = 0.95$, and consider batching with batch size $b_s \in \{1, 2, 5, 10\}$. Table 10 shows that the observed PCS can be as low as 67% for the extreme case where we use raw 0-1 observations ($b_s = 1$). However, if we take batch means with a small batch size as basic observations, the observed PCS increases significantly. With batch size 5, the observed PCS is already over 90% for both $\mathcal{AK}$ and $\mathcal{AK}+$, and with batch size 10, the observed PCS is in all cases over the nominal level 95%.

On the basis of numerical results in this section, we conclude that our procedures are robust to non-normality. For problems with severe non-normality and difficult feasibility determination and system comparison, batching with a small batch size helps greatly.

7. CONCLUSIONS

In this article, we formulate the constrained R&S problem of finding the simulated system with the largest expected primary performance measure in the presence of a stochastic constraint on the mean of a single real-valued secondary performance measure. We provide indifference-zone frameworks guaranteeing the construction of statistically valid sequentially and simultaneously running procedures. Note that sequentially running procedures are simpler to implement, whereas simultaneously running procedures introduce

a new paradigm of constructing combined R&S procedures. Example procedures of each type are presented along with an extensive numerical study. Consequently, our article demonstrates that R&S procedures can be extended to solve discrete optimization problems with constraints rigorously, though consistent with other indifference-zone approaches, they are often quite conservative. Interesting future research directions include extension of our comparison and feasibility detection procedures to handle multiple stochastic constraints. We are also interested in exploring whether variance updating as in Kim and Nelson [19] or Malone et al. [22] could further improve the efficiency of the proposed procedures.

APPENDIX: PROOF OF THEOREM 1

To prove the validity of Algorithm $\mathcal{F}$, we need the following results due to Fabian [11] and Jennison et al. [16]:

LEMMA 3 ([11]): Consider a standard Brownian motion process with drift $\mathcal{W}(t,\Delta)$ with $\Delta > 0$ and $t \geq 0$. Let $h(t) = a - \gamma t$ for some $a > 0$ and $\gamma \geq 0$. Let $H(t)$ denote the interval $(-h(t), h(t))$ (so that $H(t) = \emptyset$ when $-h(t) \geq h(t)$), and let $T = \min\{t : \mathcal{W}(t,\Delta) \notin H(t)\}$ be the first time $\mathcal{W}(t,\Delta)$ does not fall in the triangular continuation region defined by $(t, H(t))$. Finally, let $\mathcal{E}$ be the event $\{\mathcal{W}(T,\Delta) \leq -h(T)$ and $H(T) \neq \emptyset$, or $\mathcal{W}(T,\Delta) \leq 0$ and $H(T) = \emptyset\}$. If $\gamma = \Delta/(2c)$ for any positive integer c, then

$$\Pr\{\mathcal{E}\} = \sum_{\ell=1}^{c} (-1)^{\ell+1} \left(1 - \frac{1}{2}\mathcal{I}(\ell = c)\right) \exp\{-2a\gamma(2c-\ell)\ell\}.$$

LEMMA 4 ([16]): Suppose that a continuation region $H(t)$ is $(-h(t), h(t))$ given by a non-negative function $h(t), t \geq 0$. 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 $\{t_i : i = 1, 2, \ldots\}$ taking values in a given countable set. Let $T_C = \inf\{t > 0 : \mathcal{W}(t,\Delta) \notin H(t)\}$ and $T_D = \inf\{t_i : \mathcal{W}(t_i,\Delta) \notin H(t_i)\}$, and assume that $T_D < \infty$ almost surely. Note that $T_D \geq T_C$. The error probabilities are

$$\Pr\{\mathcal{E}_C\} \equiv \Pr\{\mathcal{W}(T_C,\Delta) \leq -h(T_C)\} = \Pr\{\mathcal{W}(T_C,\Delta) < 0\},$$
$$\Pr\{\mathcal{E}_D\} \equiv \Pr\{\mathcal{W}(T_D,\Delta) \leq -h(T_D)\} = \Pr\{\mathcal{W}(T_D,\Delta) < 0\}.$$

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

Lemma 3 gives the probability of an incorrect decision about the sign of the drift Δ for a continuous $\mathcal{W}(t,\Delta)$. When the observations are IID normally distributed with mean Δ and variance one, the distributions of the partial sums of the observations match that of $\mathcal{W}(t,\Delta)$ at each integer point. Lemma 4 states that under very general conditions, the probability of an incorrect decision does not increase when the Brownian motion process is observed at discrete times rather than continuously. Therefore, procedures designed for $\mathcal{W}(t,\Delta)$ provide an upper bound on the probability of an incorrect decision for a corresponding discrete process.

PROOF OF THEOREM 1: We begin by considering the case of only one system, i. Let $0 < \beta < 1$ and select η_1 such that $g(\eta_1) = \beta$. For a general output process $\mathbf{G}_i = \{G_{ij}, j = 1, 2, \ldots\}$, let

$$T_{\mathbf{G}_i} = \min\left\{r : r \geq n_0 \text{ and } -R(r;\cdot) < \sum_{j=1}^{r} G_{ij} < R(r;\cdot) \text{ is violated}\right\}. \tag{A1}$$

Therefore, $T_{\mathbf{G}_i}$ represents the stage at which $\sum_{j=1}^{r} G_{ij}$ exits the triangular region defined by $R(r;\cdot)$ for the first time after n_0. Let ICD_i denote the event that system i is assigned to an incorrect set when system i is considered in isolation. We consider three cases.

First consider the case where system i is in the desirable region. That is, $i \in S_D$ and, therefore, $y_i \leq Q^-$ or $y_i \leq q - \epsilon$. Then

$$\begin{aligned}
&\Pr\{\text{ICD}_i | y_i \leq q - \epsilon\} \\
&\quad = \Pr\{\text{system } i \text{ is declared to be infeasible}\} \\
&\quad = \Pr\left\{\sum_{j=1}^{T_{\mathbf{Y}_i}} (Y_{ij} - q) \geq +R(T_{\mathbf{Y}_i}; \epsilon, h_1^2, S_i^2)\right\} \\
&\quad = \Pr\left\{\sum_{j=1}^{T_{\mathbf{Y}_i}} (q - Y_{ij}) \leq \min\left\{0, -\frac{h_1^2 S_i^2}{2c\epsilon} + \frac{\epsilon T_{\mathbf{Y}_i}}{2c}\right\}\right\} \\
&\quad = \mathrm{E}\left[\Pr\left\{\sum_{j=1}^{T_{\mathbf{Y}_i}} (q - Y_{ij}) \leq \min\left\{0, -\frac{h_1^2 S_i^2}{2c\epsilon} + \frac{\epsilon T_{\mathbf{Y}_i}}{2c}\right\} \middle| S_i^2\right\}\right].
\end{aligned}$$

Now, we define $Z_{ij} \equiv (q - Y_{ij}) - (q - \epsilon - y_i)$, where $y_i = \mathrm{E}[Y_{ij}]$. Notice that $q - \epsilon - y_i \geq 0$ and hence $Z_{ij} \leq q - Y_{ij}$ due to the assumption that $i \in S_D$. This implies that $\sum_{j=1}^{r} Z_{ij}$ is more likely to exit a given continuation region through a lower boundary than $\sum_{j=1}^{r} (q - Y_{ij})$. Therefore,

$$\Pr\{\text{ICD}_i | y_i \leq q - \epsilon\} \leq \mathrm{E}\left[\Pr\left\{\sum_{j=1}^{T_{\mathbf{Z}_i}} \frac{Z_{ij}}{\sigma_i} \leq \min\left\{0, -\frac{h_1^2 S_i^2}{2c\epsilon\sigma_i} + \frac{\epsilon T_{\mathbf{Z}_i}}{2c\sigma_i}\right\} \middle| S_i^2\right\}\right], \tag{A2}$$

where $\sigma_i^2 = \mathrm{Var}(Y_{ij}) = \mathrm{Var}(Z_{ij})$. Notice that Z_{ij}/σ_i are IID $\mathrm{N}(\Delta, 1)$ with $\Delta = \epsilon/\sigma_i$. Let

$$a = \frac{h_1^2 S_i^2}{2c\epsilon\sigma_i} = \frac{\eta_1 (n_0 - 1) S_i^2}{\epsilon\sigma_i} > 0$$

and $\gamma = \epsilon/(2c\sigma_i) = \Delta/(2c)$. The sum of Z_{ij}, $j = 1, 2, \ldots, n_0$, is independent of S_i^2, the sample variance of Y_{ij}, $j = 1, \ldots, n_0$, and the observations we take after n_0 do not depend on S_i^2 as we assume that the Y_{ij} are IID; that is, the infinite sample path after n_0 does not depend on S_i^2. Also, notice that the distribution of $\sum_{j=1}^{r} Z_{ij}/\sigma_i$ is identical to that of $\mathcal{W}(t,\Delta)$ for $t = r \in \{n_0, n_0 + 1, \ldots\}$. Then, by Lemmas 3 and 4,

$$\begin{aligned}
&\Pr\{\text{ICD}_i | y_i \leq q - \epsilon\} \leq \mathrm{E}\left[\Pr\left\{\mathcal{W}(T_C,\Delta) < 0\right\} \middle| S_i^2\right] \\
&= \mathrm{E}\left[\sum_{\ell=1}^{c} (-1)^{\ell+1} \left(1 - \frac{1}{2}\mathcal{I}(\ell = c)\right) \times \exp\left\{-\frac{\eta_1 (2c - \ell)\ell}{c} \frac{(n_0 - 1) S_i^2}{\sigma_i^2}\right\}\right].
\end{aligned}$$

Now, consider the case that system i has $y_i \geq Q^+$ or $y_i \geq q + \epsilon$, so that $i \in S_U$. Then

$$\Pr\{\mathrm{ICD}_i | y_i \geq q+\epsilon\} = \Pr\{\text{system } i \text{ is declared to be feasible}\}$$
$$= \Pr\left\{\sum_{j=1}^{T_{\mathbf{Y}_i}} (Y_{ij}-q) \leq -R\left(T_{\mathbf{Y}_i}; \epsilon, h_1^2, S_i^2\right)\right\}$$
$$= \Pr\left\{\sum_{j=1}^{T_{\mathbf{Y}_i}} (Y_{ij}-q) \leq \min\left\{0, \frac{-h_1^2 S_i^2}{2c\epsilon} + \frac{\epsilon T_{\mathbf{Y}_i}}{2c}\right\}\right\}.$$

This time, we define $Z'_{ij} \equiv (Y_{ij}-q)-(y_i-q-\epsilon)$. Notice that $Z'_{ij} \leq Y_{ij}-q$. By applying similar arguments as those used for the case where $y_i \leq q-\epsilon$, we have

$$\Pr\{\mathrm{ICD}_i | y_i \geq q+\epsilon\} \leq \mathrm{E}\Bigg[\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1-\frac{1}{2}\mathcal{I}(\ell=c)\right) \times \exp\left\{-\frac{\eta_1(2c-\ell)\ell}{c}\frac{(n_0-1)S_i^2}{\sigma_i^2}\right\}\Bigg].$$

The conditional probabilities of ICD for both cases are bounded above by the same quantity. The third case occurs when system i is in an acceptable region, that is, $q-\epsilon < y_i < q+\epsilon$. In this case, it does not matter whether system i is declared feasible or infeasible, that is, $\Pr\{\mathrm{ICD}_i | q-\epsilon < y_i < q+\epsilon\} = 0$. Therefore, in all cases,

$$\Pr\{\mathrm{ICD}_i\} \leq \mathrm{E}\Bigg[\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1-\frac{1}{2}\mathcal{I}(\ell=c)\right) \times \exp\left\{-\frac{\eta_1(2c-\ell)\ell}{c}\frac{(n_0-1)S_i^2}{\sigma_i^2}\right\}\Bigg]. \quad \text{(A3)}$$

Observe that $(n_0-1)S_i^2/\sigma_i^2 \sim \chi^2_{n_0-1}$, a chi-squared distribution with n_0-1 degrees of freedom. To evaluate the expectation in (A3), recall that the moment generating function of a χ_ν^2 random variable satisfies $\mathrm{E}[\exp\{t\chi_\nu^2\}] = (1-2t)^{-\nu/2}$ for $t < 1/2$. Thus, the expected value of (A3) is

$$\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1-\frac{1}{2}\mathcal{I}(\ell=c)\right) \times \left(1+\frac{2\eta_1(2c-\ell)\ell}{c}\right)^{-(n_0-1)/2} = \beta,$$

where the equality follows from the way we choose η_1.

Thus, we have a bound on the probability of incorrect selection when there is only one system. Now, consider $k \geq 1$ systems, set $\beta = 1-(1-\alpha_1)^{1/k}$, and let CD_i be the event that system i is assigned to a correct set when system i is considered in isolation. Then using the facts that the systems are simulated independently and $0 < 1-\alpha_1 < 1$ implies $0 < \beta < 1$, we obtain

$$\Pr\{\mathrm{CD}\} = \Pr\left\{\bigcap_{i=1}^{k}\mathrm{CD}_i\right\} = \prod_{i=1}^{k}\Pr\{\mathrm{CD}_i\} = \prod_{i=1}^{k}(1-\Pr\{\mathrm{ICD}_i\})$$
$$\geq \prod_{i=1}^{k}\{1-(1-(1-\alpha_1)^{1/k})\} = 1-\alpha_1.$$

□

ACKNOWLEDGMENTS

The authors thank Christopher M. Healey, David Goldsman, and Barry L. Nelson for helpful comments and discussions. They also thank an anonymous Associate Editor and two anonymous referees for comments that have helped them significantly improve the article. This research was supported by the National Science Foundation under Grant Number CMMI-0400260. The first author was also supported by the National Science Foundation under Grant Numbers CMMI-0217860 and CMMI-0856600, and the second author was supported by the National Science Foundation under Grant Number CMMI-0644837.

REFERENCES

[1] S. Andradóttir, D. Goldsman, and S.-H. Kim, "Finding the best in the presence of a stochastic constraint," Proceedings of the 2005 Winter Simulation Conference, Institute of Electrical and Electronics Engineers, Piscataway, M.E. Kuhl, N.M. Steiger, F.B. Armstrong, and J.A. Joines (Editors), NJ, 2005, pp. 732–738.

[2] D. Batur and S.-H. Kim, Fully sequential selection procedures with parabolic boundary, IIE Trans 38 (2006), 749–764.

[3] R.E. Bechhofer, A single-sample multiple decision procedure for ranking means of normal populations with known variances, Ann Math Stat 25 (1954), 16–39.

[4] J. Boesel, B.L. Nelson, and S.-H. Kim, Using ranking and selection to "clean up" after simulation optimization, Oper Res 51 (2003), 814–825.

[5] J. Butler, D.J. Morrice, and P.W. Mullarkey, A multiple attribute utility theory approach to ranking and selection, Manage Sci 47 (2001), 800–816.

[6] C.H. Chen, A lower bound for the correct-selection probability and its application to discrete event simulations, IEEE Trans Automatic Control 41 (1996), 1227–1231.

[7] H. C. Chen, C.H. Chen, and E. Yücesan, Computing efforts allocation for ordinal optimization and discrete event simulation, IEEE Trans Automatic Control 45 (2000), 960–964.

[8] S.E Chick, "Selecting the best system: A decision-theoretic approach," Proceedings of the 1997 Winter Simulation Conference, Institute of Electrical and Electronics Engineers, Piscataway, S. Andradóttir, K.J. Healy, D.H. Withers, and B.L. Nelson (Editors), NJ, 1997, pp. 326–333.

[9] S.E. Chick and K. Inoue, New two-stage and sequential procedures for selecting the best simulated system, Oper Res 49 (2001), 732–743.

[10] S.E. Chick and K. Inoue, New procedures for identifying the best simulated system using common random numbers, Manage Sci 47 (2001), 1133–1149.

[11] V. Fabian, Note on Anderson's sequential procedures with triangular boundary, Ann Stat 2 (1974), 170–176.

[12] D. Goldsman, S.-H. Kim, W. Marshall, and B.L. Nelson, Ranking and selection procedures for steady-state simulation: Perspectives and procedures, INFORMS J Comput 14 (2002), 2–19.

[13] M. Hartmann, An improvement on Paulson's sequential ranking procedure, Sequent Anal 7 (1988), 363–372.

[14] M. Hartmann, An improvement on Paulson's procedure for selecting the population with the largest mean from k normal populations with a common unknown variance, Sequent Anal 10 (1991), 1–16.

[15] L.J. Hong and B.L. Nelson, The tradeoff between sampling and switching: New sequential procedures for indifference-zone selection, IIE Trans 37 (2005), 623–634.

[16] C. Jennison, I.M. Johnstone, and B.W. Turnbull, Asymptotically optimal procedures for sequential adaptive selection of the best of several normal means. Technical Report: No. 463 School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, 1980.

[17] S.-H. Kim, Comparison with a standard via fully sequential procedures, ACM Trans Model Comput Simul 15 (2005), 1–20.

[18] S.-H. Kim and B.L. Nelson, A fully sequential procedure for indifference-zone selection in simulation, ACM Trans Model Comput Simul 11 (2001), 251–273.

[19] S.-H. Kim and B.L. Nelson, On the asymptotic validity of fully sequential selection procedures for steady-state simulation, Oper Res 54 (2006), 475–488.

[20] S.-H. Kim and B.L. Nelson, "Selecting the best system," Handbooks in operations research and management science: Simulation, S.G. Henderson and B.L. Nelson (Editors), Elsevier Science, Oxford, UK, 2006, Chapter 17, pp. 501–534.

[21] A.M. Law and D. Kelton, Simulation modeling and analysis, 3rd edition, McGraw-Hill, New York, 2000.

[22] G.J. Malone, S.-H. Kim, D. Goldsman, and D. Batur, "Performance of variance updating procedures on various data," Proceedings of the 2005 Winter Simulation Conference, Institute of Electrical and Electronics Engineers, M.E. Kuhl, N.M. Steiger, F.B. Armstrong, and J.A. Joines (Editors), Piscataway, NJ, 2005, pp. 825–832.

[23] B.L. Nelson, J. Swann, D. Goldsman, and W. Song, Simple procedures for selecting the best system when the number of alternatives is large, Oper Res 49 (2001), 950–963.

[24] E. Paulson, A sequential procedure for selecting the population with the largest mean from k normal populations, Ann Math Stat 35 (1964), 174–180.

[25] J. Pichitlamken, B.L. Nelson, and L.J. Hong, A sequential procedure for neighborhood selection-of-the-best in optimization via simulation, Eur J Oper Res 173 (2006), 283–298.

[26] Y. Rinott, On two-stage selection procedures and related probability-inequalities, Commun Stat A7 (1978), 799–811.

[27] T.J. Santner and A.C. Tamhane, "Designing experiments for selecting a normal population with a large mean and a small variance," Design of Experiments – Ranking and Selection: Essays in Honor of Robert E. Bechhofer, T.J. Santer and A.C. Tamhane (Editors), Marcel-Dekker, New York, 1984, pp. 179–198.