A minimal switching procedure for constrained ranking and selection under independent or common random numbers

CHRISTOPHER M. HEALEY¹, SIGRÚN ANDRADÓTTIR² and SEONG-HEE KIM²,∗

¹Schneider Electric, North Billerica, MA 01862-2105, USA
²H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA
E-mail: skim@isye.gatech.edu

Received April 2013 and accepted December 2014

Constrained Ranking and Selection (R&S) aims to select the best system according to a primary performance measure, while also satisfying constraints on secondary performance measures. Several procedures have been proposed for constrained R&S, but these procedures seek to minimize the number of samples required to choose the best constrained system without taking into account the setup costs incurred when switching between systems. We introduce a new procedure that minimizes the number of such switches, while still making a valid selection of the best constrained system. Analytical and experimental results show that the procedure is valid for independent systems and efficient in terms of total cost (incorporating both switching and sampling costs). We also inspect the use of the Common Random Numbers (CRN) approach to improve the efficiency of our new procedure. When implementing CRN, we see a significant decrease in the samples needed to identify the best constrained system, but this is sometimes achieved at the expense of a valid Probability of Correct Selection (PCS) due to the comparison of systems with an unequal number of samples. We propose four variance estimate modifications and show that their use within our new procedure provides good PCS under CRN at the cost of some additional observations.

Keywords: Simulation, ranking and selection, switching, constraints, common random numbers, fully sequential algorithms

1. Introduction

Ranking and Selection (R&S) procedures are statistical tools for selecting the best system out of a finite number of simulated alternatives. Commonly, the best system is the one with the highest or lowest mean performance measure. Since outputs from each system are stochastic and possibly expensive, decision makers need to be concerned with both computational efficiency and validity, as expressed by the probability of correctly selecting the best system. Many approaches exist to address this general problem, either through the indifference-zone framework (e.g., Rinott (1978); Kim and Nelson (2006)) the Bayesian framework (e.g., Chick (2006) and Frazier (2012)), or the Optimal Computing Budget Allocation (OCBA) framework (e.g., Chen et al. (2000)).

We consider a more complicated form of R&S, namely, selecting the best system that satisfies constraints on one or more secondary performance measures. This problem is known as constrained R&S. To accomplish this, we adopt the framework of Andradóttir and Kim (2010), which involves a feasibility check phase to ensure that the chosen system meets the required constraints and a comparison phase to determine the best feasible system. Within this framework, we seek to ensure a desired Probability of Correct Selection (PCS) of the best feasible system.

The problem of constrained R&S has attracted some attention lately, including the development of the fully sequential indifference-zone procedures of Andradóttir and Kim (2010) and Healey et al. (2013, 2014), the OCBA methods of Pujowidianto et al. (2009) and Lee et al. (2012), the multiple attribute theory of Morrice and Butler (2006), an indifference-zone approach by Kabirian and Olafsson (2009), which consider the probability that several stochastic constraints are feasible, and asymptotic rules for sampling within a budget by Hunter and Pasupathy (2010, 2013), Butler et al. (2001), Lee et al. (2006), Chen and Lee (2009), Lee et al. (2010), and Teng et al. (2010) ad-
address multiple objective problems in general. In addition, some research has been dedicated solely to the feasibility check, as in Szechtman and Yücesan (2008) and Batur and Kim (2010).

The previously mentioned procedures for constrained R&S aim for efficiency in terms of observations required to find the best feasible system, but there are no procedures that we know of that explicitly address the cost of switching between systems in this context. Though it is common to compare procedures based on the required number of samples to achieve a nominal PCS, the possibly high cost (in both time and storage) of stopping and restarting complex simulations should also be considered. Hong and Nelson (2005) and Osogami (2009) present two fully sequential procedures that perform valid comparisons while limiting the number of switches (for a pure R&S problem).

As pointed out by Hong and Nelson (2005) and Osogami (2009), fully sequential R&S procedures, which need a lot of switching among systems such as the \( K N \) procedure of Kim and Nelson (2001), may become inefficient if the penalties for switching are large. More specifically, Hong and Nelson (2005) state that switching from one simulated system to another usually includes:

- Storing state information about the current system (including values of the random number seeds); saving all relevant output data; swapping the executable code for the current system out of, and the code for the next system into, active memory; and restoring the state information for the next system to the values it had on the last call.

They test the range of relative cost between switches and samples from one to 1000 and point out that even a 1000 times more expensive switching cost than sampling is not excessive. These costs would also be incurred by any constrained R&S procedure utilizing similar fully sequential algorithms for comparison. In this article, we present a new fully sequential indifference-zone procedure, named the Constrained Minimal Switching (\( CMS \)) procedure, designed to reduce the total (sampling and switching) cost when switching costs are significant, while identifying the best feasible system.

In recent years, parallel computing environments have become more readily available. Simulating systems in parallel could reduce switching costs; however, this is not always advisable as parallel computing involves its own complexities (including coordination among processors). Even when parallel computing is available, the number of processors for a typical individual PC or laptop is limited and thus one processor needs to run simulations of multiple systems when comparing hundreds or thousands of systems, requiring switching. Our research would facilitate minimizing the total cost of each of these processors. Research on an efficient way of distributing systems among processors would be useful, but is outside the scope of this article.

Minimal switching procedures reduce the cost of stopping and restarting simulations but often require extra samples to ensure that the number of switches in the comparison phase does not exceed the number of systems. We investigate the use of Common Random Numbers (CRN), a variance reduction technique, to reduce the number of required samples for \( CMS \). Healey et al. (2014) study the use of CRN in constrained R&S. They prove the validity of two procedures that always compare systems with equal sample sizes but express concerns about the validity of comparing systems with unequal sample sizes under CRN. We provide experimental results that show PCS can be significantly degraded under high correlation. As unequal sample sizes commonly occur within our new minimal switching procedure (and other procedures), we present four variance estimate modifications and show that their use within \( CMS \) under CRN captures savings in the required number of observations until a decision is made, while still providing a good PCS.

This article is organized as follows. Section 2 outlines the problem of constrained R&S, details notation, and sets assumptions for the validity of our procedure. Section 3 introduces the \( CMS \) procedure and includes a proof of its validity for independently simulated systems. In Section 4, we discuss the motivations for the use of CRN, discuss its effects within \( CMS \), and propose modifications to address its challenges. Section 5 features experimental results, followed by conclusions in Section 6. An earlier version of this article was presented at the Winter Simulation Conference (Healey et al., 2010).

2. Background

The goal of constrained R&S is the selection of the best system based on a primary performance measure out of a fixed number of alternatives, \( k \), with constraints on \( s \) secondary performance measures. In this section, we outline the problem, while also introducing notation and assumptions needed for our procedure and its validity proof.

Let \( (X_{in}, Y_{1in}, \ldots, Y_{rin}) \) be the \( n \)th observation of the \( r \)th system for the primary performance measure and \( s \) secondary performance measures. The set of all possible systems is \( S = \{1, 2, \ldots, k\} \). Let \( x_i = E[X_{in}] \) and \( y_{i\ell} = E[Y_{i\ell\ell}] \) be the mean values of the primary and secondary performance measures for each system \( i \in S \) and constraint \( \ell = 1, 2, \ldots, s \). Therefore, our objective is to determine which system has the best primary performance measure, while also satisfying all constraints:

\[
\arg \max_{i \in S} x_i, \\
\text{s.t. } y_{i\ell} \leq q_{\ell} \text{ for all } \ell = 1, 2, \ldots, s.
\]

We let \( \sigma_i^2 = \text{Var}[X_{in}] \) for all \( i \) and \( \sigma_{i\ell}^2 = \text{Var}[Y_{i\ell\ell}] \) for all \( i \) and \( \ell \). Moreover, the relationship between performance measures is governed by the following assumption.
Assumption 1. For each $i = 1, 2, \ldots, k$:

$$
\begin{bmatrix}
X_{in}
Y_{1n}
\vdots
Y_{kn}
\end{bmatrix}
\overset{i.i.d.}{\sim} MN \left( \begin{bmatrix}
x_i
y_{i1}
\vdots
y_{is}
\end{bmatrix}, \Sigma_i \right)
$$

for $n = 1, 2, \ldots$

where $\sim$ denotes independent and identically distributed, $MN$ denotes multivariate normal, and $\Sigma_i$ is the $(s+1) \times (s+1)$ covariance matrix of the vector $(X_{in}, Y_{1n}, \ldots, Y_{kn})$.

The normality of data is a common assumption within R&S, achieved through within-replication averages, including batched means in steady-state simulation (Law and Kelton, 2000). When the basic observations $X_{in}$ and $Y_{1n}$ are non-normal, observations from different replications may be batched, which would of course affect the sampling cost. Furthermore, data points can be correlated across systems (due to CRN) and across performance measures.

The procedure detailed in this article utilizes the indifference-zone method for both the feasibility check and comparison phases. For all systems involved in the simulation, we designate the indifference-zone parameter, $\delta > 0$, as the smallest significant difference among systems’ primary performance measures. Thus, we are indifferent between systems that have means within $\delta$ of each other, as long as the systems have not been found infeasible.

Likewise, we consider the tolerance level $\epsilon_\ell > 0$ to be the smallest significant difference between $y_{i\ell}$ and $q_\ell$. Therefore, we can place all systems into three sets in terms of feasibility. If system $i$ is in $S_D$, the set of desirable systems, then $y_{i\ell} \leq q_\ell - \epsilon_\ell$ for all $\ell = 1, 2, \ldots, s$. $S_U$ is the set of unacceptable systems where at least one secondary performance measure, $y_{i\ell}$, is infeasible, so that $y_{i\ell} > q_\ell + \epsilon_\ell$. All systems not in $S_D$ or $S_U$ fall into $S_A$, the set of acceptable systems.

Assumption 2. For all $i \in S_D \cup S_A \setminus \{[b]\}$, $x_{[b]} \geq x_i + \delta$, where $[b]$ is the index of the best desirable system.

Under Assumption 2, let $CS$ be the correct selection event that system $[b]$ is declared feasible and all systems in $S \setminus \{[b]\}$ are eliminated. If all systems are unacceptable, then $CS$ is the event that all systems in $S$ are eliminated. Finally, if $S_D = \emptyset$ and $S_A \neq \emptyset$, then $CS$ occurs both when all systems are eliminated and also when any system in $S_A$ is selected. We desire to ensure a nominal PCS of at least $1 - \alpha$.

We need some additional notation:

- $n_0$ = the first-stage sample size;
- $S^2_{Y_{[b]}}$ = the sample variance of $\{Y_{11}, Y_{12}, \ldots, Y_{in}\}$ and $\{X_{11}, X_{12}, \ldots, X_{in}\}$;
- $S^2_{X_{[b]}}$ = the sample variance of $\{X_{11}, X_{12}, \ldots, X_{in}\}$;
- $S^2_{Y_{i}}$ = the sample variance of $\{Y_{i1}, Y_{i2}, \ldots, Y_{in}\}$;
- $Y_{n} = (Y_{1n}, Y_{2n}, \ldots, Y_{kn})^T$, where $T$ denotes the transpose;
- $R(r; a, b, d) = \max\{0, \frac{bd}{2a} - \frac{r}{2}\}$, for $a, b, d \in \mathbb{R}^+$ and $a \neq 0$;
- $CS$ = the event that system $i$ is eliminated in pairwise comparison of systems $i$ and $[b]$, for any $i \in S$ with $x_{[b]} \geq x_i + \delta$;
- $CD_{i}$ = the event that correct decision is made on the feasibility of system $i \in S$ (when $i \in S_A$, any feasibility decision is $CD_{i}$);
- $\beta_1$ = the error of an individual feasibility check for one performance measure of one system;
- $\beta_2$ = the error of a pairwise comparison between two systems.

Remark 1. Note that $R(r; a, b, d)$ involves a decreasing linear function of the number of observations $r$, and originates from Fabian (1974). Our procedures require a so-called continuation region, defined as ($-R(r; a, b, d), R(r; a, b, d)$). Sampling continues when an observed statistic stays within the region and an elimination occurs when the statistic exits the region.

With this notation, we now present assumptions that govern good feasibility check and comparison phases. Assumptions 3 and 4 ensure that feasibility check and comparison are handled in a valid manner when systems are independently simulated. Assumptions 5 and 6 similarly ensure a valid feasibility check and comparison under CRN. Therefore, Assumptions 3 to 6 will not all be required simultaneously.

Assumption 3. The feasibility check phase guarantees

$$
\Pr(\bigcap_{i \in S} CD_{i}) \geq (1 - s\beta_1)^t
$$

for any $1 \leq t \leq k$ and any subset $S' \subseteq S$ with cardinality $t$ (i.e., $|S'| = t$) under $s$ constraints.

Assumption 4. The comparison phase guarantees

$$
\Pr(\bigcap_{i \in S} CS) \geq (1 - t\beta_2)^k
$$

for any $1 \leq t \leq k-1$ and any subset $S' \subseteq S$ with cardinality $t$ under $s$ constraints.

Assumption 5. The feasibility check phase guarantees

$$
\Pr(\bigcap_{i \in S} CD_{i}) \geq (1 - s\beta_1)^t
$$

for any $1 \leq t \leq k$ and any subset $S' \subseteq S$ with cardinality $t$ under $s$ constraints.

Assumption 6. The comparison phase guarantees

$$
\Pr(\bigcap_{i \in S} CS) \geq (1 - t\beta_2)^k
$$

for any $1 \leq t \leq k-1$ and any subset $S' \subseteq S$ with cardinality $t$.

Assumptions 3 to 6 can be verified for a certain class of procedures in which (i) the feasibility check of each constraint for each system only depends on parameters of the constraint of the system and (ii) a comparison is performed pairwise with parameters dependent only on the two systems involved. We refer to Hong and Nelson (2005),
Pichitlamken et al. (2006), Osogami (2009), and Batur and Kim (2010) for results that show how these assumptions may be verified for particular procedures when Assumption 1 is true.

For constrained R&S in the presence of multiple stochastic constraints, we consider three procedures as competitors to our new procedures: $H_{AK}$, $H_{AK}+$, and $MD_R$. Healey et al. (2014) present $H_{AK}$ as a sequentially running procedure and $H_{AK}+$ as a simultaneously running procedure. More specifically, $H_{AK}$ performs a complete feasibility check of all systems and then a comparison on the systems found feasible. This procedure is most efficient when the feasibility check is quick to finish. $H_{AK}+$ performs both a feasibility check and comparison on all systems remaining in contention after each stage of sampling. Healey et al. (2013) propose another simultaneously running procedure, $MD_R$. It is an improved version of $H_{AK}+$ where a system is allowed to go dormant, halting sampling, if it does not seem promising. These three procedures are compared to our proposed procedures in the experimental study.

3. The CMS procedure

In this section, we present a new method for constrained R&S, which we call the CMS procedure. It minimizes the cost of switching from one system to another, a cost that is often not factored into R&S studies but can comprise a large portion of the computation time. Our new constrained R&S approach will utilize existing fully sequential feasibility check and comparison procedures but requires additional steps to ensure minimal switching.

The outline of this section is as follows. We first describe our approach in general terms in Section 3.1 and provide results that are useful for proving the statistical validity of our proposed procedures. Then we present our CMS approach in detail and prove its validity in Section 3.2.

3.1. General approach

Our minimal switching procedure for constrained selection consists of two steps, finding an incumbent best feasible system and then eliminating systems by comparing each remaining, available system with the incumbent best until one system remains (while also performing feasibility checks on the available systems). It resembles a procedure by Hong and Nelson (2005), but our procedure is designed to solve constrained R&S problems. We now outline the general approach.

The procedure will visit each system at most once after the first stage. To achieve this limited number of switches, at least one system must receive a large number of samples, the maximum necessary to complete comparison against all other systems. Therefore, we expect this framework to be conservative in terms of number of observations but a good choice if switching costs are high. Since feasibility check of system $A$ and the comparison of systems $A$ and $B$ occur at the same step, this framework falls within the set of simultaneously running constrained R&S procedures. (A simultaneously running procedure implements both feasibility check and comparison procedures at the same time but eliminates a system only when it is declared infeasible or inferior to a system already declared to be feasible; see Andradóttir and Kim (2010).)

To prove the validity of the General Constrained Switching Approach, we note the following two lemmas. Lemma 1 allows us to present the main result of Section 3.2, a proof of the validity of our CMS procedure for independently simulated systems. Lemma 2 provides a method to ensure the validity of simultaneously running procedures under correlated systems.

Lemma 1. (Healey et al. (2014, Lemma 4.2)). When the systems are simulated independently and Assumptions 2, 3, and 4 hold, a simultaneously running procedure guarantees $\Pr(\text{CS}) \geq (1 - s\beta_1)^j + (1 - s\beta_1) + (1 - \beta)^{k-j-1} - 2$ when the number of unacceptable systems $j$ is less than $k$ and $\Pr(\text{CS}) \geq (1 - s\beta_1)^k$ when the number of unacceptable systems is equal to $k$. 

### General Constrained Switching Approach

**Setup and Initialization:** Set necessary parameters and take first-stage samples.

**Finding a Feasible System:** Sort systems based on primary performance measure sample means. Perform feasibility check on systems according to their sorted order to find the initial guess for the best feasible system ($B$). Once a feasible system is found, take additional samples (if necessary) for the system to be able to compare it against all remaining systems and go to Feasibility and Comparison of $A$ with $B$.

**Feasibility and Comparison of $A$ with $B$:** Find the next best available system ($A$). While sampling only from system $A$, perform both feasibility check and comparison against $B$. If $A$ is feasible and superior to $B$, then eliminate system $B$, take additional samples (if necessary) from $A$ to be able to compare it to all remaining systems, replace $B$ with $A$, and find the next $A$. If $A$ is either infeasible or inferior to $B$, eliminate $A$ and find the next $A$. This repeats until all systems (except for $B$) are eliminated.
setup: select the overall confidence level 1

3.2. Detailed procedure

In this section, we present and analyze our CMS procedure for solving constrained R&S problems. We choose to feature fully sequential procedures for the feasibility check and comparison phases in CMS. Fully sequential procedures have been shown to be efficient in many configurations, as the comparison and feasibility check can be re-evaluated after every stage of sampling, possibly with as little as one additional observation.

The feasibility check phase of CMS is performed by the \( \mathcal{F}_b \) procedure of Batvar and Kim (2010), a fully sequential and valid method for determining feasibility of multiple constrained performance measures. Whereas Hong and Nelson (2005) propose two fully sequential R&S procedures that minimize the number of switches, we choose the MSS procedure for implementation of comparison in our CMS, modified as described in their Remark 3. The MSS procedure first takes the maximum number of samples from a potential best system, \( N_B \) in Equation (1), and compares the potential best with only one alternative by obtaining one sample at a time from the current alternative system. When one of the two systems is eliminated, another system comes into comparison. The \( \mathcal{F}_b \) procedure identifies a set of feasible or near-feasible systems by utilizing a triangular-shaped continuation region defined by \( R_r(a, b, d) \). Note that their function \( R(a, b, d) \) for the continuation region has one more parameter \( c \), and our \( R(a, b, d) \) is a special case of it when \( c = 1 \).

CMS Procedure for Multiple Constraints

Setup: Select the overall confidence level \( 1/k \leq 1 - \alpha < 1 \) and first-stage sample size, \( n_0 \geq 2 \). Choose \( \delta, \epsilon, t \), and \( q \) for \( t = 1, 2, \ldots, s \). Let \( \epsilon_1 = (1/2)((2\beta_1)^{1/2} - \delta - 1) \) and \( \epsilon_2 = (1/2)((2\beta_2)^{1/2} - \delta - 1) \), where \( \beta_1 = \beta_2/s \) and \( \beta_2 \) is the unique solution to the equation \( \beta_2 + 2[1 - (1 - \beta_2)(k-1)^2] = \alpha \).

Initialization: Let \( h_1^2 = 2\epsilon_1(n_0 - 1) \) and \( h_2^2 = 2\epsilon_2(n_0 - 1) \). Obtain \( n_0 \) observations \( X_i \) and \( Y_i \) from each system \( i \in S \). For all \( i \) and \( t \), compute \( S_{ij}^2 \). Similarly, for all \( i \) and \( j \neq i \), compute \( S_{ij}^2 \). Let \( S_i = \emptyset \) be the set of systems inferior to system \( i \in S \) in terms of the primary performance measure. Let \( K_i = \emptyset \) be the set of constraints found to be feasible for system \( i \in S \) and let the set of competing systems include all systems, \( M = S \). Set the observation counters \( r_i = n_0 \) for all \( i \in S \).

Finding a Feasible System:

Initial Feasibility Check: For system \( i \in M \) and \( t = 1, 2, \ldots, s \), if \( \sum_{n=1}^{n_0} (Y_i - q_i) x \geq R(n_0; \epsilon_1, h_1^2, S_{ij}^2) \), declare \( i \) to be infeasible, removing \( i \) from \( M \).

Initial Screening for Comparison: Compare all systems \( i \neq j \) in \( M \) based on \( n_0 \) samples. If \( \sum_{n=1}^{n_0} (X_i - q_i) x \geq \sum_{n=1}^{n_0} (X_j - q_j) x \), then add \( j \) to \( S_i \).

Initial Sorting: Sort the systems in \( M \) based on the first-stage sample means \( \bar{X}_i = (1/n_0) \sum_{n=1}^{n_0} X_i \). Let \( B \) and \( A \) be the best system in \( M \) with the best and second-best first stage sample means. Let

\[
N_B = \max \limits_{j 
\in \mathcal{M}(\mathcal{S}_A \cup \{B\})} \left( \frac{h_j^2 S_j^2}{\delta^2} \right),
\]

where \( \lceil \cdot \rceil \) is the ceiling function and \( \max \delta = -\infty \). Then \( N_B \) is the maximum number of samples required for system \( B \) to complete comparison with all systems remaining in contention.

Feasibility Check for Best: For system \( B, t \neq K_B \), and \( r_{B} > n_0 \), if \( \sum_{n=1}^{n_0} (Y_{B} - q_{B}) x \geq R(r_{B}; \epsilon, h_1^2, S_{B}^2) \), declare \( B \) to be infeasible. Else if \( t \neq K_B \) and \( \sum_{n=1}^{n_0} (Y_{B} - q_{B}) x \leq -R(r_{B}; \epsilon, h_1^2, S_{B}^2) \), add \( t \) to \( K_B \). If \( |K_B| = s \), declare \( B \) to be feasible, remove all systems in \( S_B \) from \( M \), and update \( A, A \neq M \) and \( |M| > 1 \).

Stopping Rule 1: If \( B \) is feasible and \( |M| = 1 \), declare \( B \) as the best feasible system. If \( B \) is infeasible and \( |M| = 1 \), then no feasible systems exist. If \( B \) is feasible and \( |M| > 1 \), proceed to Feasibility and Comparison of \( A \) with \( B \). If \( B \) is infeasible and \( |M| > 1 \), then remove \( B \) from \( M \), set \( B = A \), compute \( N_B \) using Equation (1), let \( A \) be the best system in \( M \); \( |B| \) based on the first-stage sample means \( |M| > 0 \), and proceed to Feasibility Check for Best. Otherwise, take an additional sample from system \( B, X_{B,a+1}, Y_{B,a+1} \), and \( Y_{B,a+1} \), and update the observation counter \( r_{B} = r_{B} + 1 \). If \( r_{B} = N_{B} \), store \( S_{a+1} \). Go to Feasibility Check for Best.

Feasibility and Comparison of \( A \) with \( B \):

Sampling for Comparison: Find \( N_A \) using Equation (1) with the index \( A \) replacing the index \( B \). If \( r_{B} < N_{B} \), take an additional \( N_{B} - r_{B} \) observations from system \( B \) and set \( r_{B} = N_{B} \).

Comparison: If \( B \notin S_A \) and

\[
\frac{r_A - n_0}{N_B - n_0} \sum_{S_{a+1} \geq X_{B,n}} X_{B,n} + \sum_{S_{a+1} < X_{B,n}} X_{B,n} \geq \sum_{S_{a+1} > X_{B,n}} X_{B,n}
\]

then remove \( A \) from \( M \) and go to Stopping Rule 2. If \( B \notin S_A \):
Minimal switching procedure for constrained R&S

\[
\frac{r_A - n_0}{N_B - n_0} \sum_{n=n_0+1}^{n_B} X_{Bn} + \sum_{n=1}^{n_0} X_{Bn} \leq \sum_{n=1}^{r_A} X_{An}
\]

- \( R(r_A; \delta, h^2, S_{X,b}^2) \),

and \( A \) is feasible, then remove \( B \) from \( M \). If \( B \notin SI_A \), Equation (2) is true, and \( A \)'s feasibility is undetermined, add \( B \) to \( SI_A \).

Feasibility: If the feasibility of \( A \) is unknown, use the same procedure as Feasibility Check for Best, except use \( A \) as the index instead of \( B \) (and do not update \( A \)). If \( A \) is infeasible, eliminate system \( A \) from \( M \).

Stopping Rule 2: If \( |M| = 1 \), stop and declare the remaining system as the best. If \( B \notin M \), then set \( B = A \), \( N_B = N_A \), update \( A \), and go to Sampling for Comparison. If \( A \notin M \), update \( A \) and go to Sampling for Comparison. Otherwise, take an additional sample from system \( A \), \( X_{A,r+1} \) and \( Y_{A,r+1} \), and update the observation counter \( r_A = r_A + 1 \). If \( r_A = N_A \), store \( \sum_{n=n_0+1}^{N_A} X_{An} \). Go to Comparison.

Taking \( N_B \) samples for the current guess for the best feasible system allows the procedure to make statistically valid decisions, while minimizing the number of switches. Each system is sampled at most twice, once for first-stage sampling and sorting and once for feasibility check and comparison. The procedure utilizes only \( N_B \) samples for comparison, even if more samples are obtained in a long feasibility check. This is desirable because Healey et al. (2014) show that primary performance measure sample means may be biased at the completion of feasibility check if primary and secondary performance measures are correlated, so observations past \( N_B \) may be harmful.

We now present the main theorem for \( CMS \).

**Theorem 1.** When the systems are simulated independently and Assumptions 2, 3, and 4 hold, \( CMS \) guarantees \( \Pr(CS) \geq 1 - \alpha \) when \( 2(1 - \beta^2)k(k-1)/2 - \beta_2 - 1 = 1 - \alpha \).

The proof of Theorem 1 is given in the Online Supplement. Note that for fixed \( k \) and \( \alpha \), \( 2(1 - \beta^2)k(k-1)/2 - \beta_2 - 1 \) monotonically decreases from one to \(-2\) as \( \beta_2 \) increases from zero to one, guaranteeing a unique solution to the equation in Theorem 1.

4. CRNs and two-sample comparison

In our new \( CMS \) procedure, we require the current best system \( B \) to take \( N_B \) samples, the maximum samples necessary to make a decision against any remaining system. To reduce this large number of observations, we turn to a popular variance technique, namely, the CRN approach, which Nelson and Matejcik (1995), Kim and Nelson (2001), and Healey et al. (2014), among others, show can be used to improve the efficiency of both unconstrained and constrained R&S procedures.

Proper implementation of CRN can result in quicker decisions by inducing a positive correlation across systems, which can significantly reduce the value of \( S_{X,b}^2 \), the sample variance of the difference of paired samples from systems \( i \) and \( j \). In procedures that compare systems at balanced sample sizes, such as \( KN \) of Kim and Nelson (2001), only a simple parameter adjustment to \( \beta_2 \) is needed to make a valid selection under CRN. Unfortunately, we cannot make valid decisions under the CRN method for two-sample procedures that compare systems with unequal sample sizes.

We show how variance estimates in \( CMS \) under CRN negatively impact the validity of the procedure in Section 4.1 and propose four modifications to improve the estimates in Section 4.2.

4.1. Comparison with positive correlation

Two-sample procedures that estimate variance with \( S_{X,b}^2 \) can underestimate the variability of the screening process observed if data points across systems are correlated (due to the use of CRN). In our minimal switching procedure, screening is performed and a decision to continue sampling is based on a function of \( (r_A - n_0)/(N_B - n_0))\sum_{n=n_0+1}^{N_B} X_{Bn} + \sum_{n=1}^{n_0} X_{Bn}, \sum_{n=1}^{n_A} X_{An} \), and an estimate of the variance of the difference of these two sums. Note that sample variances \( S_{X,b}^2 \) are only computed after the initial \( n_0 \) observations from all systems.

If \( r_B = r_A = r \), then \( rS_{X,b,4}^2 \) is an estimator of

\[
\text{Var} \left[ \sum_{n=1}^{r} X_{Bn} - \sum_{n=1}^{r} X_{An} \right] = r \left( \sigma_{xb}^2 + \sigma_{xa}^2 - 2\rho_x\sqrt{\sigma_{xb}^2\sigma_{xa}^2} \right),
\]

where \( \rho_x \) is defined as the correlation between \( X_{Bn} \) and \( X_{An} \).

However, if we assume that \( r_A \ll r_B = N_B \), we create a situation commonly encountered in the \( CMS \) procedure. Here, the sum for system \( B \) is computed at the random time \( N_B \), a sample size many times larger than \( r_A \). Thus, as we increment \( r_A \):

\[
\text{Var} \left[ \frac{r_A - n_0}{N_B - n_0} \sum_{n=n_0+1}^{N_B} X_{Bn} + \sum_{n=1}^{n_0} X_{Bn} - \sum_{n=1}^{r_A} X_{An} \right] \\
\approx \text{Var} \left[ \sum_{n=1}^{n_0} (X_{Bn} - X_{An}) \right] + \text{Var} \left[ \sum_{n=n_0+1}^{r_A} X_{An} \right] \\
= n_0(\sigma_{xb}^2 + \sigma_{xa}^2 - 2\rho_x\sqrt{\sigma_{xb}^2\sigma_{xa}^2}) + (r_A - n_0)\sigma_{xa}^2 .
\]

When \( \rho_x \) is large, the quantity estimated by \( rAS_{X,b}^2 \) in Equation (3) could be smaller than the variability observed by the process in Equation (4); see also Levneski et al. (2007) for a similar argument. This underestimation of variability can
cause a premature decision when the best system appears to be inferior by chance, hurting PCS. The biased variance estimate creates a continuation region, \( R(\rho; \delta, h^2, S^2_{X_{\delta,4}}) \), that is too small to make a valid decision.

We present an empirical study where we compare two systems, separated by the distance of the indifference-zone, \( \delta \), with system 1 being the preferable choice. Table 1 shows the observed \( \text{Pr}(C|S_2) \) of a two-sample comparison under varying correlation, \( \rho_x \), after 10000 replications. PCS below 1 – \( \alpha \) = 0.95 marked in bold.

| \( r_1 \) | 0.0 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|---|
| 20 | 0.976 | 0.978 | 0.981 | 0.981 | 0.985 | 0.991 |
| 30 | 0.977 | 0.978 | 0.978 | 0.974 | 0.969 | 0.965 |
| 45 | 0.976 | 0.976 | 0.976 | 0.974 | 0.970 | 0.951 |
| 70 | 0.979 | 0.977 | 0.974 | 0.963 | 0.935 | 0.894 |
| 120 | 0.981 | 0.978 | 0.972 | 0.955 | 0.923 | 0.879 |
| 200 | 0.988 | 0.981 | 0.968 | 0.948 | 0.918 | 0.870 |
| 300 | 0.990 | 0.979 | 0.967 | 0.947 | 0.913 | 0.868 |
| 500 | 0.994 | 0.981 | 0.969 | 0.945 | 0.910 | 0.867 |

The first modification \( TS_1 \) describes a simple but conservative modification in the sense that \( S^2_X + S^2_{\bar{X}} \) bounds \( S^2_X \) in expectation. It restores the observed PCS but will perform similarly to the case when systems are simulated independently.

Our second modification \( TS_2 \) benefits from the reduction of variance CRN can provide but only when sample sizes are equal. This idea would be a natural choice to preserve PCS and utilize correlation but, unfortunately, when CMS is implemented, \( r_i \neq r_j \) for almost all samples. Therefore, the results obtained for the first two modifications, applied within CMS, will be virtually identical. However, for other procedures such as \( H_{AK} \), this may still be a desirable, simple modification.

The discussion in Section 4.1 suggests that the continuation region is corrupted when \( S^2_X < S^2_{\bar{X}} \) and \( r_i < r_j \). Instead of reverting to the conservative estimate of variability when sample sizes are not equal, the third modification \( TS_3 \) uses \( S^2_X \) as a bound on variability, when \( r_i < r_j \).

The last modification \( TS_4 \) is a more aggressive one. We note that when \( r_j \) is large and \( r_i \approx 0, S^2_{\bar{X}} \) would dominate the variability of the process in Equation (4). However, when \( r_i \) and \( r_j \) are close, we would see variance closer to \( S^2_X \). This suggests that when \( r_i < r_j \), the variance estimate \( S^2_{\bar{X}_{r_i}} \) should be close to \( S^2_{X} \) for small \( r_i \), and \( S^2_{\bar{X}_{r_i}} \) should approach \( S^2_X \) as \( r_i \) approaches \( r_j \).

Of all of the proposed modifications, it is reasonable to expect \( TS_4 \) to approximate the variability of Equation (4) the best, making it the most promising heuristic. Healey (2010) empirically confirms this point and thus we present the results only with \( TS_4 \) in Section 5.2.2.

Table 1. Observed \( \text{Pr}(C|S_2) \) with \( x_1 = \delta = 1/\sqrt{20}, x_2 = 0, \sigma^2_{X} = \sigma^2_{\bar{X}} = 1, r_2 = 20 \), and varying correlation, \( \rho_x \), after 10000 replications. PCS below 1 – \( \alpha \) = 0.95 marked in bold.
5. Experimental results

In this section, we evaluate the performance of our new CMS procedure compared to the performance of other R&S procedures capable of handling multiple constraints, namely, $H_{AK}$, $H_{AK+}$, and $MD_R$ of Healey et al. (2013, 2014), in terms of the number of switches, number of required observations, and observed PCS. In Section 5.1, we discuss the experimental setup. In Section 5.2, we provide an analysis of CMS, with and without the heuristic modifications to incorporate CRN.

5.1. Setup

The mean and variance configurations for our experiments attempt to provide analogous results and analysis to the experimental studies of previous related fully sequential indifference-zone R&S studies, namely, Kim and Nelson (2001), Hong and Nelson (2005), Pichitlamken et al. (2006), Andradottir and Kim (2010), and Healey et al. (2013, 2014). Our experiments will test the procedures with 10000 macro-replications. In our tables, the first three digits of the number of observations and up to the 0.001th digit of PCS are meaningful. For all tests, we set $n_0 = 20$, and $\delta = e_1 = 1/\sqrt{20}$ (the sample standard deviation of an average of $n_0 = 20$ independent samples with a variance of one) for all $\ell = 1, 2, \ldots, s$. We set a nominal PCS of $1 - \alpha = 0.95$. We set the number of acceptable system in $S_A$ to be zero, as Andradottir and Kim (2010) show that the existence of acceptable systems does not significantly affect the results.

We utilize the generalized means configurations of Healey et al. (2013). We consider only one infeasible constraint ($v = 1$) because determining feasibility is more difficult when there is only one infeasible constraint. The Difficult Means (DM) configuration attempts to test the validity of the procedures by assigning system means in a challenging setup. For our experiments, the DM configuration is

$$x_i = E[X_i] = \begin{cases} 0, & i = 1, 2, \ldots, b - 1, \\ \delta, & i = b, \\ (i - 1)\delta, & i = b + 1, b + 2, \ldots, k, \end{cases}$$

and

$$y_{i\ell} = E[Y_{i\ell}] = \begin{cases} -e_\ell, & i = 1, 2, \ldots, b, \\ e_\ell, & i = b + 1, b + 2, \ldots, k \text{ and } \ell = 1, \\ -e_\ell, & i = b + 1, b + 2, \ldots, k \text{ and } \ell = 2, 3, \ldots, s. \end{cases}$$

We set the constraint levels, $q_\ell$, to zero.

Similarly, we also consider the Monotonically Increasing Means (MIM) configuration of Healey et al. (2013) with $v = 1$, which will allow us to determine the feasibility of clearly infeasible or feasible systems and compare substantially distant systems. The MIM configuration tested is, $x_i = E[X_ij] = (i - 1)\delta$ for $i = 1, 2, \ldots, k$ and

$$y_{i\ell} = E[Y_{i\ell}] = \begin{cases} -(b - i + 1)\epsilon_\ell, & i = 1, 2, \ldots, b, \\ (i - b)\epsilon_\ell, & i = b + 1, b + 2, \ldots, k, \text{ and } \ell = 1, \\ -(i - b)\epsilon_\ell, & i = b + 1, b + 2, \ldots, k, \text{ and } \ell = 2, 3, \ldots, s, \end{cases}$$

where again we set $q_\ell = 0$.

For the experiments, we examine a combination of variance configurations to test the procedures under different difficulty of feasibility check and comparison. We consider a similar setup to Healey et al. (2014), as we include low (L) and high (H) variances for the primary and secondary performance measures, $\sigma_\alpha^2$ and $\sigma_\beta^2$, respectively. For simplicity, all secondary performance measures $\ell = 1, 2, \ldots, s$, are assigned identical variances. High variance results in either $\sigma_\alpha^2 = 10$ or $\sigma_\beta^2 = 10$ and low variance sets $\sigma_\alpha^2 = 1$ or $\sigma_\beta^2 = 1$. We consider in total four variance configurations: L/L, L/H, H/L, and H/H. For example, L/H implies that the primary performance measure has low variances and the secondary performance measures have high variances, and thus the difficulty of the feasibility check dominates.

We perform our tests with two system sizes, either 15 systems with eight feasible or 101 systems with 51 feasible, in addition to three constraints and a combination of various mean and variance configurations. Each infeasible system violates only one of the constraints. This setup challenges the PCS of the procedures, as shown by Andradottir and Kim (2010) and Healey et al. (2013). Half of the systems must be eliminated by comparison and half must be eliminated by feasibility check. The feasibility check is also difficult, as screening must catch the single violated constraint.

As in Section 4, we let $\rho_x$ be the correlation across systems’ primary performance measure samples. We will consider both independently simulated systems and systems with induced $\rho_x > 0$, modeling CRN.

Andradottir and Kim (2010) and Healey et al. (2014) present empirical results that show correlation across primary and secondary performance measures does not have a major impact on performance. Similarly, Batur and Kim (2013) show that correlation across secondary performance measures does not largely affect the performance of the feasibility check procedure $F_B$. We expect that similar conclusions would be found here. Finally, we assume that the secondary performance measures are not correlated across systems under the CRN approach. In practice, secondary performance measures will likely be correlated across systems, but this correlation is unlikely to have a major impact on performance, since the feasibility check is performed separately for individual systems. We confirm that this is indeed the case by performing experiments with correlation across primary and secondary performance measures.
Healey et al.

Table 2. Observed PCS for procedures with $k$ independent systems, $s = 3$ constraints, and $b$ feasible systems when $1 - \alpha = 0.95$

|                | L/L | L/H | H/L | H/H | L/L | L/H | H/L | H/H |
|----------------|-----|-----|-----|-----|-----|-----|-----|-----|
| DM($\sigma^2_x / \sigma^2_{yi}$) |     |     |     |     |     |     |     |     |
| $k = 15, b = 8$ |     |     |     |     |     |     |     |     |
| $\mathcal{H}AK$ | 0.985 | 0.998 | 0.977 | 0.985 | 0.997 | 0.998 | 0.995 | 0.996 |
| $\mathcal{H}AK^+$ | 0.978 | 0.980 | 0.977 | 0.977 | 0.994 | 0.993 | 0.993 | 0.994 |
| $MD_R$ | 0.978 | 0.983 | 0.977 | 0.977 | 0.994 | 0.993 | 0.993 | 0.994 |
| CMS | 0.995 | 0.995 | 0.995 | 0.994 | 0.996 | 0.996 | 0.996 | 0.996 |
| $k = 101, b = 51$ |     |     |     |     |     |     |     |     |
| $\mathcal{H}AK$ | 0.991 | 1.000 | 0.984 | 0.990 | 1.000 | 1.000 | 0.999 | 0.999 |
| $\mathcal{H}AK^+$ | 0.983 | 0.989 | 0.984 | 0.984 | 0.999 | 0.999 | 0.999 | 0.999 |
| $MD_R$ | 0.983 | 0.992 | 0.984 | 0.984 | 0.999 | 0.999 | 0.999 | 0.999 |
| CMS | 0.999 | 0.999 | 1.000 | 0.999 | 1.000 | 0.999 | 0.999 | 0.999 |

ranging from $-0.75$ to 0.75 with $\rho_x = 0, 0.25, 0.5, 0.9$. Hence, we present experimental results with (i) independent primary and secondary performance measure samples and (ii) independent secondary performance measure samples in this section.

5.2. Results

In our experimental results, we display the effectiveness of constrained R&S procedures, with respect to observed PCS, average number of required samples, and average number of switches. We define a switch to be the change of sampling from one system to another system. A two-stage procedure for $k$ systems (all feasible) requires at most $2k - 1$ switches, two sets of sampling for each system (one to gather first-stage samples and one to complete comparison). Fully sequential procedures register a switch after each stage of sampling for every system remaining in contention.

In Section 5.2.1, we consider independent systems. Section 5.2.2 discusses how the use of CRN affects the performance of $\mathcal{H}AK$ and CMS and provides an analysis of how our heuristic modifications can produce good PCS, even under high correlation. The performance of $MD_R$ (the best performer among the three considered competitors) is compared to that of CMS under correlated systems in Section 5.2.3.

5.2.1. Systems under independent sampling

To evaluate the performance of CMS under independent sampling of systems, we compare it to three procedures for constrained R&S, namely, the $\mathcal{H}AK$, $\mathcal{H}AK^+$, and $MD_R$ procedures.

We operate the four procedures under similar setups. For example, we choose $\alpha_1 = \alpha_2$ in $\mathcal{H}AK$ and $\beta_1/s = \beta_2$ in $\mathcal{H}AK^+, MD_R$, and CMS, so that error is allocated equally between feasibility check and comparison. Healey et al. (2014) experimentally show that this allocation is a robust choice. For consistency with CMS, all procedures are implemented with the feasibility check procedure $F_1^I B$, although there are other methods that could be utilized (see, e.g., Batur and Kim (2010)). We provide results for all combinations of the DM and MIM means and L/L, L/H, H/L, and H/H variance configurations. Tables 2, 3, and 4 display the observed PCS, average number of observations, and average number of switches, respectively.

The performance in Table 2 is expected to be better than the nominal 0.95. We observe this to be true in all cases.

Table 3. Average number of required samples for procedures with $k$ independent systems, $s = 3$ constraints, and $b$ feasible systems

|                | L/L | L/H | H/L | H/H | L/L | L/H | H/L | H/H |
|----------------|-----|-----|-----|-----|-----|-----|-----|-----|
| DM($\sigma^2_x / \sigma^2_{yi}$) |     |     |     |     |     |     |     |     |
| $k = 15, b = 8$ |     |     |     |     |     |     |     |     |
| $\mathcal{H}AK$ | 2679 | 23712 | 14673 | 26604 | 1165 | 9779 | 7163 | 11658 |
| $\mathcal{H}AK^+$ | 2408 | 17432 | 14737 | 23877 | 1275 | 8303 | 7221 | 12669 |
| $MD_R$ | 2233 | 10788 | 14736 | 22132 | 1071 | 5620 | 7217 | 10621 |
| CMS | 3490 | 12529 | 32129 | 39570 | 1822 | 6061 | 19432 | 22673 |
| $k = 101, b = 51$ |     |     |     |     |     |     |     |     |
| $\mathcal{H}AK$ | 24643 | 22492 | 124066 | 245786 | 3577 | 25557 | 17631 | 29480 |
| $\mathcal{H}AK^+$ | 21620 | 152764 | 122480 | 215130 | 4007 | 22051 | 17890 | 34480 |
| $MD_R$ | 20171 | 99528 | 122410 | 200696 | 3447 | 13869 | 17557 | 26532 |
| CMS | 30449 | 111633 | 283968 | 359540 | 4945 | 14483 | 43260 | 51661 |
Table 4. Average number of required switches for procedures with $k$ independent systems, $s = 3$ constraints, and $b$ feasible systems

|              | DM($\sigma_x^2/\sigma_{yi}^2$) | MIM($\sigma_x^2/\sigma_{yi}^2$) |
|--------------|---------------------------------|---------------------------------|
|              | L/L    | L/H    | H/L   | H/H   | L/L    | L/H    | H/L   | H/H   |
| $k = 15, b = 8$ |        |        |       |       |        |        |       |       |
| $\mathcal{H}_{AK}$ | 321    | 29     | 12321 | 2931  | 215    | 29     | 6212  | 1913  |
| $\mathcal{H}_{AK^+}$ | 2122   | 17146  | 14451 | 23591 | 989    | 8017   | 6935  | 12383 |
| $\mathcal{M}DR$     | 1741   | 2425   | 14442 | 19722 | 711    | 1571   | 6930  | 9522  |
| $CMS$             | 29     | 29     | 29    | 29    | 28     | 29     | 28    | 29    |
| $k = 101, b = 51$  |        |        |       |       |        |        |       |       |
| $\mathcal{H}_{AK}$ | 2292   | 201    | 101690| 21207 | 457    | 201    | 14514 | 4095  |
| $\mathcal{H}_{AK^+}$ | 19700  | 150844 | 120560| 213210| 2087   | 20131  | 15970 | 32560 |
| $\mathcal{M}DR$     | 16257  | 25007  | 119809| 177751| 1407   | 5562   | 15603 | 22834 |
| $CMS$             | 200    | 201    | 201   | 201   | 130    | 169    | 162   | 201   |

Moreover, $CMS$ commonly provides a higher PCS than the other procedures, which is a result of the extra samples needed to limit switches during the procedure’s comparison phase. The observed PCS is higher for $k = 101$ than when $k = 15$ (also with MIM as opposed to DM) for all four procedures.

The comparison phase of $CMS$ can make this procedure less attractive than the other procedures in terms of the number of required observations. Still, Table 3 shows this is not always the case. When only the feasibility check is difficult (L/H), $CMS$ can be relatively efficient, bettering the totals of all procedures except $M DR$. By deter-

Fig. 1. Average total cost for procedures under DM (top) and MIM (bottom) with the L/L and L/H variance configurations as a function of relative cost between switches and samples when there are $k = 101$ independent systems with $s = 3$ constraints and $b = 51$ feasible systems.
mining feasibility only for systems with the most attractive primary performance measures, our switching procedure CMS spends fewer observations on the feasibility check. Otherwise, CMS requires as little as 50% more observations than MD_R. Only under a hard comparison (H/L) does this extra percentage rise to about 160%. The results under \( k = 15 \) and \( k = 101 \) are similar, except the larger number of required observations for \( k = 101 \). Moreover, Table 3 indicates that more observations are required for DM than for MIM, as expected.

Table 4 shows why CMS is a competitive procedure when the cost of switches is counted. This cost is incurred when systems are simulated one at a time, not in parallel. For every configuration, CMS requires \( 2k - 1 \) or less switches when simulating \( k \) systems. The other simultaneously running procedures, HAK+ and MD_R, can require thousands of switches, as every stage of sampling consists of as little as one observation from each system in contention. HAK is a special exception. When feasibility check is difficult and no additional samples are needed to complete comparison (L/H), HAK can achieve as little as \( k - 1 \) switches. However, this performance is not seen in hard comparison configurations, where CMS clearly outperforms HAK. Not surprisingly, Table 4 also indicates that more switches are needed for a larger number of systems \( k \) and in the DM configuration.

Figure 1 provides the combined cost of sampling and switching for our systems under the L/L and L/H variance configurations, whereas Fig. A.1 in the Online Supplement provides the cost under the H/L and H/H variance configurations for \( k = 101 \) independent systems. Similarly, Figs. A.2 and A.3 in the Online Supplement provide the combined cost under various variance configurations for \( k = 15 \) systems. Tables A.1 and A.2 in the Online Supplement also present the combined cost of sampling and switching when the factors are one and 10, respectively, for \( k = 15 \) and 101 systems.

Hong and Nelson (2005) perform an analysis of total costs when switching costs are a factor of 1, 10, 100, and 1000 times larger than the sampling costs per observation. We feature experimental results for a factor from zero (no switching cost) to 10 (switching takes 10 times more time than sampling); a factor larger than 10 will yield results that are more favorable to CMS. Figure 1 shows that the relative efficiency of CMS improves compared with the other procedures when the cost of switching increases. In fact, even in the extreme case of factor = 1, our switching procedure is the best performer for the L/H variance configuration when feasibility check is hard. As switching costs become more expensive, the results substantially favor our switching procedure, as the average total cost linearly increases with a slope that is equal to the number of switches. Thus, under a large factor, say, 10, CMS is clearly the efficient choice for all mean and variance configurations, significantly improving on the other procedures in all cases, and featuring as little as a quarter of the combined sampling and switching costs in the best case (H/L). Even when HAK requires \( 2k - 1 \) switches, we still find CMS to be the best performer, as CMS requires fewer samples in these cases. As the switching costs are multiplied by an even larger factor (e.g., 100 or 1000; see Hong and Nelson (2005)), we expect to see an even wider advantage in using CMS with any number of systems.

5.2.2. Systems under CRN

In this section, we examine the performance of procedures under CRN that compare systems with uneven sample sizes, specifically HAK and CMS, with our new modified variance estimates. HAK+ and MD_R are not considered in this section because they are statistically valid under CRN.

### Table 5. Observed PCS for procedures with \( k = 15 \) systems, \( s = 3 \) constraints, and \( b = 8 \) feasible systems with induced correlation (\( \rho_c \)). PCS below \( 1 - \alpha = 0.95 \) marked in bold.

| Procedure | L/L | L/H | H/L | H/H | DM(\( \sigma^2_{\delta}/\sigma^2_{\nu+} \)) | L/L | L/H | H/L | H/H | MIM(\( \sigma^2_{\delta}/\sigma^2_{\nu+} \)) |
|-----------|-----|-----|-----|-----|-----------------|-----|-----|-----|-----|-----------------|
| HAK       | 0.985 | 0.967 | 0.996 | 0.982 | 0.997 | 0.997 | 0.995 | 0.996 |
| HAK(0)    | 0.985 | 0.967 | 0.996 | 0.982 | 0.997 | 0.997 | 0.995 | 0.996 |
| HAK(0.1)  | 0.989 | 0.968 | 0.996 | 0.982 | 0.997 | 0.997 | 0.995 | 0.996 |
| HAK(0.25) | 0.989 | 0.968 | 0.996 | 0.982 | 0.997 | 0.997 | 0.995 | 0.996 |
| HAK(0.5)  | 0.993 | 0.968 | 0.996 | 0.982 | 0.997 | 0.997 | 0.995 | 0.996 |
| HAK(0.75) | 0.995 | 0.968 | 0.996 | 0.982 | 0.997 | 0.997 | 0.995 | 0.996 |
| HAK(0.9)  | 0.997 | 0.968 | 0.996 | 0.982 | 0.997 | 0.997 | 0.995 | 0.996 |
| CMS       | 0.993 | 0.982 | 0.997 | 0.982 | 0.992 | 0.992 | 0.992 | 0.992 |
| CMS(0)    | 0.994 | 0.982 | 0.997 | 0.982 | 0.992 | 0.992 | 0.992 | 0.992 |
| CMS(0.1)  | 0.995 | 0.982 | 0.997 | 0.982 | 0.992 | 0.992 | 0.992 | 0.992 |
| CMS(0.25) | 0.995 | 0.982 | 0.997 | 0.982 | 0.992 | 0.992 | 0.992 | 0.992 |
| CMS(0.5)  | 0.995 | 0.982 | 0.997 | 0.982 | 0.992 | 0.992 | 0.992 | 0.992 |
| CMS(0.75) | 0.995 | 0.982 | 0.997 | 0.982 | 0.992 | 0.992 | 0.992 | 0.992 |
| CMS(0.9)  | 0.995 | 0.982 | 0.997 | 0.982 | 0.992 | 0.992 | 0.992 | 0.992 |
and thus do not need heuristic modifications on variance estimates.

Tables 5 and 6 display the observed PCS and the average number of required observations for $HAK$ and CMS when $k = 15$ for a similar system setup as in Section 5.2.1 but with differing levels of induced correlation. For the results when $k = 101$, see Tables A.3 and A.4 in the Online Supplement.

We compare $HAK$ and CMS applied to independent systems, with versions of $HAK$ and CMS modified for correlated systems with induced correlation $\rho_s \in \{0, 0.1, 0.25, 0.5, 0.75, 0.9\}$. We denote the procedures with these parameters as $HAK(\rho_s)$ and $CMS(\rho_s)$. In these procedures, we utilize adjusted parameters $\beta_2 = s\beta_1 = \alpha/k$. These parameters guarantee valid selection of the best feasible system under correlation in similar procedures (see Lemma 2) but not in $HAK$ and CMS, as we show in the following tables. The parameter adjustment means that $\rho_s = 0$ produces a slightly higher PCS and number of required observations than the independent case.

Similar to the results of the SSM comparison procedure in Table 1, we see that PCS suffers in configurations with correlation over 0.9 in $HAK$ and over 0.75 in CMS in most variance configurations. It is also noteworthy that PCS is much lower when $\sigma^2$ is high. This is because for a given $\rho_s$, the underestimation of variability becomes more pronounced when $\sigma^2$ is large, as one can see from Equations (3) and (4).

In terms of sampling, Table 6 shows that the use of CRN significantly reduces the number of observations needed. The new values of $\beta_1$ and $\beta_2$ used for correlated systems cause the procedures to perform slightly worse when applied to truly independent systems than procedures designed for independent systems. As correlation increases, the procedures exploiting CRN will require fewer observations. Even at modest levels of correlation, we can see significant improvement over the independent case. Savings due to CRN are restricted to the comparison phase, so (L/H) configurations feature only a small advantage for implementing CRN, whereas other variance configurations display larger savings. Not surprisingly, H/L configurations feature the largest savings. Difficult feasibility check configurations also require higher levels of correla-

Table 6. Average number of required samples for procedures with $k = 15$ systems, $s = 3$ constraints, and $b = 8$ feasible systems with induced correlation ($\rho_s$).

|                  | $DM(\sigma^2_n/\sigma^2_{\bar{x}_n})$ | $MIM(\sigma^2_n/\sigma^2_{\bar{x}_n})$ |
|------------------|--------------------------------------|--------------------------------------|
|                  | L/L        | L/H        | H/L    | H/H    | L/L        | L/H        | H/L    | H/H    |
| $HAK$            | 2679       | 23712      | 14673  | 26604  | 1165       | 9779       | 7163   | 11658  |
| $HAK(0.0)$       | 2686       | 23766      | 14677  | 26666  | 1168       | 9800       | 7164   | 11690  |
| $HAK(0.1)$       | 2598       | 23766      | 13327  | 25874  | 1120       | 9800       | 6499   | 11196  |
| $HAK(0.25)$      | 2500       | 23766      | 11223  | 24901  | 1060       | 9800       | 5470   | 10588  |
| $HAK(0.5)$       | 2412       | 23766      | 7820   | 23939  | 999        | 9800       | 3783   | 9968   |
| $HAK(0.75)$      | 2392       | 23766      | 4424   | 23760  | 983        | 9800       | 2090   | 9801   |
| $HAK(0.9)$       | 2391       | 23766      | 2637   | 23766  | 983        | 9800       | 1147   | 9800   |
| CMS              | 3490       | 12529      | 32129  | 39570  | 1822       | 6061       | 19432  | 22673  |
| CMS(0.0)         | 3496       | 12592      | 32236  | 39722  | 1841       | 6077       | 19485  | 22672  |
| CMS(0.1)         | 3194       | 12411      | 29203  | 36334  | 1683       | 5995       | 17458  | 20570  |
| CMS(0.25)        | 2745       | 12110      | 24138  | 31434  | 1438       | 5871       | 14216  | 17416  |
| CMS(0.5)         | 2046       | 11519      | 15931  | 23288  | 1071       | 5646       | 9415   | 12361  |
| CMS(0.75)        | 1408       | 10882      | 8195   | 15786  | 751        | 5399       | 4413   | 7733   |
| CMS(0.9)         | 1175       | 10328      | 3669   | 13730  | 666        | 5269       | 1885   | 6069   |

Table 7. Observed PCS for procedures with $k = 15$ systems, $s = 3$ constraints, and $b = 8$ feasible systems with induced correlation $\rho_s = 0.9$. PCS below $1 - \alpha = 0.95$ marked in bold.

|                  | $DM(\sigma^2_n/\sigma^2_{\bar{x}_n})$ | $MIM(\sigma^2_n/\sigma^2_{\bar{x}_n})$ |
|------------------|--------------------------------------|--------------------------------------|
|                  | L/L        | L/H        | H/L    | H/H    | L/L        | L/H        | H/L    | H/H    |
| $HAK$            | 0.985      | 0.998      | 0.977  | 0.985  | 0.997      | 0.998      | 0.995  | 0.996  |
| $HAK(0.9)$       | 0.997      | 0.998      | 0.933  | 0.996  | 0.994      | 0.998      | 0.975  | 0.994  |
| $HAK(0.9) + TS_l$| 0.998      | 0.998      | 0.962  | 0.997  | 0.998      | 0.998      | 0.991  | 0.997  |
| CMS              | 0.995      | 0.995      | 0.995  | 0.994  | 0.996      | 0.996      | 0.996  | 0.996  |
| CMS(0.9)         | 0.930      | 0.930      | 0.729  | 0.703  | 0.983      | 0.982      | 0.902  | 0.888  |
| CMS(0.9) + TS_l  | 0.987      | 0.991      | 0.968  | 0.979  | 0.995      | 0.995      | 0.989  | 0.992  |
Table 8. Average number of observations for procedures with $k = 15$ systems, $s = 3$ constraints, and $b = 8$ feasible systems with induced correlation $\rho_x = 0.9$.

|               | DM($\sigma_z^2 / \sigma_{y,s}^2$) | MIM($\sigma_z^2 / \sigma_{y,s}^2$) |
|---------------|----------------------------------|----------------------------------|
|               | L/L | L/H | H/L | H/H | L/L | L/H | H/L | H/H |
| $\mathcal{H}$ |      |      |      |      |      |      |      |      |
| $\mathcal{H}$K | 2679 | 23 712 | 14 673 | 26 604 | 1165 | 9779 | 7163 | 11 658 |
| $\mathcal{H}$K(0.9) | 2391 | 23 766 | 2637 | 23 766 | 983 | 9800 | 1147 | 9800 |
| $\mathcal{H}$K(0.9) + $TS_4$ | 2391 | 23 766 | 2676 | 23 785 | 984 | 9800 | 1281 | 9817 |
| $\mathcal{C}$M$S$ | 3490 | 12 529 | 32 129 | 39 570 | 1822 | 6061 | 19 432 | 22 673 |
| $\mathcal{C}$M$S$(0.9) | 1175 | 10 328 | 3669 | 13 730 | 666 | 5269 | 1885 | 6069 |
| $\mathcal{C}$M$S$(0.9) + $TS_4$ | 1215 | 10 073 | 4808 | 14 819 | 675 | 5291 | 2826 | 7402 |

tion to see improvements. Percentage savings are similar for $k = 15$ and $k = 101$ systems. Since comparison dominates sampling for $\mathcal{C}$M$S$, we observe larger percentage gains under CRN for $\mathcal{C}$M$S$ than for $\mathcal{H}$AK.

Even with large possible savings under CRN, the PCS degradation for $\mathcal{H}$AK and $\mathcal{C}$M$S$ may still cause concern. Tables 7 and 8 present the effectiveness of our heuristic variance modification $TS_4$ under $\rho_x = 0.9$ for $k = 15$ systems, with $\mathcal{H}$AK($\rho_x$) + $TS_4$ and $\mathcal{C}$M$S$($\rho_x$) + $TS_4$ denoting an implementation of $\mathcal{H}$AK and $\mathcal{C}$M$S$ with the variance modification $TS_4$ and induced correlation $\rho_x$. For the sake of brevity, we feature only configurations with $k = 15$ systems as most of the conclusions reached in Tables 5, 6, A.3, and A.4 were similar for $k = 15$ and $k = 101$.

Table 7 displays the observed PCS for our procedures with and without the heuristic modification $TS_4$, for $\rho_x = 0.9$. Although we do not report the results for the other three heuristic modifications, we observe that all four modifications display a marked improvement in PCS, raising observed values above 0.988 in all configurations. The

![Fig. 2. Average total cost for procedures under DM (top) and MIM (bottom) with the L/L and L/H variance configurations as a function of relative cost between switches and samples when there are $k = 15$ systems with $s = 3$ constraints, $b = 8$ feasible systems, and induced correlation $\rho_x = 0.25$.](image)
5.2.3. Comparison of $MD_R$ and $CMS + TS_4$ under CRN

In this section, we compare $MD_R$ and $CMS$ with $TS_4$ under the CRN approach. Although we do not report PCS due to page constraints, we observe that both procedures achieve PCS over the nominal level for all cases tested. Average numbers of required samples and switches are reported in Tables A.5 and A.6 of the Online Supplement for $k = 15$ systems with induced correlation $\rho_x$. For all $\rho_x$ tested, $MD_R$ always spends fewer average number of required samples but $CMS + TS_4$ requires a significantly smaller number of switches compared to $MD_R$.

Figure 2 shows combined costs of switching and sampling for $MD_R$ and $CMS + TS_4$ as a function of relative cost between switches and samples when there are $k = 15$ systems with induced correlation $\rho_x = 0.25$ under the L/L and L/H variance configurations. See Figure A.4 in the Online Supplement for combined costs under H/L and H/H variances with $\rho_x = 0.25$. Table A.7 in the Online Supplement reports the combined cost when switches are 10 times as costly as samples for various levels of $\rho_x$. $CMS + TS_4$ achieves smaller average total costs than $MD_R$ with negligible degradation in PCS when the cost of switching is relatively larger than that of sampling.

6. Conclusions

We present a procedure, $CMS$, for constrained R&S that minimizes the number of switches between simulated systems while finding the best constrained system. This is desirable, as the cost of switching can be expensive. We prove the validity of this procedure, guaranteeing a nominal probability of selecting the best feasible system for independently sampled systems.

To improve the efficiency of the procedure, we also wish to utilize CRN to reduce variance within comparisons. We show how strong postive correlation can adversely affect the PCS for procedures, such as $CMS$, that use two-sample comparison, because of the underestimation of the variance during the comparison. To achieve the nominal PCS while still increasing efficiency, we propose four variance modifications.

Our experiments show that $CMS$ is an efficient option, if the cost of switching is larger than the cost of sampling or the feasibility check phase is difficult, and that the savings in total costs brought by $CMS$ can be huge. If the cost of switching is not significant compared with the cost of sampling, $MD_R$ usually performs best under both independent and correlated systems. Ensuring a minimal number of switches requires extra observations, but CRN can reduce the number of necessary samples. Our experiments show that the heuristic variance modifications provide good PCS, and some of them also preserve most of the savings due to CRN.

Funding

This research was supported by the National Science Foundation under grants CMMI-0400260 and CMMI-0644837. The second author was also supported by the National Science Foundation under grant CMMI-0856600.

Supplemental Material

Supplemental data for this article can be accessed on the publisher’s website at www.tandfonline.com/uiie.

References

Andradóttir, S. and Kim, S.H. (2010) Fully sequential procedures for comparing constrained systems via simulation. Naval Research Logistics, 57(5), 403–421.

Batur, D. and Kim, S.H. (2010) Finding feasible systems in the presence of constraints on multiple performance measures. ACM Transactions on Modeling and Computer Simulation, 20(13), 1–26.

Butler, J.C., Morrice, D.J. and Mullarkey, P.W. (2001) A multiple attribute utility theory approach to ranking and selection. Management Science, 47(6), 800–816.

Chen, E.J. and Lee, L.H. (2009) A multi-objective selection procedure of determining a Pareto set. Computers & Operations Research, 36, 1872–1879.

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

Chick, S.E. (2006) Subjective probability and Bayesian methodology, in Handbooks in Operations Research & Management Science: Simulation, Henderson, S.G. and Nelson, B.L. (eds), North-Holland, Amsterdam, The Netherlands, pp. 225–257.

Fabian, V. (1974) Note on Anderson’s sequential procedures with triangular boundary. Annals of Statistics, 2, 170–176.

Frazier, P.I. (2012) Tutorial: optimization via simulation with Bayesian statistics and dynamic programming. in Proceedings of the 2012 Winter Simulation Conference, Larouque, C., Himmelspach, J., Pasupathy, R., Rose, O. and Uhrmacher, A.M. (eds), Institute of Electrical and Electronics Engineers, Piscataway, NJ, pp. 1–16.
Healey, C.M. (2010) Advances in ranking and selection: variance estimation and constraints, Ph.D. dissertation, Georgia Institute of Technology, Atlanta, GA.

Healey, C.M., Andradottir, S. and Kim, S.-H. (2010) A minimal switching procedure for constrained ranking and selection. In *Proceedings of the 2010 Winter Simulation Conference*, Johansson, B., Jain, S., Montoya-Torres, J., Hugan, J. and Yücesan, E. (eds), Institute of Electrical and Electronics Engineers, Piscataway, NJ, pp. 1145–1151.

Healey, C.M., Andradottir, S. and Kim, S.-H. (2013) Efficient comparison of constrained systems using dormancy. *European Journal of Operational Research, 224*(2), 340–352.

Healey, C.M., Andradottir, S. and Kim, S.-H. (2014) Selection procedures for simulations with multiple constraints under independent and correlated sampling. *ACM Transactions on Modeling and Computer Simulation, 24*(3), article no. 14.

Hong, L.J. and Nelson, B.L. (2005) The tradeoff between sampling and switching. *IEEE Transactions, 37*, 623–634.

Hunter, S.R. and Pasupathy, R. (2010) Large-deviation sampling laws for constrained simulation optimization on finite sets, in *Proceedings of the 2010 Winter Simulation Conference*, Johansson, B., Jain, S., Montoya-Torres, J., Hugan, J. and Yücesan, E. (eds), Institute of Electrical and Electronics Engineers, Piscataway, NJ, pp. 995–1002.

Hunter, S.R. and Pasupathy, R. (2013) Optimal sampling laws for constrained simulation optimization on finite sets. *INFORMS Journal on Computing, 25*(3), 527–542.

Kabirian, A. and Olafsson, S. (2009) Selection of the best with stochastic constraints, in *Proceedings of the 2009 Winter Simulation Conference*, Rossetti, M.D., Hill, R.R., Johansson, B., Dunkin, A. and Ingalls, R.G. (eds), Institute of Electrical and Electronics Engineers, Piscataway, NJ, pp. 584–589.

Kim, S.-H. and Nelson, B.L. (2001) A fully sequential procedure for indifference-zone selection in simulation. *ACM Transactions on Modeling and Computer Simulation, 11*, 251–273.

Kim, S.-H. and Nelson, B.L. (2006) Selecting the best system, in *Handbooks in Operations Research & Management Science: Simulation*, Henderson, S.G and Nelson, B.L. (eds), North-Holland, Amsterdam, The Netherlands, pp. 501–534.

Law, A. and Kelton, D. (2000) *Simulation Modeling and Analysis*, third edition, McGraw Hill, New York, NY.

Lee, L.H., Chew, E.P. and Teng, S.Y. (2006) Integration of statistical selection with search mechanism for solving multi-objective simulation-optimization problems, in *Proceedings of the 2006 Winter Simulation Conference*, Perrone, L.F., Wieland, F.P., Liu, J., Lawson, B.G., Nicol, D.M. and Fujimoto, R.M. (eds), Institute of Electrical and Electronics Engineers, Piscataway, NJ, pp. 294–303.

Lee, L.H., Chew, E.P., Teng, S.Y. and Goldsman, D. (2010) Finding the non-dominated Pareto set for multi-objective simulation models. *IEEE Transactions, 42*(9), 656–674.

Lee, L.H., Pujowidianto, N.A., Li, L.-W., Chen, C.H. and Yap, C.M. (2012) Approximate simulation budget allocation for selecting the best design in the presence of stochastic constraints. *IEEE Transactions on Automatic Control, 57*(11), 2940–2945.

Lesnevski, V., Nelson, B.L. and Staum, J. (2007) Simulation of coherent risk measures based on generalized scenarios. *Management Science, 53*(11), 1756–1769.

Morrice, D.J. and Butler, J.C. (2006) Ranking and selection with multiple ‘targets’, in *Proceedings of the 2006 Winter Simulation Conference*, Perrone, L.F., Wieland, F.P., Liu, J., Lawson, B.G., Nicol, D.M., and Fujimoto, R.M. (eds), Institute of Electrical and Electronics Engineers, Piscataway, NJ, pp. 222–230.

Nelson, B.L. and Matejeik, F.J. (1995) Using common random numbers for indifference-zone selection and multiple comparisons in simulation. *Management Science, 41*, 1935–1945.

Ogusami, T. (2009) Finding probably best systems quickly via simulations. *ACM Transactions on Modeling and Computer Simulation, 19*(3), article no. 12.

Pichilamken, J., Nelson, B.L. and Hong, L.J. (2006) A sequential procedure for neighborhood selection of the best in optimization via simulation. *European Journal of Operational Research, 173*, 283–298.

Pujowidianto, N.A., Lee, L.H., Chen, C.-H. and Yap, C.M. (2009) Optimal computing budget allocation for constrained optimization, in *Proceedings of the 2009 Winter Simulation Conference*, Rossetti, M.D., Hill, R.R., Johansson, B., Dunkin, A. and Ingalls, R.G. (eds), Institute of Electrical and Electronics Engineers, Piscataway, NJ, pp. 574–583.

Rinott, Y. (1978) On two-stage selection procedures and related probability-inequalities. *Communications in Statistics—Theory and Methods, A7*, 799–811.

Szechtman, R. and Yücesan, E. (2008) A new perspective on feasibility determination, in *Proceedings of the 2008 Winter Simulation Conference*, Mason, S.J., Hill, R.R., Mönch, L., Rose, O., Jefferson, T. and Fowler, J.W. (eds), Institute of Electrical and Electronics Engineers, Piscataway, NJ, pp. 273–280.

Teng, S., Lee, L.H. and Chew, E.P. (2010) Integration of indifference-zone with multi-objective computing budget allocation. *European Journal of Operational Research, 203*(2), 419–429.

**Biographies**

Christopher M. Healey is a senior engineer at Schneider Electric. He received his Ph.D. from the H. Milton Stewart School of Industrial and Systems Engineering at the Georgia Institute of Technology. His research interests lie in ranking and selection, simulation output analysis, and applications to data center energy and cooling efficiency. His e-mail address is chris.healey@schneider-electric.com.

Sigrún Andradóttir is a Professor in the H. Milton Stewart School of Industrial and Systems Engineering at the Georgia Institute of Technology. Prior to joining Georgia Tech, she was an Assistant Professor and later an Associate Professor in the Departments of Industrial Engineering, Mathematics, and Computer Sciences at the University of Wisconsin – Madison. She received her Ph.D. in Operations Research from Stanford University. Her research interests include simulation, applied probability, and stochastic optimization. Her e-mail address is sa@gatech.edu.

Seong-Hee Kim is a Professor in the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Institute of Technology. She received her Ph.D. in Industrial Engineering and Management Sciences from Northwestern University. Her research interest centers on ranking and selection procedures for stochastic simulation, optimization via simulation, statistical output analysis, quality control, and applications of simulation methods to environmental management. Her e-mail address is skim@isce.gatech.edu.