*Proceedings of the 2015 Winter Simulation Conference*
*L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.*

# COMPUTATIONAL IMPROVEMENTS IN BOOTSTRAP RANKING & SELECTION PROCEDURES VIA MULTIPLE COMPARISON WITH THE BEST

Soonhui Lee

School of Business Administration
UNIST
Ulsan, REPUBLIC OF KOREA

Barry L. Nelson

Dept. of Ind. Engr. & Mgmt. Sci.
Northwestern University
Evanston, IL 60208-3119, USA

## ABSTRACT

General-purpose ranking and selection (R&S) procedures using bootstrapping were investigated by Lee and Nelson in WSC '14; their work provides the seminal idea for this study. Here we present bootstrap R&S procedures that achieve significant computational savings by exploiting multiple comparison with the best inference. We establish the asymptotic probability of correct selection for the new procedures, and report some experiment results to illustrate small-sample performance, both in attained probability of correct selection and computational efficiency relative to the procedures in Lee and Nelson.

## 1 INTRODUCTION

In Lee and Nelson (2014a), we presented general-purpose ranking and selection (R&S) procedures based on bootstrapping. The strength of the procedures is that they are valid for many types of performance measures (e.g., means, probabilities, variances or quantiles) and types of data (discrete- or continuous-valued and almost arbitrary distributions). In fact, they can be applied to situations in which not all systems even have the same output distribution family. Therefore, these procedures need not be tailored to the specific performance measure of interest or an assumed distribution of the simulation output, unlike virtually every other R&S procedure; see, for instance, Bechhofer, Goldsman, and Santner (1995) and Kim and Nelson (2006). Thus, their strength is their generality—two procedures cover virtually all R&S problems—and they are most valuable when there is no reason to be comfortable with specific distributional assumptions on the simulation output.

However, the generality of the procedures of Lee and Nelson (2014a) comes at a price: They are based on simultaneous confidence intervals for *all pairs* of differences of the performance measures of interest, which is stronger inference than required for selecting the best, resulting in a larger sample size than is really needed. Thus, one of the remaining challenges is to reduce the computational overhead to implement bootstrap R&S.

Our goal in this study is to derive less-conservative procedures while retaining the advantages stated above. To do this we introduce a bootstrap-based approach that exploits a connection between multiple comparison with best (MCB) confidence intervals and R&S. MCB attacks the problem of selecting the best system by forming simultaneous confidence intervals for the difference between the performance of each system and the best of the others; $k$ confidence intervals rather than $k(k-1)/2$ for all-pairwise multiple comparisons (MCA). MCB has been adapted for R&S problems in computer simulation before, including Nelson and Matejcik (1995), who established a connection between indifference-zone selection and MCB, and Damerdji and Nakayama (1999), who proposed MCB procedures in the context of steady-state simulation. More general discussion about MCB procedures can be found in Hsu (1996) and Hochberg and Tamhane (1987).

978-1-4673-9743-8/15/$31.00 ©2015 IEEE 

We propose two new versions of the procedures that parallel those in Lee and Nelson (2014a): with or without common random numbers (CRN). Our procedures are based on the bootstrap approach, exploiting intense computation rather than distributional information. Under mild assumptions on the output data we prove the asymptotic validity of the procedures. We also present results obtained from some experiments to evaluate their small-sample behavior. The experiment results include comparisons to the previously proposed bootstrap R&S procedures in Lee and Nelson (2014a).

## 2 RANKING & SELECTION AND MCB

First, we introduce the key notation. Let $X_{ij}$ represent the $j$th observed output of system $i$, for $i = 1, 2, \ldots, k$, so that $\mathbf{X}_j = (X_{1j}, X_{2j}, \ldots, X_{kj})^T$ is a $k \times 1$ vector representing the $j$th observed output across all systems. Throughout the paper, we assume that $X_{i1}, X_{i2}, \ldots$ are independent and identically distributed (i.i.d.) with marginal distribution $F_i(x) = \Pr\{X_{ij} \leq x\}$. When we employ common random numbers (CRN), we can consider $\mathbf{X}_1, \mathbf{X}_2, \ldots$ as i.i.d. with common joint distribution function $F(\mathbf{x}) = \Pr\{X_{1j} \leqslant x_1, \ldots, X_{kj} \leqslant x_k\}$, $\mathbf{x} = (x_1, \ldots, x_k)^T \in \mathbb{R}^k$. Let $\Theta = (\theta_1, \theta_2, \ldots, \theta_k)^T$ be a vector whose $i$th element is a statistical property of the marginal distribution $F_i$, such as its mean, a quantile, or a probability. Throughout the paper, we also assume that larger $\theta_i$ is better and $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_{k-1} < \theta_k$ (unknown to us).

We are interested in finding the sample size that allows us to select the system with the largest value of $\theta_i$ with a specified probability of correct selection (PCS) by choosing the one with the largest empirical estimate $\widehat{\theta}_i$ of it. The procedures in Lee and Nelson (2014a) build simultaneous fixed-width $\delta$ confidence intervals for all $\theta_i - \theta_j$, $i \neq j$ with simultaneous coverage $1 - \alpha$, and guarantee

$$\Pr\{\text{select } k | \theta_k - \theta_{k-1} \geq \delta\} \geq 1 - \alpha$$

asymptotically where $\delta$ is the indifference-zone parameter. However, less is required to provide a PCS guarantee. The following lemma establishes sufficient conditions under which (constrained) MCB intervals can be formed, which has been proven in Hsu (1996).

**Lemma 1** If

$$\Pr\{\widehat{\theta}_i - \widehat{\theta}_k - (\theta_i - \theta_k) \leq \delta, \quad \forall i \neq k\} \geq 1 - \alpha \tag{1}$$

then with probability greater than or equal to $1 - \alpha$

$$\theta_i - \max_{j \neq i} \theta_j \in [-(\widehat{\theta}_i - \max_{j \neq i} \widehat{\theta}_j - \delta)^-, (\widehat{\theta}_i - \max_{j \neq i} \widehat{\theta}_j + \delta)^+] \tag{2}$$

for $i = 1, 2, \ldots, k$, where $-x^- = \min\{0, x\}$ and $x^+ = \max\{0, x\}$.

With this lemma, we now approach the problem more directly by building MCB confidence intervals. Let $(1), (2), \ldots, (k)$ be indices such that $\widehat{\theta}_{(1)} \leq \widehat{\theta}_{(2)} \leq \cdots \leq \widehat{\theta}_{(k)}$. Lemma 1 implies that if we find sample size $n$ such that (1) holds, and we select the system with the largest performance estimate $\widehat{\theta}_{(k)}$, then the MCB confidence lower bounds guarantee with probability $\geq 1 - \alpha$ that

$$\theta_{(k)} - \max_{j \neq (k)} \theta_j \geq -(\widehat{\theta}_{(k)} - \max_{j \neq (k)} \widehat{\theta}_j - \delta)^- \geq -\delta$$

where the last inequality follows from the definition of $(k)$. This establishes that the selected system is guaranteed to be the best system or within $\delta$ of the best, with probability $\geq 1 - \alpha$. Furthermore, if $\theta_k - \theta_{k-1} > \delta$, then the selected system is guaranteed to be the best system with probability $\geq 1 - \alpha$.

On the other hand, from the MCB upper bounds we can conclude that with probability $\geq 1 - \alpha$

$$\theta_i - \max_{j \neq i} \theta_j \leq (\widehat{\theta}_i - \max_{j \neq i} \widehat{\theta}_j + \delta)^+.$$

This means if the upper bound is zero, we can infer that system $i$ is no better than the best; therefore, since $\theta_k - \max_{j\neq k}\theta_j = \theta_k - \theta_{k-1} > 0$, system $k$ will not be identified as one of inferior systems with probability at least $1-\alpha$. Thus, procedures based on (1) allow us to select the best system, or a system within $\delta$ of the best system, with a guaranteed probability.

## 3 PROCEDURES

In this section we describe algorithms for performing bootstrap R&S with a different method of selecting the best system from Lee and Nelson (2014a). We guarantee the PCS more directly through (1), that is, we form fixed-width $\delta$ simultaneous CIs on the difference between each system and the best of the other systems, resulting $k$ CIs, whereas the algorithms presented Lee and Nelson (2014a) were based on fixed-width simultaneous CIs on all pairwise differences, resulting in $k(k-1)/2$ CIs. Typically, the more fixed-width CIs we require to be simultaneously correct, the more observations are required to guarantee it. We refer the reader to Swanepoel et al. (1983) and Lee and Nelson (2014a) for more details on how bootstrapping is performed to generate a single (or multiple) fixed-width CI(s) for R&S for $k \geq 2$ systems.

Let $\underline{\mathrm{X}}_{in} = \{X_{i1}, X_{i2}, \ldots, X_{in}\}$ be a sample of size $n$ from a system with output distribution $F_i$ having distributional property $\theta_i$, and $\widehat{F}_{in}$ the empirical distribution function (ecdf) based on $\underline{\mathrm{X}}_{in}$ for system $i = 1,2,\ldots,k$. Let $\widehat{\theta}(\underline{\mathrm{X}}_{in})$ be an estimate of $\theta_i$ based on $\underline{\mathrm{X}}_{in}$ for $i=1,2,\ldots,k$, and $\widehat{\theta}_{i(k)}(\underline{\mathbf{X}}_n) = \widehat{\theta}(\underline{\mathrm{X}}_{in}) - \widehat{\theta}(\underline{\mathrm{X}}_{(k)n})$ for all $i \neq (k)$. Let $\underline{\mathrm{X}}^*_{inb} = \{X^*_{i1b}, X^*_{i2b}, \ldots, X^*_{inb}\}$, $b = 1,2,\ldots,B$ be $B$ random samples of size $n$ from $\widehat{F}_{in}$. Let $\widehat{\theta}(\underline{\mathrm{X}}^*_{inb})$ be an estimate of $\widehat{\theta}(\underline{\mathrm{X}}_{in})$ based on $\underline{\mathrm{X}}^*_{inb}$ and $\widehat{\theta}_{i[k]}(\underline{\mathbf{X}}^*_{nb}) = \widehat{\theta}(\underline{\mathrm{X}}^*_{inb}) - \widehat{\theta}(\underline{\mathrm{X}}^*_{[k]nb})$ for all $i \neq [k]$ where $[k]$ is the index of the system with the largest sample statistic, $\widehat{\theta}(\underline{\mathrm{X}}_{(k)n})$. We need the additional notation because we use $(k)$ to indicate the sample best, but $\widehat{\theta}(\underline{\mathrm{X}}^*_{[k]nb})$ may not be the largest among the $k$ bootstrap estimates.

We want to build simultaneous fixed-width $\delta$ confidence intervals for the differences between each system and the best of the rest, $\theta_i - \max_{j\neq i}\theta_j$ for $i = 1,2,\ldots,k$ (i.e., MCB intervals) by finding $n$ such that

$$\Pr\left\{\widehat{\theta}_i - \widehat{\theta}_k - (\theta_i - \theta_k) \leq \delta,\ \forall i \neq k\right\} \geqslant 1-\alpha. \tag{3}$$

We call a procedure for selecting the system with the largest performance estimate based on (3) *R&S via MCB*.

In the bootstrap version of R&S via MCB, given $\underline{X}_{in}$ and $\underline{X}^*_{inb}$ for $i = 1,2,\ldots,k;\ b = 1,2,\ldots,B$, the estimated coverage probability in (3) using bootstrapping is given by

$$P^*_{nB} = \frac{1}{B}\sum_{b=1}^{B}\prod_{i\neq(k)}\mathbf{I}\left\{\widehat{\theta}_{i[k]}(\underline{\mathbf{X}}^*_{nb}) - \widehat{\theta}_{i(k)}(\underline{\mathbf{X}}_n) \leq \delta\right\}. \tag{4}$$

Then the value of $n$ will be the smallest one for which the estimated bootstrap coverage probability $P^*_{nB}$ in (4) is at least $1-\alpha$. Notice that in the bootstrap version of (3), simultaneous CIs for the differences between each system and the sample best, instead of the true best, are formed as the sample best $(k)$ is estimated but $\theta_1 \leq \cdots \leq \theta_{k-1} < \theta_k$ is unknown.

We now present algorithms for performing bootstrap R&S via MCB in two versions, with CRN and without CRN as in Lee and Nelson (2014a). We first describe the procedure without CRN.

**Bootstrap R&S procedure via MCB without CRN**

1. Specify $N = n_0$, set $1/k < 1-\alpha < 1$, $\delta > 0$, and $\Delta n \geq 1$.
2. Obtain $\underline{\mathrm{X}}_{iN} = \{X_{i1}, X_{i2}, \ldots, X_{iN}\}$ a sample of size $N$ from the distribution $F_i$ for $i = 1,2,\ldots,k$.
3. Compute $\widehat{\theta}_{i(k)}(\underline{\mathbf{X}}_N) = \widehat{\theta}(\underline{\mathrm{X}}_{iN}) - \widehat{\theta}(\underline{\mathrm{X}}_{(k)N})$ for all $i \neq (k)$ where $\theta_i$ is a distributional property of $F_i$, $\widehat{\theta}(\underline{\mathrm{X}}_{iN})$ is an estimate of $\theta_i$ based on $\underline{\mathrm{X}}_{iN}$, and $(k) = \arg\max_{i=1,\ldots,k}\widehat{\theta}(\underline{\mathrm{X}}_{iN})$. Form the ecdf $\widehat{F}_{iN}$ of $F_i$ for system $i = 1,2,\ldots,k$.

4. Obtain $B$ bootstrap samples of size $N$ from $\widehat{F}_{iN} : \underline{X}^*_{iN1}, \ldots, \underline{X}^*_{iNB}, i = 1,2,\ldots,k$.
5. Compute $\widehat{\theta}_{i[k]}(\underline{\mathbf{X}}^*_{Nb}) = \widehat{\theta}(\underline{X}^*_{iNb}) - \widehat{\theta}(\underline{X}^*_{[k]Nb}),\ b = 1,2,\ldots,B$ for all $i \neq (k)$.
6. Estimate the PCS as

$$P^*_{NB} = \frac{1}{B}\sum_{b=1}^{B}\prod_{i\neq(k)} \mathbf{I}\left\{\widehat{\theta}_{i[k]}(\underline{\mathbf{X}}^*_{Nb}) - \widehat{\theta}_{i(k)}(\underline{\mathbf{X}}_N) \leq \delta\right\}.$$

7. If $P^*_{NB} \geqslant 1-\alpha$, report $\arg\max_{i=1,\ldots,k}\widehat{\theta}(\underline{X}_{iN})$ as the best.
   Else
      Obtain $\underline{X}_{i\Delta n}$ a sample of size $\Delta n$ from the distribution $F_i$ for $i = 1,2,\ldots,k$.
      Set $\underline{X}_{iN} = \underline{X}_{iN} \cup \underline{X}_{i\Delta n}$ for $i = 1,2,\ldots,k$ and $N = N + \Delta n$.
      Go to Step 3.
   End If

We also present the procedure with CRN below. In the algorithm with CRN, a sample will be taken from each of the $k$ systems using CRN across systems to induce a joint distribution on $\{F_1, F_2, \ldots, F_k\}$, denoted by $F$. Below we list only the steps that change from the **Bootstrap R&S procedure via MCB without CRN**.

**Bootstrap R&S procedure via MCB with CRN**

2. Obtain a sample $\mathbf{X}_j = (X_{1j}, X_{2j}, \ldots, X_{kj})^T\ \ j = 1,2,\ldots,N$ from the joint distribution $F$.
3. Compute $\widehat{\theta}_{i(k)}(\underline{\mathbf{X}}_N) = \widehat{\theta}(\underline{X}_{iN}) - \widehat{\theta}(\underline{X}_{(k)N})$ for all $i \neq j$ where $\theta_i$ is a distributional property of $F_i$, and $\widehat{\theta}(\underline{X}_{iN})$ is an estimate of $\theta_i$ based on $\underline{X}_{iN}$. Form the ecdf $\widehat{F}_N$ based on $\underline{\mathbf{X}}_N = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N\}$ as

$$\widehat{F}_N(\mathbf{x}) = \frac{1}{N}\sum_{j=1}^{N}\mathbf{I}\{X_{1j} \leqslant x_1,\ X_{2j} \leqslant x_2,\ \ldots,\ X_{kj} \leqslant x_k\}.$$

4. Obtain $B$ bootstrap samples of size $N$ from $\widehat{F}_N : \{\mathbf{X}^*_{1b}, \mathbf{X}^*_{2b}, \ldots, \mathbf{X}^*_{Nb}\}$ for $b = 1,2,\ldots,B$, where $\mathbf{X}^*_{jb} = (X^*_{1jb}, X^*_{2jb}, \ldots, X^*_{kjb})^T$ for $j = 1,2,\ldots,N$.
7. If $P^*_{NB} \geqslant 1-\alpha$, report $\arg\max_{i=1,\ldots,k}\widehat{\theta}(\underline{X}_{iN})$ as the best.
   Else
      Obtain $\underline{\mathbf{X}}_{\Delta n} = \{\mathbf{X}_j,\ j = 1,2,\ldots,\Delta n\}$ a sample of size $\Delta n$ from the distribution $F$.
      Set $\underline{\mathbf{X}}_N = \underline{\mathbf{X}}_N \cup \underline{\mathbf{X}}_{\Delta n}$ and $N = N + \Delta n$.
      Go to Step 3.
   End If

The sample-size increment on each iteration, $\Delta n$, will be generated adaptively as in Lee and Nelson (2014b); $\Delta n$ is large when the bootstrap coverage probability is far from the desired PCS, and $\Delta n$ is small when the bootstrap coverage probability is close to the desired PCS. Here we briefly describe the jump-ahead algorithm in Lee and Nelson (2014b) adjusted to the MCB procedure.

Given an observed $(N, P^*_{NB})$ pair, we fit a simplified normal-theory approximation for PCS as a function of $N$. Specifically, we compute the target sample size $\widehat{N}^*$ to obtain the desired PCS value ($\geq 1-\alpha$) as

$$\widehat{N}^* \geq N \times \left(\frac{\Omega^{-1}_{\mathbf{AIA}^\top}(1-\alpha)}{\Omega^{-1}_{\mathbf{AIA}^\top}(P^*_{NB})}\right)^2$$

and set $\Delta n = c(\widehat{N}^* - N)$, where $0 < c < 1$ and $\Omega$ and $\mathbf{A}$ are described in Definition 1, and (7), respectively, in the next section, and $\mathbf{I}$ is the $k \times k$ identity matrix. We do this provided $\Omega^{-1}_{\mathbf{AIA}^\top}(P^*_{NB}) > \varepsilon$ for some $\varepsilon > 0$;

otherwise, $\Delta n = 2N$. That is, we double the sample size if the bootstrap coverage probability $P^*_{NB}$ is very close to zero since fitting a simplified normal-theory approximation for PCS as a function of $N$ will not be applicable for such cases.

## 4 ASYMPTOTIC VALIDITY

The asymptotic results presented in this section support R&S via MCB using bootstrapping for $k \geq 2$ systems when the performance measures are means or quantiles. Proofs can be found in Lee and Nelson (2015). Our key notation is reviewed before stating the results.

Let $\underline{\mathbf{X}}_n = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\}$ be a random sample of size $n$ from distribution $F$ (in $\mathbb{R}^k$) with a $k \times 1$ vector of marginal distribution properties $\Theta$, where $\mathbf{X}_j = (X_{1j}, X_{2j}, \ldots, X_{kj})^T,\ j = 1, 2, \ldots, n$. Further, let $\underline{\mathbf{X}}^*_n = \{\mathbf{X}^*_1, \mathbf{X}^*_2, \ldots, \mathbf{X}^*_n\}$ denote a random sample of size $n$ from $\widehat{F}_n$ where $\widehat{F}_n$ is the ecdf based on $\underline{\mathbf{X}}_n$ defined in one of two different ways as described in Step 3 of the algorithms with CRN and without CRN in Section 3.

Consider $\Theta = \mathrm{E}[\mathbf{X}]$ for the mean case. Then $\widehat{\Theta}(\underline{\mathbf{X}}_n)$ and $\widehat{\Theta}(\underline{\mathbf{X}}^*_n)$ are the sample mean vectors based on $\underline{\mathbf{X}}_n$ and $\underline{\mathbf{X}}^*_n$, respectively. That is $\widehat{\Theta}(\underline{\mathbf{X}}_n) = \bar{\mathbf{X}}_n = \sum_{j=1}^n \mathbf{X}_j/n$ and $\widehat{\Theta}(\underline{\mathbf{X}}^*_n) = \bar{\mathbf{X}}^*_n = \sum_{j=1}^n \mathbf{X}^*_j/n$.

For the quantile case, consider $\Theta$ being a set of specific quantiles of the $k$ marginal distributions where the $i$th element is defined as

$$\theta_i = F_i^{-1}(q) = \inf\{x : F_i(x) \geqslant q\}, \qquad 0 < q < 1 \qquad i = 1, 2, \ldots, k.$$

Then $\widehat{\Theta}(\underline{\mathbf{X}}_n)$ and $\widehat{\Theta}(\underline{\mathbf{X}}^*_n)$ are the sample $q$th quantiles based on $\underline{\mathbf{X}}_n$ and $\underline{\mathbf{X}}^*_n$, respectively, where the $i$th element of $\widehat{\Theta}(\underline{\mathbf{X}}_n)$ is the sample $q$th quantile of $X_{i1}, X_{i2}, \ldots, X_{in}$ and the $i$th element of $\widehat{\Theta}(\underline{\mathbf{X}}^*_n)$ is the sample $q$th quantile of $X^*_{i1}, X^*_{i2}, \ldots, X^*_{in}$. Let Pr and Pr$^*$ denote probabilities under $F$ and $\widehat{F}_n$, respectively.

The bootstrap stopping time used in our procedure is given by

$$N^* = \inf\left\{n \geqslant n_0 : \Pr{}^*\{\mathbf{A}_n\widehat{\Theta}(\underline{\mathbf{X}}^*_n) - \mathbf{A}_n\widehat{\Theta}(\underline{\mathbf{X}}_n) \leqslant \delta \cdot \mathbf{1}_{k-1}\} \geqslant 1 - \alpha\right\} \tag{5}$$

where the linear transformation $\mathbf{A}_n$ is defined as follows:

$$\mathbf{A}_n = [a_{ij}],\ i = 1, 2, \ldots, k-1;\ j = 1, 2, \ldots, k \tag{6}$$

where

$$a_{ij} = \begin{cases} 1, & i = j \ \text{ for } j < (k); i+1 = j \ \text{ for } j > (k) \\ -1, & 1 \leqslant i \leqslant k-1;\ j = (k) \\ 0, & \text{otherwise} \end{cases}$$

and $(\cdot)$ denotes an index such that $\widehat{\theta}(\underline{\mathbf{X}}_{(1)n}) \leq \widehat{\theta}(\underline{\mathbf{X}}_{(2)n}) \leq \cdots \leq \widehat{\theta}(\underline{\mathbf{X}}_{(k)n})$. We also need to define the linear transformation $\mathbf{A}$ to state our theoretical results as follows:

$$\mathbf{A} = [a_{ij}],\ i = 1, 2, \ldots, k-1;\ j = 1, 2, \ldots, k \tag{7}$$

where

$$a_{ij} = \begin{cases} 1, & 1 \leqslant i = j \leqslant k-1 \\ -1, & 1 \leqslant i \leqslant k-1;\ j = k \\ 0, & \text{otherwise.} \end{cases}$$

To state our asymptotic results, we also need the following definitions:

**Definition 1**

**(a)** For any $c \in \mathbb{R}$ and positive definite covariance matrix $\Sigma$, define $\Omega_\Sigma : \mathbb{R} \mapsto [0,1]$ as

$$\Omega_\Sigma(c) \quad = \quad \int_{(-\infty,c]^k} (2\pi)^{-k/2} |\Sigma|^{-1/2} e^{-\mathbf{y}^T \Sigma^{-1} \mathbf{y}/2} \, d\mathbf{y}. \tag{8}$$

**(b)** For $\eta \in (0,1)$, define $a_\eta = \Omega_\Sigma^{-1}(\eta)$ such that $\Omega_\Sigma(a_\eta) = \eta$.

We now state the asymptotic validity of our generic R&S procedures in Theorems 1 and 2. These results show two things: As the smallest difference we care to detect $\delta \to 0$, the sample-size stopping time $N^*$ grows as $O(1/\delta^2)$, and the coverage of the MCB intervals (which guarantees PCS) converges to $1-\alpha$. The theorems assume the number of bootstrap resamples $B = \infty$, which in practice means that $B$ is large enough that the bootstrap PCS estimate is reasonably precise.

Theorem 1 considers the performance measure being the mean, i.e., $\Theta = \mathrm{E}[\mathbf{X}]$; this includes variances and probabilities.

**Theorem 1** Let $\Theta = \mathrm{E}[\mathbf{X}]$. Suppose that $\mathrm{E}[|\mathbf{X}-\Theta|^3] < \infty$ and $\Sigma = \mathrm{E}\left[(\mathbf{X}-\Theta)(\mathbf{X}-\Theta)^T\right]$ is a positive definite matrix. Consider $N^*$ as defined in (5).

(a) As $\delta \downarrow 0$, we have

$$\delta^2 N^* \to a_{1-\alpha}^2 \qquad a.s.$$

where $a_{1-\alpha} = \Omega^{-1}_{\mathbf{A}\Sigma\mathbf{A}^T}(1-\alpha)$.

(b) As $\delta \downarrow 0$, we have

$$\Pr\left\{\mathbf{A}\widehat{\Theta}(\underline{\mathbf{X}}_{N^*}) - \mathbf{A}\Theta \le \delta \cdot \mathbf{1}_{k-1}\right\} \to 1-\alpha.$$

The theorem below considers the quantile case; $\Theta$ is a set of specific quantiles of the $k$ marginal distributions.

**Theorem 2** Let $F_i$ be twice continuously differentiable in a neighborhood of $\theta_i$ and $\zeta_i = f_i(\theta_i) > 0$, for $i = 1,2,\ldots,k$, where $f_i$ is the density associated with $F_i$. Further, let $F_{ij}$ be $(i,j)$th bivariate marginal distribution function. Consider $N^*$ as defined in (5).

(a) As $\delta \downarrow 0$, we have

$$\delta^2 N^* \to a_{1-\alpha}^2 \qquad a.s.$$

where $a = \Omega^{-1}_{\mathbf{A}\Sigma\mathbf{A}^T}(1-\alpha)$ with covariance matrix

$$\Sigma = \begin{pmatrix} \frac{q(1-q)}{\zeta_1^2} & \frac{\sigma_{12}}{\zeta_1\zeta_2} & \cdots & \frac{\sigma_{1k}}{\zeta_1\zeta_k} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\sigma_{k1}}{\zeta_k\zeta_1} & \frac{\sigma_{k2}}{\zeta_k\zeta_2} & \cdots & \frac{q(1-q)}{\zeta_k^2} \end{pmatrix}$$

with

$$\sigma_{ij} = F_{ij}(\theta_i,\theta_j) - q^2.$$

(b) As $\delta \downarrow 0$, we have

$$\Pr\left\{\mathbf{A}\widehat{\Theta}(\underline{\mathbf{X}}_{N^*}) - \mathbf{A}\Theta \le \delta \cdot \mathbf{1}_{k-1}\right\} \to 1-\alpha.$$

## 5 NUMERICAL RESULTS

This section illustrates computational improvements achieved by R&S via MCB compared to the MCA-based procedure presented in Lee and Nelson (2014a). We also demonstrate that the new procedure is robust to the distribution of the output data, unlike a normal-theory procedure due to Rinott (1978).

To establish the computational improvement, we consider systems with outputs having normal marginal distributions when the performance measure $\theta$ is the mean. All results presented here are averaged over 100 macroreplications of the entire experiment. To speed up the procedure, the jump-ahead modification introduced in Section 3 is incorporated for the sample-size increment $\Delta n$.

In Tables 1–2, the true mean vector is the slippage configuration $(5,5,\ldots,5,5+\delta)^\top$. In Tables 3–4, the true mean vector is the monotone increasing means configuration with $\theta_i = 5+(i-1)(0.05)$, $i = 1,2,\ldots,k$. Variances of the normally distributed output are all 1, and CRN are used inducing a correlation of 0.9. We vary the initial sample size $n_0$ and the number of systems of interest. The indifference-zone parameter is $\delta = 0.1$ or 0.05, the number of bootstrap resamples is $B = 200$, and the nominal confidence level is $1-\alpha = 0.95$ for all experiments.

Tables 1 and 3 report the results obtained from applying R&S via MCB, while Tables 2 and 4 report the corresponding results obtained from the procedure based on fixed-width CIs for all-pairwise comparisons (R&S via MCA) presented in Lee and Nelson (2014a). We distinguish "correct selection" (CS) from "good selection" (GS): PCS refers to the probability of choosing the true best system, while PGS refers to the probability of choosing the best, or a system within $\delta$ of the best system. As stated in Section 2, our procedures provide a PGS guarantee. When the true difference between the best and the rest is $\geq \delta$, then CS and GS are the same event.

Since both R&S via MCB and MCA are based on confidence intervals, we also evaluate their coverage. In Tables 1 and 3, the estimated coverage probability is $P^*_{NB}$ from Step 6 of the algorithm, and the true coverage probability is computed as the fraction of macroreplications on which all $k$ (constrained) MCB intervals simultaneously cover $\theta_i - \max_{j\neq i}\theta_j$ for $i = 1,2,\ldots,k$. In Tables 2 and 4, the estimated coverage probability is computed based on Step 6 of the algorithm in Lee and Nelson (2014a), and the true coverage probability is computed as the fraction of macroreplications on which all $k(k-1)/2$ CIs simultaneously cover $\theta_i - \theta_j$ for all $i \neq j$. We know that the estimated coverage will always be $\geq 0.95$ because we do not terminate the algorithm until it is; the true coverage results show that we do not overshoot by much.

In all cases the required sample size in R&S via MCB (Tables 1 and 3) is smaller than the corresponding sample size of R&S via MCA (Tables 2 and 4). The PCS values in Tables 2 and 4 are larger than the nominal PCS, while those in Tables 1 and 3 as much closer. For example, when $k = 20$ and $n_0 = 100$, the sample size in R&S via MCA is 257 and its PCS is 1.00, while the sample size in R&S via MCB is 141 and its PCS is 0.96. The sample size savings achieved by R&S via MCB increases as the number of systems $k$ increases and $\delta$ decreases. Therefore, as the number of systems increases, the new procedure yields more sample-size savings; in addition, the bootstrap computational effort is greatly reduced relative to R&S via MCA. In summary, R&S via MCB is clearly superior in these experiments.

If we knew the output distributions of our simulation were normal, then there would be no need for our bootstrap R&S procedures. Their value comes from the fact that we never know the distributions for certain in practice. To illustrate that R&S via MCB is robust, Table 5 reports results in the slippage configuration, with or without using CRN, for problems with a mix of output distributions. Specifically, half of the systems have normally distributed output and the other half of the systems have exponentially distributed output. When employing CRN for this case, we used the NORTA method as described in Nelson (2013), with the resulting correlation approximately 0.9. In Table 5, "n" and "e" indicate that the best system has normally or exponentially distributed outputs, respectively. Notice that there are two cases where the nominal PCS is not attained; the estimated PCS of 0.92 and 0.90 are around two standard errors from the nominal.

To illustrate that a robust R&S procedure is actually needed for this problem, we also applied Rinott's procedure to the $n_0 = 50$ and $k = 10$ cases without CRN; these cases are indicated by the † in Table 5. Rinott's procedure is a two-stage, normal-theory procedure for which the sample size from each system

Table 1: Empirical results of R&S via MCB from 100 macroreplications for normal distributions with CRN under the slippage configuration.

| $k$ | $n_0$ | $\delta$ | Average $N^*$ | PCS | PGS | Est. Coverage | True Coverage | Ave Jumps |
|---|---|---|---|---|---|---|---|---|
| 10 | 50 | 0.1 | 119 | 0.94 | 0.94 | 0.96 | 0.94 | 3.29 |
| 10 | 100 | 0.1 | 119 | 0.96 | 0.96 | 0.96 | 0.96 | 2.38 |
| 20 | 50 | 0.1 | 143 | 0.92 | 0.92 | 0.96 | 0.92 | 3.61 |
| 20 | 100 | 0.1 | 141 | 0.96 | 0.96 | 0.96 | 0.96 | 2.98 |

Table 2: Empirical results of R&S via MCA from 100 macroreplications for normal distributions with CRN under the slippage configuration.

| $k$ | $n_0$ | $\delta$ | Average $N^*$ | PCS | PGS | Est. Coverage | True Coverage | Ave Jumps |
|---|---|---|---|---|---|---|---|---|
| 10 | 50 | 0.1 | 205 | 0.98 | 0.98 | 0.96 | 0.96 | 4.12 |
| 10 | 100 | 0.1 | 205 | 1.00 | 1.00 | 0.96 | 0.98 | 3.72 |
| 20 | 50 | 0.1 | 260 | 1.00 | 1.00 | 0.96 | 0.98 | 4.74 |
| 20 | 100 | 0.1 | 257 | 1.00 | 1.00 | 0.96 | 0.98 | 3.99 |

Table 3: Empirical results of R&S via MCB from 100 macroreplications for normal distributions with CRN under the monotone increasing means configuration.

| $k$ | $n_0$ | $\delta$ | Average $N^*$ | PCS | PGS | Est. Coverage | True Coverage | Ave Jumps |
|---|---|---|---|---|---|---|---|---|
| 10 | 50 | 0.05 | 494 | 1.00 | 1.00 | 0.96 | 0.98 | 3.54 |
| 10 | 50 | 0.1 | 121 | 0.84 | 1.00 | 0.96 | 0.93 | 3.68 |
| 20 | 50 | 0.05 | 581 | 1.00 | 1.00 | 0.96 | 0.97 | 3.49 |
| 20 | 50 | 0.1 | 143 | 0.92 | 1.00 | 0.96 | 0.98 | 3.61 |

Table 4: Empirical results of R&S via MCA from 100 macroreplications for normal distributions with CRN under the monotone increasing means configuration.

| $k$ | $n_0$ | $\delta$ | Average $N^*$ | PCS | PGS | Est. Coverage | True Coverage | Ave Jumps |
|---|---|---|---|---|---|---|---|---|
| 10 | 50 | 0.05 | 814 | 1.00 | 1.00 | 0.96 | 0.94 | 4.82 |
| 10 | 50 | 0.1 | 205 | 0.97 | 1.00 | 0.96 | 0.96 | 4.12 |
| 20 | 50 | 0.05 | 1024 | 1.00 | 1.00 | 0.96 | 0.96 | 5.86 |
| 20 | 50 | 0.1 | 260 | 0.97 | 1.00 | 0.96 | 0.98 | 4.74 |

Table 5: Empirical results of R&S via MCB from 100 macroreplications for half exponential and half normal distributions under the slippage configuration.

| $k$ | $n_0$ | $\delta$ | Average $N^*$ | PCS | PGS | Est. Coverage | True Coverage | Ave Jumps | CRN |
|---|---|---|---|---|---|---|---|---|---|
| 10 (n)† | 50 | 0.1 | 19207 | 1.00 | 1.00 | 0.97 | 1.00 | 4.37 | N |
| 10 (e)† | 50 | 0.1 | 25191 | 0.98 | 0.98 | 0.96 | 0.98 | 6.01 | N |
| 10 (n) | 50 | 0.1 | 7133 | 0.90 | 0.90 | 0.96 | 0.90 | 6.21 | Y |
| 10 (e) | 50 | 0.1 | 5632 | 0.97 | 0.97 | 0.96 | 0.97 | 6.11 | Y |
| 20 (n) | 50 | 0.1 | 27065 | 1.00 | 1.00 | 0.97 | 1.00 | 4.28 | N |
| 20 (e) | 50 | 0.1 | 29779 | 0.95 | 0.95 | 0.96 | 0.95 | 5.71 | N |
| 20 (n) | 50 | 0.1 | 7977 | 0.92 | 0.92 | 0.96 | 0.92 | 6.87 | Y |
| 20 (e) | 50 | 0.1 | 6041 | 0.99 | 0.99 | 0.96 | 0.99 | 5.99 | Y |

is proportional to its sample variance; it cannot exploit CRN. The average sample size per system for Rinott's procedure based on 100 macroreplications was 17,767 and 18,185, with corresponding PCS values of 0.35 and 0.46, when the best system is exponentially distributed and normally distributed, respectively. The corresponding sample size per system for our R&S via MCB was 19,207 and 25,191, and their PCS values were 1.00 and 0.98, respectively. Clearly Rinott's procedure—which assumes normally distributed output—obtains too few samples, resulting in a PCS that is far from the nominal, while our R&S via MCB attains at least the desired PCS.

## ACKNOWLEDGMENTS

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning 2014 R1A1A3049955.

## REFERENCES

Bechhofer, R. E., D. M. Goldsman, and T. J. Santner. 1995. *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons*. New York: Wiley.

Damerdji, H., and M. K. Nakayama. 1999. "Two-Stage Multiple-Comparison Procedures for Steady-State Simulations". *ACM Transactions on Modeling and Computer Simulation* 9 (1): 1–30.

Hochberg, Y., and A. C. Tamhane. 1987. *Multiple Comparison Procedures*. New York: Wiley.

Hsu, J. C. 1996. *Multiple Comparisons: Theory and Methods*. Boca Raton, FL: Chapman & Hall/CRC.

Kim, S.-H., and B. L. Nelson. 2006. "Selecting the Best System". In *Handbooks in Operations Research and Management Science: Simulation*, edited by S. G. Henderson and B. L. Nelson, Volume 13, Chapter 17, 501–534. New York: Elsevier.

Lee, S., and B. L. Nelson. 2014a. "Bootstrap Ranking & Selection Revisited". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 3857–3868. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lee, S., and B. L. Nelson. 2014b. "Bootstrap Ranking and Selection for Simulation". Technical report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL.

Lee, S., and B. L. Nelson. 2015. "Direct PCS Procedures in Bootstrap Ranking and Selection". Technical report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL.

Nelson, B. L. 2013. *Foundations and Methods of Stochastic Simulation: A First Course*. New York: Springer.

Nelson, B. L., and F. J. Matejcik. 1995. "Using Common Random Numbers for Indifference-Zone Selection and Multiple Comparisons in Simulation". *Management Science* 41 (12): 1935–1945.

Rinott, Y. 1978. "On Two-Stage Selection Procedures and Related Probability-Inequalities". *Communications in Statistics-Theory and methods* 7 (8): 799–811.

Swanepoel, J., J. V. Wyk, and J. Venter. 1983. "Fixed Width Confidence Intervals Based on Bootstrap Procedures". *Communications in Statistics-Sequential Analysis* 2 (4): 289–310.

## AUTHOR BIOGRAPHIES

**SOONHUI LEE** is an Assistant Professor in the School of Business Administration at UNIST. She received her B.S. at KAIST, M.S. at Georgia Institute of Technology and Ph.D. in Industrial Engineering and Management Sciences at Northwestern University. Her research interests include stochastic optimization, and its application. Her email address is shlee@unist.ac.kr.

**BARRY L. NELSON** is the Walter P. Murphy Professor in the Department of Industrial Engineering and Management Sciences at Northwestern University. He is a Fellow of INFORMS and IIE. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems, and he is the author of *Foundations and Methods of Stochastic Simulation: A First Course*, from Springer. His e-mail address is nelsonb@northwestern.edu.