IIE Transactions

ISSN: 0740-817X (Print) 1545-8830 (Online) Journal homepage: http://www.tandfonline.com/loi/uiie20

General-purpose Ranking & Selection for Computer Simulation

Soonhui Lee & Barry L. Nelson

To cite this article: Soonhui Lee & Barry L. Nelson (2016): General-purpose Ranking & Selection for Computer Simulation, IIE Transactions, DOI: 10.1080/0740817X.2015.1125043

To link to this article: http://dx.doi.org/10.1080/0740817X.2015.1125043

General-purpose Ranking & Selection for Computer Simulation

Soonhui Lee[1] and Barry L. Nelson[2]

[1]shlee@unist.ac.kr, [2]nelsonb@northwestern.edu

[1]School of Business Administration

Ulsan National Institute of Science and Technology

Ulsan, REPUBLIC OF KOREA

[2]Department of Industrial Engineering and Management Sciences

Northwestern University

Evanston, IL 60208-3119, USA

January 4, 2016

Abstract

Many indifference-zone ranking-and-selection (R&S) procedures have been invented for choosing the best simulated system. To obtain the desired probability of correct selection (PCS), existing procedures exploit knowledge about the particular combination of system performance measure (e.g., mean, probability, variance, quantile) and assumed output distribution (e.g., normal, exponential, Poisson). In this paper we take a step toward general-purpose R&S procedures that work for many types of performance measures and output distributions, including situations in which different simulated alternatives

have entirely different output distribution families. There are only two versions of our procedure: with and without the use of common random numbers. To obtain the required PCS we exploit intense computation via bootstrapping, and to mitigate the computational effort we create an adaptive sample-allocation scheme that guides the procedure to quickly reach the necessary sample size. We establish the asymptotic PCS of these procedures under very mild conditions, and provide a finite-sample empirical evaluation of them as well.

1 Introduction

Although invented in the 1950's and 60's for biostatistics problems, the statistical methods of ranking and selection (R&S) have been embraced by the stochastic simulation community as a standard tool for selecting the best of a finite (and relatively small) number of alternative system designs. In fact, R&S procedures are featured in several commercial simulation products. A common characteristic of selection-of-the-best procedures that have been used extensively in simulation is that "best" is defined to be smallest or largest *mean* performance. Further, whether Bayesian or frequentist in philosophy, these procedures typically assume that the simulation output data are normally distributed; even procedures that are shown to be asymptotically valid under more general assumptions are derived based on normality. Later we provide a realistic illustration in which the output data are highly non-normal.

There is also a literature on R&S problems that differ from the normal-mean case. For instance, there are procedures that define "best" to be the largest or smallest probability, variance or qth quantile. And there are also procedures designed for output data that are known to be non-normal, including Poisson, Bernoulli and exponential. Procedures for these situations are customized for the particular performance measure or type of data, exploiting mathematical-statistical properties of the relevant estimator or distribution.

Suppose, however, that the alternatives under consideration involve distinct technologies: manual vs.

automated; in-house vs. outsourced; or synthetic vs. biological. The simulation output data may not only be non-normal, but there is no a priori reason to believe that the alternatives even generate output data from the same distribution family. The R&S literature provides no direct approach for such settings other than batching the output data to achieve normality.

In this paper we take a big step toward general-purpose R&S, by which we mean procedures that work 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); and not all systems need to have the same output distribution family. We exploit intense computation—via bootstrapping—instead of clever mathematical analysis. To do this we employ a connection between fixed-width confidence intervals (CIs) and the desired probability of correct selection (PCS). Our approach is sequential, frequentist in philosophy and incorporates an indifference zone. Finally, our procedures are fixed precision (as measured by PCS) not fixed budget. Therefore, our first concern is to be able to deliver the desired PCS on virtually any problem, and a secondary (but still important) concern is being able to do so efficiently. This means we simulate until a specified PCS is achieved and stop, rather than trying to smartly allocate a fixed budget to achieve the best PCS we can attain with it.

Because we substitute computation for analysis, our generic procedure will not be competitive when simulation output data are so computationally cheap that we can simulate each alternative system until its point estimator has effectively zero variance. We also will not beat procedures that directly exploit (correct) distributional information; for instance, if we know our output data really are Poisson, then we expect that a procedure based on that knowledge should be more efficient than ours, although in our empirical studies we are surprisingly close (Lee and Nelson, 2014). On the other hand, we make only very mild assumptions about the output data, and there are only two versions of our procedure: with or without common random numbers (CRN). When CRN are employed we require no conservative inequalities to establish the validity

of our procedure. The PCS guarantees are proven asymptotically, but because they bootstrap the actual simulation output data our procedures work well in finite samples across a variety of situations.

Preliminary versions of the procedures described in this paper and a small empirical evaluation of them were presented in Lee and Nelson (2014). Here we prove their asymptotic validity and supplement the empirical results in Lee and Nelson (2014) with a more carefully chosen set of experiments. Also, we introduce a significant computational speed-up that allows our methods to be applied very rapidly to small problems (say 10 or fewer alternatives) and to be computationally feasible for larger problems.

The paper is organized as follows. In Section 2 we describe the related literature that supports our work. Sections 3–4 present our R&S procedures and prove their asymptotic validity. A method for significantly reducing the number of sequential steps required to reach a correct selection is introduced in Section 5. Sections 6–7 contain experiment results. Conclusions are offered in the final section.

2 Background

Comprehensive treatments of R&S outside of computer simulation can be found in Bechhofer et al. (1995) and Gupta and Panchapakesan (1979). Simulation focused surveys are provided by Kim and Nelson (2006b) and Kim (2013). While this background is relevant, our approach to creating a R&S procedure is different in that we exploit a connection between indifference-zone-δ selection of the best and fixed-width-δ CIs.

Let X_{ij} represent the jth observed output of system i, for $i = 1,2,\ldots,k$, so that $\mathbf{X}_j = (X_{1j}, X_{2j}, \ldots, X_{kj})^\top$ is a $k \times 1$ vector representing the jth 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 CRN it will be useful to think of $\mathbf{X}_1, \mathbf{X}_2, \ldots$ as i.i.d. with common joint distribution function $F(\mathbf{x}) = \Pr\{X_{1j} \leqslant x_1, \ldots, X_{kj} \leqslant x_k\}$, $\mathbf{x} = (x_1, \ldots, x_k)^\top \in \mathbb{R}^k$. We neither assume nor fit any specific

distribution to the simulation output. We use boldface type as in $\mathbf{X}_j$ to indicate a vector of observations across all k systems, and underlining as in $\underline{X}_{in}$ to indicate a vector of n observations from system i. A $\widehat{\ }$ denotes an estimator, and we append a $*$ to quantities defined by bootstrapping.

Let $\Theta = (\theta_1, \theta_2, \ldots, \theta_k)^\top$ be a vector whose ith element is a statistical property of the marginal distribution F_i, such as its mean, variance, a quantile, or a probability. We are interested in finding the sample size that allows us to select the system with the largest value of θ_i with a specified PCS by choosing the one with the largest empirical estimate $\widehat{\theta}_i$ of it. Notice that the objective is finding an (ideally small) sample size that satisfies the PCS constraint.

For $k \geq 2$ systems, suppose we can build fixed-width-δ CIs for all $\theta_i - \theta_j, i \neq j$ with simultaneous coverage $1-\alpha$. As shown in Hsu (1996), if we have

$$\Pr\{\widehat{\theta}_i - \widehat{\theta}_j - (\theta_i - \theta_j) \leq \delta, \ \ \forall i \neq j\} \geqslant 1-\alpha$$

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

$$\theta_i - \max_{j \neq i} \theta_j \in \left[\widehat{\theta}_i - \max_{j \neq i} \widehat{\theta}_j - \delta, \widehat{\theta}_i - \max_{j \neq i} \widehat{\theta}_j + \delta\right] \tag{1}$$

for $i = 1, 2, \ldots, k$. If M is the index of the system with the largest performance estimate, i.e., $\widehat{\theta}_M \geqslant \widehat{\theta}_i$ for all $i \neq M$, then it follows from (1) that with probability at least $1-\alpha$

$$\theta_M - \max_{j \neq M} \theta_j \geqslant \widehat{\theta}_M - \max_{j \neq M} \widehat{\theta}_j - \delta \geqslant -\delta$$

since $\widehat{\theta}_M - \max_{j \neq M} \widehat{\theta}_j \geqslant 0$. This result implies that if we select the system with the largest performance estimate $\widehat{\theta}_M$ as the best system, the selected system will be the best system or a system within δ of the best system with probability at least $1-\alpha$. Moreover, if the difference between the largest and the second-largest parameter value is strictly greater than δ, then the selected system is the best system with probability at least $1-\alpha$. This is exactly the desired inference for an indifference-zone R&S procedure. In R&S, δ is a user-specified parameter corresponding to the smallest practically significant difference worth detecting. *Thus,*

if we have a procedure to create fixed-width-δ CIs for $\theta_i - \theta_j$ with overall coverage $\geqslant 1-\alpha$, then we also have a selection-of-the-best procedure.

Swanepoel et al. (1983) describe a sequential bootstrapping procedure for generating a single fixed-width CI with a specified coverage probability $1-\alpha$ when θ is either the mean or median. Given an i.i.d. sample of size n, denoted $\underline{X}_n = \{X_1, X_2, \ldots, X_n\}$, from a population with marginal distribution F having a distribution property θ, let $\widehat{F}_n$ denote the empirical cumulative distribution function (ecdf) of $\underline{X}_n$ defined as

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{I}\{X_j \leqslant x\}.$$

Let $\widehat{\theta}_n$ be the corresponding distributional property of $\widehat{F}_n$. Further, let $\underline{X}_n^* = \{X_1^*, X_2^*, \ldots, X_n^*\}$ denote a random sample of size n from $\widehat{F}_n$, $\widehat{F}_n^*$ the implied ecdf, and $\widehat{\theta}(\underline{X}_n^*)$ (also denoted by $\widehat{\theta}_n^*$) the corresponding distributional property of $\widehat{F}_n^*$. The bootstrap estimator of the probability that θ is contained within the interval $[\widehat{\theta}_n - \delta, \widehat{\theta}_n + \delta]$ is

$$P_n^* = \Pr\left\{\widehat{\theta}_n \in \left[\widehat{\theta}_n^* - \delta, \widehat{\theta}_n^* + \delta\right]\right\}. \tag{2}$$

Exact computation of P_n^* is often difficult, but (2) can be estimated given B random samples of size n from $\widehat{F}_n$, say $\underline{X}_{nb}^* = \{X_{1b}^*, X_{2b}^*, \ldots, X_{nb}^*\}$, $b = 1, 2, \ldots, B$, by using

$$P_{nB}^* = \frac{1}{B}\sum_{b=1}^{B} \mathbf{I}\left\{\widehat{\theta}_n \in \left[\widehat{\theta}_{nb}^* - \delta,\ \widehat{\theta}_{nb}^* + \delta\right]\right\} \tag{3}$$

where $\widehat{\theta}_{nb}^*$, $b = 1, 2, \ldots, B$, is the estimate of the distributional property of interest from the bth bootstrap sample.

In their procedure, Swanepoel et al. (1983) sequentially increase the number of observations of X until the stopping time $N^* = \inf\{n \geqslant n_0 : P_n^* \geqslant 1-\alpha\}$, when the desired bootstrap coverage probability is $1-\alpha$. The asymptotic properties of N^* were shown when θ is the mean or median of X, as stated in the following theorem:

Theorem (Swanepoel et al. (1983))**.** *Under some mild assumptions, as* $\delta \downarrow 0$*,*

(a) $\delta^2 N^* \to c$ *a.s.*

(b) $\Pr\left\{|\widehat{\theta}_{N^*} - \theta| \leq \delta\right\} \to 1-\alpha$.

The limit c depends on the distributional property of interest: $c = E(X-\theta)^2 z^2_{1-\alpha/2}$ when θ is the mean and $c = z^2_{1-\alpha/2}/(4f(\theta)^2)$ when θ is the median, where f is the density function of X and $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution.

Building from this foundation, we establish similar, but significantly more difficult, asymptotic results for fixed-width-δ simultaneous CIs for $k(k-1)/2$ pairs of differences $\theta_i - \theta_j$. This in turn provides an indifference-zone R&S procedure. Further, we propose a method to more quickly jump to the stopping time N^*, avoiding many costly evaluations of P^*_{nB}.

There are, of course, parametric (in particular normal-theory) methods to construct fixed-width CIs. See, for instance, Hochberg and Tamhane (1987). More closely related to this research, Aerts and Gijbels (1993) and Hlávka (2003) proposed three-stage procedures that use bootstrapping to estimate the critical constant that determines sample size (e.g., z-value in the normal-theory case): the first stage provides a crude estimate; the second stage refines it; and the third stage jumps to the stopping time N^*. These methods do not directly estimate the PCS, but more importantly seem difficult to generalize beyond a single CI.

3 Procedures

In this section we describe algorithms for performing R&S for $k \geqslant 2$ systems using the bootstrap-based fixed-width confidence interval approach. We present two versions of the algorithm, one that exploits CRN and one without CRN. The algorithm without using CRN has been presented in Bekki et al. (2010) when the sample size is incremented one at a time. We generalize the algorithm here to allow $\Delta n \geq 1$ additional

observations on each iteration; this has the effect of speeding up the algorithm at the possible cost of taking more observations than necessary to guarantee a correct selection. Later we present a method for adaptively choosing Δn that provides a substantial computational speed-up without noticeable overshoot of PCS.

First we describe the procedure without using CRN. Let $\underline{X}_{in} = \{X_{i1}, X_{i2}, \ldots, X_{in}\}$ be a sample of size n from a system with output distribution F_i having distribution property θ_i, and $\widehat{F}_{in}$ the ecdf based on $\underline{X}_{in}$ for system $i = 1, 2, \ldots, k$. Let $\widehat{\theta}(\underline{X}_{in})$ be an estimate of θ_i based on $\underline{X}_{in}$ for $i = 1, 2, \ldots, k$ and $\widehat{\theta}_{ij}(\underline{\mathbf{X}}_n) = \widehat{\theta}(\underline{X}_{in}) - \widehat{\theta}(\underline{X}_{jn})$ for all $i \neq j$. We want to build simultaneous fixed-width-δ confidence intervals for all pairs of differences $\theta_i - \theta_j$ for $i \neq j$ by finding n such that

$$\Pr\left\{\theta_{ij} \in \left[\widehat{\theta}_{ij}(\underline{\mathbf{X}}_n) - \delta, \widehat{\theta}_{ij}(\underline{\mathbf{X}}_n) + \delta\right], \forall i \neq j\right\} \geq 1 - \alpha \tag{4}$$

where $\theta_{ij} = \theta_i - \theta_j$. The value of n will be the smallest one for which the estimated coverage probability using bootstrapping is at least $1 - \alpha$. Specifically, given B random samples of size N from $\widehat{F}_{iN}$, $\underline{X}^*_{iNb} = \{X^*_{i1b}, X^*_{i2b}, \ldots, X^*_{iNb}\}$, $b = 1, 2, \ldots, B$, the bootstrap coverage probability is estimated by

$$P^*_{NB} = \frac{1}{B}\sum_{b=1}^{B} \prod_{(i,j|i\neq j)} \mathbf{I}\left\{\widehat{\theta}_{ij}(\underline{\mathbf{X}}_N) \in \left[\widehat{\theta}_{ij}(\underline{\mathbf{X}}^*_{Nb}) - \delta,\ \widehat{\theta}_{ij}(\underline{\mathbf{X}}^*_{Nb}) + \delta\right]\right\} \tag{5}$$

where $\widehat{\theta}(\underline{X}^*_{iNb})$ is an estimate of $\widehat{\theta}(\underline{X}_{iN})$ based on $\underline{X}^*_{iNb}$, and $\widehat{\theta}_{ij}(\underline{\mathbf{X}}^*_{Nb}) = \widehat{\theta}(\underline{X}^*_{iNb}) - \widehat{\theta}(\underline{X}^*_{jNb})$ for all $i \neq j$. The procedure without CRN described below starts with a sample of size $N = n_0$ from each system $i = 1, 2, \ldots, k$, a desired PCS $1 - \alpha$, a half width (indifference-zone parameter) δ for the CIs, and a sample-size increment Δn.

Bootstrap R&S procedure without CRN

1. Specify $N = n_0$, set $1/k < 1 - \alpha < 1$, $\delta > 0$, and $\Delta n \geq 1$.

2. Obtain $\underline{X}_{iN} = \{X_{i1}, X_{i2}, \ldots, X_{iN}\}$ a sample of size N from the distribution F_i for $i = 1, 2, \ldots, k$.

3. Compute $\widehat{\theta}_{ij}(\underline{\mathbf{X}}_N) = \widehat{\theta}(\underline{\mathrm{X}}_{iN}) - \widehat{\theta}(\underline{\mathrm{X}}_{jN})$ for all $i \neq j$ where θ_i is a distributional property of F_i and $\widehat{\theta}(\underline{\mathrm{X}}_{iN})$ is an estimate of θ_i based on $\underline{\mathrm{X}}_{iN}$; and 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 $\widehat{F}_{iN} : \underline{\mathrm{X}}^*_{iN1}, \ldots, \underline{\mathrm{X}}^*_{iNB}, i = 1, 2, \ldots, k$.

5. Compute $\widehat{\theta}_{ij}(\underline{\mathbf{X}}^*_{Nb}) = \widehat{\theta}(\underline{\mathrm{X}}^*_{iNb}) - \widehat{\theta}(\underline{\mathrm{X}}^*_{jNb}),\ b = 1, 2, \ldots, B$ for all $i \neq j$.

6. Estimate the PCS as

$$P^*_{NB} = \frac{1}{B} \sum_{b=1}^{B} \prod_{(i,j|i\neq j)} \mathbf{I}\left\{ |\widehat{\theta}_{ij}(\underline{\mathbf{X}}^*_{Nb}) - \widehat{\theta}_{ij}(\underline{\mathbf{X}}_N)| \leqslant \delta \right\}.$$

7. If $P^*_{NB} \geqslant 1 - \alpha$, report $\arg\max_{i=1,\ldots,k} \widehat{\theta}(\underline{\mathrm{X}}_{iN})$ as the best system.

   Else

   Obtain $\underline{\mathrm{X}}_{i\Delta n}$ a sample of size Δn from the distribution F_i for $i = 1, 2, \ldots, k$.

   Set $\underline{\mathrm{X}}_{iN} = \underline{\mathrm{X}}_{iN} \cup \underline{\mathrm{X}}_{i\Delta n}$ for $i = 1, 2, \ldots, k$ and $N = N + \Delta n$.

   Go to Step 3.

   End If

We next present the bootstrap R&S procedure that exploits CRN. The goal of CRN is to induce a positive covariance between $\widehat{\theta}(\underline{\mathrm{X}}_{iN})$ and $\widehat{\theta}(\underline{\mathrm{X}}_{jN})$, which reduces the variance of the difference $\widehat{\theta}_{ij}(\underline{\mathbf{X}}_N)$. This in turn tends to reduce the sample size N required to achieve a given PCS. Therefore, the sample size required to attain the desired PCS when employing CRN is expected to be reduced relative to independent sampling, as shown later in Section 6. We recommend using CRN wherever it can be effectively employed.

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 $\{F_1, F_2, \ldots, F_k\}$; we denote that distribution by F. Correspondingly, we draw bootstrap samples from the empirical *joint* ecdf $\widehat{F}_N$, rather than from each marginal ecdf $\widehat{F}_{iN}$. Below we list only the steps that change from the **Bootstrap R&S procedure without CRN**:

Bootstrap R&S procedure with CRN

2. Obtain N samples $\mathbf{X}_j = (X_{1j}, X_{2j}, \ldots, X_{kj})^\top$ $j = 1, 2, \ldots, N$ from the joint distribution F.

3. Compute $\widehat{\theta}_{ij}(\underline{\mathbf{X}}_N) = \widehat{\theta}(\underline{\mathrm{X}}_{iN}) - \widehat{\theta}(\underline{\mathrm{X}}_{jN})$ for all $i \neq j$ where θ_i is a distributional property of F_i, and $\widehat{\theta}(\underline{\mathrm{X}}_{iN})$ is an estimate of θ_i based on $\underline{\mathrm{X}}_{iN}$; and 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 $\mathbf{X}^*_{jb} = (X^*_{1jb}, X^*_{2jb}, \ldots, X^*_{kjb})^\top$ for $j = 1, 2, \ldots, N$.

7. If $P^*_{NB} \geqslant 1 - \alpha$, report $\arg\max_{i=1,\ldots,k} \widehat{\theta}(\underline{\mathrm{X}}_{iN})$ as the best system.

   Else

   Obtain $\underline{\mathbf{X}}_{\Delta n} = \{\mathbf{X}_j,\ j = 1, 2, \ldots, \Delta n\}$ a sample of size Δn 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

Remark. *Although our algorithms state that we form the ecdfs of the output data, we do not need to actually create the ecdfs since bootstrapping simply requires random sampling from the data with replacement. However, the ecdfs are needed for the proofs in Section 4 below.*

4 Asymptotic Analysis

This section provides theoretical support for the R&S procedures introduced in the previous section. Recall that our goal is to achieve the desired PCS on virtually any problem, taking only as many observations as

needed. For an indifference-zone R&S procedure, the most difficult case is when the true differences (which are unknown in practice) are small, and we demand to be able to detect small differences. Recall that δ is a fixed quantity representing the smallest difference that is practically important to the user. Therefore, we employ an asymptotic regime in which we let the indifference-zone parameter δ go to 0. This drives the required sample size N^* (obtained from the procedures) to infinity, and hence, we show that the procedures achieve the desired PCS asymptotically.

The theorems stated below, and proved in the Appendix, extend Swanepoel et al. (1983) from a CI for a univariate mean or median, to simultaneous CIs for the means or *any* quantile from $k \geq 2$ systems. We then show in the corollaries that these results justify our use of bootstrap R&S for either mean or quantile performance measures by providing simultaneous CIs for all pairwise differences.

We first review the key notation. 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 Θ, where $\mathbf{X}_j = (X_{1j}, X_{2j}, \ldots, X_{kj})^\top$, $j = 1, 2, \ldots, n$. Further, let $\widehat{F}_n(\mathbf{x})$ be the ecdf based on $\underline{\mathbf{X}}_n$ defined in two different ways for use in the procedure without CRN, as in (6), and with CRN, as in (7):

$$\widehat{F}_n(\mathbf{x}) = \prod_{i=1}^{k} \left(\frac{1}{n} \sum_{j=1}^{n} \mathbf{I}\left\{X_{ij} \leqslant x_i\right\} \right) \tag{6}$$

$$\widehat{F}_n(\mathbf{x}) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{I}\left\{X_{1j} \leqslant x_1,\ X_{2j} \leqslant x_2,\ \ldots,\ X_{kj} \leqslant x_k\right\} \tag{7}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_k)^\top \in \mathbb{R}^k$. The results below are valid in either case.

Let $\underline{\mathbf{X}}_n^* = \{\mathbf{X}_1^*, \mathbf{X}_2^*, \ldots, \mathbf{X}_n^*\}$ denote a random sample of size n from $\widehat{F}_n$. Let Pr and Pr* denote probabilities under F and $\widehat{F}_n$; and E and E* denote expectations under F and $\widehat{F}_n$, respectively. Probabilities and expectations under $*$ are equivalent to estimating them using $B \to \infty$ bootstrap samples.

The bootstrap stopping variable N^* is given by

$$N^* = \inf\left\{ n \geqslant n_0 : \Pr^*\left\{ |\widehat{\Theta}(\underline{\mathbf{X}}_n^*) - \widehat{\Theta}(\underline{\mathbf{X}}_n)| \leqslant \delta \cdot \mathbf{1}_k \right\} \geqslant 1 - \alpha \right\} \tag{8}$$

where $\mathbf{1}_k$ is the $k \times 1$ column vector of ones. The inequality between vectors used in (8) is defined as $\mathbf{x} \leqslant \mathbf{y}$ if $x_i \leqslant y_i$ for $i = 1,2,\ldots,k$, where $\mathbf{x} = (x_1, x_2, \ldots, x_k)^\top$ and $\mathbf{y} = (y_1, y_2, \ldots, y_k)^\top$. When $\Theta = \mathrm{E}(\mathbf{X})$, 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 $\widehat{\Theta}(\underline{\mathbf{X}}_n^*) = \bar{\mathbf{X}}_n^* = \sum_{j=1}^n \mathbf{X}_j^*/n$. Notice that the "mean" case includes probabilities as they are expected values of indicator outputs. We assume that n_0 in (8) grows as $\delta \downarrow 0$, as in Swanepoel et al. (1983).

Before we state our theorems, we introduce the following definitions:

Definition 1. **(a)** *For any $c > 0$ and positive-definite covariance matrix Σ, let $\Gamma_\Sigma : \mathbb{R}^+ \mapsto (0,1]$ be*

$$\Gamma_\Sigma(c) \quad = \quad \int_{[-c,c]^k} (2\pi)^{-k/2} |\Sigma|^{-1/2} e^{-\mathbf{y}^\top \Sigma^{-1} \mathbf{y}/2} \, d\mathbf{y}.$$

(b) *For $\eta \in (0,1)$, let $a_\eta = \Gamma_\Sigma^{-1}(\eta)$; that is*

$$\Gamma_\Sigma(a_\eta) = \eta.$$

From here on, whenever we raise a vector or matrix to a power, as in $\mathbf{X}^2$, we mean element-by- element exponentiation. Matrix multiplications are written explicitly, as in $\mathbf{X}\mathbf{X}^\top$.

Theorem 1. *Let $\Theta = \mathrm{E}[\boldsymbol{X}]$. Suppose that $\mathrm{E}[|\boldsymbol{X} - \Theta|^3] < \infty$ and that $\Sigma = \mathrm{E}\left[(\boldsymbol{X} - \Theta)(\boldsymbol{X} - \Theta)^\top\right]$ is a positive definite matrix. Consider N^* as defined in (8).*

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

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

where $a_{1-\alpha} = \Gamma_\Sigma^{-1}(1-\alpha)$.

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

$$\Pr\left\{|\bar{\boldsymbol{X}}_{N^*} - \Theta| \leq \delta \cdot \mathbf{1}_k\right\} \to 1 - \alpha.$$

Proof See Appendix A.2. □

Notice that the assumptions are very weak when $\Theta = \mathrm{E}[\mathbf{X}]$ and the statistic is a sample average of i.i.d. outputs. Suppose next that Θ is the set of variances of the k marginal distributions, where the ith element is $\theta_i = \mathrm{E}(X_i - \mu_i)^2$, $i = 1, 2, \ldots, k$. Unless μ_i is known, θ_i is not the expected value of the average of i.i.d. observations. Instead, $\widehat{\Theta}(\underline{\mathbf{X}}_n)$ and $\widehat{\Theta}(\underline{\mathbf{X}}_n^*)$ are the sample variances based on $\underline{\mathbf{X}}_n$ and $\underline{\mathbf{X}}_n^*$, respectively, where the ith element of $\widehat{\Theta}(\underline{\mathbf{X}}_n)$ (denoted by S_{in}^2) is the sample variance of $X_{i1}, X_{i2}, \ldots, X_{in}$ and the ith element of $\widehat{\Theta}(\underline{\mathbf{X}}_n^*)$ (denoted by S_{in}^{*2}) is the sample variance of $X_{i1}^*, X_{i2}^*, \ldots, X_{in}^*$; i.e.,

$$S_{in}^2 = \frac{\sum_{j=1}^n (X_{ij} - \bar{X}_i)^2}{n-1} \quad \text{and} \quad S_{in}^{*2} = \frac{\sum_{j=1}^n (X_{ij}^* - \bar{X}_i^*)^2}{n-1}.$$

Nevertheless, we can establish analogous asymptotic results for the sample variance in a manner very similar to Theorem 1, so we omit the proof.

Corollary 1. *Let* $\Theta = [\mathrm{Var}(X_1), \mathrm{Var}(X_2), \ldots, \mathrm{Var}(X_k)]^\top$. *Suppose* $\mathrm{E}[|\mathbf{X} - \mathrm{E}(\mathbf{X})|^4] < \infty$, *and let* N^* *be as defined in (8).*

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

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

where $a_{1-\alpha} = \Gamma_\Sigma^{-1}(1-\alpha)$ *with covariance matrix*

$$\Sigma = \mathrm{E}\left[\left((\mathbf{X} - \mathrm{E}[\mathbf{X}])^2 - \Theta \right) \left((\mathbf{X} - \mathrm{E}[\mathbf{X}])^2 - \Theta \right)^\top \right].$$

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

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

Next let Θ be a set of specific quantiles of the k marginal distributions where the ith element is

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

Then $\widehat{\Theta}(\underline{\mathbf{X}}_n)$ and $\widehat{\Theta}(\underline{\mathbf{X}}_n^*)$ are the sample qth quantiles based on $\underline{\mathbf{X}}_n$ and $\underline{\mathbf{X}}_n^*$, respectively, where the ith element of $\widehat{\Theta}(\underline{\mathbf{X}}_n)$ is the sample qth quantile of $X_{i1}, X_{i2}, \ldots, X_{in}$ and the ith element of $\widehat{\Theta}(\underline{\mathbf{X}}_n^*)$ is the sample qth quantile of $X_{i1}^*, X_{i2}^*, \ldots, X_{in}^*$. Stronger assumptions on the distributions F_i are required to obtain similar asymptotic results for quantiles.

Theorem 2. *Let F_i be twice continuously differentiable in a neighborhood of θ_i and $\xi_i = f_i(\theta_i) > 0$, 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 (8). Suppose that the covariance matrix*

$$\Sigma = \begin{pmatrix} \dfrac{q(1-q)}{\xi_1^2} & \dfrac{\sigma_{12}}{\xi_1 \xi_2} & \cdots & \dfrac{\sigma_{1k}}{\xi_1 \xi_k} \\ \vdots & \vdots & \cdots & \vdots \\ \dfrac{\sigma_{k1}}{\xi_k \xi_1} & \dfrac{\sigma_{k2}}{\xi_k \xi_2} & \cdots & \dfrac{q(1-q)}{\xi_k^2} \end{pmatrix}$$

with

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

is positive definite.

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

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

where $a_{1-\alpha} = \Gamma_\Sigma^{-1}(1-\alpha)$.

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

$$\Pr\left\{ |\widehat{\Theta}(\underline{\boldsymbol{X}}_{N^*}) - \Theta| \leq \delta \cdot \mathbf{1}_k \right\} \to 1 - \alpha.$$

Proof See Appendix A.3. □

The results above justify simultaneous fixed-width-δ CIs for k means or quantiles based on bootstrapping. They establish the asymptotic validity of our generic R&S procedure when we extend them to all pairs of difference estimates using the linear transformation $\mathbf{A}$ defined as

$$\mathbf{A} = [a_{ij}],\ i = 1,2,\ldots,\binom{k}{2};\ j = 1,2,\ldots,k \tag{9}$$

where

$$a_{ij} = \begin{cases} 1, & (j-1)\left(k-\frac{j}{2}\right)+1 \leqslant i \leqslant j\left(k-\frac{j+1}{2}\right);\ j = 1,2,\ldots,k-1 \\ -1, & i = hk-\frac{h(h+1)}{2}-(k-j);\ j = 2,3,\ldots,k;\ 1 \leqslant h \leqslant j-1 \\ 0, & \text{otherwise.} \end{cases}$$

The linear transformation $\mathbf{A}$ transforms vectors in $\mathbb{R}^k$ into vectors that correspond to all pairs of differences in $\mathbb{R}^{\binom{k}{2}}$. We are now prepared to state the asymptotic validity of our generic R&S procedures in Corollaries 2 and 3. The stopping time used in our procedures is

$$N^*_{\mathbf{A}} = \inf\left\{n \geqslant n_0 : \Pr^*\{|(\mathbf{A}\widehat{\Theta}(\underline{\mathbf{X}}^*_n) - \mathbf{A}\widehat{\Theta}(\underline{\mathbf{X}}_n)| \leqslant \delta \cdot \mathbf{1}_{\binom{k}{2}}\} \geqslant 1-\alpha\right\}. \tag{10}$$

Corollary 2. *Under the same assumptions as in Theorem 1, consider $N^*_{\mathbf{A}}$ as defined in (10).*

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

$$\delta^2 N^*_{\mathbf{A}} \to a^2_{1-\alpha} \quad a.s.$$

where $a_{1-\alpha} = \Gamma^{-1}_{\mathbf{A}\Sigma\mathbf{A}^\top}(1-\alpha)$ and Σ is defined as in Theorem 1.

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

$$\Pr\left\{|\mathbf{A}\widehat{\Theta}(\underline{\mathbf{X}}^*_{N^*_{\mathbf{A}}}) - \mathbf{A}\Theta| \leq \delta \cdot \mathbf{1}_{\binom{k}{2}}\right\} \to 1-\alpha.$$

Proof See Appendix A.4 □

Next let $\widehat{\Theta}$ be the sample quantiles defined in Theorem 2.

Corollary 3. *Under the same assumptions as in Theorem 2, consider $N_{\mathbf{A}}^*$ as defined in (10).*

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

$$\delta^2 N_{\mathbf{A}}^* \to a^2 \quad a.s.$$

where $a_{1-\alpha} = \Gamma^{-1}_{\mathbf{A}\Sigma\mathbf{A}^\top}(1-\alpha)$ and Σ is defined as in Theorem 2.

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

$$\Pr\left\{|\mathbf{A}\widehat{\Theta}(\underline{\mathbf{X}}^*_{N^*_{\mathbf{A}}}) - \mathbf{A}\Theta| \le \delta \cdot \mathbf{1}_{\binom{k}{2}}\right\} \to 1-\alpha.$$

Proof See Appendix A.5 □

The assumptions that support our theorems are comparatively weaker than those that support other $\delta \to 0$ analysis in R&S. For instance, Robbins et al. (1968) assumed that output data from each system are independent and have common, finite variance. Damerdji et al. (1996) also assumed that output data from each system are independent, and further that their variances are known, along with other conditions on the distribution mean, variance, and absolute centered moments. Similarly, Kim and Nelson (2006a) assumed that the output data from each system are mutually independent, and that the standardized partial sum of each system's output converges to a Brownian motion process.

Remark. *Notice that means (which includes probabilities), variances and quantiles cover much of what is used in practice. One possible extension of our asymptotic results is when the performance measure θ is a*

function, like a cost, associated with (say) the mean performance measure: $\theta = g(\mu)$*, and we can estimate* μ*. Then our procedure can still be shown to be asymptotically valid for continuously differentiable g using a delta-method argument.*

5 Jump-ahead Procedure

In the generic procedures described in Section 3, the number of observations N obtained from each simulated system is increased until the estimated PCS is at least $1-\alpha$. The sample-size increment on each iteration, Δn, can be made larger than 1 to speed up the procedure, and this is important because estimating the PCS is a non-trivial calculation for large k. In Lee and Nelson (2014) we suggested $\Delta n = 10$, but this could be too aggressive in some problems, and unnecessarily conservative (small) when the stopping time N^* is in the hundreds or thousands. Recall our primary goal is to provide a PCS guarantee, but we also want to attain it efficiently in the sense of keeping both the required sample size N^* and the computational overhead from bootstrapping small. Bootstrapping occurs after we collect each increment Δn of additional output data. When variances are unknown, the minimal number of stages or "jumps" that an indifference-zone R&S procedure can have is two, and this is only possible if we know the distributions of the outputs. Since we assume no distributional information, the best we can hope for is an approximation that keeps the number of jumps small.

Here we outline an adaptive method for selecting Δn that is aggressive when P^*_{NB} is far from $1-\alpha$, but cautious as we approach it. The idea is straightforward: From any intermediate observed $(N, P^*_{NB} < 1-\alpha)$ pair, fit a simplified normal-theory approximation for PCS as a function of N and use it to project what the stopping time N^* will be; call this projection $\widehat{N}^*$. Then set $\Delta n = c(\widehat{N}^* - N)$, where c is a fraction such as 0.8, to avoid overshoot. Now simulate the additional observations, estimate the new PCS and new projection,

and repeat until $P^*_{NB} \geq 1-\alpha$.

To simplify notation, let $\widehat{\theta}_i = \widehat{\theta}(\underline{X}_{iN})$ be the point estimator of θ_i, $i = 1,2,\ldots,k$, where the sample size N will be clear from the context. Our approximation will assume that $(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_k)$ are jointly normally distributed with means $(\theta_1, \theta_2, \ldots, \theta_k)$, common variance σ^2/N and common correlation ρ (which is 0 unless CRN are employed). We would like to approximate the sample size N^* for which

$$\Pr\{|\widehat{\theta}_i - \widehat{\theta}_j - (\theta_i - \theta_j)| \leq \delta, \forall i \neq j\} \geq 1-\alpha$$

by fitting this simplified model to the observed PCS.

Let $Z_{ij} = (\widehat{\theta}_i - \widehat{\theta}_j - (\theta_i - \theta_j))$. Then $\mathbf{Z} = \{Z_{ij}, \forall i \neq j\}$, appropriately organized and under our simplified model, has a multivariate normal distribution with mean vector 0 and variance-covariance matrix

$$\frac{\sigma^2(1-\rho)}{N}\mathbf{A}\mathbf{I}\mathbf{A}^\top$$

where $\mathbf{A}$ is as defined in (9) and $\mathbf{I}$ is the $k \times k$ identity matrix. Notice that only the term $\sigma^2(1-\rho)$ is unknown, since $\mathbf{A}$ and $\mathbf{I}$ depend only on k, and N is given. For $0 < \beta < 1$, let $h(\beta) = \Gamma^{-1}_{\mathbf{A}\mathbf{I}\mathbf{A}^\top}(\beta)$; $h(\beta)$ is the β quantile of the maximum of the absolute values of the components of a mean-zero multivariate normal distribution with covariance matrix $\mathbf{A}\mathbf{I}\mathbf{A}^\top$. Given β, $h(\beta)$ is also only a function of k. Then for any such β,

$$\Pr\left\{|Z_{ij}| \leq \sqrt{\frac{\sigma^2(1-\rho)}{N}}h(\beta), \forall i \neq j\right\} = \beta.$$

Suppose we have run our procedure for N observations and obtained estimated PCS $P^*_{NB} < 1-\alpha$. Recall that P^*_{NB} is also the estimated coverage probability; that is, for the current value of N,

$$\Pr\{|Z_{ij}| \leq \delta, \forall i \neq j\} \approx P^*_{NB}.$$

We can calibrate our approximation to this observed probability by selecting $\widehat{\sigma^2(1-\rho)}$ so that

$$\sqrt{\frac{\widehat{\sigma^2(1-\rho)}}{N}}h(P^*_{NB}) = \delta.$$

Here we are treating the entire term $\widehat{\sigma^2(1-\rho)}$ as a single parameter to match the observed PCS, and not worrying about separating σ^2 and ρ.

Given this parameter, the target sample size to obtain PCS of at least $1-\alpha$ satisfies

$$\widehat{N}^* \geq \left(\frac{\sqrt{\widehat{\sigma^2(1-\rho)}} \cdot h(1-\alpha)}{\delta} \right)^2 = N \cdot \left(\frac{h(1-\alpha)}{h(P^*_{NB})} \right)^2$$

and the $\widehat{\sigma^2(1-\rho)}$ terms cancel. The intuition is that the critical value $h(1-\alpha)^2$ is the scale-free sample-size multiplier to attain PCS $1-\alpha$. Thus, the ratio $(h(P^*_{NB})/h(1-\alpha))^2$ is the fraction of the curve we have climbed if the current sample size N corresponds to a PCS of only P^*_{NB}.

Unfortunately, calculating $h(\beta) = \Gamma^{-1}_{\mathbf{AIA}^\top}(\beta)$ is a difficult root-finding problem for large k, but we have straightforward empirical solution: Let $M = \max_{i \neq j} |Z'_{ij}|$, where the $\{Z'_{ij}, \forall i \neq j\}$ have a mean-zero multivariate normal distribution with variance $\mathbf{AIA}^\top$. Then $h(\beta)$ is the β quantile of M, whose distribution depends only on k. Therefore, from an i.i.d. Monte Carlo sample $M_1, M_2, \ldots, M_t$, $h(\beta)$ can be estimated by the order statistic $M_{(\lceil \beta t \rceil)}$, and only one sample is needed for any fixed k since $M_{(\lceil \beta t \rceil)}$ is a function of β once the data are sorted.

The jump-ahead approximation moves our procedure from theoretically interesting to practically useful because, as we illustrate in Section 6, it allows us to find N^* in a small number of jumps, thereby avoiding the computational overhead of repeatedly evaluating P^*_{NB}, while not overshooting and being inefficient.

6 Empirical Evaluation

In this section we present an empirical evaluation of the algorithms described in Section 3 combined with the jump-ahead procedure of the previous section. We consider $k = 2, 4, 10$ and 20 systems. For one set of results, the output data have a mix of exponential and normal marginal distributions, and the performance

measures include the mean and 0.8 quantile. For a second set of experiments the output data have negative binomial distributions and the performance measure is the mean.

In preliminary work, favorable empirical results for a single distribution family (normal or Poisson), θ being the mean or various quantiles, and k up to 10, were presented in Bekki et al. (2010) and Lee and Nelson (2014). Lee and Nelson (2014) also contains results when varying the initial sample size n_0, the indifference-zone parameter δ, and the number of bootstrap resamples B; they also compared the efficiency of the algorithm with Poisson outputs to procedures designed specifically for Poisson data. In all cases the algorithms performed very well. Therefore, we focus here on the mixed distribution case, on a second discrete case (both highly-skewed and less-skewed instances), larger numbers of systems k, and the use of the jump-ahead procedure. We show below and in Appendix A.6 that the algorithms achieve the desired PCS in 5–10 jumps provided n_0 is large enough. They can be more or less efficient than the normal-theory procedure of Rinott (1978), depending on the initial sample size, but they retain their PCS guarantee for nonnormal data while Rinott's procedure may not.

6.1 Selecting the Best Mean

All results presented here are averaged over 100 macro-replications of the entire experiment. Tables 1–3 contain results for selecting the system with the largest mean when using of CRN with induced correlation $\rho = 0.9$. We varied the initial sample size n_0, the number of systems k and the configuration of the means. The number of bootstrap samples is $B = 200$, the desired confidence level is $1 - \alpha = 0.95$ and the indifference-zone parameter is $\delta = 0.1$ for all experiments. We present only CRN results because we recommend using CRN whenever possible. Corresponding results without CRN are presented in Tables A.1–A.3, respectively, of Appendix A.6, and we compare them to the results with CRN at the end of this section.

In Tables 1–2, the simulation outputs have a mix of distributions; specifically, half of the systems have

normally distributed output and the other half have exponentially distributed output with the means either in a slippage configuration or a monotone decreasing means configuration. The true mean vector in Table 1 is the slippage configuration $\Theta = (5,5,\ldots,5,5+\delta)^\top$, so system k is the best. The true mean vector used in Table 2 has $\theta_i = 5+(i-1)\delta$ for $i = 1,2,\ldots,k$, so system k is again the best system. In the monotone decreasing means configuration, we alternated exponentially distributed systems and normally distributed systems; for example, when $k = 4$, systems 1 and 3 have normally distributed outputs, while systems 2 and 4 have exponentially distributed outputs; we then reverse the assignments. The normally distributed outputs have variance 1, while the exponentially distributed outputs have variance θ_i^2. In the tables, "n" or "e" indicate that the best system has normally or exponentially distributed outputs, respectively.

In Table 3 negative binomial output data are used; there are two parameters, the target number of successes r and the probability of success p; these imply that the mean is $\theta = r(1-p)/p$. We used $p = 0.5$ (therefore, $\theta = r$) for all experiments. The true mean vector used in Table 3 is the slippage configuration $(\theta,\theta,\ldots,\theta,\theta+\delta)^\top$ with $\theta = 1$ or 10 implying system k is the best system. The negative binomial distribution is highly skewed when $\theta = 1$ whereas it is less-skewed when $\theta = 10$.

To induce correlation representing the effect of CRN for the mixed-distribution case, we used the NORTA method described in Nelson (2013), inducing a common correlation of 0.9 between all pairs of systems. To represent CRN for negative binomial (NB) distributions, we set $X_i = Z + W_i$ where the W_i's are independent NB(β_i, p) for $i = 1,2,\ldots,k$, and Z is NB(β_0, p). Then the correlation between X_i and X_j for $i \neq j$ is

$$\mathrm{Corr}(X_i, X_j) = \frac{\beta_0}{\sqrt{\beta_0+\beta_i}\sqrt{\beta_0+\beta_j}}.$$

With $p = 0.5$, when $\theta_i = \theta_j = 1$, we have $\mathrm{Corr}(X_i, X_j) = \beta_0$; when $\theta_i = \theta_j = 10$, we have $\mathrm{Corr}(X_i, X_j) = \beta_0/10$. In Table 3, $\beta_0 = 0.9$ is used to induce a common correlation 0.9 when $\theta = 1$; similarly we set $\beta_0 = 9$ when $\theta = 10$.

The PCS is the fraction of the 100 macroreplications in which a system whose mean is within δ of the true best mean was selected. The estimated coverage probability is P^*_{NB} from Step 6 of the algorithm, and the true coverage probability is the fraction of the macroreplications in which the $\binom{k}{2}$ CIs simultaneously cover all pairwise differences $\theta_i - \theta_j$ for all $i \neq j$. The average number of jumps that the jump-ahead procedure generates until $P^*_{NB} \geq 1-\alpha$ is denoted by "Ave jumps".

We discovered quickly that P^*_{NB} is often 0 when N is small and k is large, such as when $N = n_0$ and $k \geq 10$, which provides no information to the jump-ahead procedure. Therefore, we adjusted the procedure to use $\max\{1/k, P^*_{NB}\}$, since pure guessing provides PCS $1/k$. Figure 1 shows the sample sizes generated by the jump-ahead procedure with this adjustment for a case with $k = 20$ systems in the slippage configuration and having a mix of distribution types; the left-hand figure is the result of one macro replication, while the right-hand figure is the result of 10 macro replications. Notice that the observed PCS values are zero initially, but after a couple of jumps the procedure is able to make rapid progress without significant overshoot. For the left-hand plot in Figure 1, each dot denotes a pair of sample size N (starting from $N = n_0$) and its corresponding estimated PCS P^*_{NB} at each jump; for this example, the procedure stops at the 7th jump as $P^*_{NB} \geq 0.95$, which results in N^*=50,824. Recall that in the best case with unknown variance R&S requires 2 jumps, and that fully sequential procedure uses $\Delta n = 1$ (i.e., N^* jumps). We observe that 5–10 jumps were made for $c = 0.8$ in all experiments; however, if the user wants to make the procedure less conservative (fewer jumps), the constant c can be set closer to 1.

The results from all tables show that the number of jumps and the required terminal sample size N^* increases as the number of systems increases, as expected. The desired PCS is achieved in all but one case, and the desired CI coverage is attained except in a few cases when $n_0 = 50$. It seems clear that using CRN implies that we should start with a larger initial sample size n_0. This makes sense because when we employ CRN we bootstrap entire vectors, rather than each marginal ecdf individually, and 50 vectors is a

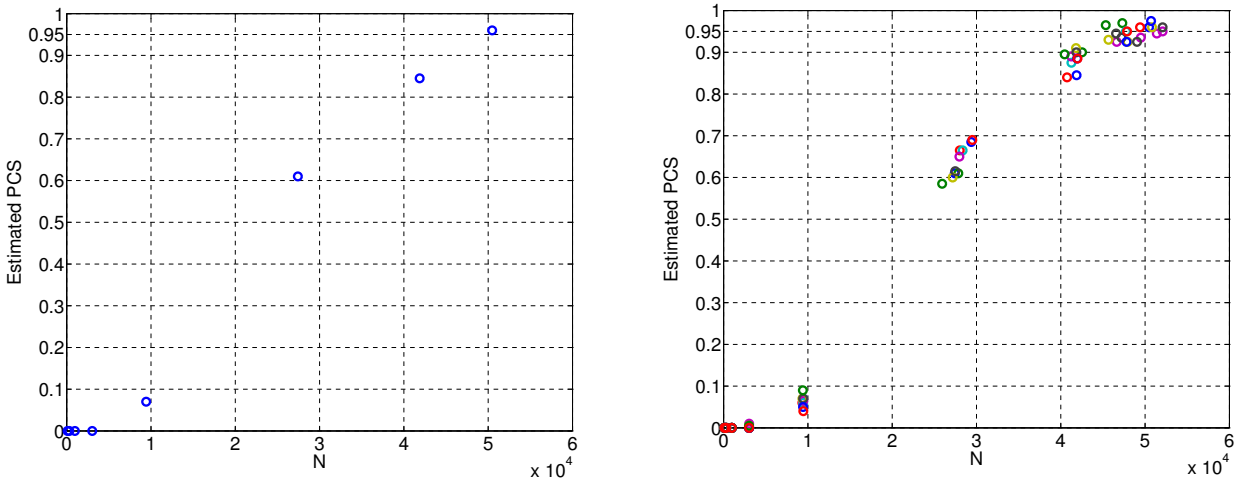

Figure 1: $\widehat{N}^*$ vs. estimated PCS for one macroreplication (left) and 10 macroreplications (right) of the jump-ahead procedure for $k = 20$ systems in the slippage configuration.

small number to represent a joint distribution, particularly when k is large. As we have observed in previous papers, n_0 should be larger for bootstrap R&S than for distribution-specific procedures because we need the bootstrap distribution, even at n_0, to be a good representation of the true distribution.

Notice that when the best system has exponentially distributed outputs the required sample size N^* tends to be larger than when the best system has normally distributed outputs because the variances of the exponentially distributed outputs are much higher than the normal case. The required sample size N^* for the monotone decreasing means configuration is larger than the slippage configurations for the same reason.

For both output distribution cases, the use of CRN greatly reduces the required sample size, with savings of as much as 90%. As a specific instance, in the case of half-normal and half-exponential distributions when $k = 20$ and the best system is exponentially distributed, the required sample size obtained from the algorithm with CRN is 11,437 (Table 1) while the required sample size without CRN is 50,257 (Table A.1 in Appendix

A.6); when $k = 2$, the required sample size obtained from the algorithm with CRN is 7466 (Table 1) while the required sample size without CRN is 11,061 (Table A.1 in Appendix A.6). For the case of negative binomial distributions, the sample size savings obtained from the use of CRN are all close to 90% (compare Tables 3 and A.3).

6.2 Selecting the Best 0.8 Quantile

Next we presents results for selecting the system with the largest 0.8-quantile, with or without CRN when we have a mix of output distributions. In all cases we used the slippage configuration.

The best 0.8 quantile was set as the 0.8 quantile of a normal distribution with mean $5+\delta$ and variance 1; that is, $\theta_k = 5+\delta+\Phi^{-1}(0.8)$. The inferior systems had 0.8 quantile $\theta_i = 5+\Phi^{-1}(0.8)$. When the distribution was exponential, we solved for the distribution's mean so that it had the desired value θ_k or θ_i for its 0.8 quantile. To incorporate CRN, we used the NORTA method cited earlier. The results with CRN (with induced correlation $\rho = 0.9$) are in Table 4, and are without CRN in Table A.4 of Appendix A.6.

The PCS values are all greater than or equal to 0.95, and the CI simultaneous coverage probabilities are mostly greater than or equal to 0.95. As expected, CRN greatly reduces the required sample size.

6.3 Comparison with Rinott's Procedure

In this section we consider four different output distributions, normal, Poisson, negative binomial, and a mix of half-normal & half-exponential distributions to compare the PCS and efficiency of our bootstrap R&S algorithm without CRN to Rinott's procedure. Rinott's procedure is a two-stage, normal-theory procedure for which the sample size from each system is proportional to its sample variance; it does not exploit CRN, but does guarantee the desired PCS when the output data are normally distributed.

The results in Table 5 show the required sample sizes denoted by N_B and N_R obtained from our bootstrap

Table 1: Empirical results for means from 100 macroreplications for half exponential and half normal distributions in slippage configuration with CRN ($\rho = 0.9$).

k	n_0	Ave N^*	PCS	Est Coverage	True Coverage	Ave Jumps
2(n)	100	7186	0.96	0.96	0.93	4.83
2(e)	100	7466	0.95	0.96	0.94	4.86
4(n)	100	8307	0.97	0.96	0.94	6.64
4(e)	100	8447	0.99	0.96	0.95	6.54
10(n)	100	9974	0.98	0.96	0.96	8.92
10(e)	100	9994	0.95	0.96	0.92	8.56
20(n)	100	11221	0.97	0.96	0.91	10.05
20(e)	100	11437	1.00	0.96	0.96	10.13

Table 2: Empirical results for means from 100 macroreplications for half exponential and half normal distributions in the monotone decreasing means configuration with CRN ($\rho = 0.9$).

k	n_0	Ave N^*	PCS	Est Coverage	True Coverage	Ave Jumps
4(n)	100	8647	0.98	0.96	0.95	6.53
4(e)	100	9079	1.00	0.96	0.97	6.55
10(n)	100	12086	1.00	0.96	0.96	8.84
10(e)	100	12575	1.00	0.96	0.93	8.98
20(n)	100	17745	0.99	0.96	0.96	10.82
20(e)	100	18421	1.00	0.96	0.95	10.80

Table 3: Empirical results for means from 100 macroreplications for negative binomial distributions in the slippage configuration with CRN ($\rho = 0.9$).

k	θ	β_0	Average N^*	PCS	Est. Coverage	True Coverage	Ave Jumps
2	1	0.9	240	0.94	0.97	0.93	3.21
2	10	9	1680	0.99	0.96	0.98	4.27
4	1	0.9	336	0.93	0.96	0.93	4.28
4	10	9	2815	1.00	0.96	0.94	5.36
10	1	0.9	482	1.00	0.96	0.98	5.12
10	10	9	4144	1.00	0.96	0.95	6.27
20	1	0.9	590	0.99	0.96	0.97	5.48
20	10	9	5147	0.99	0.96	0.95	7.02

Table 4: Empirical results for quantiles ($q = 0.8$) from 100 macroreplications for half exponential and half normal distributions in the slippage configuration with CRN ($\rho = 0.9$).

k	n_0	Ave N^*	PCS	Est. Coverage	True Coverage	Ave jumps
2(n)	100	16936	0.97	0.97	0.94	5.24
2(e)	100	17584	0.96	0.96	0.95	5.12
4(n)	100	22408	0.99	0.96	0.97	6.60
4(e)	100	22904	0.97	0.96	0.91	6.48
10(n)	100	30454	1.00	0.96	1.00	8.03
10(e)	100	30703	0.99	0.96	0.95	8.32
20(n)	100	38102	0.99	0.96	0.95	9.37
20(e)	100	38553	0.99	0.96	0.96	9.36

R&S algorithm and Rinott's procedure, respectively, and their corresponding estimated PCS values denoted by PCS_B and PCS_R for each output distribution. For this table, we consider $k = 10$ systems with the slippage configuration of the means $(5, 5, \ldots, 5, 5.1)$ for all distributions. The results show that the attained PCS values (0.33 and 0.27) from Rinott's procedure are far from the desired PCS 0.95 for half-normal and half-exponential distributions; the sample sizes 12,619 and 12,354 from Rinott's procedure are much smaller than the sample sizes 37,383 and 37,745 obtained from our procedure, showing that Rinott's procedure under-samples in this case.

For the case of negative binomial distributions, the attained PCS from Rinott's procedure is 0.83 when $n_0 = 10$, which is significantly under the desired PCS of 0.95, while our bootstrap R&S algorithms achieved the desired PCS for all cases.

Rinott's procedure is more sensitive to the initial sample size than our bootstrap R&S procedure; for example, when $n_0 = 10$ and 50, $N_R < N_B$ for all distributions, whereas $N_R > N_B$ when $n_0 = 5$, even when the output data are normal as assumed by Rinott's procedure. For instance, the sample size obtained from applying our procedures for normal distributions when $n_0 = 5$ is 2030, which is much smaller than the sample size 3494 from Rinott's procedure. Moreover, our procedure achieves the desired PCS in this case. This is because Rinott's procedure is a two-stage procedure, and hence with a small initial sample size in the first stage, the procedure becomes conservative in the second stage to guarantee the desired PCS. However, our jump-ahead procedure takes several stages to estimate the PCS, and stops as soon as it is achieved.

7 Illustration

This section illustrates using our procedure in a difficult R&S problem that Goldsman et al. (1991) employed to demonstrate methods for selecting the best system. The goal is to find the airline-reservation system

Table 5: Comparison results from 100 macroreplications for four different output distributions in the slippage configuration $(5,5,\ldots,5,5.1)$ without CRN for $k = 10$ systems; $1-\alpha = 0.95$.

Distributions	n_0	N_R	PCS_R	N_B	PCS_B
Normal	5	3494	1.00	2030	1.00
Normal	10	1835	0.94	2029	1.00
Normal	50	1364	0.97	2042	1.00
Poisson	5	17558	0.94	10129	1.00
Poisson	10	9280	0.97	10183	0.99
Poisson	50	6714	0.96	10130	1.00
Negative binomial	10	55916	0.83	61106	1.00
Negative binomial	50	41155	0.98	61315	1.00
Half-normal&half-exponential (n)	50	12619	0.33	37383	1.00
Half-normal&half-exponential (e)	50	12354	0.27	37745	0.99

Table 6: Empirical results from 100 macroreplications for the TTF example without and with CRN; the results with CRN are indicated by †.

$\boldsymbol{\mu}^\top$	n_0	Ave N^*	PCS	Est Coverage	True Coverage	Ave Jumps
(19,19,19,20)	50	6131	0.98	0.96	0.97	5.47
(19,19,19,20)	100	5992	0.98	0.96	0.96	5.11
(19,19,19,20)	200	6027	0.98	0.96	0.96	4.44
(19,19,19,20)†	50	122	0.73	0.98	0.91	2.35
(19,19,19,20)†	100	158	1.00	0.97	1.00	2.16
(19,19,19,20)†	200	233	1.00	0.98	1.00	1.56

with the largest expected time to failure (TTF). In the experimental setup, each system works if either of two computers works. Assuming that the two computers are identical, E[TTF] will be dependent on two parameters, a failure rate λ and a repair rate μ. In our experiment $k = 4$ airline-reservation systems are considered, $\lambda = 1$ for all systems but we vary the repair rates μ as shown in Table 6. Clearly the system with the fastest repair rate will be the best, but we apply the procedures as if we do not know this.

Note that computer failure is rare, repair times are fast, and hence the resulting E[TTF] is large. The replication outputs are highly variable and non-normal (closer to exponential), so much so that Goldsman et al. (1991) batched the data before applying the two-stage R&S procedure created for normally distributed output data from Rinott (1978). We apply our procedure directly without any pre-processing of the data.

Table 6 shows the results without and with CRN; the CRN cases are indicated by †. Notice that the desired PCS of 0.95 is attained, provided n_0 is not too small, and the required sample sizes are reduced significantly by using CRN.

8 Conclusions

In this paper we have provided essentially generic R&S procedures for computer simulation. Under very mild conditions on the simulation output data they provide asymptotically correct selection for all of the typical performance measures: means, probabilities, variances and quantiles. The distribution families of the simulation outputs need not be known, or even the same across systems, and there are only two versions of the procedure: without or with CRN.

We achieve this generality by using bootstrapping to estimate the PCS at the current sample size, and we sequentially increase the sample size until the bootstrap PCS is $\geq 1-\alpha$. Thus, the generality comes at the cost of data storage (we need to retain the output data to bootstrap it) and calculation of the estimated PCS. To mitigate the computational expense, particularly as the number of systems k increases, we provide a procedure that jumps to the required sample size N^* quickly.

To derive and prove the correctness of our procedures we exploit a connection between PCS and the coverage of simultaneous CIs for all pairs to differences. Since coverage of all pairwise differences is a more demanding objective than correct selection of the best (see, for instance, Hsu (1996)), there is clearly room to make the procedures more efficient. And while our procedure is sequential, it is not an eliminating procedure, meaning that all k systems receive N^* replications. This provides another opportunity for savings as reducing k reduces both the simulation and bootstrap computation effort required.

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 2014R1A1A3049955. Portions of this work were published in the *Proceedings of the 2014 Winter Simulation Conference* as Lee

and Nelson (2014). The authors thank the associate editor, and two referees for their feedback.

References

Aerts, M. and I. Gijbels (1993). A three-stage procedure based on bootstrap critical points. *Sequential Analysis 12*, 93–113.

Anscombe, F. J. (1952). Large-sample theory of sequential estimation. *Mathematical Proceedings of the Cambridge Philosophical Society 48*, 600–607.

Babu, G. J. and C. R. Rao (1988). Joint asymptotic distribution of marginal quantiles and quantile functions in samples from a multivariate population. *Journal of Multivariate Analysis 27*, 15–23.

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

Bekki, J. M., B. L. Nelson, and J. W. Fowler (2010). Bootstrapping-based fixed-width confidence intervals for ranking and selection. In B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan (Eds.), *Proceedings of the 2010 Winter Simulation Conference*, Piscataway, New Jersey, pp. 1024–1033. Institute of Electrical and Electronics Engineers, Inc.

Bergström, H. (1945). On the central limit theorem in the space R_k, $k > 1$. *Scandinavian Actuarial Journal 1–2*, 106–127.

Bergström, H. (1969). On the central limit theorem in R_k. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 14*(2), 113–126.

Bickel, P. J. and D. A. Freedman (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics 9*, 1196–1217.

Billingsley, P. (1995). *Probability and Measure* (3rd. ed.). New York: Wiley-Intersicence.

Chow, Y. S. and H. Teicher (1988). *Probability Theory*. New York: Springer.

Damerdji, H., P. W. Glynn, M. K. Nakayama, and J. R. Wilson (1996). Selecting the best system in transient simulations with variances known. In *Proceedings of the 28th conference on Winter simulation*, pp. 281–286. IEEE Computer Society.

Goldsman, D., B. L. Nelson, and B. Schmeiser (1991). Methods for selecting the best system. In B. L. Nelson, W. D. Kelton, and G. M. Clark (Eds.), *Proceedings of the 1991 Winter Simulation Conference*, Piscataway, New Jersey, pp. 177–186. Institute of Electrical and Electronics Engineers, Inc.

Gotze, F. (1991). On the rate of convergence in the multivariate CLT. *The Annals of Probability 19*, 724–739.

Gupta, S. S. and S. Panchapakesan (1979). *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations*, Volume 44. SIAM.

Hlávka, Z. (2003). Asymptotic properties of robust three-stage procedure based on bootstrap for *m*-estimator. *Journal of Statistical Planning and Inference 115*, 637–656.

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

Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods*. London: Chapman & Hall.

Kim, S.-H. (2013). Statistical ranking and selection. In *Encyclopedia of Operations Research and Management Science*, pp. 1459–1469. New York: Springer.

Kim, S.-H. and B. L. Nelson (2006a). On the asymptotic validity of fully sequential selection procedures for steady-state simulation. *Operations Research 54*(3), 475–488.

Kim, S.-H. and B. L. Nelson (2006b). Selecting the best system. In S. G. Henderson and B. L. Nelson (Eds.), *Handbooks in Operations Research and Management Science: Simulation*, Volume 13, Chapter 17, pp. 501–534. New York: Elsevier.

Lee, S. and B. L. Nelson (2014). Bootstrap ranking & selection revisited. In A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller (Eds.), *Proceedings of the 2014 Winter Simulation Conference*, Piscataway, New Jersey, pp. 3857–3868. Institute of Electrical and Electronics Engineers, Inc.

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

Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics*. New York: Springer.

Rinott, Y. (1978). On two-stage selection procedures and related probability-inequalities. *Communications in Statistics-Theory and methods 7*(8), 799–811.

Robbins, H., M. Sobel, and N. Starr (1968). A sequential procedure for selecting the largest of k means. *The Annals of Mathematical Statistics*, 88–92.

Sazonov, V. V. (1968). On the multi-dimensional central limit theorem. *Sankhyā: The Indian Journal of Statistics, Series A 30*, 181–204.

Swanepoel, J., J. V. Wyk, and J. Venter (1983). Fixed width confidence intervals based on bootstrap procedures. *Communications in Statistics–Sequential Analysis 2*, 289–310.

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.

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.

A Online Supplement for "General-Purpose Ranking & Selection for Computer Simulation"

Soonhui Lee[1] and Barry L. Nelson[2]

[1]shlee@unist.ac.kr, [2]nelsonb@northwestern.edu

[1]School of Business Administration

Ulsan National Institute of Science and Technology

Ulsan, REPUBLIC OF KOREA

[2]Department of Industrial Engineering and Management Sciences

Northwestern University

Evanston, IL 60208-3119, USA

This appendix provides the mathematical proofs of the main theorems and the experiment results of the algorithms in the paper.

A.1 Auxiliary Results

The following theorem in Bergström (1945) provides a Berry-Esseen type bound in $\mathbb{R}^k$.

Theorem A.1. *(Bergström (1945)) Let $F(\mathbf{x})$ be a distribution function in $\mathbb{R}^k$ with the mean values $\underline{\mu}$, finite third moments and the translated second order moments σ_{ij}, such that $Q = \sum_{i,j=1,\ldots,k} \sigma_{ij} t_i t_j$ is a positive definite form with the determinant Δ. Further let $\Phi_\Sigma(\mathbf{x})$ be a normal distribution function which has the mean values 0 and the same translated second order moments as $F(\mathbf{x})$. Then we have for all $\mathbf{x}$,*

$$\left| \Pr\left\{ n^{1/2}(\bar{\mathbf{X}}_n - \underline{\mu}) \leqslant \mathbf{x} \right\} - \Phi_\Sigma(\mathbf{x}) \right| \leqslant C(k)\rho n^{-1/2} \max_{i,j=1,\ldots,k} |[\Sigma^{-1}]_{(i,j)}|^{3/2}$$

where $C(k)$ is a constant only depending on k, $\rho = \sum_{i=1}^{k} \rho^{(i)}$ with $\rho^{(i)} = \int_{\mathbb{R}} |x_i - \mathrm{E}(X_i)|^3 dF_i(x)$, $\Sigma = \mathrm{E}(\mathbf{X} - \mathrm{E}\mathbf{X})(\mathbf{X} - \mathrm{E}\mathbf{X})^\top$.

This type of the uniform bound also holds on the probability of any measurable convex set in $\mathbb{R}^k$, as shown in Gotze (1991), Sazonov (1968), and Bergström (1969). Although they present different bounds, they imply the following which we state as Theorem A.2.

Theorem A.2. *For all $x \in \mathbb{R}^+$,*

$$\left| \Pr\left\{ n^{1/2} |\bar{\mathbf{X}}_n - \underline{\mu}| \leqslant x \cdot \mathbf{1}_k \right\} - \Gamma_\Sigma(x) \right| \leqslant C_b(k) \rho_F f_1(n) \tag{A.1}$$

where $C_b(k)$ is a constant only depending on k, ρ_F is a function of several moments of the distribution F, and $f_1(n)$ is a function depending only on n, which tends to zero as $n \to \infty$.

Note that ρ_F in Gotze (1991), Sazonov (1968), and Bergström (1969) is finite when the third moment of F is finite.

Theorem A.3. *Pólya's Theorem (Chow and Teicher (1988)). If X_n converges in distribution to X, and F_X is everywhere continuous, then*

$$\sup_x |F_{X_n}(x) - F_X(x)| \to 0.$$

Definition A.1. *Uniform continuity in probability of $\{Y_n\}$ (Anscombe (1952)). Given any small positive ε and η, there is a large ν and small positive c such that, for any $n > \nu$,*

$$\Pr\{|Y_{n'} - Y_n| < \varepsilon w_n \quad \text{for all } n' \text{ such that } |n - n'| < cn\} > 1 - \eta \tag{A.2}$$

where $\{w_n\}$ is a sequence of positive numbers.

Remark. *We use Definition A.1 with $w_n = n^{-1/2}$ throughout this paper.*

Lemma A.1. *Suppose that $\hat{\theta}_{1n}, \hat{\theta}_{2n}, \ldots, \hat{\theta}_{kn}$ are k estimators that are individually uniformly continuous in probability. Then $Y_n = \max_i \hat{\theta}_{in}$ and $Y'_n = \max_i |\hat{\theta}_{in}|$ are also uniformly continuous in probability.*

Proof If $\hat{\theta}_{1n}, \hat{\theta}_{2n}, \ldots, \hat{\theta}_{kn}$ are k estimators that are individually uniformly continuous in probability, then given ε and η, we can find ν_i and small positive c_i such that

$$\Pr\left\{|\hat{\theta}_{in'} - \hat{\theta}_{in}| < \varepsilon w_n \quad \text{for all } n' \text{ such that } |n - n'| < c_i n\right\} > 1 - \eta/(k+1)$$

for any $n > \nu_i$, $i = 1, \ldots, k$. Let $\nu = \max_i \nu_i$ and $c = \min_i c_i$. Then we have

$$\begin{aligned}
&\Pr\left\{|Y_{n'} - Y_n| < \varepsilon w_n \quad \text{for all } n' \text{ such that } |n - n'| < cn\right\} \\
&\geqslant \Pr\left\{|\hat{\theta}_{in'} - \hat{\theta}_{in}| < \varepsilon w_n \quad \text{for all } n' \text{ such that } |n - n'| < c_i n, \quad i = 1, 2, \ldots, k\right\} \\
&\geqslant 1 - k\eta/(k+1) > 1 - \eta.
\end{aligned}$$

The first inequality is easily verified using the definition of Y_n and the second inequality holds by the Bonferroni inequality. Therefore, Y_n is uniformly continuous in probability. Now let ν_i, ν, c_i, and c be defined as above. Then we have

$$\begin{aligned}
&\Pr\left\{|Y'_{n'} - Y'_n| < \varepsilon w_n \quad \text{for all } n' \text{ such that } |n - n'| < cn\right\} \\
&\geqslant \Pr\left\{\left||\hat{\theta}_{in'}| - |\hat{\theta}_{in}|\right| < \varepsilon w_n \quad \text{for all } n' \text{ such that } |n - n'| < c_i n, \quad i = 1, 2, \ldots, k\right\} \\
&\geqslant \Pr\left\{\left|\hat{\theta}_{in'} - \hat{\theta}_{in}\right| < \varepsilon w_n \quad \text{for all } n' \text{ such that } |n - n'| < c_i n, \quad i = 1, 2, \ldots, k\right\} \\
&\geqslant 1 - k\eta/(k+1) > 1 - \eta.
\end{aligned}$$

□

Remark. *Anscombe (1952) shows that $\hat{\theta}_{in}$ is individually uniformly continuous in probability when $\hat{\theta}_{in}$ is a sample quantile or average.*

Proposition A.1. *Under assumptions stated in Theorem 2, given $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, as $n \to \infty$,*

$$\sup_{\mathbf{y}} |\Pr^*\{\mathbf{G}_n(q) \leq \mathbf{y}\} - \Phi_\Sigma(\mathbf{y})| \to 0$$

where

$$\mathbf{G}_n(q) = \left(\sqrt{n}(\widehat{F}_{1n}^{*-1}(q) - \widehat{F}_{1n}^{-1}(q)), \ldots, \sqrt{n}(\widehat{F}_{kn}^{*-1}(q) - \widehat{F}_{kn}^{-1}(q))\right)^\top$$

and Σ *is defined as in Theorem 2.*

Proof We first consider the univariate case and extend it to the multivariate case. WLOG, suppose that

$$\frac{r-1}{n} < q \leqslant \frac{r}{n}$$

which implies that $r = \lceil nq \rceil$. To simplify notation, let $F_n = \widehat{F}_n$ and $F_n^* = \widehat{F}_n^*$. For the univariate case, given $X_1, X_2, \ldots, X_n$ ($\in \mathbb{R}$), we want to show

$$\Pr^*\left\{\sqrt{n}(F_n^{*-1}(q) - F_n{}^{-1}(q)) \leqslant y\right\} \to \Phi_\Sigma(y) \tag{A.3}$$

where $\Sigma = q(1-q)/(f(F^{-1}(q)))^2$. Note we have

$$\Pr^*\left\{\sqrt{n}(F_n^{*-1}(q) - F_n{}^{-1}(q)) \leqslant y\right\} = \Pr^*\left\{F_n^{*-1}(q) \leq \frac{y}{\sqrt{n}} + F_n^{-1}(q)\right\}$$

and let

$$x_n = \frac{y}{\sqrt{n}} + F_n^{-1}(q).$$

On the other hand,

$$\begin{aligned} \Pr^*\left\{F_n^{*-1}(q) \leqslant x\right\} &= \Pr^*\left\{\sum_{i=1}^n \mathbf{I}\{X_i^* \leqslant x\} \geqslant r\right\} \\ &= \Pr^*\left\{\sqrt{n}\,(F_n^*(x) - F_n(x)) \geqslant \sqrt{n}\left(\frac{r}{n} - F_n(x)\right)\right\}. \end{aligned} \tag{A.4}$$

It can be shown from Theorem 2.1 in Bickel and Freedman (1981) that $\sqrt{n}\,(F_n^*(x) - F_n(x))$ converges in distribution to $N(0, \Sigma')$ as $n \to \infty$ where $\Sigma' = F(x)(1-F(x))$. After plugging $x = x_n$ into (A.4), we have

$$\Pr^*\left\{F_n^{*-1}(q) \leqslant x_n\right\} = \Pr^*\left\{\sqrt{n}\,(F_n^*(x_n) - F_n(x_n)) \geqslant \sqrt{n}\left(\frac{r}{n} - F_n(x_n)\right)\right\}$$

where

$$\sqrt{n}\left(\frac{r}{n}-F_n(x_n)\right)=\frac{\sqrt{n}\left(F_n(F_n^{-1}(q)))-F_n(\frac{y}{\sqrt{n}}+F_n^{-1}(q))\right)y}{y}$$

noting that $F_n(F_n^{-1}(q))=r/n$. In Reiss (1989), it has been shown that

$$\frac{F_n(F_n^{-1}(q))-F_n(\frac{y}{\sqrt{n}}+F_n^{-1}(q))}{\frac{y}{\sqrt{n}}}\to -f(F^{-1}(q))$$

a.s. for any $y\neq 0$ ($\in\mathbb{R}$). As $n\to\infty$, $F(x_n)(1-F(x_n))$ converges to $q(1-q)$ a.s. This completes the proof for the univariate case. Note that (A.3) has been presented by Bickel and Freedman (1981).

We now extend this result to the multivariate case. Let

$$\mathbf{Y}_j=\left(\mathbf{I}_{\{X_{1j}\leq x_1\}},\ldots,\mathbf{I}_{\{X_{kj}\leq x_k\}}\right)^\top \tag{A.5}$$

and

$$\mathbf{Y}_j^*=\left(\mathbf{I}_{\{X_{1j}^*\leq x_1\}},\ldots,\mathbf{I}_{\{X_{kj}^*\leq x_k\}}\right)^\top$$

defined for $\mathbf{x}\in\mathbb{R}^k$ such that $\mathrm{Var}(\mathbf{Y}_j)$ is positive definite, which implies that $x_i, i=1,2,\ldots,k$ are not all too small or large. By defining $Z_j=\underline{\lambda}^\top\mathbf{Y}_j$ for non-zero $\underline{\lambda}\in\mathbb{R}^k$, we have $\underline{\lambda}^\top\sqrt{n}(\bar{\mathbf{Y}}^*-\bar{\mathbf{Y}})=\sqrt{n}(\bar{Z}^*-\bar{Z})$ where we can apply Theorem 2.1 in Bickel and Freedman (1981). Therefore, $\sqrt{n}(\bar{Z}^*-\bar{Z})$ converges in distribution to $N(0,\underline{\lambda}^\top\boldsymbol{\tau}_{\mathbf{x}}\underline{\lambda})$ where $\tau_{\mathbf{x}}(i,j)=F_{ij}(x_i,x_j)-F_i(x_i)F_j(x_j)$. As this holds for each $\underline{\lambda}\in\mathbb{R}^k$, $\sqrt{n}(\bar{\mathbf{Y}}^*-\bar{\mathbf{Y}})$ converges in distribution to $N_k(0,\boldsymbol{\tau}_{\mathbf{x}})$ by the Cramér-Wold theorem (Billingsley (1995)) as $n\to\infty$. Let

$$\mathbf{H}_n(\mathbf{x})=(\sqrt{n}(F_{1n}^*(x_1)-F_{1n}(x_1)),\ldots,\sqrt{n}(F_{kn}^*(x_k)-F_{kn}(x_k)))^\top,$$

$$\mathbf{t}_n(\mathbf{x})=\sqrt{n}\left(\frac{r}{n}-F_{1n}(x_1),\ldots,\frac{r}{n}-F_{kn}(x_k)\right)^\top,$$

$\mathbf{y}=(y_1,\ldots,y_k)\in\mathbb{R}^k$ and $\mathbf{x}=(x_1,\ldots,x_k)\in\mathbb{R}^k$. Then by the result shown above, we know that $\mathbf{H}_n(\mathbf{x})$ converges in distribution to $N_k(0,\boldsymbol{\tau}_{\mathbf{x}})$ as $n\to\infty$. From this definition, note that we have

$$\Pr^*\{\mathbf{G}_n(q)\leq\mathbf{y}\}=\Pr^*\{\mathbf{H}_n(\mathbf{x}_n)\geq\mathbf{t}(\mathbf{x}_n)\}$$

where $\mathbf{x}_n = (y_1/\sqrt{n} + F_{1n}^{-1}(q), \ldots, y_k/\sqrt{n} + F_{kn}^{-1}(q))^\top$. Let $\tilde{\mathbf{x}} = (F_1^{-1}(q), \ldots, F_k^{-1}(q))^\top$ and $\mathbf{Y}_j$ in (A.5) defined on $\tilde{\mathbf{x}}$. Then its covariance matrix is $\boldsymbol{\tau}_{\tilde{\mathbf{x}}}$ where $\boldsymbol{\tau}_{\tilde{\mathbf{x}}}(i,j) = F_{ij}(F_i^{-1}(q), F_j^{-1}(q)) - F_i(F_i^{-1}(q))F_j(F_j^{-1}(q)) = F_{ij}(\theta_i, \theta_j) - q^2 = \sigma_{ij}$ as defined in Theorem 2. Therefore, as $n \to \infty$, $\mathbf{H}_n(\mathbf{x}_n)$ converges in distribution to $N_k(0, \boldsymbol{\tau}_{\tilde{\mathbf{x}}})$ as $\boldsymbol{\tau}_{\mathbf{x}_n}$ converges to $\boldsymbol{\tau}_{\tilde{\mathbf{x}}}$ a.s. And as $n \to \infty$, $\mathbf{t}(\mathbf{x}_n)$ converges to $(-f_1(F_1^{-1}(q))y_1$ $\ldots, -f_k(F_k^{-1}(q))y_k)^\top$ a.s. since marginal convergence implies joint convergence in this case. Now it follows that as $n \to \infty$,

$$\Pr^*\{\mathbf{G}_n(q) \le \mathbf{y}\} \to \Phi_\Sigma(\mathbf{y}).$$

Thus as $n \to \infty$, we have

$$\sup_{\mathbf{y}} |\Pr^*\{\mathbf{G}_n(q) \le \mathbf{y}\} - \Phi_\Sigma(\mathbf{y})| \to 0$$

by Theorem A.3. □

Lemma A.2. *Let $F_{\mathbf{X}_k}$ be the cdf of $\mathbf{X}_k$ where $\mathbf{X}_k = (X_1, X_2, \ldots, X_k)^\top$ is a random vector in $\mathbb{R}^k$; for $\mathbf{x}_k = (x_1, x_2, \ldots, x_k)^\top$ and $x_{k+1} \in \mathbb{R}$, let*

$$F_{\mathbf{X}_k}(\mathbf{x}_k) = \Pr\{\mathbf{X}_k \le \mathbf{x}_k\}$$

and

$$F_{\mathbf{X}_k, X_{k+1}}(\mathbf{x}_{k+1}) = \Pr\{\mathbf{X}_k \le \mathbf{x}_k, X_{k+1} \le x_{k+1}\}.$$

Let $\mathbf{L} = (x_{11}, x_{21}, \ldots, x_{k1})^\top$ and $\mathbf{U} = (x_{12}, x_{22}, \ldots, x_{k2})^\top$. Then for $x_{i1} < x_{i2}$, $i = 1, 2, \ldots, k$,

$$\Pr\{\mathbf{L} < \mathbf{X}_k \le \mathbf{U}\} = \sum_{i_1=1}^{2} \cdots \sum_{i_k=1}^{2} (-1)^{i_1 + \cdots + i_k} F_{\mathbf{X}_k}(x_{1i_1}, x_{2i_2}, \ldots, x_{ki_k}). \tag{A.6}$$

Proof For $k = 1$, $\Pr\{x_{11} < X_1 \le x_{12}\} = F_{X_1}(x_{12}) - F_{X_1}(x_{11})$. If (A.6) holds for $k = t$, we show that it holds for $k = t + 1$ in the following:

$$\Pr\{(x_{11}, \ldots, x_{t1})^\top < (X_1, \ldots, X_t)^\top \le (x_{12}, \ldots, x_{t2})^\top, x_{t+1,1} < X_{t+1} \le x_{t+1,2}\}$$

$$
\begin{aligned}
&= \sum_{i_1=1}^{2} \cdots \sum_{i_t=1}^{2} (-1)^{i_1+\cdots+i_t} F_{\mathbf{X}_t, X_{t+1}}(x_{1i_1}, \ldots, x_{ti_t}, x_{t+1,2}) \\
&\quad - \sum_{i_1=1}^{2} \cdots \sum_{i_t=1}^{2} (-1)^{i_1+\cdots+i_t} F_{\mathbf{X}_t, X_{t+1}}(x_{1i_1}, \ldots, x_{ti_t}, x_{t+1,1}) \\
&= \sum_{i_1=1}^{2} \cdots \sum_{i_t=1}^{2} \sum_{i_{t+1}=1}^{2} (-1)^{i_1+\cdots+i_t+i_{t+1}} F_{\mathbf{X}_t, X_{t+1}}(x_{1i_1}, \ldots, x_{ti_t}, x_{t+1,i_{t+1}}).
\end{aligned}
$$

□

Note that when $\mathbf{X}$ is a continuous random vector, $\Pr\{\mathbf{L} < \mathbf{X} \leq \mathbf{U}\} = \Pr\{\mathbf{L} \leq \mathbf{X} \leq \mathbf{U}\}$. If $\mathbf{X}$ is discrete valued, then there exists a vector $\boldsymbol{\varepsilon} > 0$ small enough that $\Pr\{\mathbf{L} \leq \mathbf{X} \leq \mathbf{U}\} = \Pr\{\mathbf{L} - \boldsymbol{\varepsilon} < \mathbf{X} \leq \mathbf{U}\}$. Then we can apply (A.6) with $\mathbf{L}' = \mathbf{L} - \boldsymbol{\varepsilon}$ replacing $\mathbf{L}$. Therefore, for both cases, $\Pr\{\mathbf{L} \leq \mathbf{X} \leq \mathbf{U}\}$ can be written as a linear combination of a finite number of orthant probabilities.

A.2 Proof of Theorem 1

Part (**a**): It follows from Theorem A.2 for all $x \in \mathbb{R}^+$ that

$$
|\Pr^*\{n^{1/2}(\bar{\mathbf{X}}_n^* - \bar{\mathbf{X}}_n) \leq x \cdot \mathbf{1}_k\} - \Gamma_{\Sigma_n}(x)| \quad \leqslant \quad C_b(k)\rho_{F_n} f_1(n) \tag{A.7}
$$

where $C_b(k)$ is a constant only depending on k, $\Sigma_n = \sum_{j=1}^{n}(\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^\top / n$, and $\rho_{F_n} < \infty$ when $\mathrm{E}^*|\bar{\mathbf{X}}_n^* - \bar{\mathbf{X}}_n|^3 < \infty$ with large enough n. Let P_n^* denote the probability used in the bootstrap stopping variable as follows,

$$
P_n^* = \Pr^*\{|\bar{\mathbf{X}}_n^* - \bar{\mathbf{X}}_n| \leqslant \delta \cdot \mathbf{1}_k\}.
$$

Applying (A.7) with $x = \delta n^{1/2}$ we obtain

$$
P_n^* \quad = \quad \Gamma_{\Sigma_n}(\delta n^{1/2}) + o(1) \tag{A.8}
$$

a.s. as $n \to \infty$. The inequalities $P^*_{N^*} \geqslant 1-\alpha$ and $P^*_{N^*-1} < 1-\alpha$ hold with the bootstrap stopping variable N^* given by

$$N^* \quad = \quad \inf\{n \geqslant n_0 : P^*_n \geqslant 1-\alpha\}. \tag{A.9}$$

Note that as $\delta \to 0$, we have $N^* \to \infty$ since $N^* \geqslant n_0$ and n_0 grows as $\delta \to 0$. It follows from (A.8), (A.9), and the fact that $N^* \to \infty$ as $\delta \to 0$ that

$$\begin{aligned} P^*_{N^*} &= \Gamma_{\Sigma_{N^*}}(\delta {N^*}^{1/2}) + o(1) \geqslant 1-\alpha \\ P^*_{N^*-1} &= \Gamma_{\Sigma_{N^*-1}}(\delta (N^*-1)^{1/2}) + o(1) < 1-\alpha \end{aligned}$$

a.s. as $\delta \to 0$; equivalently,

$$\delta {N^*}^{1/2} \quad \geqslant \quad \Gamma^{-1}_{\Sigma_{N^*}}(1-\alpha+o(1)), \tag{A.10}$$

$$\delta (N^*-1)^{1/2} \quad < \quad \Gamma^{-1}_{\Sigma_{N^*-1}}(1-\alpha+o(1)). \tag{A.11}$$

Since $\Sigma_{N^*} \to \Sigma$ as $\delta \to 0$, it follows from (A.10), (A.11) and the continuous mapping theorem (Billingsley (1995)) that we have

$$\delta {N^*}^{1/2} \to a_{1-\alpha} = \Gamma^{-1}_{\Sigma}(1-\alpha) \quad a.s.$$

as $\delta \to 0$.

Part(**b**): Under our assumptions about $\mathbf{X}$, as $n \to \infty$, we have

$$\Pr\left\{ n^{1/2}(\bar{\mathbf{X}}_n - \Theta) \leqslant x \cdot \mathbf{1}_k \right\} \to \Phi_\Sigma(x \cdot \mathbf{1}_k)$$

from the multivariate central limit theorem. Let $Y_n = \max_i |\bar{X}_{in} - \theta_i|$. Then by continuous mapping theorem, we have

$$n^{1/2} Y_n \to Z \text{ in distribution} \tag{A.12}$$

where $Z = \max\{|Z_1|, \ldots, |Z_k|\}$ and the distribution of $\mathbf{Z} = (Z_1, \ldots, Z_k)^\top$ is Φ_Σ. As $\delta \to 0$, we are interested in $\Pr\{Y_{N^*} \leqslant \delta\}$ and show that it can be stated in the form of Theorem 1 of Anscombe (1952) as follows:

$$\begin{aligned} \Pr\{Y_{N^*} \leqslant \delta\} &= \Pr\left\{ \sqrt{\frac{a_{1-\alpha}^2}{\delta^2}} \cdot Y_{N^*} \leq \sqrt{\frac{a_{1-\alpha}^2}{\delta^2}} \cdot \delta \right\} \\ &= \Pr\left\{ \frac{a_{1-\alpha}}{\delta} \cdot Y_{N^*} \leq a_{1-\alpha} \right\} \to \Gamma_\Sigma(a_{1-\alpha}) \end{aligned}$$

where $\lceil a_{1-\alpha}^2/\delta^2 \rceil = n_r$, $N^* = N_r$, and $w_{n_r} = \delta/a_{1-\alpha}$ in Anscombe's result. Then it follows from Lemma A.1 and (A.12) that

$$\Pr\left\{ |\bar{\mathbf{X}}_{N^*} - \Theta| \leqslant \delta \cdot \mathbf{1}_k \right\} \to 1 - \alpha \tag{A.13}$$

as $\delta \downarrow 0$. □

A.3 Proof of Theorem 2

Part (a): Let P_n^* denote the probability used in the bootstrap stopping variable as follows,

$$P_n^* = \Pr^*\{|\widehat{\Theta}(\underline{\mathbf{X}}_n^*) - \widehat{\Theta}(\underline{\mathbf{X}}_n)| \leqslant \delta \cdot \mathbf{1}_k\}.$$

where $\widehat{\Theta}(\underline{\mathbf{X}}_n^*)$ and $\widehat{\Theta}(\underline{\mathbf{X}}_n)$ are the sample qth quantiles based on $\underline{\mathbf{X}}_n^*$ and $\underline{\mathbf{X}}_n$, respectively. Analogous to the Proof of Theorem 1, it follows from Lemma A.2 and Proposition A.1 that

$$\sup_x |\Pr^*(\mathbf{G}_n(q) \leqslant x \cdot \mathbf{1}_k) - \Phi_\Sigma(x \cdot \mathbf{1}_k)| \to 0.$$

By letting $x = \delta n^{1/2}$, we obtain

$$P_n^* = \Gamma_\Sigma(\delta n^{1/2}) + o(1) \tag{A.14}$$

a.s. as $n \to \infty$. From here on, the argument follows that of Theorem 1 (a). The inequalities $P_{N^*}^* \geqslant 1 - \alpha$ and $P_{N^*-1}^* < 1 - \alpha$ hold with the bootstrap stopping variable N^* given by

$$N^* = \inf\{n \geqslant n_0 : P_n^* \geqslant 1 - \alpha\}. \tag{A.15}$$

It follows from (A.14), (A.15), and the fact that $N^* \to \infty$ as $\delta \to 0$ that

$$\begin{aligned} P^*_{N^*} &= \Gamma_\Sigma(\delta {N^*}^{1/2}) + o(1) \geqslant 1-\alpha \\ P^*_{N^*-1} &= \Gamma_\Sigma(\delta (N^*-1)^{1/2}) + o(1) < 1-\alpha \end{aligned}$$

a.s. as $\delta \to 0$; equivalently,

$$\delta {N^*}^{1/2} \geqslant \Gamma_\Sigma^{-1}(1-\alpha+o(1)), \tag{A.16}$$

$$\delta (N^*-1)^{1/2} < \Gamma_\Sigma^{-1}(1-\alpha+o(1)). \tag{A.17}$$

As $\delta \to 0$, it follows from (A.16), (A.17) that

$$\delta {N^*}^{1/2} \to a_{1-\alpha} = \Gamma_\Sigma^{-1}(1-\alpha) \quad a.s.$$

as $\delta \to 0$.

Part (b): As $n \to \infty$, we have

$$\Pr\{n^{1/2}(\widehat{\Theta}(\underline{\mathbf{X}}_n) - \Theta) \leq x \cdot \mathbf{1}_k\} \to \Phi_\Sigma(x \cdot \mathbf{1}_k),$$

as shown in Babu and Rao (1988). Let $Y_n = \max_i |\widehat{\theta}(\underline{X}_{in}) - \theta_i|$. Then by the continuous mapping theorem, we have

$$n^{1/2} Y_n \to Z \text{ in distribution} \tag{A.18}$$

where $Z = \max\{|Z_1|, \ldots, |Z_k|\}$ and the distribution function of $\mathbf{Z} = (Z_1, \ldots, Z_k)^\top$ is Φ_Σ. Then using the same argument as in the proof of Part (b) of Theorem 1, we have

$$\Pr\{Y_{N^*} \leqslant \delta\} \to \Gamma_\Sigma(a_{1-\alpha})$$

as $\delta \downarrow 0$. Then it follows from Lemma A.1 and (A.18) that

$$\Pr\left\{|\widehat{\Theta}(\underline{\mathbf{X}}_{N^*}) - \Theta| \leqslant \delta \cdot \mathbf{1}_k\right\} \to 1-\alpha \tag{A.19}$$

as $\delta \downarrow 0$. □

A.4 Proof of Corollary 2

Let $\mathbf{Y}_j = \mathbf{A}\mathbf{X}_j$ and P_n^* denote the probability used in the bootstrap stopping variable in (10) as follows,

$$P_n^* = \Pr^*\left\{|\mathbf{A}\bar{\mathbf{X}}_n^* - \mathbf{A}\bar{\mathbf{X}}_n| \leqslant \delta \cdot \mathbf{1}_{\binom{k}{2}}\right\} = \Pr^*\left\{|\bar{\mathbf{Y}}_n^* - \bar{\mathbf{Y}}_n| \leqslant \delta \cdot \mathbf{1}_{\binom{k}{2}}\right\}.$$

It follows from Theorem A.2 for all $x \in \mathbb{R}^+$ that

$$\left|\Pr^*\left\{n^{1/2}(\bar{\mathbf{Y}}_n^* - \bar{\mathbf{Y}}_n) \leq x \cdot \mathbf{1}_{\binom{k}{2}}\right\} - \Gamma_{\mathbf{A}\Sigma_n\mathbf{A}^\top}(x)\right| \quad \leqslant \quad C_b(k)\rho_{F_n} f_1(n) \tag{A.20}$$

where $C_b(k)$ is a constant only depending on k, $\Sigma_n = \sum_{j=1}^n (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^\top / n$, and $\rho_{F_n} < \infty$ when $\mathrm{E}^*|\bar{\mathbf{Y}}_n^* - \bar{\mathbf{Y}}_n|^3 < \infty$ with large enough n. Applying (A.20) with $x = \delta n^{1/2}$ we obtain

$$P_n^* \quad = \quad \Gamma_{\mathbf{A}\Sigma_n\mathbf{A}^\top}(\delta n^{1/2}) + o(1)$$

a.s. as $n \to \infty$. The inequalities $P_{N_\mathbf{A}^*}^* \geqslant 1-\alpha$ and $P_{N_\mathbf{A}^*-1}^* < 1-\alpha$ hold with the bootstrap stopping variable $N_\mathbf{A}^*$ given by

$$N_\mathbf{A}^* \quad = \quad \inf\{n \geqslant n_0 : P_n^* \geqslant 1-\alpha\}.$$

Using the same arguments as in the proof of Theorem 1 (a) and (b), we have

$$\delta^2 N_\mathbf{A}^* \to a_{1-\alpha}^2 \quad a.s.$$

where $a_{1-\alpha} = \Gamma^{-1}_{\mathbf{A}\Sigma\mathbf{A}^\top}(1-\alpha)$ and

$$\Pr\left\{|\mathbf{A}\widehat{\Theta}(\underline{\mathbf{X}}_{N_A^*}^*) - \mathbf{A}\Theta| \leq \delta \cdot \mathbf{1}_{\binom{k}{2}}\right\} \to 1-\alpha$$

as $\delta \downarrow 0$. □

A.5 Proof of Corollary 3

Proof As shown in Proposition A.1, we have

$$\Pr^*\{n^{1/2}(\widehat{\Theta}(\underline{\mathbf{X}}_n^*) - \widehat{\Theta}(\underline{\mathbf{X}}_n)) \leqslant \mathbf{x}\} \to \Phi_\Sigma(\mathbf{x})$$

for all continuity points $\mathbf{x}$ as $n \to \infty$. It follows that

$$\Pr^*\{n^{1/2}(\mathbf{A}\widehat{\Theta}(\underline{\mathbf{X}}_n^*) - \mathbf{A}\widehat{\Theta}(\underline{\mathbf{X}}_n)) \leqslant \mathbf{x}\} \to \Phi_{\mathbf{A}\Sigma\mathbf{A}^\top}(\mathbf{x})$$

by the continuous mapping theorem. Therefore, as $n \to \infty$

$$\sup_{\mathbf{x}} |\Pr^*\{\mathbf{A}\widehat{\Theta}(\underline{\mathbf{X}}_n^*) - \mathbf{A}\widehat{\Theta}(\underline{\mathbf{X}}_n) \leq \mathbf{x}\} - \Phi_{\mathbf{A}\Sigma\mathbf{A}^\top}(\mathbf{x})| \to 0$$

by Theorem A.3. We redefine P_n^* as follows,

$$P_n^* = \Pr^*\left\{|\mathbf{A}\widehat{\Theta}(\underline{\mathbf{X}}_n^*) - \mathbf{A}\widehat{\Theta}(\underline{\mathbf{X}}_n)| \leqslant \delta \cdot \mathbf{1}_{\binom{k}{2}}\right\}.$$

Using the same arguments as in the proof of Theorem 2 (a) and (b), we have

$$\delta^2 N_{\mathbf{A}}^* \to a_{1-\alpha}^2 \quad a.s.$$

where $a_{1-\alpha} = \Gamma^{-1}_{\mathbf{A}\Sigma\mathbf{A}^\top}(1-\alpha)$ and

$$\Pr\left\{|\mathbf{A}\widehat{\Theta}(\underline{\mathbf{X}}_{N_{\mathbf{A}}^*}^*) - \mathbf{A}\Theta| \leq \delta \cdot \mathbf{1}_{\binom{k}{2}}\right\} \to 1-\alpha$$

as $\delta \downarrow 0$. □

A.6 Results for Selecting the Best Mean or the Best 0.8 quantile without CRN

This section contains the experiment results of the bootstrap R&S procedure without CRN combined with the jump-ahead procedure, which have been referred in Section 6. The corresponding results with CRN are presented in Tables 1–4 of Section 6.

Table A.1: Empirical results for means from 100 macroreplications for half exponential and half normal distributions in the slippage configuration without CRN.

k	n_0	Ave N^*	PCS	Est Coverage	True Coverage	Ave Jumps
2(n)	50	10550	0.99	0.96	0.92	5.18
2(n)	100	10306	0.94	0.96	0.92	5.11
2(e)	50	10971	1.00	0.97	0.99	5.22
2(e)	100	11061	1.00	0.96	0.99	5.27
4(n)	50	20138	1.00	0.96	0.89	7.11
4(n)	100	20501	1.00	0.96	0.94	6.88
4(e)	50	20752	0.97	0.96	0.93	7.31
4(e)	100	20723	0.97	0.96	0.95	6.94
10(n)	50	37383	1.00	0.96	0.98	9.03
10(n)	100	37510	1.00	0.96	0.94	8.62
10(e)	50	37745	0.99	0.96	0.96	8.91
10(e)	100	37717	1.00	0.96	0.95	8.58
20(n)	50	50217	1.00	0.96	0.93	10.2
20(n)	100	50029	1.00	0.96	0.97	9.21
20(e)	100	50257	1.00	0.96	0.97	9.36
20(e)	50	50229	1.00	0.96	0.95	9.99

Table A.2: Empirical results for means from 100 macroreplications for half exponential and half normal distributions in the monotone decreasing means configuration without CRN.

k	n_0	Ave N^*	PCS	Est Coverage	True Coverage	Ave Jumps
4(n)	50	21276	1.00	0.96	0.92	7.54
4(n)	100	21239	1.00	0.96	0.96	6.95
4(e)	50	21974	1.00	0.96	0.95	7.38
4(e)	100	21760	1.00	0.96	0.94	6.76
10(n)	50	43913	1.00	0.96	0.98	0.35
10(n)	100	44056	1.00	0.96	0.97	8.68
10(e)	50	45917	1.00	0.96	0.95	9.38
10(e)	100	45349	1.00	0.96	0.98	8.51
20(n)	50	72426	1.00	0.96	0.96	10.43
20(n)	100	72471	1.00	0.96	0.90	10.03
20(e)	50	74808	1.00	0.96	0.95	10.72
20(e)	100	73802	1.00	0.96	0.98	9.69

Table A.3: Empirical results for means from 100 macroreplications for negative binomial distributions in the slippage configuration without CRN.

k	θ	Average N^*	PCS	Est. Coverage	True Coverage	Ave Jumps
2	1	1729	0.98	0.97	0.97	4.47
2	10	15842	0.99	0.96	0.97	5.24
4	1	2757	0.96	0.96	0.93	5.36
4	10	26980	1.00	0.96	0.97	6.29
10	1	4064	0.99	0.96	0.95	6.09
10	10	40837	0.99	0.96	0.95	7.72
20	1	5125	1.00	0.96	0.93	6.80
20	10	51189	1.00	0.96	0.97	8.98

Table A.4: Empirical results for quantiles ($q = 0.8$) from 100 macroreplications for half exponential and half normal distributions in the slippage configuration without CRN.

k	n_0	Ave N^*	PCS	Est Coverage	True Coverage	Ave Jumps
2(n)	50	22355	0.97	0.97	0.92	5.72
2(n)	100	23156	0.99	0.97	0.95	5.01
2(e)	50	22304	0.97	0.96	0.94	5.82
2(e)	100	24103	0.96	0.96	0.95	5.14
4(n)	50	43351	1.00	0.96	0.95	7.72
4(n)	100	42708	1.00	0.96	0.95	7.31
4(e)	50	44194	0.95	0.96	0.93	7.84
4(e)	100	43824	0.99	0.96	0.99	7.38
10(n)	50	80487	1.00	0.96	0.97	9.56
10(n)	100	79610	1.00	0.96	0.98	9.14
10(e)	50	79611	1.00	0.96	0.97	9.48
10(e)	100	79276	1.00	0.96	0.97	8.84
20(n)	50	107071	1.00	0.96	0.98	10.61
20(n)	100	106991	1.00	0.96	0.97	10.07
20(e)	50	108670	1.00	0.96	0.98	10.93
20(e)	100	107608	1.00	0.96	0.94	9.98