Two-Stage Multiple-Comparison Procedures For Steady-State Simulations

Halim Damerdji
Department of Industrial Engineering
North Carolina State University
Raleigh, NC 27695-7906, U.S.A.
e-mail: damerdji@eos.ncsu.edu

Marvin K. Nakayama
Department of Computer and Information Science
New Jersey Institute of Technology
Newark, NJ 07102, U.S.A.
e-mail: marvin@cis.njit.edu

July 15, 1998

Abstract

Procedures for multiple comparisons with the best are investigated in the context of steady-state simulation, whereby a number k of different systems (stochastic processes) are compared based upon their (asymptotic) means μ_i $(i = 1, 2, \ldots, k)$. The variances of these (asymptotically stationary) processes are assumed to be unknown and possibly unequal. We consider the problem of constructing simultaneous confidence intervals for $\mu_i - \max_{j \neq i} \mu_j$ $(i = 1, 2, \ldots, k)$, which is known as multiple comparisons with the best (MCB). Our intervals will be constrained to contain 0, and so they are so-called constrained MCB intervals. In particular, two-stage procedures for construction of absolute- and relative-width confidence intervals are presented. Their validity is addressed by showing that the confidence intervals cover the parameters with probability at least some user-specified threshold value, as the confidence intervals' width parameter shrinks down to 0. The general assumption on the processes is that they satisfy a functional central limit theorem. The simulation output analysis procedures used are based on the method of standardized time series, of which the method of batch means is a special case. The techniques developed here extend to other multiple-comparison procedures such as unconstrained MCB, multiple comparisons with a control and all-pairwise comparisons. Although simulation is the context in this paper, the results naturally apply to (asymptotically) stationary time series.

Key words and phrases. Stochastic simulation; steady-state output analysis; standardized time series; multiple comparisons; two-stage procedures.

1 Introduction

Suppose that k systems (ergodic stochastic processes) are to be compared, based upon the processes' asymptotic (or steady-state) means, where larger is assumed to be better. (The case where smaller is better can be similarly treated.) In our setting, the process means are to be estimated via independent steady-state simulations. We do not assume that variances are known nor that they are equal. For example, the different systems could represent stable but analytically intractable queueing systems, with certain factors (e.g., the number and type of machines in a production context) set at various levels. The best system here might be the one with the largest throughput. Procedures for multiple comparisons with the best (MCB) are considered. Let μ_i denote the asymptotic mean of system i $(i = 1, 2, \ldots, k)$. Simultaneous confidence intervals are constructed for the parameters $\mu_i - \max_{j \neq i} \mu_j$ $(i = 1, 2, \ldots, k)$. Multiple-comparison procedures have the advantage over classical ranking-and-selection procedures in that the problem is treated as an inference problem; the confidence intervals provide information on how close the systems may be to one another. This is useful when deciding in the context, say, of selecting the best system, which alternative to actually implement since the system with the largest mean ultimately may not be picked when other factors not directly reflected in the performance measure are taken into account (e.g., capital investment).

In this paper, we present procedures for construction of simultaneous, absolute and relative, fixed-width confidence intervals for MCB in the context of ergodic processes. Asymptotic validity of these procedures is addressed, in the sense that it is shown that the confidence intervals simultaneously cover the parameters of interest with probability at least some prespecified threshold value (as the confidence-interval-width parameter δ shrinks down to zero). We assume throughout that the k processes satisfy a functional central limit theorem. Continuous-time stochastic processes are considered, but extension to the discrete-time case is straightforward, as will be discussed in Section 2. Matejcik and Nelson (1995) and Nakayama (1994) contain the seminal ideas used in this paper. The procedures proposed are two-stage. (Single-stage multiple-comparison procedures for steady-state simulations are studied in Nakayama 1997.) In the first stage, system i $(i = 1, 2, \ldots, k)$ is simulated up to some (large) time T. Its sample path is then divided up into a fixed number m of non-overlapping batches, each of length T/m. We then apply a standardized time series methodology to "estimate" each system's process variance. (The double quotes are to stress that the quantity constructed is *not* a consistent estimator of the process variance; see Glynn and Whitt 1991.) The required total number $N_i(\delta)$ of batches for system i is then computed. For $i = 1, 2, \ldots, k$, if $N_i(\delta) \geq m + 1$, simulation of system i must be continued until time $N_i(\delta)T/m$. Simultaneous confidence intervals are then formed.

Constrained MCB confidence intervals must include the origin, i.e., the origin is the left point, the right point, or an interior point of the confidence interval; see the first paragraph of Section 3 for an explanation. (Unconstrained MCB confidence intervals may or may not include the origin.) In the steady-state context, Matejcik and Nelson (1995) propose a (constrained) MCB two-stage procedure based on Rinott's procedure for selection of the best system under an indifference-zone formulation for independent and identically distributed (i.i.d.) populations with a normal distribution, using the batch means method of steady-state output analysis. Matecjik and Nelson (1995) show that their procedure is exact under the assumption that the batch means are i.i.d. with a normal distribution. In our work, we relax this condition and make rigorous the procedure in the general dependent case, extend the output-analysis method that can be used to the class of standardized time series methods (of which batch means is a special case), and, in addition to absolute-width, also include relative-width confidence intervals in the theory. We also propose related procedures with shorter second-stage simulations. The techniques developed in this paper extend to unconstrained MCB, multiple comparisons with a control (e.g., the control may be an existing system), and all-pairwise comparisons (where the systems are all compared to one another directly); see Damerdji and Nakayama (1996).

We have chosen to focus in this paper on MCB methods over MCA and MCC methods because the differences $\mu_i - \max_{j \neq i} \mu_j$ $(i = 1, 2, \ldots, k)$ in system performance are natural to look at when selecting the best system among a set of k alternatives since they indicate how close each system's performance is to the best of the rest; MCB methods provide *direct* inference on these quantities. We have favored *constrained* MCB over *unconstrained* MCB because they are sharper in the sense that the confidence intervals are narrower for the same sample size. They are also more "user friendly"; e.g., it is immediate to see whether the origin is a right-end, left-end, or middle point of an interval. See Hsu (1996, p. 82) for a discussion of the merits of constrained MCB.

There has been additional work on MCB for use in simulation besides Matejcik and Nelson (1995). Nelson and Matecjik (1995) modified the approach of Matejcik and Nelson (1995) to apply when the normal random variables from the different systems are no longer independent but instead are generated using the variance-reduction technique of common random numbers. In Goldsman and Nelson (1990), a heuristic single-stage batch-means procedure for MCB is presented. Goldsman and Nelson pick the batch sizes in an interesting way so as to get comparable variances across systems. Yuan and Nelson (1993) investigated MCB procedures under the assumption that the simulation output can be modeled as an autoregressive process. None of these studies investigated relative-width confidence intervals, as is done in this paper.

The rest of the paper has the following organization. In Section 2 we define our notation, state the assumption on the processes being simulated, and provide an overview of the steady-state

output analysis method of standardized time series. We present the two-stage MCB algorithms for the absolute-width (respectively, relative-width) case and the relevant results in Section 3 (resp., Section 4). Section 5 contains some numerical experiments. Finally, all of the proofs are given in Section 6.

2 Preliminaries

2.1 Assumptions on the processes

Suppose that there are k systems, labeled $1, 2, \ldots, k$, that we would like to compare based on the (asymptotic or steady-state) means of the ergodic processes $\mathbf{Y}_i \equiv \{Y_i(s) : s \geq 0\}$ (for $i = 1, 2, \ldots, k$). The asymptotic mean of $\mathbf{Y}_i$, or simply the mean if the process is stationary or weakly stationary (see Anderson 1971 for a definition), is to be estimated from a simulation. The sample paths of the processes are taken to be in $D[0, \infty)$, the space of real-valued functions on $[0, \infty)$ that are right continuous and have left limits. A large number of processes considered in engineering can be taken as such. (See Chapter 3 of Ethier and Kurtz 1986 for more details on the space $D[0, \infty)$.) We will be working in continuous time, but the theory also extends to discrete-time processes $\{Y_i(n) : n = 0, 1, 2, \ldots\}$ by taking $Y_i(s) = Y_i(\lfloor s \rfloor)$, where $\lfloor \beta \rfloor$ denotes the greatest integer less than or equal to $\beta \in \Re$.

We assume that the processes $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_k$ are mutually independent. In practice, this means that the simulations of systems $1, 2, \ldots, k$ are generated using mutually non-overlapping streams of uniform random numbers. In a study, it is possible that some of the design points considered by the analyst lead to non-ergodic processes (e.g., an (s, S) inventory system may turn out to be transient for some of the (s, S)-design points). The analyst would typically not know this until a pilot simulation has been undertaken. It is assumed here that the k processes are *all* ergodic. This is quite reasonable if, for example, the analyst is interested in stable systems only; the number k would then be the number of design points, out of all tried, that lead to a process that is ergodic. The major assumption on the processes is that they obey a functional central limit theorem (FCLT). More specifically, letting "$\Rightarrow$" denote weak convergence (see p. 7 of Billingsley 1968 for a mathematical definition), we assume the following:

A1 *For each system $i = 1, 2, \ldots, k$, there exist finite constants μ_i and $\sigma_i > 0$ such that*

$$X_{i,T} \quad \Rightarrow \quad \sigma_i B_i$$

as $T \to \infty$, where T is deterministic, B_i is a standard 1-dimensional Brownian motion, and

$$X_{i,T}(t) \quad \equiv \quad \sqrt{T}\,(\bar{Y}_{i,T}(t) \; - \; \mu_i t)$$

with

$$\bar{Y}_{i,T}(t) \equiv \frac{1}{T}\int_0^{tT} Y_i(s)ds,$$

for $T > 0$, $t \geq 0$.

Since the Y_i's are independent, it follows (see Billingsley 1968, Theorem 3.2 and pp. 26–27) that Assumption A1 is equivalent to requiring

$$X_T \Rightarrow \Gamma B \tag{1}$$

as $T \to \infty$, where $X_T = (X_{1,T}, X_{2,T}, \ldots, X_{k,T})$, $B = (B_1, B_2, \ldots, B_k)$ is a standard k-dimensional Brownian motion, and $\Gamma = \text{diag}(\sigma_i : i = 1, 2, \ldots, k)$. (Throughout this paper, we make the abuse of notation whereby a vector $x = (x_1, x_2, \ldots, x_k)$ is meant to denote a column vector. Also, we use the notation $\text{diag}(\sigma_i : i = 1, 2, \ldots, k)$ to denote a $k \times k$-matrix with diagonal elements σ_1,σ_2,...,σ_k, and zero off-diagonal elements.) By definition, $B_1, B_2, \ldots, B_k$ are mutually independent. Also, we use the notation $\mu = (\mu_1, \mu_2, \ldots, \mu_k)$.

Let $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_k)$ and $Y(s) = (Y_1(s), Y_2(s), \ldots, Y_k(s))$. Both the processes X_T and $\bar{Y}_T = (\bar{Y}_{1,T}, \bar{Y}_{2,T}, \ldots, \bar{Y}_{k,T})$ lie in $C[0,\infty)$, the space of continuous $\Re^k$-valued functions on $[0,\infty)$; for further details on the space $C[0,\infty)$, see Section 3.10 of Ethier and Kurtz (1986). Assumption A1 implies that for system i, $(1/T)\int_0^T Y_i(s)ds - \mu_i \Rightarrow 0$, as $T \to \infty$, and so the constant μ_i appearing in A1 is precisely the asymptotic mean of the process $\mathbf{Y}_i$. It also follows from A1 that for system i, as $T \to \infty$, $\sqrt{T}\left((1/T)\int_0^T Y_i(s)ds - \mu_i\right) \Rightarrow \sigma_i B_i(1)$. Since $B_i(1)$ has a standard normal distribution, the constant σ_i^2, whose square-root appears in the above central limit theorem, is then the process variance parameter of $\mathbf{Y}_i$. (If the process is stationary or weakly stationary, the process variance parameter, under a certain regularity condition, is equal to the sum of the covariances at all lags; see p. 367 of Durrett 1991.) Note that we do not assume that the variances are all equal nor that their values are known.

The FCLT generally holds for processes that are *weakly dependent*. These processes are such that events far apart in time are almost independent. Conditions for the FCLT to hold typically involve rates on the degree of dependency (decay rates) and boundedness of some moments. Examples of weakly dependent processes include regenerative (Glynn and Whitt 1987), ϕ-mixing (Billingsley 1968, p. 174), strong mixing (Durrett 1991, p. 381), moving averages (Durrett 1991, p. 385), martingale (Durrett 1991, p. 374), and associated (Newman and Wright 1981) processes. The cited references contain precise definitions of these processes and conditions under which they obey a FCLT.

2.2 Standardized time series

The class of standardized time series in the context of steady-state output analysis was introduced in Schruben (1983). Glynn and Iglehart (1990) is another important reference. A major advantage for constructing a confidence interval via a standardized time series method is that consistent estimation of the process variance parameter is not necessary. Instead, a function, called g say, that satisfies Properties (ii)–(v) of Assumption A2 below is applied to the normalized process $X_{i,T}$. Glynn and Iglehart (1990) show that a cancellation of the (square-root of the) process variance occurs in the ratio $B_i(1)/g(B_i)$, which is the limiting value of both $X_{i,T}(1)/g(X_{i,T})$ and $(\bar{Y}_{i,T}(1) - \mu_i)/g(\bar{Y}_{i,T})$. Hence, the ratio $B_i(1)/g(B_i)$ is free of the process variance parameter, and there is, then, no need to estimate this parameter in order to construct a confidence interval for μ_i. Because of this cancellation of the variance parameter, standardized time series methods are also called cancellation methods. To be practical, the function g must also be such that the distribution of the limiting ratio $B_i(1)/g(B_i)$ is known. The method of batch means, where the number of batches is fixed, is a standardized time series method, as shown in Glynn and Iglehart (1990). There are also other standardized time series methods: the area and maximum methods (Schruben 1983), the weighted area method (Goldsman, Meketon, and Schruben 1990), the Cramér-von Mises method (Goldsman, Kang, and Seila 1993), and the L_p-norm methods (Tokol, Goldsman, Ockerman, and Swain 1998).

Let $C_1[0,\infty)$ be the space of continuous real-valued functions on $[0,\infty)$, and let B_1 be a standard one-dimensional Brownian motion process. For a (measurable) function $h : C_1[0,\infty) \to \Re$, let $D(h)$ be its set of discontinuities (see Section 6 for a precise definition). We can now present the class of "g functions" comprising standardized time series.

A2. The (measurable) function $g : C_1[0,\infty) \to \Re$ satisfies the following conditions:

- **(i)** For every $x \in C_1[0,\infty)$, the quantity $g(x)$ depends only on $\{x(s) : 0 \le s \le 1\}$. Thus, if $x, y \in C_1[0,\infty)$ and $x(s) = y(s)$ for all $0 \le s \le 1$, then $g(x) = g(y)$.
- **(ii)** $g(\alpha x) = \alpha g(x)$ for all $\alpha > 0$ and $x \in C_1[0,\infty)$.
- **(iii)** $g(x - \beta e_1) = g(x)$ for all $\beta \in \Re$, $x \in C_1[0,\infty)$, and where $e_1(t) = t$.
- **(iv)** $P\{g(B_1) > 0\} = 1$.
- **(v)** $P\{B_1 \in D(g)\} = 0$.
- **(vi)** $P\{g(B_1) \in A\} = 0$ for countable sets A.

As already mentioned, a function g must satisfy conditions (ii)–(v) to qualify as a standardized time series method (Glynn and Iglehart 1990). (And if the distribution of $B_1(1)/g(B_1)$ is known, a

confidence interval for the mean can then be constructed.) Nakayama (1994) needed conditions (i) and (vi) to extend standardized time series to the two-stage setting. (Actually, our condition (vi) is slightly weaker than that used in Nakayama 1994.) Condition (vi) is not restrictive in practice since it is satisfied by all the known "g functions," i.e., the g functions available in the literature that lead to the construction of confidence intervals.

The methods proposed in this paper are in the spirit of Nakayama (1994), who introduced Stein-like two-stage procedures for construction of (asymptotically valid) absolute- and relative-width confidence intervals for the process mean. For system i, the first stage consists of a pilot simulation of (large) run length T, which is divided up into m batches so as to compute $g(X_{i,T})$, usually called an estimate of the process variance (although it is not truly an estimator in the sense that it is not consistent). The required total number of batches is computed, and the simulation continued if necessary. The validity of the procedures is shown by appealing to the continuous mapping theorem. This was the approach in Nakayama (1994), which we will adjust here in the context of multiple comparisons for steady-state simulation.

To close the section, the "g function" for the batch means method is given; consider

$$g(x) = \left[\frac{m}{m-1} \sum_{j=1}^{m} \left(\Delta_m x\left(\frac{j}{m}\right) - \frac{x(1)}{m} \right)^2 \right]^{1/2},$$

where $\Delta_d x(t) = x(t) - x\,(t - 1/d)$. Glynn and Iglehart (1990), who introduced this function, also showed that $B_1(1)/g(B_1)$ follows a Student t-statistic with $m-1$ degrees of freedom. In the context of this paper, we have that for each system i,

$$g(X_{i,T}) = \left(\frac{T}{m}\right)^{1/2} \left[\frac{1}{m-1} \sum_{j=1}^{m} \left(Z_{i,j}(T) - \frac{1}{m} \sum_{\ell=1}^{m} Z_{i,\ell}(T) \right)^2 \right]^{1/2}, \tag{2}$$

where, for $j = 1, 2, \ldots,$

$$Z_{i,j}(T) = \frac{1}{T/m} \int_{(j-1)T/m}^{jT/m} Y_i(s)\ ds, \tag{3}$$

and so $Z_{i,j}(T)$ is the sample mean of the jth batch (which is of size T/m). Note that $S_i^2(T) \equiv g^2(X_{i,T})m/T$ is equal to the sample variance of the m first-stage batch means of system i. Conditions (i)–(vi) of Assumption A2 are easily verified.

3 Absolute-width simultaneous confidence intervals

Following Hsu (1984), we are interested in constructing joint confidence intervals for the parameters $\mu_i - \max_{j \neq i} \mu_j$, for $i = 1, 2, \ldots, k$. Let $1 - \alpha$ be the joint confidence level desired. The confidence intervals will be constrained to contain zero (i.e., zero will be the confidence interval's left end-point or right end-point or be strictly within). If the confidence interval for $\mu_i - \max_{j \neq i} \mu_j$ contains

zero as its left end-point, then system i is declared to be the best in the sense that the analyst is confident that system i has the largest mean. If zero is the right end-point of the interval, then, with a high likelihood, there is at least one other system that is better than system i. If the confidence interval of $\mu_i - \max_{j \neq i} \mu_j$ contains zero in its interior, then, with a high likelihood, either system i is the best, or it is very close to the best, with means less than δ units away in the case of absolute-width confidence intervals, where δ is the required half-width. In this section, a procedure for constructing absolute-width simultaneous confidence intervals is presented, and their (asymptotic) validity established.

Before explicitly stating our algorithms, we need to make the following definition. In the spirit of Rinott (1978), for a given confidence level $1 - \alpha$, function g satisfying A2, first-stage number of batches m (the function g depends upon m), and number of systems k, compute the constant $\gamma \equiv \gamma(k, \alpha, g, m)$ such that

$$E\left[\prod_{i=1}^{k-1} \Phi\left(\frac{\gamma}{(1/g^2(B_i) + 1/g^2(B_k))^{1/2}}\right)\right] = 1 - \alpha, \tag{4}$$

where Φ is the distribution function of a standard normal random variable (r.v.) and the B_i, $i = 1, 2, \ldots, k$, are independent standard one-dimensional Brownian motion processes. Since Φ is continuous and strictly increasing between zero and one, it follows from Assumptions A2(iv)–(vi) that γ exists and is unique. For the batch mean methods, Glynn and Iglehart (1990) showed that $(m-1)g^2(B_1)$ has a chi-square distribution with $(m-1)$ degrees of freedom (denoted χ^2_{m-1}). Therefore, in the context of batch means, (4) is equivalent to equation (13) of Rinott (1978). The constant γ can then be obtained from a table (Wilcox 1984, or Bechhofer, Santner, and Goldsman 1995), or, for cases not covered by the tables, from available computer programs such as `PROGRAM RINOTT` in Bechhofer, Santner, and Goldsman (1995).

Define $(1), (2), \ldots, (k)$ such that $\mu_{(1)} \leq \mu_{(2)} \leq \cdots \leq \mu_{(k)}$. In other words, system (j) has the jth smallest asymptotic mean, and we are interested in determining the value of (k). The two-stage algorithm for absolute-width simultaneous confidence intervals presented next is an extension of Matejcik and Nelson (1995), where we allow any standardized time series method (including batch means) and consider continuous-time processes.

Procedure A

1. Specify the desired (approximate) absolute half-width of the MCB confidence intervals δ, desired confidence level $1 - \alpha$, first-stage number of batches m, function g satisfying A2, and number of systems k. Let γ solve (4) with m, k, α, and g.

2. Independently simulate each system i $(i = 1, 2, \ldots, k)$ with a (large) deterministic run length

T and group the output of each system into m (non-overlapping) batches. Let

$$\tilde{\mu}_i(T) \;=\; \frac{1}{m}\sum_{j=1}^{m} Z_{i,j}(T) \;=\; \frac{1}{T}\int_0^T Y_i(s)ds,$$

be the first-stage sample mean, where $Z_{i,j}(T)$ is defined in (3).

3. For each system $i = 1, 2, \ldots, k$, compute

$$S_i^2(T) = g^2(X_{i,T})m/T,$$

which, when divided by m, is an "estimate" of the variance of the first-stage sample mean $\tilde{\mu}_i(T)$. (An explicit formula for calculating $g(X_{i,T})$ for the batch means method is given by (2).)

4. For each system $i = 1, 2, \ldots, k$, compute the total number of batches to collect as

$$N_{a,i}(\delta) = \max\left\{m, \left\lceil\left(\frac{\gamma S_i(T)}{\delta}\right)^2\right\rceil\right\}, \tag{5}$$

where $\lceil\beta\rceil$ denotes the smallest integer greater than or equal to $\beta \in \Re$.

5. For each $i = 1, 2, \ldots, k$, if $N_{a,i}(\delta) \geq m + 1$, independently resume the simulation of system i, and collect the additional (non-overlapping) batches $m + 1, \ldots, N_{a,i}(\delta)$, each of size T/m. Compute the batch means $Z_{i,m+1}(T), Z_{i,m+2}(T), \ldots, Z_{i,N_{a,i}(\delta)}(T)$ as defined in (3).

6. For each system $i = 1, 2, \ldots, k$, compute the overall sample mean as

$$\hat{\mu}_{a,i}(\delta) = \frac{1}{N_{a,i}(\delta)}\sum_{j=1}^{N_{a,i}(\delta)} Z_{i,j}(T).$$

(If $N_{a,i}(\delta) = m$, then $\hat{\mu}_{a,i}(\delta) = \tilde{\mu}_i(T)$.)

7. Form the (simultaneous) constrained MCB (absolute-width) confidence intervals

$$I_{a,i}(\delta) = \left[\left(\hat{\mu}_{a,i}(\delta) - \max_{\ell\neq i}\hat{\mu}_{a,\ell}(\delta) - \delta\right)^-,\; \left(\hat{\mu}_{a,i}(\delta) - \max_{\ell\neq i}\hat{\mu}_{a,\ell}(\delta) + \delta\right)^+\right] \tag{6}$$

for $\mu_i - \max_{j\neq i}\mu_j$, $i = 1, 2, \ldots, k$, where $(\beta)^- = \min(\beta, 0)$ and $(\beta)^+ = \max(\beta, 0)$ for $\beta \in \Re$.

As previously mentioned, these confidence intervals are constrained to contain zero. We call the parameter δ the *approximate* half-width of the confidence intervals (because δ would have been the exact half-width if the lower and upper bounds of the confidence intervals had been constructed without the positive and negative functions). The main result of the section is now stated. Its proof is given in Section 6. A definition is needed: a r.v. is said to be *proper* if it is finite with probability one.

Theorem 1 *Assume Assumptions A1 and A2 hold, and Procedure A is used with first-stage run length* $T = \zeta/\delta^2$ *for system* i $(i = 1, 2, \ldots, k)$, $\zeta > 0$ *arbitrary but fixed. Then,*

(i) $N_a(\delta) \equiv (N_{a,1}(\delta), N_{a,2}(\delta), \ldots, N_{a,k}(\delta)) \Rightarrow N_a$ *as* $\delta \to 0$, *where* N_a *is some proper random vector;*

(ii) $\lim_{\delta \to 0} P\{\mu_i - \max_{j \neq i} \mu_j \in I_{a,i}(\delta),\ i = 1, 2, \ldots, k\} > 1 - \alpha$.

Part (i) of the theorem states that the total number of batches collected for each system is finite with probability 1. Under the extra assumption that $E[g^2(B_1)] < \infty$, which holds for all the known "g functions," one gets that the expected total number of batches for the limiting Brownian motion process is finite; see the remark after the proof of Theorem 1. This of course implies that N_a is proper. (Our study is asymptotic, however, and it is the batch size that gets large as δ shrinks down to 0.) Part (ii) shows that asymptotically (as the approximate absolute half-width of the confidence intervals $\delta \to 0$), the coverage probability of the MCB simultaneous confidence intervals is greater than $1 - \alpha$. Our proof of the theorem requires that the same batch size (ζ/δ^2) be used for all populations. In practice, it may be desirable to use shorter batches for systems that might be known to have relatively small autocorrelation, and this can be viewed as a drawback of the procedure.

In practice, once δ is fixed, one must choose the proportionality constant ζ large enough so that the batch means have a distribution that is close to normal and are almost uncorrelated. However, ζ should not be chosen too large because the first-stage run length may then be longer than necessary, as the following argument shows. From the proof of Theorem 1, we get that for $N_a = (N_{a,1}, N_{a,2}, \ldots, N_{a,k})$,

$$N_{a,i} \;=\; \max\left\{ m\ ,\ \left\lceil \frac{m\sigma_i^2 g^2(B_i)\gamma^2}{\zeta} \right\rceil \right\}. \tag{7}$$

If ζ is so large that $\sigma_i^2 g^2(B_i)\gamma^2/\zeta$ is smaller than 1 for most sample realizations, then $N_{a,i} = m$ for most sample realizations. The first-stage simulation run lengths are then longer than necessary for most realizations. The choice of a good proportionality constant for the batch sizes is still an open research problem. We empirically study the effect of the choice of ζ in Section 5.

Since the parameters of interest here are $\mu_i - \max_{j \neq i} \mu_j$, $i = 1, 2, \ldots, k$, *simultaneous* confidence intervals are constructed *directly* for these parameters. One could have of course constructed k confidence intervals for the parameters μ_i $(i = 1, 2, \ldots, k)$, and then deduced confidence intervals for the parameters of interest (i.e., the $\mu_i - \max_{j \neq i} \mu_j$). However, longer simulations would have been necessary. See Hsu (1996, pp. 21 and 26) for a discussion of direct inference versus deduced inference. Also, for the same confidence level, construction of direct simultaneous confidence inter-

vals requires shorter simulation run lengths over methods based on probabilistic inequalities, such as the Bonferroni inequality.

In practice, when using Procedure A, one may perform the complete simulation of one system, storing the final overall sample mean, before starting the simulation of the next system. Or, one may run all the k first-stage simulations, before starting the second-stage simulations. However, the latter approach requires one to must keep a "snapshot" of each system at the end of its first-stage simulation so as to be able to resume the second-stage simulations. This involves additional work on the part of the analyst and the use of more computer memory (main or disk).

The presence of the ceiling function in (5) ensures that the last batch is of the same size as the others when the second stage is needed. However, because of this, the simulation is run longer than necessary. To address this issue, the procedure can be modified in the following way. At the end of the first-stage simulation of system i, determine the total run length as

$$\bar{T}_i = \max\left\{T, \frac{T}{m}\left(\frac{\gamma S_i(T)}{\delta}\right)^2\right\}.$$

If $\bar{T}_i > T$, resume the simulation of system i from time T up to time $\bar{T}_i$. Compute the overall sample mean of each system i as $(1/\bar{T}_i)\int_0^{\bar{T}_i} Y_i(s)\,ds$, and construct the confidence intervals as in (6). When using this procedure, we no longer require the function g to satisfy A2(vi). Also, on the event $\{N_{a,i}(\delta) > m\}$, we avoid running the simulation of each system i from time $(T/m)\,(\gamma S_i(T)/\delta)^2$ to $(T/m)\left\lceil(\gamma S_i(T)/\delta)^2\right\rceil$ by using $\bar{T}_i$ to determine the total run length rather than using $N_{a,i}(\delta)$ to determine the total number of batches. This could lead to sizable computational savings when δ is small or ζ is large.

4 Relative-width simultaneous confidence intervals

The previous procedures are for the case when we want MCB confidence intervals having approximate absolute half-width δ. However, in many instances, we desire confidence intervals having approximate *relative* half-width that is $\delta = 5\%$, say, of the point estimator. Hence, we now develop algorithms for constructing MCB confidence intervals having approximate *relative* half-width δ. (Below we use a subscript "r" to denote that a quantity is for the relative-width procedure.) The first relative-width confidence intervals for the parameters $\mu_i - \max_{j\neq i}\mu_j$ will be

$$I_{r,i}(\delta) = \left[\left(\hat{\mu}_{r,i}(\delta) - \max_{\ell\neq i}\hat{\mu}_{r,\ell}(\delta) - \delta\left|\hat{\mu}_{r,i}(\delta) - \max_{\ell\neq i}\hat{\mu}_{r,\ell}(\delta)\right|\right)^-,\right.$$
$$\left.\left(\hat{\mu}_{r,i}(\delta) - \max_{\ell\neq i}\hat{\mu}_{r,\ell}(\delta) + \delta\left|\hat{\mu}_{r,i}(\delta) - \max_{\ell\neq i}\hat{\mu}_{r,\ell}(\delta)\right|\right)^+\right]. \tag{8}$$

For system i, this confidence interval's width is relative to $\hat{\mu}_{r,i}(\delta) - \max_{\ell \neq i} \hat{\mu}_{r,\ell}(\delta)$. The confidence intervals introduced next have widths relative to $\hat{\mu}_{r,i}(\delta)$ and $\max_{\ell \neq i} \hat{\mu}_{r,\ell}(\delta)$. Consider

$$J_{r,i}(\delta) = \left[\left(\hat{\mu}_{r,i}(\delta) - \max_{\ell \neq i} \hat{\mu}_{r,\ell}(\delta) - \delta \max_{\ell \neq i} |\hat{\mu}_{r,\ell}(\delta)| \right)^{-}, \left(\hat{\mu}_{r,i}(\delta) - \max_{\ell \neq i} \hat{\mu}_{r,\ell}(\delta) + \delta \, |\hat{\mu}_{r,i}(\delta)| \right)^{+} \right]. \quad (9)$$

These (non-classical) relative-width confidence intervals are wider than those of (8) when the $|\hat{\mu}_{r,i}(\delta)|$ are large. However, fewer batches would be required for the second-stage sampling, as will be seen below. Confidence intervals of the form

$$\left[\left(\hat{\mu}_{r,i}(\delta) - \max_{\ell \neq i} \hat{\mu}_{r,\ell}(\delta) - \delta \max_{j} |\hat{\mu}_{r,j}(\delta)| \right)^{-}, \left(\hat{\mu}_{r,i}(\delta) - \max_{\ell \neq i} \hat{\mu}_{r,\ell}(\delta) + \delta \max_{j} |\hat{\mu}_{r,j}(\delta)| \right)^{+} \right]$$

are theoretically (i.e., when $\delta \to 0$) equivalent to those of (9) but are wider in practice.

Some notation is needed. Let $\langle k \rangle$ and $\langle k-1 \rangle$ denote the systems with, respectively, the largest and second-largest first-stage sample means. Therefore, $\tilde{\mu}_{\langle k \rangle}(\delta) = \max_{i=1,\ldots,k} \tilde{\mu}_i(T)$ and $\tilde{\mu}_{\langle k-1 \rangle}(\delta) = \max_{i \neq \langle k \rangle} \tilde{\mu}_i(T)$. Below we will assume that $\mu_{(k-1)} < \mu_{(k)}$, and so with probability 1, $\tilde{\mu}_{\langle k \rangle}(T) \neq \tilde{\mu}_i(T)$ for $i \neq \langle k \rangle$ as $T \to \infty$.

Procedure R.1

Denote the total number of batches required for system i by $N_{r,i}(\delta)$. Replace the corresponding steps of Procedure A by:

4. For system $\langle k \rangle$, compute the total number of batches as

$$N_{r,\langle k \rangle}(\delta) = \max \left\{ m, \left\lceil \left(\frac{\gamma S_{\langle k \rangle}(T)}{\delta(\tilde{\mu}_{\langle k \rangle}(T) - \tilde{\mu}_{\langle k-1 \rangle}(T))} \right)^2 \right\rceil \right\} \quad (10)$$

   if $\tilde{\mu}_{\langle k \rangle}(T) \neq \tilde{\mu}_{\langle k-1 \rangle}(T)$, and set $N_{r,\langle k \rangle}(\delta) = m$ otherwise. For system $i \neq \langle k \rangle$, compute the total number of batches as

$$N_{r,i}(\delta) = \max \left\{ m, \left\lceil \left(\frac{\gamma S_i(T)}{\delta(\tilde{\mu}_{\langle k \rangle}(T) - \tilde{\mu}_i(T))} \right)^2 \right\rceil \right\} \quad (11)$$

   if $\tilde{\mu}_{\langle k \rangle}(T) \neq \tilde{\mu}_i(T)$, and set $N_{r,i}(\delta) = m$ otherwise.

6. Compute the overall sample mean of each system i as

$$\hat{\mu}_{r,i}(\delta) = \frac{1}{N_{r,i}(\delta)} \sum_{j=1}^{N_{r,i}(\delta)} Z_{i,j}(T).$$

7. From (8), form the simultaneous relative-width confidence intervals $I_{r,i}(\delta)$ $(i = 1, 2, \ldots, k)$.

Asymptotic validity of this procedure is now addressed.

Theorem 2 *Assume Assumptions A1 and A2 hold, and* $\mu_{(k-1)} < \mu_{(k)}$. *If Procedure R.1 is used with first-stage run length* $T = \zeta/\delta^2$ *for system* i $(i = 1, 2, \ldots, k)$, $\zeta > 0$ *arbitrary but fixed, then,*

(i) $N_r(\delta) \equiv (N_{r,1}(\delta), N_{r,2}(\delta), \ldots, N_{r,k}(\delta)) \Rightarrow N_r$ *as* $\delta \to 0$, *where* N_r *is some proper random vector;*

(ii) $\lim_{\delta \to 0} P\left\{\mu_i - \max_{j \neq i} \mu_j \in I_{r,i}(\delta),\ i = 1, 2, \ldots, k\right\} > 1 - \alpha$.

Remarks similar to those following Theorem 1 apply here as well.

When the process-mean parameters are close to one another, and so close to the largest one, the total number of batches required in Procedure R.1 may be large (see (10) and (11)). The procedure given next does not require such long simulations when $|\tilde{\mu}_{\langle k \rangle}(T)| > |\tilde{\mu}_{\langle k \rangle}(T) - \tilde{\mu}_i(T)|$ for all i, which is often the case in practice. (However, the confidence intervals will typically be wider.) We will assume now that $\mu_{(k)} \neq 0$.

Procedure R.2

Denote the total number of batches required for system i by $N'_{r,i}(\delta)$. Replace the corresponding steps of Procedure A by:

4. Compute the total number of batches to collect for system i $(i = 1, 2, \ldots, k)$ as

$$N'_{r,i}(\delta) = \max\left\{m, \left\lceil \left(\frac{\gamma S_i(T)}{\delta \tilde{\mu}_{\langle k \rangle}(T)} \right)^2 \right\rceil \right\} \tag{12}$$

   if $\tilde{\mu}_{\langle k \rangle}(T) \neq 0$, and set $N'_{r,i}(\delta) = m$ otherwise.

7. From (9), form the simultaneous relative-width confidence intervals $J_{r,i}(\delta)$ $(i = 1, 2, \ldots, k)$.

Asymptotic validity of the procedure is given next. The proof is omitted here, but can be found in Damerdji and Nakayama (1996).

Theorem 3 *Assume Assumptions A1 and A2 hold, and* $\mu_{(k)} \neq 0$. *If Procedure R.2 is used with first-stage run length* $T = \zeta/\delta^2$ *for system* i $(i = 1, 2, \ldots, k)$, $\zeta > 0$ *arbitrary but fixed, then,*

(i) $N'_r(\delta) \equiv (N'_{r,1}(\delta), N'_{r,2}(\delta), \ldots, N'_{r,k}(\delta)) \Rightarrow N'_r$ *as* $\delta \to 0$, *where* N'_r *is some proper random vector;*

(ii) $\lim_{\delta \to 0} P\left\{\mu_i - \max_{j \neq i} \mu_j \in J_{r,i}(\delta),\ i = 1, 2, \ldots, k\right\} > 1 - \alpha$.

The total numbers of batches in Procedures R.1 and R.2 (i.e., (10) and (11) for R.1, and (12) for R.2) involve all k first-stage sample means. Thus, unlike in the absolute-width case, when implementing these relative-width procedures in practice, one must keep a "snapshot" of each system at the end of its first-stage simulation so as to be able to resume the second-stage simulations.

5 Numerical experiments

The theoretical coverages of the procedures discussed in this paper are, in the limit, larger than $1-\alpha$. Here, we briefly investigate the *actual* coverage, and its relationship with the width parameter δ and first-stage run lengths T. We also examine the total number of batches needed, which according to our theory, should be finite with probability one; see Theorem 1. To carry out our studies, we used Procedure A to construct absolute-width confidence intervals for MCB.

We compared simple stochastic processes with known theoretical means so that the actual coverage for the procedures can be estimated. We considered $k = 4$ systems, namely, $M/M/c$ queues with $c = 1, 2, 3$, and 4. These are c-server queues with exponentially distributed service and interarrival times. The service and interarrival times are all independent. We call system i the one with i servers, and consider the stochastic processes $\{L_i(s) : s \geq 0\}$, where $L_i(s)$ is the number of customers in system i $(i = 1, \ldots, 4)$ at time s. The mean μ_i is, then, the steady-state average number of customers in system i $(i = 1, \ldots, 4)$.

In our experiments, systems 1 through 4 have the same arrival rate $\lambda = 1.0$, and the service rates are $\alpha_1 = 1.250$, $\alpha_2 = 0.645$, $\alpha_3 = 0.500$, and $\alpha_4 = 0.360$, respectively. Let $a_i \equiv \lambda/\alpha_i$. If $\pi_{i,0}$ is the (stationary) probability that system i is empty, one has that $\pi_{i,0} = (\sum_{j=0}^{i-1}(a_i^j/j!) + (ia_i^i/(i!(i - a_i))))^{-1}$, and $\mu_i = a_i + \pi_{i,0}((a_i^{i+1}/i)/(i!(1 - (a_i/i))^2))$; see Gross and Harris (1985, pp. 86–88). For the above rates, we obtain $\mu_1 = 4.000$, $\mu_2 \approx 3.884$, $\mu_3 \approx 2.888$, and $\mu_4 \approx 3.732$. Therefore, if larger is better, it follows that (4)=1, (3)=2, (2)=4, and (1)=3.

We constructed confidence intervals at level $\alpha = 10\%$ and estimated the actual coverage by the *observed* coverage, obtained from 200 independent replications, thereby yielding a theoretical standard error of $\sqrt{(1-\alpha)\alpha/200} \approx 2.2\%$. The four systems were simulated independently, and we used batch means as our standardized time series method. In the context of fixed-sample-size simulation of a single system, Schmeiser (1982) recommends a number of batches between 10 and 30 for this method, while for a two-stage procedure applied to one system, Nakayama (1994) suggests a first-stage number of batches between 5 and 15. In the MCB context, Matecjik and Nelson (1995) recommend an initial number of batches between 10 and 40. We picked $m = 10$. The parameter γ is Rinott's constant and can be obtained from Table 2.13 of Bechhofer, Santner, and Goldsman (1995): using their notation, $P^* = 0.90$, $n_1 = 10$, and $t = 4$, and so $\gamma = 2.913$. To avoid contamination from the initial bias problem, we started the simulations in steady-state (the stationary distributions of these simple stochastic processes are known). In practice, this is of course not feasible, and one should expect a further degradation of the observed coverages.

We considered the absolute half-widths $\delta = 0.05$, 0.1, and 0.2. Theorems 1–3 hold with the first-stage simulation run-length $T = \zeta/\delta^2$ for any $\zeta > 0$, and in our experiments, we varied the

Table 1: Observed coverage for first-stage run lengths T. Standard error is 2.2%. Theoretical coverage is larger than 90% .

MCB	T			
δ	$4000\delta^{-2}$	$1500\delta^{-2}$	$500\delta^{-2}$	$100\delta^{-2}$
0.05	94.0%	92.0%	91.5%	91.5%
0.10	97.5%	94.0%	92.0%	85.0%
0.20	96.5%	89.5%	94.0%	86.0%

proportionality constant ζ over the values 100, 500, 1500, and 4000. The results from these 12 experiments are given in Table 1.

Note that for the smallest value of δ (0.05), the observed coverages are all roughly the same for different values of ζ. By examining the proof of Theorem 1 (specifically, the last step when applying Lemma 8), we see that the theoretical limiting coverage probability, which is greater than $1 - \alpha$, does not depend on the value of ζ, and so the experiments seem to agree with the theory. The degree to which the observed coverages are over the nominal value of 90% for small δ may indicate that our methods are somewhat conservative.

We now study the effect of varying δ for fixed ζ. For the largest value of ζ (4000), the coverages for large δ are close to 1, and they seem to decrease as δ gets smaller. The reason for this is as follows. If one chooses a very large first-stage run length and a fairly large desired half-width δ, then the variance estimator from the first-stage will probably be small, and so a second stage most likely is not needed. Since the half-width of our resulting confidence intervals is always δ (regardless of the variance estimate), the intervals probably will be wider than necessary, and so the resulting coverages should be much larger than the nominal value. On the other hand, for the smallest value of ζ (100), the coverage for large δ is below the nominal level, and seems to increase as δ becomes small to what appears to be approximately the limiting value. In both these cases, we see that the asymptotics do not seem to be in effect when $\delta = 0.20$, but they do appear to hold for $\delta = 0.05$. Now for $\zeta = 500$ and 1500, it seems that the asymptotics already hold for $\delta = 0.20$. Hence, it seems that ζ affects the rate of convergence of the coverage probability to its limiting value.

Thus, we make the following recommendations for applying our methods. When simulating queueing systems, one may be able to use some of Whitt's (1989) calculations to try to determine a value for ζ that will hopefully result in a "good" convergence rate; see Section 6 of Nakayama (1994) for a related discussion. If one is not in the queueing setting or one cannot apply Whitt's formulas, then we recommend that the user chooses, once δ is fixed, a first-stage run length (i.e., ζ) such that the user feels the assumptions underlying the batch means method (approximate normality and independence of the batch means) would be in force for the k systems.

Table 2: Observed coverages for first-stage run lengths T of different orders relative to δ. Standard error is 2.2%. Theoretical coverage is larger than 90% .

MCB	T		
δ	$500/\delta$	$500/\delta^2$	$500/\delta^3$
0.05	87.5%	91.5%	98%
0.10	88.0%	92.0%	97.0%
0.20	87.0%	94.0%	91.0%

Recall that our theory required the first-stage run length T to be proportional to $1/\delta^2$ when the desired half-width is δ. This makes sense since the central limit theorem implies that in general the half-width of a confidence interval is asymptotically proportional to the inverse of the square root of the sample size. We now analyze, for a given desired half-width δ, the effect of *not* choosing T proportional to $1/\delta^2$. First, if T is proportional to $1/\delta$, then the coverage should degrade as $\delta \to 0$ because one would be using a confidence-interval approximate half-width δ in lieu of $\sqrt{\delta}$ as required by the central limit theorem (and $\delta < \sqrt{\delta}$ for $\delta < 1$). If the first-stage simulation run length is of order $1/\delta^3$ instead, the confidence-interval half-width should be of order $\delta^{3/2}$, but we are using δ. Since $\delta^{3/2} < \delta$ for $\delta < 1$, the actual coverage will increase to 1. We now examine this empirically.

To do this, we considered first-stage simulation run lengths $T = 500/\delta$, $500/\delta^2$, and $500/\delta^3$ for $\delta = 0.05$, 0.1, and 0.2. We ran 200 replications for the six new cases (the case $T = 500/\delta^2$ was considered previously), and Table 2 contains the observed coverages. Results for the cases $\delta = 0.05$ and $\delta = 0.1$ agree with our discussion above. On the other hand, the results may not appear to be consistent when $\delta = 0.2$; however, since the standard error is of order 2.2%, the discrepancy that $94 \not< 91$ can be attributed to statistical variability. For the case $\delta = 0.2$, we then ran 1000 replications, yielding a standard error of less than 1%. The coverages obtained were then 85.3% for $T = 500/\delta$, 92.6% for $T = 500/\delta^2$, and 94.2% for $T = 500/\delta^3$, completely in line with the theory. Because of prohibitive computing times, it was impractical to run a large number of replications for the case $\delta = 0.05$.

We then studied empirically the average total run lengths (ARL) and total number of batches collected (ANB) by Procedure A, and the results are given in Table 3. By examining the ARL columns, it appears that halving the absolute half-width when δ is small results in a quadrupling of the total run length, which agrees with the central limit theorem. Also, for each system, the total number of batches for each ζ seems to be fairly stable as δ decreases, in line with part (i) of Theorem 1.

Table 3: Average total run lengths (i:ARL) and average total of batches (i:ANB) for system i with first-stage run length $T = \zeta/\delta^2$ for all systems. Times are in multiples of 10^5 time units. Averages are over 200 replications.

δ	ζ	1:ANB	1:ARL	2:ANB	2:ARL	3:ANB	3:ARL	4:ANB	4:ARL
0.05	4000	30.1	48.16	19	30.40	10	16.00	10	16.00
	1500	80.9	48.54	47.6	28.56	10.5	6.30	13.7	8.22
	500	249	49.80	150	30.00	21.2	4.24	36.5	7.30
	100	1260	50.40	674	26.96	107	4.28	166	6.64
0.1	4000	30.6	12.24	19.2	7.68	10	4.00	10	4.00
	1500	83.3	12.50	47.9	7.19	10.6	1.59	13.3	2.00
	500	241	12.05	143	7.15	22	1.10	33.5	1.68
	100	1210	12.10	660	6.60	110	1.10	160	1.60
0.2	4000	30.1	3.01	19.7	1.97	10	1.00	10.1	1.01
	1500	84	3.15	54	2.03	10.7	0.40	13.4	0.50
	500	242	3.02	130	1.62	23.1	0.29	34.5	0.44
	100	1060	2.65	512	1.28	104	0.26	101	0.25

6 Proofs

First, some introductory results are stated. The continuous mapping theorem will be used repeatedly, and for easy reference, is given below in the context of interest to us (where the processes are in $C[0,\infty)$ and the functional is $\Re^k$-valued). A definition is needed. Let $D(h)$ be the set of discontinuities of the functional $h : C[0,\infty) \to \Re^k$, i.e., $D(h) = \{x \in C[0,\infty) : \exists\{x_T\} \subset C[0,\infty)$ such that, as $T \to \infty$, we have that $d(x_T, x) \to 0$ but $\|h(x_T) - h(x)\| \not\to 0\}$, where d is a metric in $C[0,\infty)$, e.g., the metric induced by the uniform norm on compact subsets, and $\|\cdot\|$ is the Euclidean norm in $\Re^k$.

Proposition 4 *Suppose $X_T, X \in C[0,\infty)$ are such that $X_T \Rightarrow X$ as $T \to \infty$. Consider a (measurable) function $h : C[0,\infty) \to \Re^k$. If $P\{X \in D(h)\} = 0$, then $h(X_T) \Rightarrow h(X)$ as $T \to \infty$.*

We will also be using the continuous mapping theorem for real-valued functionals. See Ethier and Kurtz (1986, p. 103) for a statement of the continuous mapping theorem in the general case. The so-called Slepian's inequality (Slepian 1962) is given next.

Lemma 5 *Let $W = (W_1, W_2, \ldots, W_{k-1})$ be a multivariate normal random vector with mean vector 0 and covariance matrix $\Theta = (\Theta_{i,j} : i, j = 1, 2, \ldots, k-1)$. Suppose that $\Theta_{i,i} = 1$ for all $i = 1, 2, \ldots, k-1$, and that $\Theta_{i,j} > 0$ for all $i, j = 1, 2, \ldots, k-1$. Then, for any $(\beta_1, \beta_2, \ldots, \beta_{k-1}) \in \Re^{k-1}$,*

$$P\{W_i \le \beta_i,\ i = 1, 2, \ldots, k-1\} \quad > \quad \prod_{i=1}^{k-1} P\{W_i \le \beta_i\}.$$

Some more notation is needed. Recalling from Section 2 the definition of the "g functions," let $\mathcal{F}_g$ denote the σ-field generated by $(g(B_1), g(B_2), \ldots, g(B_k))$. The next lemma is proven in Nakayama (1994) in the case of a single system (i.e., $k = 1$). The (conditional) independence will follow from the fact that the Brownian motion processes $B_1, B_2, \ldots, B_k$ are independent.

Lemma 6 *Let $R = (R_1, R_2, \ldots, R_k)$ be an $\mathcal{F}_g$-measurable random vector such that $1 \leq R_i < \infty$ a.s. for $i = 1, 2, \ldots, k$, and the R_i are independent. Then, given $(g(B_1), g(B_2), \ldots, g(B_k))$, $\sigma_i B_i(R_i)/R_i$, $i = 1, 2, \ldots, k$, are normally distributed with mean 0 and variance σ_i^2/R_i and are (conditionally) mutually independent.*

Lemma 7 *Let $R = (R_1, R_2, \ldots, R_k)$ be an $\mathcal{F}_g$-measurable random vector such that $1 \leq R_i < \infty$ a.s. for $i = 1, 2, \ldots, k$, and the R_i are independent. Define the random vector $U = (U_1, U_2, \ldots, U_{k-1})$, where, for $i = 1, 2, \ldots, k-1$,*

$$U_i = \frac{\sigma_{(i)} B_{(i)}(R_{(i)})}{R_{(i)}} - \frac{\sigma_{(k)} B_{(k)}(R_{(k)})}{R_{(k)}}.$$

Then, U has a continuous distribution function.

Proof. We show that for each $i = 1, 2, \ldots, k-1$, the distribution function H_i of U_i is continuous. Start by defining

$$H_i(x) = E\left[P\left\{ \frac{\sigma_{(i)} B_{(i)}(R_{(i)})}{R_{(i)}} - \frac{\sigma_{(k)} B_{(k)}(R_{(k)})}{R_{(k)}} \leq x \;\middle|\; \mathcal{F}_g \right\} \right].$$

By Lemma 6, it follows that given $(g(B_1), g(B_2), \ldots, g(B_k))$, $\sigma_i B_i(R_i)/R_i$, $i = 1, 2, \ldots, k$, are normally distributed with mean 0 and variance σ_i^2/R_i and (conditionally) mutually independent. Consequently, letting Φ_β denote the distribution function of a normal r.v. with mean 0 and variance β, we see that

$$H_i(x) = E\left[\int_{-\infty}^{\infty} \Phi_{\sigma_{(i)}^2/R_{(i)}}(x + y)\ \Phi_{\sigma_{(k)}^2/R_{(k)}}(dy) \right].$$

The continuity of Φ_β and the bounded convergence theorem (Ethier and Kurtz 1986) imply that $H_i(\,\cdot\,)$ is continuous. ∎

The following lemma is in the spirit of Rinott (1978).

Lemma 8 *Let $R = (R_1, R_2, \ldots, R_k)$ be an $\mathcal{F}_g$-measurable random vector such that $1 \leq R_i < \infty$ a.s. for $i = 1, 2, \ldots, k$, and the R_i are independent. Then,*

$$P\left\{\xi_i \ < \ \eta_i,\ i = 1, 2, \ldots, k-1\right\} \quad > \quad 1 - \alpha,$$

where, for $i = 1, 2, \ldots, k-1$,

$$\xi_i \quad \equiv \quad \frac{\sigma_{(i)} B_{(i)}(R_{(i)})/R_{(i)} - \sigma_{(k)} B_{(k)}(R_{(k)})/R_{(k)}}{(\sigma_{(i)}^2/R_{(i)} + \sigma_{(k)}^2/R_{(k)})^{1/2}}$$

and

$$\eta_i \equiv \frac{\gamma}{(1/g^2(B_{(i)}) + 1/g^2(B_{(k)}))^{1/2}}.$$

Proof. Let $q \equiv P\{\xi_i < \eta_i,\ i = 1, 2, \ldots, k-1\}$. We have that $q = E[\ P\{\xi_i < \eta_i,\ i = 1, 2, \ldots, k-1 \mid \mathcal{F}_g\}\ ]$. From Lemma 6, the $\sigma_i B_i(R_i)/R_i$ (i = 1,2, ..., k), given $(g(B_1), g(B_2), \ldots, g(B_k))$, are normally distributed with mean 0 and variance σ_i^2/R_i and (conditionally) mutually independent. Thus, given $(g(B_1), g(B_2), \ldots, g(B_k))$, $\xi = (\xi_1, \xi_2, \ldots, \xi_{k-1})$ has a multivariate normal distribution with mean 0 and covariance matrix $\Theta = (\Theta_{i,j} : i, j = 1, 2, \ldots, k-1)$, where $\Theta_{i,i} = 1$ for all i and

$$\Theta_{i,j} = \frac{\sigma^2_{(k)}/R_{(k)}}{\left[\left(\sigma^2_{(i)}/R_{(i)}\right) + \left(\sigma^2_{(k)}/R_{(k)}\right)\right]^{1/2}\left[\left(\sigma^2_{(j)}/R_{(j)}\right) + \left(\sigma^2_{(k)}/R_{(k)}\right)\right]^{1/2}}$$

for $i \neq j$. Note that $\Theta_{i,j} > 0$ for all i and j, and so Lemma 5 implies that

$$q \ > \ E\left[\prod_{i=1}^{k-1} P\{\xi_i < \eta_i \mid \mathcal{F}_g\}\right] \ = \ E\left[\prod_{i=1}^{k-1} \Phi(\eta_i)\right] \ = \ 1 - \alpha$$

by the definition of γ in (4). ∎

Proof of Theorem 1. We first prove part (i). Define the function $h_a : C[0, \infty) \to \Re^k$ as $h_a = (h_{a,1}, h_{a,2}, \ldots, h_{a,k})$ with

$$h_{a,i}(x) = \frac{1}{m} \max\left\{m, \left\lceil \frac{m g^2(x_i)\gamma^2}{\zeta} \right\rceil\right\} \tag{13}$$

for $i = 1, 2, \ldots, k$, where $\zeta > 0$ is deterministic and is the proportionality constant for the first-stage run length (see Theorem 1). Note that $h_{a,i}(X_{\zeta/\delta^2}) = N_{a,i}(\delta)/m$ and $D(h_{a,i}) \subset D(g) \cup \{x : mg^2(x_i)\gamma^2/\zeta \in \{m, m+1, m+2, \ldots\}\ \}$. Assumption A2(v) implies $P\{\sigma_i B_i \in D(g)\} = 0$ since σ_i is finite. Also, using Assumptions A2(ii) and A2(vi), we have

$$\begin{aligned} & P\{m g^2(\sigma_i B_i)\gamma^2/\zeta \in \{m, m+1, m+2, \ldots\}\ \} \\ & \quad = \ P\left\{g(B_i) \in \left\{\frac{(m\zeta)^{1/2}}{m^{1/2}\sigma_i\gamma}, \frac{((m+1)\zeta)^{1/2}}{m^{1/2}\sigma_i\gamma}, \frac{((m+2)\zeta)^{1/2}}{m^{1/2}\sigma_i\gamma}, \ldots\right\}\right\} \\ & \quad = \ 0 \end{aligned}$$

for $i = 1, 2, \ldots, k$. Thus, $P\{\Gamma B \in D(h_{a,i})\} = 0$ for all $i = 1, 2, \ldots, k$, and so $P\{\Gamma B \in D(h_a)\} = 0$. It then follows from (1) and the continuous mapping principle that $h_a(X_{\zeta/\delta^2}) \Rightarrow h_a(\Gamma B)$ as $\delta \to 0$. Hence, we have shown that $N_a(\delta) \Rightarrow N_a$ as $\delta \to 0$, where $N_a = m h_a(\Gamma B)$.

Now we have to show that $P\{N_{a,i} < \infty,\ i = 1, 2, \ldots, k\} = 1$. Since $0 < \sigma_i < \infty$, $\{g(\sigma_i B_i) < \infty\} = \{g(B_i) < \infty\}$ by Assumption A2(ii). Assumption A2(v) states that g is continuous at B_i with probability 1, and so $P\{g(B_i) < \infty\} = 1$ (since a continuous function must be finite). Therefore, $P\{h_{a,i}(\Gamma B) < \infty\} = 1$ for each i, and N_a is a proper random vector, thereby establishing part (i).

We now prove part (ii). We will accomplish this by constructing an event $G_a(\delta)$ such that

$$G_a(\delta) \subset \left\{ \mu_i - \max_{j \neq i} \mu_j \in I_{a,i}(\delta),\ i = 1, 2, \ldots, k \right\}, \tag{14}$$

and showing that

$$\lim_{\delta \to 0} P\{G_a(\delta)\} > 1 - \alpha. \tag{15}$$

We first define the event $G_a(\delta)$ as

$$G_a(\delta) = \{u_a(X_{\zeta/\delta^2}) < \mathbf{1}\}, \tag{16}$$

where $\mathbf{1}$ is the $(k-1)$-dimensional unit vector and the function $u_a : C[0, \infty) \to \Re^{k-1}$ is defined as $u_a = (u_{a,1}, u_{a,2}, \ldots, u_{a,k-1})$ with

$$u_{a,i}(x) = \frac{x_{(i)}(h_{a,(i)}(x))}{h_{a,(i)}(x)\sqrt{\zeta}} - \frac{x_{(k)}(h_{a,(k)}(x))}{h_{a,(k)}(x)\sqrt{\zeta}}, \tag{17}$$

for $i = 1, 2, \ldots, k-1$. Note that

$$G_a(\delta) = \{\hat{\mu}_{a,(k)}(\delta) - \hat{\mu}_{a,(i)}(\delta) - (\mu_{(k)} - \mu_{(i)}) > -\delta,\ i = 1, 2, \ldots, k-1\}.$$

We will establish (14) by following a line of reasoning developed by Hsu (1984) (also see pages 150–151 of Hochberg and Tamhane 1987). Define the sets

$$\begin{aligned} E_{a,1}(\delta) &= \left\{ \mu_i - \max_{j \neq i} \mu_j \leq \left(\hat{\mu}_{a,i}(\delta) - \max_{\ell \neq i} \hat{\mu}_{a,\ell}(\delta) + \delta \right)^+,\ i = 1, 2, \ldots, k \right\}, \\ E_{a,2}(\delta) &= \left\{ \mu_i - \max_{j \neq i} \mu_j \geq \left(\hat{\mu}_{a,i}(\delta) - \max_{\ell \neq i} \hat{\mu}_{a,\ell}(\delta) - \delta \right)^-,\ i = 1, 2, \ldots, k \right\}, \end{aligned}$$

and note that

$$E_{a,1}(\delta) \cap E_{a,2}(\delta) = \left\{ \mu_i - \max_{j \neq i} \mu_j \in I_{a,i}(\delta),\ i = 1, 2, \ldots, k \right\}.$$

In order to show that $G_a(\delta) \subset E_{a,1}(\delta) \cap E_{a,2}(\delta)$, we first prove that $G_a(\delta) \subset E_{a,1}(\delta)$:

$$\begin{aligned} G_a(\delta) &\subset \{\mu_{(k)} - \mu_{(k-1)} \leq \hat{\mu}_{a,(k)}(\delta) - \hat{\mu}_{a,i}(\delta) + \delta, \quad i \neq (k)\} \\ &= \left\{ \mu_{(k)} - \mu_{(k-1)} \leq \hat{\mu}_{a,(k)}(\delta) - \max_{\ell \neq (k)} \hat{\mu}_{a,\ell}(\delta) + \delta \right\} \\ &= \left\{ \mu_{(k)} - \mu_{(k-1)} \leq \left(\hat{\mu}_{a,(k)}(\delta) - \max_{\ell \neq (k)} \hat{\mu}_{a,\ell}(\delta) + \delta \right)^+, \right\} \\ &= E_{a,1}(\delta), \end{aligned}$$

where the penultimate step follows since $(\ \cdot\ )^+ \geq 0$ and $\mu_i - \mu_{(k-1)} \leq 0$ for all $i \neq (k)$. We now show that $G_a(\delta) \subset E_{a,2}(\delta)$:

$$G_a(\delta) = \{\mu_{(i)} - \mu_{(k)} > \hat{\mu}_{a,(i)}(\delta) - \hat{\mu}_{a,(k)}(\delta) - \delta, \quad i = 1, 2, \ldots, k-1\}$$

$$
\begin{aligned}
&\subset \quad \{\mu_i - \mu_{(k)} \geq \hat{\mu}_{a,i}(\delta) - \max_{\ell \neq i} \hat{\mu}_{a,\ell}(\delta) - \delta, \quad i \neq (k)\} \\
&= \quad \{\mu_i - \mu_{(k)} \geq \left(\hat{\mu}_{a,i}(\delta) - \max_{\ell \neq i} \hat{\mu}_{a,\ell}(\delta) - \delta\right)^{-}, \quad i \neq (k)\} \\
&= \quad E_{a,2}(\delta),
\end{aligned}
$$

where the penultimate step follows since $(\,\cdot\,)^{-} \leq 0$ and $\mu_i - \mu_{(k)} \leq 0$ for $i \neq (k)$. Hence, we have established the validity of (14).

Now to show (15) is true, we will prove that $P\{u_a(X_{\zeta/\delta^2}) < \mathbf{1}\} \to P\{u_a(\Gamma B) < \mathbf{1}\}$ as $\delta \to 0$ and that $P\{u_a(\Gamma B) < \mathbf{1}\} > 1-\alpha$. We first show that $u_a(X_{\zeta/\delta^2}) \Rightarrow u_a(\Gamma B)$ as $\delta \to 0$ by appealing to the continuous mapping theorem. Note that $h_{a,i}(x) \geq 1$ for all x by (13). Then, since each x_i is assumed to be continuous, we have that $D(u_{a,i}) \subset D(h_{a,(i)}) \cup D(h_{a,(k)})$ for all $i = 1, 2, \ldots, k-1$. From the proof of part (i), $P\{\Gamma B \in D(h_{a,i})\} = 0$ for all $i = 1, 2, \ldots, k$. Therefore, $P\{\Gamma B \in D(u_a)\} = 0$, so $u_a(X_{\zeta/\delta^2}) \Rightarrow u_a(\sigma B)$ as $\delta \to 0$ by (1) and the continuous mapping principle.

Now observe that

$$
u_{a,i}(\Gamma B) = \frac{\sigma_{(i)} B_{(i)}(h_{a,(i)}(\Gamma B))}{h_{a,(i)}(\Gamma B)\sqrt{\zeta}} - \frac{\sigma_{(k)} B_{(k)}(h_{a,(k)}(\Gamma B))}{h_{a,(k)}(\Gamma B)\sqrt{\zeta}}
$$

for $i = 1, 2, \ldots, k-1$. Also, note that $h_{a,i}(\Gamma B)$ is $\mathcal{F}_g$-measurable for all $i = 1, 2, \ldots, k$, and that we showed in the proof of part (i) that $1 \leq h_{a,i}(\Gamma B) < \infty$ a.s. It then follows from Lemma 7 that $u_a(\Gamma B)$ has a continuous distribution function, and so Theorem 2.1 of Billingsley (1968) implies that

$$
P\{u_a(X_{\zeta/\delta^2}) < \mathbf{1}\} \to P\{u_a(\Gamma B) < \mathbf{1}\} \tag{18}
$$

as $\delta \to 0$. Now it remains to show that

$$
P\{u_a(\Gamma B) < \mathbf{1}\} > 1 - \alpha. \tag{19}
$$

To do this, we first define

$$
\xi_{a,i} = \frac{\sigma_{(i)} B_{(i)}(h_{a,(i)}(\Gamma B))/(h_{a,(i)}(\Gamma B)\sqrt{\zeta}) - \sigma_{(k)} B_{(k)}(h_{a,(k)}(\Gamma B))/(h_{a,(k)}(\Gamma B)\sqrt{\zeta})}{(\sigma^2_{(i)}/h_{a,(i)}(\Gamma B) + \sigma^2_{(k)}/h_{a,(k)}(\Gamma B))^{1/2}}
$$

and

$$
\kappa_{a,i} = \frac{1}{(\sigma^2_{(i)}/h_{a,(i)}(\Gamma B) + \sigma^2_{(k)}/h_{a,(k)}(\Gamma B))^{1/2}}
$$

for $i = 1, 2, \ldots, k-1$. Notice that

$$
\begin{aligned}
P\{u_a(\Gamma B) < \mathbf{1}\} &= P\left\{ \frac{\sigma_{(i)} B_{(i)}(h_{a,(i)}(\Gamma B))}{h_{a,(i)}(\Gamma B)\sqrt{\zeta}} - \frac{\sigma_{(k)} B_{(k)}(h_{a,(k)}(\Gamma B))}{h_{a,(k)}(\Gamma B)\sqrt{\zeta}} < 1, \ i = 1, 2, \ldots, k-1 \right\} \\
&= P\left\{ \xi_{a,i} < \kappa_{a,i}, \ i = 1, 2, \ldots, k-1 \right\}.
\end{aligned}
$$

Also, since $h_{a,i}(\Gamma B) \geq g^2(\sigma_i B_i)\gamma^2/\zeta$ for all i by (13), we obtain

$$\begin{aligned} \kappa_{a,i} &\geq \frac{\gamma/\sqrt{\zeta}}{(\sigma^2_{(i)}/g^2(\sigma_{(i)}B_{(i)}) + \sigma^2_{(k)}/g^2(\sigma_{(k)}B_{(k)}))^{1/2}} \\ &= \frac{\gamma/\sqrt{\zeta}}{(1/g^2(B_{(i)}) + 1/g^2(B_{(k)}))^{1/2}} \equiv \eta_i \end{aligned} \tag{20}$$

for $i = 1, 2, \ldots, k-1$, by Assumption A2(ii). It then follows from Lemma 8 that

$$P\{u_a(\Gamma B) < \mathbf{1}\} \geq P\left\{\xi_{a,i} < \eta_i,\ i = 1, 2, \ldots, k-1\right\} > 1 - \alpha,$$

thereby proving (19). Putting this together with (18) establishes (15), and so the proof of part (ii) is complete. ∎

Remark 1 In the previous proof, in the last step before applying Lemma 8, the $1/\sqrt{\zeta}$ terms cancel out; hence, the limiting coverage probability is independent of ζ. The same remark applies to the proof of Theorem 2 below.

Remark 2 In the proof of Theorem 1, it follows from the bound $\max\{m, \lceil mg^2(\sigma_i B_i)\gamma^2\rceil\} \leq m + mg^2(\sigma_i B_i)\gamma^2 + 1$, that $E[N_{a,i}] \leq m + 1 + m\gamma^2\sigma_i^2 E[g^2(B_i)]$. If we assume that $E[g^2(B_i)] < \infty$, which holds, for example, for the batch means and area methods, it is obtained that the expected total number of batches $E[N_{a,i}]$ is finite.

Proof of Theorem 2. First, we prove part (i). Define the function $e \in C[0,\infty)$ as $e(t) = (t, \ldots, t)$, and let M be a diagonal $k \times k$-matrix with diagonal elements $(\mu_1, \mu_2, \ldots, \mu_k)$. By (1), $\bar{Y}_{\zeta/\delta^2} - Me = (\delta/\sqrt{\zeta})X_{\zeta/\delta^2} \Rightarrow 0$ as $\delta \to 0$. Consequently, $\bar{Y}_{\zeta/\delta^2} \Rightarrow Me$ as $\delta \to 0$. Since Me is deterministic, $(X_{\zeta/\delta^2}, \bar{Y}_{\zeta/\delta^2}) \Rightarrow (\Gamma B, Me)$ as $\delta \to 0$ by Theorem 4.4 of Billingsley (1968).

Now define the functions v_1 and v_2 from $C[0,\infty)$ into $\Re$ such that, for $y = (y_1, y_2, \ldots, y_k)$, $v_1(y)$ and $v_2(y)$ are the largest and second largest values, respectively, of $(y_1(1), y_2(1), \ldots, y_k(1))$. Since the functions y_i are continuous, it follows that the functions v_1 and v_2 are continuous because of the continuity of the projection and maximum mappings. Because of the relationship between the first-stage run length (T for system i) and the width parameter δ, we write $\tilde{\mu}_i(\delta)$ in lieu of $\tilde{\mu}_i(T)$. Note that $v_1(\bar{Y}_{\zeta/\delta^2}) = \tilde{\mu}_{\langle k\rangle}(\delta)$ and $v_2(\bar{Y}_{\zeta/\delta^2}) = \tilde{\mu}_{\langle k-1\rangle}(\delta)$ are, respectively, the largest and second-largest first-stage sample means. Consider now the function $w : C[0,\infty) \to \Re^k$ defined for $w = (w_1, w_2, \ldots, w_k)$ as

$$w_i(y) = \begin{cases} v_1(y) - y_i(1) & \text{if } y_i(1) < v_1(y) \\ v_1(y) - v_2(y) & \text{if } y_i(1) = v_1(y) \end{cases}. \tag{21}$$

The functions w_i are continuous. Also consider the function $h_r = (h_{r,1}, h_{r,2}, \ldots, h_{r,k}) : C[0,\infty) \times C[0,\infty) \to \Re^k$ defined as

$$h_{r,i}(x,y) = \frac{1}{m}\max\left\{m, \left\lceil \frac{mg^2(x_i)\gamma^2}{\zeta w_i^2(y)} \right\rceil\right\}$$

if $w_i(y) \neq 0$, and $h_{r,i}(x,y) = 1$ if $w_i(y) = 0$. Note that $h_{r,i}(X_{\zeta/\delta^2}, \bar{Y}_{\zeta/\delta^2}) = N_{r,i}(\delta)/m$.

To apply the continuous mapping theorem, we need to verify that $P\{(\Gamma B, Me) \in D(h_r)\} = 0$. It suffices to show that $P\{(\Gamma B, Me) \in D(h_{r,i})\} = 0$ for each i. We have that $D(h_{r,i}) \subset A_{1,i} \cup A_{2,i} \cup A_{3,i}$, where $A_{1,i} = \{(x,y) : x_i \in D(g)\}$, $A_{2,i} = \{(x,y) : w_i(y) = 0\}$, and $A_{3,i} = \{(x,y) : mg^2(x_i)\gamma^2/(\zeta w_i^2(y)) \in \{m,\, m+1, \ldots\}\ \}$. Assumption A2(v) implies $P\{\sigma_i B_i \in D(g)\} = 0$, and so $P\{(\Gamma B, Me) \in A_{1,i}\} = 0$. Since $w_i(Me) = \mu_{(k)} - \mu_i$ for $i \neq (k)$, $w_{(k)}(Me) = \mu_{(k)} - \mu_{(k-1)}$, and it is assumed that $\mu_{(k-1)} < \mu_{(k)}$, it follows that $P\{(\Gamma B, Me) \in A_{2,i}\} = 0$. Finally, we have from Assumptions A2(ii) and A2(vi) that, for $i \neq (k)$,

$$P\left\{g(B_i) \in \left\{\frac{|\mu_{(k)} - \mu_i|(m\zeta)^{1/2}}{m^{1/2}\sigma_i\gamma}, \frac{|\mu_{(k)} - \mu_i|((m+1)\zeta)^{1/2}}{m^{1/2}\sigma_i\gamma}, \ldots\right\}\right\} \;=\; 0.$$

A similar relation holds for $g(B_{(k)})$, but with $|\mu_{(k)} - \mu_{(k-1)}|$ replacing $|\mu_{(k)} - \mu_i|$.

Hence, $P\{(\Gamma B, Me) \in D(h_{r,i})\} = 0$ for each i, and so $P\{(\Gamma B, Me) \in D(h_r)\} = 0$. Consequently, $h_r(X_{\zeta/\delta^2}, \bar{Y}_{\zeta/\delta^2}) \Rightarrow h_r(\Gamma B, Me)$ as $\delta \to 0$, by the continuous mapping principle, where, for $i \neq (k)$,

$$mh_{r,i}(\Gamma B, Me) \;=\; \max\left\{m, \left\lceil \frac{mg^2(\sigma_i B_i)\gamma^2}{\zeta(\mu_{(k)} - \mu_i)^2} \right\rceil\right\},$$

and

$$mh_{r,(k)}(\Gamma B, Me) \;=\; \max\left\{m, \left\lceil \frac{mg^2(\sigma_{(k)} B_{(k)})\gamma^2}{\zeta(\mu_{(k)} - \mu_{(k-1)})^2} \right\rceil\right\}.$$

We have thus shown that $N_r(\delta) \Rightarrow N_r$ as $\delta \to 0$, where $N_r = mh_r(\Gamma B, Me)$. Using similar arguments to those used in the proof of Theorem 1, it follows that $P\{N_{r,i} < \infty, i = 1, 2, \ldots, k\} = 1$.

Let us now turn to (ii). Consider the sets

$$\begin{aligned} E_{r,1}(\delta) &= \left\{\mu_i - \max_{j\neq i}\mu_j \leq \left(\hat{\mu}_{r,i}(\delta) - \max_{\ell\neq i}\hat{\mu}_{r,\ell}(\delta) + \delta|\hat{\mu}_{r,i}(\delta) - \max_{\ell\neq i}\hat{\mu}_{r,\ell}(\delta)|\right)^+ ,\ i = 1, 2, \ldots, k\right\} \\ &= \left\{\mu_{(k)} - \mu_{(k-1)} \leq \hat{\mu}_{r,(k)}(\delta) - \max_{\ell\neq(k)}\hat{\mu}_{r,\ell}(\delta) + \delta|\hat{\mu}_{r,(k)}(\delta) - \max_{\ell\neq(k)}\hat{\mu}_{r,\ell}(\delta)|\right\} \end{aligned}$$

and

$$\begin{aligned} E_{r,2}(\delta) &= \left\{\mu_i - \max_{j\neq i}\mu_j \geq \left(\hat{\mu}_{r,i}(\delta) - \max_{\ell\neq i}\hat{\mu}_{r,\ell}(\delta) - \delta|\hat{\mu}_{r,i}(\delta) - \max_{\ell\neq i}\hat{\mu}_{r,\ell}(\delta)|\right)^- ,\ i = 1, 2, \ldots, k\right\} \\ &= \left\{\mu_i - \mu_{(k)} \geq \hat{\mu}_{r,i}(\delta) - \max_{\ell\neq i}\hat{\mu}_{r,\ell}(\delta) - \delta|\hat{\mu}_{r,i}(\delta) - \max_{\ell\neq i}\hat{\mu}_{r,\ell}(\delta)|\ ,\ \ i \neq (k)\right\}, \end{aligned}$$

constructed so that $E_{r,1}(\delta) \cap E_{r,2}(\delta) = \{\mu_i - \max_{j\neq i}\mu_j \in I_{r,i}(\delta)\ ,\ i = 1, 2, \ldots, k\}$. We will proceed as in the proof of Theorem 1: we construct a set $G_r(\delta)$ such that $G_r(\delta) \subset E_{r,1}(\delta) \cap E_{r,2}(\delta)$, and then use the continuous mapping theorem to show that $P\{G_r(\delta)\}$ converges, as $\delta \to 0$, to a probability that is larger than $1 - \alpha$. Consider

$$\begin{aligned} G_r(\delta) = \Big\{&\mu_{(k)} - \mu_i \leq \hat{\mu}_{r,(k)}(\delta) - \hat{\mu}_{r,i}(\delta) + \delta(\hat{\mu}_{r,(k)}(\delta) - \hat{\mu}_{r,i}(\delta))\,,\ \hat{\mu}_{r,(k)}(\delta) > \hat{\mu}_{r,i}(\delta)\,,\ i \neq (k)\,; \\ &\langle k\rangle = (k)\,;\ \langle k-1\rangle = (k-1)\Big\}. \end{aligned}$$

The event $\{\langle k\rangle = (k)\,;\ \langle k-1\rangle = (k-1)\}$ is equivalent to $\{v_1(\bar{Y}_{\zeta/\delta^2}) = \tilde{\mu}_{r,(k)}(\delta)\,;\ v_2(\bar{Y}_{\zeta/\delta^2}) = \tilde{\mu}_{r,(k-1)}(\delta)\}$. The latter notation will be used when we apply the continuous mapping theorem.

To show that $G_r(\delta) \subset E_{r,1}(\delta) \cap E_{r,2}(\delta)$, we first show that $G_r(\delta) \subset E_{r,1}(\delta)$. We have that

$$\begin{aligned}
G_r(\delta) &\subset \left\{\mu_{(k)} - \mu_i \le \hat{\mu}_{r,(k)}(\delta) - \hat{\mu}_{r,i}(\delta) + \delta(\hat{\mu}_{r,(k)}(\delta) - \hat{\mu}_{r,i}(\delta))\,,\ \hat{\mu}_{r,(k)}(\delta) > \hat{\mu}_{r,i}(\delta)\ \ i \neq (k)\right\} \\
&\subset \left\{\mu_{(k)} - \mu_{(k-1)} \le (1+\delta)\hat{\mu}_{r,(k)}(\delta) - (1+\delta)\hat{\mu}_{r,i}(\delta)\ \ \hat{\mu}_{r,(k)}(\delta) > \hat{\mu}_{r,i}(\delta)\ \ i \neq (k)\right\} \\
&= \left\{\mu_{(k)} - \mu_{(k-1)} \le (1+\delta)\hat{\mu}_{r,(k)}(\delta) - (1+\delta)\max_{\ell\neq(k)}\hat{\mu}_{r,\ell}(\delta)\,;\ \hat{\mu}_{r,(k)}(\delta) > \hat{\mu}_{r,i}(\delta)\,,\ i \neq (k)\right\} \\
&= \left\{\mu_{(k)} - \mu_{(k-1)} \le \hat{\mu}_{r,(k)}(\delta) - \max_{\ell\neq(k)}\hat{\mu}_{r,\ell}(\delta) + \delta|\hat{\mu}_{r,(k)}(\delta) - \max_{\ell\neq(k)}\hat{\mu}_{r,\ell}(\delta)|;\right. \\
&\qquad\qquad \left.\hat{\mu}_{r,(k)}(\delta) > \hat{\mu}_{r,i}(\delta)\,,\ \ i \neq (k)\right\} \\
&\subset \left\{\mu_{(k)} - \mu_{(k-1)} \le \hat{\mu}_{r,(k)}(\delta) - \max_{\ell\neq(k)}\hat{\mu}_{r,\ell}(\delta) + \delta|\hat{\mu}_{r,(k)}(\delta) - \max_{\ell\neq(k)}\hat{\mu}_{r,\ell}(\delta)|\right\} \\
&= E_{r,1}(\delta).
\end{aligned}$$

We now show that $G_r(\delta) \subset E_{r,2}(\delta)$. We have that

$$\begin{aligned}
G_r(\delta) &\subset \left\{\mu_i - \mu_{(k)} \ge (1+\delta)\hat{\mu}_{r,i}(\delta) - (1+\delta)\hat{\mu}_{r,(k)}(\delta),\ \ \hat{\mu}_{r,(k)}(\delta) > \hat{\mu}_{r,i}(\delta),\ \ i \neq (k)\right\} \\
&\subset \left\{\mu_i - \mu_{(k)} \ge (1+\delta)\hat{\mu}_{r,i}(\delta) - (1+\delta)\max_{\ell\neq i}\hat{\mu}_{r,\ell}(\delta),\ \ \hat{\mu}_{r,(k)}(\delta) > \hat{\mu}_{r,i}(\delta),\ \ i \neq (k)\right\} \\
&= \left\{\mu_i - \mu_{(k)} \ge \hat{\mu}_{r,i}(\delta) - \max_{\ell\neq i}\hat{\mu}_{r,\ell}(\delta) - \delta(\max_{\ell\neq i}\hat{\mu}_{r,\ell}(\delta) - \hat{\mu}_{r,i}(\delta)),\ \ \hat{\mu}_{r,(k)}(\delta) > \hat{\mu}_{r,i}(\delta),\ \ i \neq (k)\right\} \\
&= \left\{\mu_i - \mu_{(k)} \ge \hat{\mu}_{r,i}(\delta) - \max_{\ell\neq i}\hat{\mu}_{r,\ell}(\delta) - \delta|\hat{\mu}_{r,i}(\delta) - \max_{\ell\neq i}\hat{\mu}_{r,\ell}(\delta)|,\ \ \hat{\mu}_{r,(k)}(\delta) > \hat{\mu}_{r,i}(\delta),\ \ i \neq (k)\right\} \\
&\subset \left\{\mu_i - \mu_{(k)} \ge \hat{\mu}_{r,i}(\delta) - \max_{\ell\neq i}\hat{\mu}_{r,\ell}(\delta) - \delta|\hat{\mu}_{r,i}(\delta) - \max_{\ell\neq i}\hat{\mu}_{r,\ell}(\delta)|,\ \ i \neq (k)\right\} \\
&= E_{r,2}(\delta).
\end{aligned}$$

It follows, then, that $G_r(\delta) \subset E_{r,1}(\delta) \cap E_{r,2}(\delta)$.

We can rewrite

$$\begin{aligned}
&G_r(\delta) \\
&= \left\{\frac{1}{\delta}\left((\hat{\mu}_{r,i}(\delta) - \mu_i) - (\hat{\mu}_{r,(k)}(\delta) - \mu_{(k)})\right) - (\hat{\mu}_{r,(k)}(\delta) - \hat{\mu}_{r,i}(\delta)) \le 0,\right. \\
&\quad \left.\hat{\mu}_{r,(k)}(\delta) - \hat{\mu}_{r,i}(\delta) > 0\,,\ i \neq (k)\,;\ \langle k\rangle = (k)\ ;\ \langle k-1\rangle = (k-1)\right\} \\
&= \left\{\frac{X_{i,\zeta/\delta^2}(N_{r,i}(\delta)/m)}{\sqrt{\zeta}N_{r,i}(\delta)/m} - \frac{X_{(k),\zeta/\delta^2}(N_{r,(k)}(\delta)/m)}{\sqrt{\zeta}N_{r,(k)}(\delta)/m} - \left(\frac{\bar{Y}_{(k),\zeta/\delta^2}(N_{r,(k)}(\delta)/m)}{N_{r,(k)}(\delta)/m} - \frac{\bar{Y}_{i,\zeta/\delta^2}(N_{r,i}(\delta)/m)}{N_{r,i}(\delta)/m}\right) \le 0,\right. \\
&\quad \frac{\bar{Y}_{(k),\zeta/\delta^2}(N_{r,(k)}(\delta)/m)}{N_{r,(k)}(\delta)/m} - \frac{\bar{Y}_{i,\zeta/\delta^2}(N_{r,i}(\delta)/m)}{N_{r,i}(\delta)/m} > 0\,,\ i \neq (k)\,; \\
&\quad \left. v_1(\bar{Y}_{\zeta/\delta^2}) = \tilde{\mu}_{(k)}(\delta)\,;\ v_2(\bar{Y}_{\zeta/\delta^2}) = \tilde{\mu}_{(k-1)}(\delta)\right\}.
\end{aligned}$$

Consider the functions $u_{1,i}(x,y)$, $u_{2,i}(x,y)$, $u_3(x,y)$, and $u_4(x,y)$ (for $i \neq (k)$), defined as

$$u_{1,i}(x,y) = \frac{x_i(h_{r,i}(x,y))}{\sqrt{\zeta}h_{r,i}(x,y)} - \frac{x_{(k)}(h_{r,(k)}(x,y))}{\sqrt{\zeta}h_{r,(k)}(x,y)} - \left(\frac{y_{(k)}(h_{r,(k)}(x,y))}{h_{r,(k)}(x,y)} - \frac{y_i(h_{r,i}(x,y))}{h_{r,i}(x,y)}\right),$$

$$u_{2,i}(x,y) = \frac{y_i(h_{r,i}(x,y))}{h_{r,i}(x,y)} - \frac{y_{(k)}(h_{r,(k)}(x,y))}{h_{r,(k)}(x,y)},$$

$$u_3(x,y) = v_1(y) - y_{(k)}(1),$$

and

$$u_4(x,y) = v_2(y) - y_{(k-1)}(1).$$

Consider the set

$$A \equiv (-\infty, 0]^{k-1} \times (-\infty, 0)^{k-1} \times \{0\} \times \{0\}.$$

We have that $G_r(\delta)$ is equivalent to

$$\left\{ \left(u_{1,i}(X_{\zeta/\delta^2}, \bar{Y}_{\zeta/\delta^2}),\ i \neq (k);\ u_{2,i}(X_{\zeta/\delta^2}, \bar{Y}_{\zeta/\delta^2}),\ i \neq (k);\ u_3(X_{\zeta/\delta^2}, \bar{Y}_{\zeta/\delta^2});\ u_4(X_{\zeta/\delta^2}, \bar{Y}_{\zeta/\delta^2})\right) \in A \right\}.$$

Note that $h_{r,i}(x,y) \geq 1$ for all x and y. Since x and y are assumed to be continuous, $D(u_{j,i}) \subset D(h_{r,i}) \cup D(h_{r,(k)})$ for $j = 1, 2$ and $i \neq (k)$. In the proof of part (i), we showed that $P\{(\Gamma B, Me) \in D(h_r)\} = 0$, which implies that $P\{(\Gamma B, Me) \in D(u_{j,i})\} = 0$ for $j = 1, 2$ and $i \neq (k)$. The continuity of the maximum and projection mappings ensures that u_3 and u_4 are continuous. Also, Lemma 7 implies that $P\{u_{1,i}(\Gamma B, Me) = 0,\ i \neq (k)\} = 0$, and so the boundary of the set A has probability 0 under the limiting measure. Thus, from the continuous mapping theorem and Theorem 2.1 of Billingsley (1968), we get that, as $\delta \to 0$,

$$P\{G_r(\delta)\} \to P\{\ (u_{1,i}(\Gamma B, Me),\ i \neq (k);\ u_{2,i}(\Gamma B, Me),\ i \neq (k);\ u_3(\Gamma B, Me); u_4(\Gamma B, Me)) \in A\ \}.$$

We have that $u_3(\Gamma B, Me) = \mu_{(k)} - \mu_{(k)} = 0$ and $u_4(\Gamma B, Me) = \mu_{(k-1)} - \mu_{(k-1)} = 0$. Therefore, the above limiting probability is equal to

$$\begin{aligned}
&P\{u_{1,i}(\Gamma B, Me) \leq 0,\ \ u_{2,i}(\Gamma B, Me) < 0,\ \ i \neq (k)\} \\
&= P\left\{ \frac{\sigma_i B_i(h_{r,i}(\Gamma B, Me))}{\sqrt{\zeta}h_{r,i}(\Gamma B, Me)} - \frac{\sigma_{(k)} B_{(k)}(h_{r,(k)}(\Gamma B, Me))}{\sqrt{\zeta}h_{r,(k)}(\Gamma B, Me)} - (\mu_{(k)} - \mu_i) \leq 0, \right. \\
&\qquad \left. \mu_i - \mu_{(k)} < 0,\ i \neq (k) \right\} \\
&= P\left\{ \frac{\sigma_i B_i(h_{r,i}(\Gamma B, Me))}{\sqrt{\zeta}h_{r,i}(\Gamma B, Me)} - \frac{\sigma_{(k)} B_{(k)}(h_{r,(k)}(\Gamma B, Me))}{\sqrt{\zeta}h_{r,(k)}(\Gamma B, Me)} \leq \mu_{(k)} - \mu_i,\ \ i \neq (k) \right\} \\
&= P\left\{ \frac{\sigma_i B_i(h_{r,i}(\Gamma B, Me))/(\sqrt{\zeta}h_{r,i}(\Gamma B, Me)) - \sigma_{(k)} B_{(k)}(h_{r,(k)}(\Gamma B, Me))/(\sqrt{\zeta}h_{r,(k)}(\Gamma B, Me))}{\left(\sigma_i^2/h_{r,i}(\Gamma B, Me) + \sigma_{(k)}^2/h_{r,(k)}(\Gamma B, Me)\right)^{1/2}} \leq \right. \\
&\qquad \left. \frac{\mu_{(k)} - \mu_i}{\left(\sigma_i^2/h_{r,i}(\Gamma B, Me) + \sigma_{(k)}^2/h_{r,(k)}(\Gamma B, Me)\right)^{1/2}},\ i \neq (k) \right\}, \qquad (22)
\end{aligned}$$

where the second equality follows from our assumption that $\mu_{(k-1)} < \mu_{(k)}$. Call $\xi_{r,i}$ the left-hand side of the inequality in (22). By the definition of the function h_r and by Assumption A2(ii), we have that $mh_{r,i}(\Gamma B, Me) \geq (m\sigma_i^2 g^2(B_i)\gamma^2)/(\zeta(\mu_{(k)} - \mu_i)^2)$ for $i \neq (k)$, and $mh_{r,(k)}(\Gamma B, Me) \geq (m\sigma_{(k)}^2 g^2(B_{(k)})\gamma^2)/(\zeta(\mu_{(k)} - \mu_{(k-1)})^2)$. Hence, for $i \neq (k)$,

$$\begin{aligned}
\frac{\sigma_i^2}{h_{r,i}(\Gamma B, Me)} + \frac{\sigma_{(k)}^2}{h_{r,(k)}(\Gamma B, Me)} &\leq \frac{(\mu_{(k)} - \mu_i)^2}{g^2(B_i)\gamma^2} + \frac{\zeta(\mu_{(k)} - \mu_{(k-1)})^2}{g^2(B_{(k)})\gamma^2} \\
&\leq \frac{\zeta(\mu_{(k)} - \mu_i)^2}{\gamma^2}\left(\frac{1}{g^2(B_i)} + \frac{1}{g^2(B_{(k)})}\right).
\end{aligned}$$

Consequently,

$$\begin{aligned}
P\left\{\xi_{r,i} \leq \frac{\mu_{(k)} - \mu_i}{\left(\sigma_i^2/h_{r,i}(\Gamma B, Me) + \sigma_{(k)}^2/h_{r,(k)}(\Gamma B, Me)\right)^{1/2}}, \quad i \neq (k)\right\} \\
\geq P\left\{\xi_{r,i} \leq \frac{\gamma/\sqrt{\zeta}}{\left(1/g^2(B_i) + 1/g^2(B_{(k)}\right)^{1/2}}, \quad i \neq (k)\right\} \\
> 1 - \alpha
\end{aligned}$$

by Lemma 8, which completes the proof. ■

Acknowledgments

The authors would like to thank the Associate Editor, the referees, and the Area Editor whose insightful reviews helped improve both the content and presentation of the paper. The work of the second author was supported in part by the National Science Foundation under grant DMI-9624469 and by NJIT SBR grant 421180.

References

Anderson, T. W. 1971. *Statistical Analysis of Time Series.* Wiley, New York.

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

Billingsley, P. 1968. *Convergence of Probability Measures.* Wiley, New York.

Damerdji, H. and M. K. Nakayama. 1996. Two-stage multiple-comparison procedures for steady-state simulations. New Jersey Institute of Technology Research Report CIS-96-30.

Durrett, R. 1991. *Probability: Theory and Examples.* Wadsworth and Brooks/Cole, Pacific Grove, California.

Ethier, S. N. and T. G. Kurtz. 1986. *Markov Processes: Characterization and Convergence*. Wiley, New York.

Glynn, P. W. and D. L. Iglehart. 1990. Simulation output analysis using standardized time series. *Mathematics of Operations Research* 15:1–16.

Glynn, P. W. and W. Whitt. 1987. Sufficient conditions for functional-limit-theorem versions of $L = \lambda W$. *Queueing Systems: Theory and Applications* 1:279–287.

Glynn, P. W. and W. Whitt. 1991. Estimating the asymptotic variance with batch means. *Operations Research Letters* 10:431–435.

Goldsman, D., K. Kang, and A. F. Seila. 1993. Cramér-von Mises variance estimators for simulations. Technical Report, School of Industrial Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia.

Goldsman, D., M. S. Meketon, and L. W. Schruben. 1990. Properties of standardized time series weighted area variance estimators. *Management Science* 36:602–612.

Goldsman, L. and B. L. Nelson. 1990. Batch-size effects on simulation optimization using multiple comparisons with the best in computer simulation. In *Proceedings of the 1990 Winter Simulation Conference*, ed. O. Balci, R. P. Sadowski, and R. E. Nance, 288–293. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.

Gross, D. and C. M. Harris. 1985. *Fundamentals of Queueing Theory, Second Edition*, Wiley, New York.

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

Hsu, J. C. 1984. Constrained simultaneous confidence intervals for multiple comparisons with the best. *Annals of Statistics* 12:1136–1144.

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

Matejcik, F. J. and B. L. Nelson. 1995. Two-stage multiple comparisons with the best for computer simulation. *Operations Research* 43:633–640.

Nakayama, M. K. 1994. Two-stage stopping procedures based on standardized time series. *Management Science* 40:1189–1206.

Nakayama, M. K. 1997. Multiple-comparison procedures for steady-state simulations. *Annals of Statistics* 25:2433–2450 .

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

Newman, C. M. and A. L. Wright. 1981. An invariance principle for certain dependent sequences. *Annals of Probability* 9:671–675.

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

Schmeiser, B. 1982. Batch size effects in the analysis of simulation output. *Operations Research* 30:556–568.

Schruben, L. W. 1983. Confidence interval estimation using standardized time series. *Operations Research* 31:1090–1108.

Slepian, D. 1962. The one-sided barrier problem for Gaussian noise. *Bell System Technical Journal* 41:463–501.

Tokol, G., D. Goldsman, D. J. Ockerman, and J. J. Swain. 1998. Standardized time series L_p-norm variance estimators for simulations. To appear in *Management Science*.

Whitt, W. 1989. Planning queueing simulations. *Management Science* 35:1341–1366.

Wilcox, R. R. 1984. A table for Rinott's selection procedure. *Journal of Quality Technology* 16:97–100.

Yuan, M. and B. L. Nelson. Multiple comparisons with the best for steady-state simulation. *ACM Transactions on Modeling and Computer Simulation* 3:66–79.