Proceedings of the 2009 Winter Simulation Conference
M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, eds.

A GENERAL FRAMEWORK FOR THE ASYMPTOTIC VALIDITY OF TWO-STAGE PROCEDURES FOR SELECTION AND MULTIPLE COMPARISONS WITH CONSISTENT VARIANCE ESTIMATORS

Marvin K. Nakayma

Department of Computer Science
New Jersey Institute of Technology
Newark, NJ 07102, USA

ABSTRACT

We consider two-stage procedures for selection and multiple comparisons, where the variance parameter is estimated consistently. We examine conditions under which the procedures are asymptotically valid in a general framework. Our proofs of asymptotic validity require that the estimators at the end of the second stage are asymptotically normal, so we require a random-time-change central limit theorem. We explain how the assumptions hold for comparing means in transient simulations, steady-state simulations and quantile estimation, but the assumptions are also valid for many other problems arising in simulation studies.

1 INTRODUCTION

Simulation is often used to compare alternatives relative to a given parameter. The alternatives' parameter values are unknown, and we estimate them using simulation. For example, we may have ten alternative designs for a fault-tolerant system, and we want to compare the designs in terms of their 0.9 quantile of their failure times.

In the setting when the number k of alternatives is fairly small, say no more than 20, simulation approaches for comparing alternatives include selection procedures (Bechhofer, Santner, and Goldsman 1995) and multiple-comparison procedures (Hochberg and Tamhane 1987). The goal of a selection procedure is to identify with prespecified high probability the alternative with the largest parameter. For multiple comparisons, the aim is to construct simultaneous confidence intervals for certain functions (such as pairwise differences) of the parameters, where the intervals have a prespecified joint confidence level.

In this paper we study two-stage procedures for selection and multiple comparisons. For the multiple-comparison problem, the user prespecifies a constant $\delta > 0$, and the two-stage procedure determines a simulation run length for each alternative so that all of the constructed simultaneous confidence intervals have half width at most δ. For the selection problem we use Bechhofer's (1954) *indifference-zone* formulation, where we assume the user is indifferent between alternatives whose parameter values are within δ of each other. The two-stage selection procedure determines a simulation run length for each alternative so that the probability of correctly selecting the true best alternative at the end is at least a prespecified value.

We discuss the asymptotic validity of multiple-comparison and selection procedures as $\delta \to 0$ when the alternatives are mutually independent. We adopt a general framework, assuming only that we have parameter estimators that satisfy random-time-change central limit theorems (CLT), and that we have consistent estimators of the variance parameters appearing in the CLTs. Our framework encompasses many simulation settings arising in practice, such as comparing stochastic systems relative to their transient or steady-state means, and comparing populations relative to a quantile or some moment.

There has been some previous work developing asymptotically valid two-stage or sequential procedures for selection or multiple comparisons. Damerdji and Nakayama (1999) study two-stage multiple-comparison procedures for steady-state means based on Schruben's (1983) standardized time series (STS) methods. STS methods do *not* produce consistent variance estimators (for a fixed number of batches), so our current results do not encompass those in Damerdji and Nakayama (1999). Kim and Nelson (2006) develop a sequential screening procedure for comparing steady-state means. Mukhopadhyay and Solanky (1994) cover various selection procedures for comparing means of independent populations, using independent and identically distributed (i.i.d.) sampling within each population.

There is also other previous work that studies comparing alternatives via simulation under different frameworks than what we consider here. For example, there are Bayesian methods that determine the allocation of samples to the alternatives to maximize the posterior probability of correctly selecting the true best alternative; e.g., see Chick (2006) for an overview. Also, Glynn and Juneja (2004) adopt a large-deviations perspective for comparing alternatives.

The rest of the paper has the following organization. Section 2 lays down the mathematical framework. We describe a two-stage multiple-comparison procedure in Section 3 and an extension to make it a selection procedure in Section 4. We also discuss the asymptotic validity of the procedures in those sections. Section 5 gives conditions under which our assumptions hold for three types of performance measures often arising in simulation practice: transient means, steady-state means and quantiles. Our methods apply much more generally, though. We provide some concluding comments in Section 6. Proofs are given in Nakayama (2009).

2 FRAMEWORK

Suppose there are k alternatives $1,2,\ldots,k$, where each alternative i has an associated parameter θ_i, whose value is unknown and is to be estimated via simulation. We compare the alternatives in terms of their parameters θ_i, $i=1,\ldots,k$, with larger values being better. Corresponding to each alternative i is an estimation process $\widehat{\theta}_i = [\widehat{\theta}_i(t) : t \geq 0]$, where $\widehat{\theta}_i(t)$ is the estimator of θ_i based on a simulation of length t of alternative i. For example, suppose we have k stochastic processes $X_1,\ldots,X_k$ to compare, where the ith process $X_i = [X_i(t) : t \geq 0]$ has steady-state mean θ_i, which we use to compare the alternatives. Then we could take $\widehat{\theta}_i(t) = (1/t)\int_0^t X_i(s)\,ds$. We can also handle discrete-time estimators $\widehat{\theta}_{i,n}$ in our framework by letting $\widehat{\theta}_i(t) = \widehat{\theta}_{i,\lfloor t \rfloor}$, where $\lfloor a \rfloor$ is the greatest integer less than or equal to $a \in \Re$. For example, suppose we have k populations, where population i has distribution F_i (not necessarily normal) and mean θ_i. Let $X_{i,1}, X_{i,2},\ldots$ be i.i.d. samples from F_i, and we can define $\widehat{\theta}_{i,n} = (1/n)\sum_{j=1}^n X_{i,j}$ as the sample mean of the first n samples from F_i.

We assume the estimation processes $\widehat{\theta}_1,\ldots,\widehat{\theta}_k$ are independent. Moreover, we will assume that there exists a parameter $\eta > 0$ such that the following central limit theorem (CLT) holds for each alternative i:

$$t^{\eta}\left[\widehat{\theta}_i(t) - \theta_i\right] \Rightarrow N(0,\sigma_i^2) \tag{1}$$

as $t \to \infty$, where $0 < \sigma_i < \infty$ is some constant, $\Rightarrow$ denotes convergence in distribution (Billingsley 1999), and $N(a,b^2)$ denotes a normal distribution with mean a and variance b^2. In many settings, the parameter η assumes the canonical value of $1/2$, but we allow for other values also. The CLT in (1) implies that $\widehat{\theta}_i(t)$ is a consistent estimator of θ_i as $t \to \infty$.

We will develop two-stage procedures to compare the k alternatives. In our two-stage multiple-comparison procedure, the goal is to develop joint confidence intervals for specified functions of the parameters $\theta_1,\ldots,\theta_k$, with each interval having half-width at most δ, which the user specifies beforehand. To do this, the user runs a first stage to estimate σ_i^2 for each alternative i, and the estimate of σ_i^2 is used to determine the total run length $T_i(\delta)$ required for alternative i so the resulting confidence intervals are no wider than δ. Thus, each $T_i(\delta)$ depends on δ and the simulated first-stage of alternative i (and possibly of other alternatives $j \neq i$), so $T_i(\delta)$ is random. Our approaches require that the estimator (appropriately centered and scaled) of each θ_i at the end of the second stage $T_i(\delta)$ is asymptotically normally distributed, which will require that we strengthen the CLT in (1) to allow for a random-time change. As we will later see, it turns out that for each alternative i, the total run length satisfies

$$\delta^{1/\eta} T_i(\delta) \Rightarrow \tau_i \text{ as } \delta \to 0, \text{ where } 0 < \tau_i < \infty \text{ is a constant.} \tag{2}$$

Let $\tau_{i,\delta} = \tau_i \delta^{-1/\eta}$, and we assume the following random-time-change CLT:

Assumption 1 *There exists a constant $0 < \eta < \infty$ such that for each $i = 1,\ldots,k$, and for $T_i(\delta)$ satisfying (2),*

$$\left(\tau_{i,\delta}^{\eta}\left[\widehat{\theta}_i(T_i(\delta)) - \theta_i\right], i=1,\ldots,k\right) \Rightarrow \left(N_i(0,\sigma_i^2), i=1,\ldots,k\right) \text{ as } \delta \to 0, \tag{3}$$

where $N_i(0,\sigma_i^2)$, $i=1,\ldots,k$, are independent normals.

By replacing $T_i(\delta)$ with $\tau_{i,\delta}$ in (2) and (3), we see that the CLT in (1) is a special case of Assumption 1.

We also assume that for each alternative i, there exists another estimation process $V_i = [V_i(t) : t \geq 0]$, where $V_i(t)$ is the estimator of σ_i^2 appearing in (1) and (3) from simulating alternative i for a run length of t. We also allow for discrete-time estimators $V_{i,n}$ by taking $V_i(t) = V_{i,\lfloor t \rfloor}$. For example, in our previous example of comparing k populations relative to their means θ_i, we can take $V_{i,n} = (1/(n-1))\sum_{j=1}^{n}(X_{i,j} - \widehat{\theta}_{i,n})^2$, the sample variance. We assume that each $V_i(t)$ is a consistent estimator of σ_i^2:

Assumption 2 *For each i, $V_i(t) \Rightarrow \sigma_i^2$ as $t \to \infty$.*

We discuss in Section 5 conditions under which Assumptions 1 and 2 hold.

3 TWO-STAGE MULTIPLE COMPARISONS

We first focus on multiple comparisons with the best (MCB), where the goal is to construct simultaneous confidence intervals for $\theta_i - \max_{\ell \neq i} \theta_\ell$, $i = 1, 2, \ldots, k$; see Hsu (1984). Below we present a two-stage procedure for constructing MCB intervals having prespecified width parameter $\delta > 0$. The procedure extends one developed by Rinott (1978).

Procedure A

1. Specify the confidence level $1-\alpha$, the desired absolute-width parameter δ of the MCB confidence intervals, and the first-stage run length $T_{0,i}$ for each alternative i.
2. Independently simulate each alternative i for a run length of $T_{0,i}$.
3. For each alternative i, compute the total run length required as

$$T_i(\delta) = \max\left(T_{0,i}, \left(\frac{\gamma\sqrt{V_i(T_{0,i})}}{\delta}\right)^{1/\eta}\right), \tag{4}$$

   where the constant $\gamma \equiv \gamma(k, 1-\alpha) = \sqrt{2}z_{(1-\alpha)^{1/(k-1)}}$, with z_β satisfying $\Phi(z_\beta) = \beta$ for $0 < \beta < 1$, Φ is the distribution function of a standard (mean 0 and variance 1) normal distribution, η is as defined in Assumption 1, and $V_i(t)$ is any estimator satisfying Assumption 2.
4. For each alternative i, continue to simulate from time $T_{0,i}$ to $T_i(\delta)$, where the k alternatives are simulated independently, and form the point estimator $\widetilde{\theta}_i(\delta) = \widehat{\theta}_i(T_i(\delta))$ of θ_i.
5. Use the width parameter δ to construct simultaneous MCB confidence intervals

$$I_i(\delta) = \left[-\left(\widetilde{\theta}_i(\delta) - \max_{\ell\neq i}\widetilde{\theta}_\ell(\delta) - \delta\right)^-, \left(\widetilde{\theta}_i(\delta) - \max_{\ell\neq i}\widetilde{\theta}_\ell(\delta) + \delta\right)^+\right], \quad i = 1, \ldots, k,$$

   for $\theta_i - \max_{\ell\neq i}\theta_\ell$, $i = 1, \ldots, k$, respectively, where $-(\beta)^- = \min(\beta, 0)$ and $(\beta)^+ = \max(\beta, 0)$.

Let $\bar{\gamma} \equiv \bar{\gamma}(k, 1-\alpha) = \sqrt{2}\bar{z}_{k-1,1-\alpha}$, where $\bar{z}_{p,\beta}$ is the upper-β equicoordinate point of a p-variate standard normal distribution with unit variances and common correlation coefficient $1/2$. Table B.1 of Bechhofer, Santner, and Goldsman (1995) provides values for $\bar{z}_{p,\beta}$ for various p and β. When σ_i, $i = 1, \ldots, k$, in (1) and (3) are known, one can instead use a single-stage procedure with the total run length for alternative i being $(\bar{\gamma}\sigma_i/\delta)^{1/\eta}$ (Section 2.6 of Bechhofer, Santner, and Goldsman 1995).

Theorem 1 *Suppose that the CLT in (1) and Assumption 2 hold, and suppose that Procedure A is used with first-stage run length $T_{0,i} = \zeta_i \delta^{-\lambda}$ for each alternative i, where $\zeta_i > 0$ and $0 < \lambda \leq 1/\eta$ are any constants. Then*

(i) *the limit in equation (2) holds with*

$$\tau_i = \begin{cases} (\gamma\sigma_i)^{1/\eta} & \text{if } \lambda < 1/\eta \\ \max(\zeta_i, (\gamma\sigma_i)^{1/\eta}) & \text{if } \lambda = 1/\eta \end{cases}. \tag{5}$$

Moreover, if $\{V_i(t) : t \geq 0\}$ is uniformly integrable, then $E[\delta^{1/\eta} T_i(\delta)] \to \tau_i$ as $\delta \to 0$.

In addition, if the CLT in (1) is strengthened to Assumption 1, then the following also hold:

(ii) $\lim_{\delta\to 0} P\left\{\theta_i - \max_{\ell\neq i}\theta_\ell \in I_i(\delta), i = 1, \ldots, k\right\} > 1-\alpha$.

(iii) *If $\eta = 1/2$ in Assumption 1 and $0 < \lambda < 2$, then (i)–(ii) still hold when γ in (4) is replaced with $\overline{\gamma}$, and $\overline{\gamma} < \gamma$. Moreover, $T_i(\delta)/(\overline{\gamma}\sigma_i/\delta)^2 \Rightarrow 1$ as $\delta \to 0$, and if, in addition, $\{V_i(t) : t \geq 0\}$ is uniformly integrable, then $E[T_i(\delta)/(\overline{\gamma}\sigma_i/\delta)^2] \to 1$ as $\delta \to 0$.*

Part (i) shows that for each alternative i, the total run length $T_i(\delta)$ for small δ is roughly $\tau_i/\delta^{1/\eta}$, where τ_i in (5) is deterministic. Part (ii) establishes that the asymptotic joint coverage of the MCB intervals is greater than the nominal level $1-\alpha$. Part (iii) considers the special case when η assumes the canonical value $1/2$ and when we replace the critical point γ in (4) with $\overline{\gamma}$. In this case, Procedure A is asymptotically efficient, in the sense that the total run length for each alternative is asymptotically equivalent (in distribution and in mean) to what we would use if we knew each σ_i^2 (Chow and Robbins 1965).

Procedure A produces MCB intervals in which the half-width of the resulting intervals are at most δ. Nakayama (2009) also develops asymptotically valid two-stage MCB procedures that yield intervals having prespecified relative half-widths; i.e., the half-widths are at most $100\delta\%$ of the point estimators.

4 SELECTION PROCEDURE

Procedure A constructs asymptotically valid MCB confidence intervals. Matejcik and Nelson (1995) showed that MCB in the non-asymptotic setting is often closely related to a selection procedure under Bechhofer's (1954) indifference-zone formulation. In the indifference-zone set-up, the user is indifferent between two alternatives whose parameter values are less than δ apart. To show that we can similarly extend our asymptotic MCB procedure to also allow for selection, we first define $[1],[2],\ldots,[k]$ such that $\widetilde{\theta}_{[1]}(\delta) \leq \widetilde{\theta}_{[2]}(\delta) \leq \cdots \leq \widetilde{\theta}_{[k]}(\delta)$; i.e., alternative $[i]$ has the ith smallest estimator after simulating the second stage. Then we modify to Procedure A to get the following:

Procedure A.2

Use steps 1–5 of Procedure A, and include the additional step:

6. Select alternative $[k]$ as the best alternative.

Let $\theta = (\theta_1,\ldots,\theta_k)$ be the vector of parameter values of the alternatives, and define the event of a correct selection as

$$\mathrm{CS}_\theta(\delta) = \{\theta_{[k]} > \theta_{[i]}, i = 1,2,\ldots,k-1\},$$

which occurs when the selected alternative $[k]$ actually has the largest true parameter value. We would like to show that for Procedure A.2, the probability of correct selection (PCS) is at least $1-\alpha$ asymptotically as $\delta \to 0$. When establishing such a result, we need to be careful in formulating the problem to end up with a theoretically interesting conclusion. Specifically, suppose we fix the alternatives and their parameter values θ_i beforehand and then take the limit as $\delta \to 0$. The asymptotic set-up in Theorem 1 assumes the first stage of each alternative i in Procedure A.2 has run length $T_{0,i} = \zeta_i \delta^{-\lambda}$, for arbitrary constants $\zeta_i > 0$ and $0 < \lambda \leq 1/\eta$, and lets δ get small. Hence, as $\delta \to 0$, each first-stage length grows to infinity, so the total run lengths also do. But then typically a strong law of large numbers implies each estimator $\widetilde{\theta}_i(\delta) \to \theta_i$ almost surely as $\delta \to 0$, so $\lim_{\delta\to 0} P(\mathrm{CS}(\delta)) \to 1$. In fact, virtually any procedure that lets run lengths of every alternative grow to infinity will have asymptotic probability of correct selection equal to 1, so this asymptotic set-up is not theoretically interesting.

To obtain a non-trivial result, we need to make the problem "harder." One way of doing this is to allow the configuration of parameter values $\theta_1,\ldots,\theta_k$ to vary as δ shrinks. Specifically, define $(1),(2),\ldots,(k)$ such that $\theta_{(1)} \leq \theta_{(2)} \leq \cdots \leq \theta_{(k)}$; i.e., alternative (i) has the ith smallest parameter value. Now define

$$\Omega(\delta) = \{\theta : \theta_{(k)} > \theta_{(k-1)} + \delta\},$$

which is the *preference zone*, the configurations of parameter values in which the user prefers only alternative (k). It now becomes considerably more difficult to prove that the limit of the PCS over all configurations of parameters $\theta \in \Omega(\delta)$ is at least $1-\alpha$ as $\delta \to 0$, e.g., see Damerdji et al. (1996). Thus, we adopt a simplifying assumption:

Assumption 3 *For each alternative i, there exists a process $Y_i = [Y_i(t) : t \geq 0]$ such that $\widehat{\theta}_i(t) = \theta_i + Y_i(t)$ for all $t > 0$, where the distribution of Y_i does not depend on θ_i, and $Y_1,\ldots,Y_k$ are independent.*

When Assumption 3 holds, we can think of each θ_i as a "location" parameter and Y_i is a "noise" process added to θ_i. Kim and Nelson (2006) make a similar assumption in establishing the asymptotic validity of their selection procedure for steady-state means.

Theorem 2 *If Assumptions 1, 2 and 3 hold and Procedure A.2 is used with first-stage run length* $T_{0,i} = \zeta_i \delta^{-\lambda}$ *for each alternative i, where* $\zeta_i > 0$ *and* $0 < \lambda \le 1/\eta$ *are any constants, then*

$$\lim_{\delta \to 0} \inf_{\theta \in \Omega(\delta)} P\left\{ CS_\theta(\delta),\ \theta_i - \max_{\ell \neq i} \theta_\ell \in I_i(\delta), i = 1, \ldots, k \right\} > 1 - \alpha.$$

Thus, we see that the joint event of correct selection and simultaneous MCB interval coverage has asymptotic probability greater than $1 - \alpha$.

5 EXAMPLES

We now show that Assumptions 1 and 2 hold in a variety of simulation contexts arising in practice. We first consider the case of comparing means of independent population, with i.i.d. sampling within each population to construct the sample mean as the estimator of the true mean.

Example 1 **[Population means]** Suppose each alternative i corresponds to a population with distribution F_i, and let $X_{i,1}, X_{i,2}, \ldots$ be i.i.d. samples from F_i. Let $\theta_i = E[X_{i,1}]$, and the goal is to compare the k populations in terms of their means. Let $\widehat{\theta}_{i,n} = (1/n)\sum_{j=1}^{n} X_{i,j}$ be the sample mean of the first n samples from F_i. Let σ_i^2 be the variance of F_i, and we assume that $0 < \sigma_i^2 < \infty$. Then the ordinary CLT in (1) holds with $\eta = 1/2$ (e.g., Theorem 6.4.4 of Chung 2001). Thus, Assumption 1 holds with $\eta = 1/2$ by Theorem 7.3.2 of Chung (2001). The sample variance $V_{i,n} = (1/(n-1))\sum_{j=1}^{n}(X_{i,j} - \widehat{\theta}_{i,n})^2$ satisfies Assumption 2; e.g., see p. 73 of Serfling (1980).

Example 1 covers the setting of comparing means in transient simulations. Specifically, let X_i denote the (random) performance of alternative i over a finite (possibly random) time horizon. Then $\theta_i = E[X_i]$ is the transient mean of alternative i, and we compare the k alternatives relative to $\theta_1, \ldots, \theta_k$. We run i.i.d. replications of each alternative i to yield $X_{i,1}, X_{i,2}, \ldots$, which we use to estimate θ_i and σ_i^2 using the ordinary sample mean and sample variance, respectively. One could also use other estimators of θ_i and σ_i^2 as long as they satisfy Assumptions 1 and 2.

For Example 1, Assumption 1 immediately follows from the ordinary CLT in (1) without any additional conditions. However, this is not always the case, so we need to strengthen the ordinary CLT in (1) for Assumption 1 to hold. One such way is by also assuming *Anscombe's* (1952) *condition* below.

Condition 1 (Anscombe) *For each alternative i and for each positive* ε *and* ψ, *there exist positive* c_i *and* t_i *such that*

$$P\left\{ \sup_{s:\, |s-t| \le c_i t} t^\eta \left| \widehat{\theta}_i(s) - \widehat{\theta}_i(t) \right| > \varepsilon \right\} < \psi, \quad \text{for all } t \ge t_i.$$

Proposition 1 *The CLT in (1) and Condition 1 together imply Assumption 1.*

Next we consider a functional central limit theorem (FCLT), which states that each estimation process, when properly centered and scaled, converges in distribution to a Brownian motion. To state an FCLT precisely, let $D[0,1]$ denote the space of left-continuous functions with right limits on the unit interval. See Billingsley (1999) for details on $D[0,1]$ and Brownian motion.

Condition 2 (FCLT) *There exist finite positive constants* η, υ *and* ω *with* $2\omega - \upsilon = 2\eta$ *such that for each i,* $U_{i,n} \Rightarrow U_i$ *in* $D[0,1]$ *as* $n \to \infty$, *where* $U_{i,n} = [U_{i,n}(t) : 0 \le t \le 1]$, $U_i = [U_i(t) : 0 \le t \le 1]$, $U_{i,n}(t) = n^\eta t^\omega \left[\widehat{\theta}_i(nt) - \theta_i\right]$ *for* $0 \le t \le 1$, $U_i(t) = \sigma_i B_i(t^\upsilon)$ *for* $0 \le t \le 1$, *and* $B_i = [B_i(t) : t \ge 0]$ *a standard Brownian motion.*

The canonical case for the FCLT has $\eta = 1/2$ and $\omega = \upsilon = 1$, in which case the limiting process $U_i(t) = \sigma_i B_i(t)$. The following shows that the FCLT is stronger than the combination of the ordinary CLT in (1) and Anscombe's condition considered in Proposition 1.

Proposition 2 *If Condition 2 holds, then the CLT in (1) and Condition 1 hold, so Assumption 1 also holds.*

We now describe some other simulation settings where Assumptions 1 and 2 hold.

Example 2 **[Steady-state simulations]** Suppose each alternative i corresponds to a stochastic process $X_i = [X_i(t) : t \ge 0]$, which we assume has a steady-state mean θ_i. The goal is to compare the k processes in terms of their steady-state means. Let $\widehat{\theta}_i(t) = (1/t)\int_0^t X_i(s)\,ds$ for $t > 0$, and $\widehat{\theta}_i(0) = 0$, so $\widehat{\theta}_i(t)$ is the time-average of the process X_i over the first t time units. Under a variety of assumptions on the process X_i, the estimator $\widehat{\theta}_i$ satisfies a FCLT with $\eta = 1/2$ and $\omega = \upsilon = 1$, in which case Assumption 1 holds by Proposition 2. For example, under appropriate conditions, the FCLT holds for Markov chains, martingales, and regenerative processes; e.g., see Section 4.4 of Whitt (2002). Various methods have been developed for constructing estimators $V_i(t)$ of σ_i^2 that satisfy Assumption 2 under a variety of conditions. The techniques include the regenerative method (Glynn and Iglehart 1993), spectral methods (Damerdji 1991), autoregressive estimators (Fishman 1978, p. 252), and various batch means and batched area estimators with the number of batches growing to infinity (Damerdji 1994).

Example 3 **[Quantiles]** Suppose that each alternative i is a population with distribution function F_i. For any distribution function G and constant $0 < q < 1$, define $G^{-1}(q) = \inf\{x : G(x) \geq q\}$. For a fixed value $0 < p < 1$, let $\theta_i = F_i^{-1}(p)$, which is the pth quantile of F_i, so we are comparing the alternatives in terms of their pth quantiles. We assume that F_i is differentiable at θ_i, with $F_i'(\theta_i) > 0$, where prime denotes derivative. Let $X_{i,1}, X_{i,2}, \ldots$ be i.i.d. samples from F_i. For each $n \geq 1$, define the empirical distribution function $F_{i,n}$ based on the first n samples from F_i as $F_{i,n}(x) = (1/n)\sum_{j=1}^{n} I\{X_{i,j} \leq x\}$, where $I\{A\}$ is the indicator function of the event $\{A\}$, evaluating to 1 when $\{A\}$ occurs, and 0 otherwise. We define a discrete-time estimator $\widehat{\theta}_{i,n} = F_{i,n}^{-1}(p)$ of θ_i based on n samples from population i. Then the CLT in (1) holds for each population i with $\sigma_i^2 = p(1-p)/(F_i'(\theta_i))^2$; e.g., see p. 77 of Serfling (1980). Anscombe (1952) establishes that Condition 1 holds with $\eta = 1/2$, so Assumption 1 holds by Proposition 1. To construct an estimator $V_{i,n}$ of σ_i^2 that satisfies Assumption 2, first define $q_{i,n} = p + \sqrt{p(1-p)/n} + o(1/\sqrt{n})$, where $h(n) = o(g(n))$ means $h(n)/g(n) \to 0$ as $n \to \infty$. Then $V_{i,n} = n(F_{i,n}^{-1}(q_{i,n}) - F_{i,n}^{-1}(p))^2 \Rightarrow \sigma_i^2$ as $n \to \infty$, so Assumption 2 holds; e.g., see p. 94 of Serfling (1980). Other estimators of σ_i^2 satisfying Assumption 2 could also be used, including the bootstrap quantile variance estimator or smoothed versions of bootstrap estimators (Hall 1992, pp. 319–320).

Many other settings arising in simulation contexts also satisfy Assumptions 1 and 2. These include comparing comparing functions of means of i.i.d. random vectors and Kiefer-Wolfowitz (1952) stochastic approximation. The last case is of special interest because Assumption 1 holds with noncanonical $\eta = 1/3$.

6 CONCLUSIONS

We described a general framework for establishing the asymptotically validity of two-stage procedures for MCB and selection for comparing alternatives. The assumptions for the asymptotic validity are that the alternatives are independent, the estimators satisfy a random-time-change CLT, and we have consistent estimators for the variance parameters appearing in the CLT. Our framework encompasses a wide variety of settings that are of interest to the simulation user, including the comparison of means in transient simulations, steady-state means of stochastic processes, and quantiles of populations.

REFERENCES

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

Bechhofer, R. E. 1954. A single-sample multiple-decision procedure for ranking means of normal populations with known variances. *Annals of Mathematical Statistics* 25:16–39.

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: John Wiley & Sons.

Billingsley, P. 1999. *Convergence of probability measures*. Second ed. New York: John Wiley & Sons.

Chick, S. E. 2006. Subjective probability and bayesian methodology. In *Handbook in Operations Research and Management Science: Simulation*, ed. S. G. Henderson and B. L. Nelson, 225–257. Oxford, UK: Elsevier.

Chow, Y. S., and H. Robbins. 1965. On the asymptotic theory of fixed width sequential confidence intervals for the mean. *Annals of Mathematical Statistics* 36:457–462.

Chung, K. L. 2001. *A course in probability theory*. 3rd ed. San Diego: Academic Press.

Damerdji, H. 1991. Strong consistency and other properties of the spectral variance estimator. *Management Science* 37:1424–1440.

Damerdji, H. 1994. Strong consistency of the variance estimator in steady-state simulation output analysis. *Mathematics of Operations Research* 19:494–512.

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 1996 Winter Simulation Conference*, ed. J. Charnes, D. Morrice, D. Brunner, and J. Swain, 281–286. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Damerdji, H., and M. K. Nakayama. 1999. Two-stage multiple-comparison procedures for steady-state simulations. *ACM Transactions on Modeling and Computer Simulations* 9:1–30.

Fishman, G. S. 1978. *Principles of discrete event simulation*. New York: Wiley.

Glynn, P. W., and D. L. Iglehart. 1993. Conditions for the applicability of the regenerative method. *Management Science* 39:1108–1111.

Glynn, P. W., and S. Juneja. 2004. A large deviations perspective on ordinal optimization. In *Proceedings of the 2004 Winter Simulation Conference*, ed. R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, 577–585. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Hall, P. 1992. *The bootstrap and Edgeworth expansions*. New York: Springer.

Hochberg, Y., and A. C. Tamhane. 1987. *Multiple comparison procedures*. New York: Wiley.

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

Kiefer, J., and J. Wolfowitz. 1952. Stochastic estimation of the maximum of a regression function. *Annals of Mathematical Statistics* 23:462–466.

Kim, S.-H., and B. L. Nelson. 2006. Selecting the best system. In *Elsevier Handbooks in Operations Research and Management Science: Simulation*. Amsterdam: Elsevier.

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

Mukhopadhyay, N., and T. K. S. Solanky. 1994. *Multistage selection and ranking procedures: Second-order asymptotics*. New York: Marcel Dekker.

Nakayama, M. K. 2009. Asymptotic analysis of two-stage selection and multiple-comparion procedures for simulations. Submitted.

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

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

Serfling, R. J. 1980. *Approximation theorems of mathematical statistics*. New York: John Wiley & Sons.

Whitt, W. 2002. *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*. New York: Springer-Verlag.

AUTHOR BIOGRAPHIES

MARVIN K. NAKAYAMA is an associate professor in the Department of Computer Science at the New Jersey Institute of Technology. He received a Ph.D. in operations research from Stanford University. He won second prize in the 1992 George E. Nicholson Student Paper Competition sponsored by INFORMS and is a recipient of a CAREER Award from the National Science Foundation. He is the stochastic models area editor for *ACM Transactions on Modeling and Computer Simulation* and the simulation area editor for *INFORMS Journal on Computing*. His research interests include applied probability, statistics, simulation and modeling. His e-mail address is <marvin@njit.edu>, and his web page is <web.njit.edu/~marvin>.