OPERATIONS RESEARCH
Vol. 54, No. 3, May–June 2006, pp. 475–488
ISSN 0030-364X | EISSN 1526-5463 | 06 | 5403 | 0475

DOI 10.1287/opre.1060.0281

On the Asymptotic Validity of Fully Sequential Selection Procedures for Steady-State Simulation

Seong-Hee Kim
School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332,
skim@isye.gatech.edu

Barry L. Nelson
Department of Industrial Engineering and Management Sciences, Northwestern University,
Evanston, Illinois 60208-3119, nelsonb@northwestern.edu

We present fully sequential procedures for steady-state simulation that are designed to select the best of a finite number of simulated systems when "best" is defined by the largest or smallest long-run average performance. We also provide a framework for establishing the asymptotic validity of such procedures and prove the validity of our procedures. An example based on the $M/M/1$ queue is given.

Subject classifications: simulation: ranking and selection, steady-state simulation.
Area of review: Simulation.
History: Received July 2001; revisions received October 2002, April 2003, March 2004, January 2005; accepted April 2005.

1. Introduction

Statistical procedures have been proposed to select the simulated system with the largest or smallest long-run average performance from among a finite number of alternative systems. In this paper, we focus on the indifference-zone formulation of the selection problem, meaning that we desire procedures that are guaranteed to select the best system with high probability when the long-run average performance of the best is at least a given amount better than the rest (see Chen et al. 2000 and Chick and Inoue 2001 for alternative formulations). Many of these procedures have been constructed under the assumption that the output data generated by each system are independent and identically distributed (i.i.d.), and also marginally normal. The assumption of independence within each system's output, which is appropriate for terminating simulations, is the biggest barrier to applying these procedures directly to steady-state simulation experiments. The outputs within a single replication of a steady-state simulation are typically dependent. For instance, the delays in queue of successive parts processed at a work center may be dependent because each part must wait for the ones ahead of it.

We can apply procedures for i.i.d. normal data to steady-state simulation experiments if we make multiple replications of each alternative, use the within-replication averages as the basic observations, and make the replications long enough that the within-replication averages are approximately normally distributed. Or, we can generate a single long replication of each alternative and use batch means of many individual outputs as the basic observations. In typical simulation output processes, the batch means are much less dependent and more nearly normally distributed than the individual outputs if the batch size is large enough. See Law and Kelton (2000) for a general discussion of replication versus batching, Goldsman and Nelson (1998) for a presentation specialized to selection procedures, and Glynn and Iglehart (1990) for conditions under which the batch means method is asymptotically valid.

Unfortunately, both of these remedies for dependent data have disadvantages. If we make replications, then we have to discard the so-called warm-up period from each one; this will be very inefficient if a large number of observations need to be deleted. Batching within a replication may also be inefficient for the following reason: Selection procedures attempt to minimize the simulation data required to obtain a "correct selection" by working sequentially—meaning two or more stages of sampling—with decisions on how much, if any, additional sampling is needed made at the end of each stage. If a "stage" is defined by batch means, rather than individual observations, then the simulation effort consumed by a stage is a multiple of the batch size. When a large batch size is required to achieve approximate independence—and batch sizes of several thousand are common—then the selection procedure is forced to make decisions at long intervals, wasting observations and time.

These disadvantages have fostered efforts to develop new procedures designed specifically for steady-state simulation, procedures that can be applied to a single replication from each alternative and that use basic outputs rather than batch means. Goldsman and Marshall (1999) extended Rinott's (1978) procedure for use in steady-state simulation. Nakayama (1997) presented

single-stage multiple-comparison procedures, and Damerdji and Nakayama (1999) developed two-stage multiple-comparison procedures, to provide inference on the best system for steady-state simulation. The procedures in all three papers take a single replication of each system, keep the number of sampling stages small (one or two), and do not eliminate any alternatives until the final stage of sampling is completed. The procedures by Nakayama (1997) and Damerdji and Nakayama (1999) were shown to be asymptotically valid.

The inefficiency of batching is particularly troublesome when fully sequential procedures are employed. *Fully sequential procedures* use many stages of sampling, take only a single basic observation from each alternative still in play at each stage, and may eliminate alternatives that appear (with high probability) to be inferior (see Siegmund 1985 for a general reference on fully sequential procedures). Batching constrains these decisions to batch means, rather than individual observations, partially defeating the benefits of using a fully sequential procedure. This is unfortunate, because Kim and Nelson (2001) showed that fully sequential procedures for i.i.d. normal data can be quite efficient. We have empirical evidence that appropriate modifications of Kim and Nelson's procedure $\mathcal{KN}$ also work well in the single-replication, steady-state simulation environment (Goldsman et al. 2000, 2002). Unfortunately, proving that these procedures, or any such procedures, provide a correct-selection guarantee is hopeless unless we make assumptions that are clearly unrealistic for a single replication of a steady-state simulation. *Therefore, a goal of this paper is to provide theoretical support, in terms of an appropriate asymptotic analysis, for what we have observed empirically*. This is important because it establishes conditions under which we can expect fully sequential procedures to work, rather than just relying on anecdotal evidence that they do. *We also provide guidance for actual use, because there are implementation choices that have a substantial impact on the performance of the procedures in finite samples.*

This paper is organized as follows: In §2, we set up the key assumptions for output processes from steady-state simulation and the variance estimators we will employ. In §3, we introduce a framework for establishing the asymptotic validity of fully sequential procedures. Section 4 presents a new procedure in which variance estimators are computed once from a first-stage sample, and proves its asymptotic validity. In §5, we describe another new procedure—one that allows variance updates—and also establish its asymptotic validity. Section 6 contains an illustrative example. We conclude by offering our opinion about future research in §7.

2. Background

This section describes the key assumptions we make for output processes from steady-state simulations. First, we discuss a Functional Central Limit Theorem (FCLT) that shows how to standardize the output data from steady-state simulations to make them behave like Brownian motion processes in the limit. Because we assume unknown variances, we need to estimate the variance of each system to standardize the output. In §2.2, we discuss the properties we require of variance estimators.

2.1. Functional Central Limit Theorem

Suppose that there are k systems. For system $i = 1, 2, \ldots, k$, let $\mathbf{X}_i = \{X_{ij};\ j = 1, 2, \ldots\}$ be a discrete-time stochastic process representing the simulation output of system i (although we work in discrete time, an entirely analogous development can be done for continuous-time output processes). We assume that the processes $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k$ are mutually independent; this rules out the use of common random numbers, an issue we discuss further in §7. However, we allow the joint distribution of each $\mathbf{X}_i$ to be different, and do not require any assumption of normality.

The sample mean of the first r observations from system i is represented by $\overline{X}_i(r)$, and the standardized partial sum for system i, $C_i(t, r)$, is defined as

$$C_i(t, r) \equiv \frac{\sum_{j=1}^{\lfloor rt \rfloor} X_{ij} - rt\mu_i}{v_i\sqrt{r}}, \quad 0 \leqslant t \leqslant 1, \tag{1}$$

where $\lfloor \cdot \rfloor$ indicates truncation of any fractional part. For any $r \geqslant 1$, the random function $C_i(t, r)$ is an element of the Skorohod space $D[0, 1]$ as defined in Chapter 3 of Billingsley (1968). To establish our results, we restrict attention to processes that satisfy the following assumption, called a Functional Central Limit Theorem (see Billingsley 1968, Chapter 4).

Assumption 1. *There exist finite constants μ_i and v_i^2 such that the probability distribution of $C_i(t, r)$ over $D[0, 1]$ converges to that of a standard Brownian motion process, $\mathcal{W}(t)$, for t on the unit interval, as r increases*; *i.e.*,

$$C_i(\cdot, r) \Longrightarrow \mathcal{W}(\cdot)$$

as $r \to \infty$, where $\Longrightarrow$ denotes convergence in distribution. Further, we assume that for every $t \in [0, 1]$, the family of random variables $\{C_i^2(t, r):\ r = 1, 2, \ldots\}$ is uniformly integrable.

Under Assumption 1, μ_i can be identified as the steady-state mean, and

$$v_i^2 = \lim_{r \to \infty} r\mathrm{Var}(\overline{X}_i(r))$$

is the asymptotic variance of system i.

We take Assumption 1 as a given; establishing conditions under which it holds is not our goal. However, one set of sufficient conditions for the FCLT is that $\mathbf{X}_i$ is stationary and ϕ-mixing with appropriate mixing constants (see

p. 166 of Billingsley 1968 for the exact definition). Roughly speaking, in a ϕ-mixing process the distant future is nearly independent of the past and present. Many stochastic processes are ϕ-mixing; e.g., l-dependent processes and finite-state, irreducible, and aperiodic Markov chains.

Our procedures observe the sum of differences between two systems, however, so we need to establish that the standardized partial sum of differences between two systems also behaves like $\mathcal{W}(t)$. We can show this by defining $Z_{i\ell}(j) = X_{ij} - X_{\ell j}$ for $j = 1, 2, \ldots, r$, $i \neq \ell$, and letting

$$C_{i\ell}(t, r) = \frac{\sum_{j=1}^{\lfloor rt \rfloor} Z_{i\ell}(j) - rt(\mu_i - \mu_\ell)}{v_{i\ell}\sqrt{r}}. \tag{2}$$

LEMMA 1. *For $i \neq \ell$, if $\mathbf{X}_i$ and $\mathbf{X}_\ell$ satisfy Assumption* 1 *and are independent, then there exists a constant $v_{i\ell}^2$ such that*

$$C_{i\ell}(\cdot, r) \Longrightarrow \mathcal{W}(\cdot)$$

as $r \to \infty$.

PROOF. First, note that

$$v_{i\ell}^2 = \lim_{r\to\infty} r\mathrm{Var}(\bar{Z}_{i\ell}(r)) = v_i^2 + v_\ell^2$$

because of the independence of $\mathbf{X}_i$ and $\mathbf{X}_\ell$. Let $C_i(t, r)$ and $C_\ell(t, r)$ be the standardized partial sums of $\mathbf{X}_i$ and $\mathbf{X}_\ell$, respectively. Then,

$$\begin{aligned} C_{i\ell}(t, r) &= \frac{\sum_{j=1}^{\lfloor rt \rfloor} Z_{i\ell}(j) - rt(\mu_i - \mu_\ell)}{v_{i\ell}\sqrt{r}} \\ &= \frac{\sum_{j=1}^{\lfloor rt \rfloor} (X_{ij} - X_{\ell j}) - rt(\mu_i - \mu_\ell)}{v_{i\ell}\sqrt{r}} \\ &= \frac{\sum_{j=1}^{\lfloor rt \rfloor} X_{ij} - rt\mu_i}{v_{i\ell}\sqrt{r}} - \frac{\sum_{j=1}^{\lfloor rt \rfloor} X_{\ell j} - rt\mu_\ell}{v_{i\ell}\sqrt{r}} \\ &= \left(\frac{v_i}{v_{i\ell}}\right) C_i(t, r) - \left(\frac{v_\ell}{v_{i\ell}}\right) C_\ell(t, r). \end{aligned}$$

Because we assume that $\mathbf{X}_i$ and $\mathbf{X}_\ell$ are independent, so are $C_i(t, r)$ and $C_\ell(t, r)$. Assumption 1 implies $C_i(\cdot, r) \Longrightarrow \mathcal{W}_i(\cdot)$ and $C_\ell(\cdot, r) \Longrightarrow \mathcal{W}_\ell(\cdot)$, where $\mathcal{W}_i(\cdot)$ and $\mathcal{W}_\ell(\cdot)$ are independent standard Brownian motion processes. By Theorem 3.2 of Billingsley (1968), $(C_i(\cdot, r), C_\ell(\cdot, r)) \Longrightarrow (\mathcal{W}_i(\cdot), \mathcal{W}_\ell(\cdot))$. If $f(t)$ and $g(t)$ are elements of $D[0, 1]$, the mapping, $h(f(t), g(t)) = (v_i/v_{i\ell})f(t) - (v_\ell/v_{i\ell})g(t)$, $0 \leqslant t \leqslant 1$, from $D[0, 1] \times D[0, 1] \to D[0, 1]$, is continuous. Thus, by the Continuous Mapping Theorem (CMT) (see Theorem 5.5 of Billingsley 1968),

$$C_{i\ell}(\cdot, r) \Longrightarrow \left(\frac{v_i}{v_{i\ell}}\right) \mathcal{W}_i(\cdot) - \left(\frac{v_\ell}{v_{i\ell}}\right) \mathcal{W}_\ell(\cdot),$$

which is also a standard Brownian motion process. □

Lemma 1 is critical to establishing the validity of our procedures because it provides a way to standardize the output of a steady-state simulation so that it converges to a known process. We then use properties of Brownian motion to control the chance that our procedures incorrectly choose an inferior system as the best.

The standardized partial sum contains $v_{i\ell}^2$, which is unknown. Thus, we need an estimator of $v_{i\ell}^2$. In the next subsection, we establish the conditions that variance estimators should satisfy to be building blocks for asymptotically valid procedures.

2.2. Variance Estimators

There are several estimators for the asymptotic variance of a stationary stochastic process that are based on the concept of batching. In batching, we partition the r simulation outputs into b batches of m outputs each, so that $r = mb$. Therefore, the jth (nonoverlapping) batch from system i contains the outputs $X_{i,(j-1)m+1}, X_{i,(j-1)m+2}, \ldots, X_{i,jm}$. Although our selection procedures use individual outputs to make elimination decisions, they use batching to estimate the asymptotic variance.

Let mV^2 denote an estimator of the asymptotic variance v^2 of either a single system or the difference between two systems. To establish the validity of our procedures, we need variance estimators with either the *chi-squared property* or the *strong consistency property*. Variance estimators mV^2 with the chi-squared property satisfy the following:

ASSUMPTION 2.

$$mV^2 \Longrightarrow \frac{v^2\chi_d^2}{d} \quad \textit{as } m \to \infty \textit{ with } b \textit{ fixed},$$

where χ_d^2 denotes a chi-squared random variable with d degrees of freedom, and d is determined by the specific variance estimator used.

The batch means estimator and area estimator are known to have the chi-squared property (Schruben 1983, Chien et al. 1997, Sargent et al. 1992). The batch means estimator has degrees of freedom $d = b - 1$, while the area estimator has $d = b$. The overlapping batch means estimator does not have this property, but Meketon and Schmeiser (1984) showed that it is approximately chi-squared distributed for large m. Meketon and Schmeiser (1984) also showed that the overlapping batch means estimator has approximately 3/2 the degrees of freedom of the batch means estimator.

Some variance estimators can be shown to be strongly consistent estimators of the asymptotic variance. To explain this property, we need the concept of a *batching sequence*. A batching sequence is denoted $\{(m_r, b_r), r = 1, 2, \ldots\}$, where m_r and b_r are the batch size and the number of batches, respectively, as functions of the number of outputs r. Variance estimators with the strong consistency property satisfy the following:

ASSUMPTION 3.

$$m_r V^2 \longrightarrow v^2 \quad \textit{with probability } 1 \textit{ as } r \to \infty.$$

In the literature, conditions on (m_r, b_r) that guarantee almost sure convergence (convergence with probability 1) and mean-squared convergence have been established (Damerdji 1994, 1995; Damerdji and Goldsman 1995; Chien et al. 1997). In all cases, both m_r and b_r are nondecreasing functions of r and go to infinity as r increases.

For our procedure that estimates variances only once (denoted $\mathcal{KN}+$), we establish asymptotic validity based on variance estimators satisfying Assumption 2. The validity of our procedure with variance updates (denoted $\mathcal{KN}++$) is proved based on variance estimators satisfying Assumption 3.

3. A Framework for Establishing Asymptotic Validity

Goldsman et al. (2002) discuss different ways to evaluate the "goodness" of a ranking-and-selection procedure. For an indifference-zone procedure designed to find the best among a finite number of systems, the most important measure of performance is the probability of correct selection (PCS). If there is more than one procedure for the same purpose that satisfies the PCS requirement, then the most efficient procedure is superior. However, if there is no procedure that can *guarantee* a prespecified PCS—which is the case in our context—then a procedure that is likely to come close to the desired PCS is needed.

Mathematical analysis of finite-sample performance is the ideal way to establish the PCS. Unfortunately, if assumptions such as normality and independence of the simulation output processes are relaxed, then bounding the PCS is typically impossible and we are forced to consider other approaches. In this section, we propose a framework for *asymptotic analysis*: showing that a procedure achieves at least the desired PCS in an appropriate limit. Our definition of asymptotic PCS is similar to Dalal and Hall (1977) and Mukhopadhyay and Solanky (1994). Sections 4 and 5 apply the framework to procedures $\mathcal{KN}+$ and $\mathcal{KN}++$, while §6 and Goldsman et al. (2002) report on an empirical evaluation of the finite-sample performance of both procedures.

Our selection procedures are based on the concept of an *indifference zone*, which is defined by a parameter $\delta > 0$ that can be interpreted as the smallest difference in steady-state mean performance that is worth detecting. More precisely, we would like our procedures to provide a guaranteed PCS when the best system's steady-state mean is at least δ better (smaller or larger, depending on the problem) than the second-best system's mean. If there are systems whose means are within δ of the best, then we are indifferent as to which one of these good systems is selected.

For simplicity, we discuss the case of only two systems, denoted k and i. We assume that a larger mean is better and that $\mu_k \geqslant \mu_i + \delta$. Once we show the validity of our procedures for $k = 2$, it is not difficult to extend the result to $k \geqslant 2$ systems.

To be more specific, let $S(r) = \sum_{j=1}^{r}(X_{kj} - X_{ij})$. Our procedures are based on tracking the partial sum $S(r)$, $r = n_0, n_0 + 1, \ldots, N$, as long as it stays within a "continuation region" $R(r)$ of finite length $r = 1, 2, \ldots, N$. Depending on how $S(r)$ exits $R(r)$, we either select system k (if it exits the region going up) or system i (if it exits the region going down) as best. If the length of the continuation region is proportional to mV^2/δ^2—as it is in our procedures—then as $\delta \to 0$, we have $N \to \infty$.

To make probability statements about how $S(r)$ will exit $R(r)$, we consider how a standardized version of $S(r)$ exits a correspondingly modified version of $R(r)$. Specifically, we standardize $S(r)$ over the range $0 \leqslant r \leqslant N$ as in (2); this yields $C_{ki}(t, N) + (\mu_k - \mu_i) \cdot \Delta(N) \cdot t$ over the interval $0 \leqslant t \leqslant 1$. The "leftover" term $\Delta(N)$ is positive and a function of both δ and mV^2 (through N). The term $\mu_k - \mu_i > 0$ tends to push the process toward exiting the continuation region going up, which is the correct direction. Therefore, to bound the PCS, we replace $\mu_k - \mu_i$ by δ, which is no larger than the true difference. As $\delta \to 0$, we employ the FCLT (and other arguments that depend on the specific procedure) to show that $C_{ki}(t, N) + \delta \cdot \Delta(N) \cdot t \Longrightarrow \mathcal{W}(t) + \Delta t$, a standard Brownian motion process with positive drift Δ. We then show that the PCS is attained for the limiting process, which gives us some hope that it will be approximately attained for realistic, finite-sample-size problems (although this cannot be guaranteed).

To show the asymptotic validity of our procedures, we let the parameter that defines the indifference zone, δ, go to zero, which drives the sample size to infinity. Of course, if the true difference $\mu_k - \mu_i > 0$ is fixed, then as the sample size goes to infinity, the PCS $\to 1$ (which is clearly greater than $1 - \alpha$, as claimed, but is not very meaningful). There are at least two ways to view an asymptotic analysis where $\delta \to 0$ that provide insight into when the procedure will work well in practice:

(1) In reality, we do not know the true differences between the steady-state means. The most important case for an indifference-zone ranking-and-selection procedure is when the true differences are small and we demand to be able to detect small differences (when the true differences are small but δ is large, or the true differences are large but δ is small, we are unlikely to make a mistake). The asymptotic validity of the proposed procedures shows that if the true, unknown differences are small, even vanishingly small, then the procedures will achieve approximately the desired PCS when we also require them to detect small differences.

(2) Let

$$X_{kj} = \mu_k + \varepsilon_{kj},$$
$$X_{ij} = \mu_i + \varepsilon_{ij} \tag{3}$$

represent the output processes from systems k and i, where $\{\varepsilon_{kj},\ j = 1, 2, \ldots\}$ and $\{\varepsilon_{ij},\ j = 1, 2, \ldots\}$ are independent, mean-zero, stationary stochastic processes satisfying

Assumption 1 (this is the model adopted by Schruben 1983, for instance). Then, let $\mu_k - \mu_i = \delta$ so that as $\delta \to 0$ the *true difference* between the systems' means goes to zero also. Under this regime, the PCS does not go to one, and our analysis shows that as the problem becomes more and more difficult (the true differences in the means, and the differences we desire to distinguish, become smaller and smaller), the procedure's PCS becomes at least as large as the desired PCS. See Lehmann (1999, §3.3) for an in-depth justification of this type of analysis.

Model (3) has the convenient feature that properties of the standardized difference process $C_{ki}(t, N)$ are not functions of μ_k and μ_i because they are only location parameters. In general, properties of a stochastic process, such as its dependence structure, may be linked to its mean. Our point is not to *actually* change the means of the processes of interest, however, but rather to demonstrate that the procedures can be expected to work in an appropriate asymptotic sense, as analysis under model (3) does. To make the development unambiguous, we will assume that model (3) holds throughout the remainder of the paper.

We now present two fully sequential procedures for steady-state simulation and prove the validity of each one using the framework discussed here.

4. $\mathcal{KN}+$ Procedure

In this section, we describe $\mathcal{KN}+$, a procedure that does not update the variance estimators after the first stage of sampling. We also prove the asymptotic validity of $\mathcal{KN}+$. The procedure is formulated with the goal of finding the system with the largest steady-state mean, but can be modified in the obvious way to select the system with the smallest mean.

$\mathcal{KN}+$ Procedure

Setup. Select confidence level $1/k < 1 - \alpha < 1$, indifference-zone parameter $\delta > 0$, first-stage sample size $n_0 \geqslant 2$, and batch size $m_0 < n_0$. Calculate η and c as described below.

Initialization. Let $I = \{1, 2, \ldots, k\}$ be the set of systems still in contention, and let $h^2 = 2c\eta d$, where the degrees of freedom d is determined by which variance estimator is used.

Obtain n_0 observations X_{ij}, $j = 1, 2, \ldots, n_0$, from each system $i = 1, 2, \ldots, k$.

For all $i \neq \ell$, compute the estimator $m_0 V_{i\ell}^2$, the sample asymptotic variance of the difference between systems i and ℓ. Note that $m_0 V_{i\ell}^2$ is based only on the first n_0 observations. Let

$$N_{i\ell} = \left\lfloor \frac{h^2 m_0 V_{i\ell}^2}{\delta^2} \right\rfloor$$

and let

$$N_i = \max_{\ell \neq i} N_{i\ell}.$$

Here, $N_i + 1$ is the maximum number of observations that will be taken from system i. If $n_0 \geqslant \max_i N_i + 1$, then stop and select the system with the largest $\overline{X}_i(n_0)$ as the best.

Otherwise, set the observation counter $r = n_0$ and go to **Screening**.

Screening. Set $I^{\text{old}} = I$. Let

$$I = \{i\colon i \in I^{\text{old}} \text{ and } \overline{X}_i(r) > \overline{X}_\ell(r) - W_{i\ell}(r), \forall\, \ell \in I^{\text{old}}, \ell \neq i\},$$

where

$$W_{i\ell}(r) = \max\left\{0, \frac{\delta}{2cr}\left(\frac{h^2 m_0 V_{i\ell}^2}{\delta^2} - r\right)\right\}.$$

Stopping Rule. If $|I| = 1$, then stop and select the system whose index is in I as the best.

Otherwise, take one additional observation $X_{i,r+1}$ from each system $i \in I$, set $r = r + 1$, and go to **Screening**.

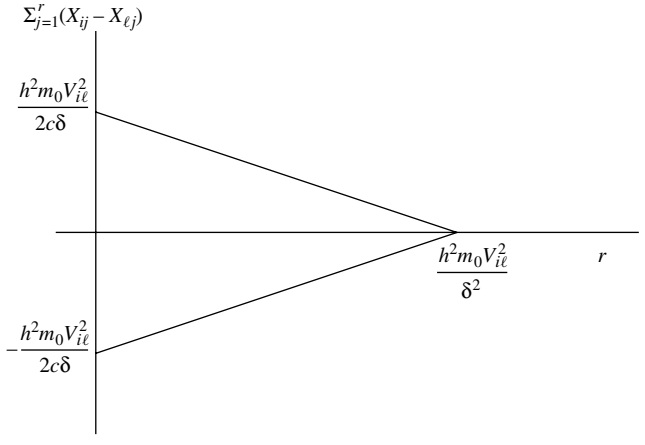
Constants. The constant c may be any nonnegative integer. The constant η is the solution to the equation

$$g(\eta) = \sum_{\ell=1}^{c} (-1)^{\ell+1}\left(1 - \frac{1}{2}\mathcal{I}(\ell = c)\right) \cdot \left(1 + \frac{2\eta(2c-\ell)\ell}{c}\right)^{-d/2} = 1 - (1-\alpha)^{1/(k-1)}, \qquad (4)$$

where $\mathcal{I}$ is the indicator function.

In our procedures, $r \cdot W_{i\ell}(r)$ defines a continuation region for the partial sum, $\sum_{j=1}^{r}(X_{ij} - X_{\ell j})$. Figure 1 shows the continuation region for $\mathcal{KN}+$. If $c < \infty$, this region is a triangle, as shown in the figure. As c increases the triangle becomes longer, but narrower, and in the limit becomes parallel lines. As long as the partial sum stays within the continuation region, sampling continues; sampling stops when it exits the region.

Figure 1. Continuation region for the fully sequential, indifference-zone procedure when $c < \infty$. The horizontal axis represents the current sample size (stage) r and the vertical axis represents $\sum_{j=1}^{r}(X_{ij} - X_{\ell j})$.



To guarantee a unique solution to (4), $c = 1$ should be chosen. Kim and Nelson (2001) suggest that $c = 1$ is the best compromise choice when the experimenter has no idea if there are a few dominant systems or a number of close competitors. Also, it is easy to compute η when $c = 1$. Goldsman et al. (2000, 2002) present empirical results when $c = 1$, and we will also use $c = 1$ in our example.

To prove the validity of the procedure, we need a result due to Fabian (1974).

LEMMA 2 (FABIAN 1974). *Let $\mathcal{W}(\cdot, \Delta)$ be a Brownian motion process on $[0, +\infty)$, with $\mathrm{E}[\mathcal{W}(t, \Delta)] = \Delta \cdot t$ and $\mathrm{Var}[\mathcal{W}(t, \Delta)] = t$, where $\Delta > 0$. Let*

$$L(t) = -\mathcal{A} + \mathcal{B}t,$$
$$U(t) = \mathcal{A} - \mathcal{B}t$$

for some $\mathcal{A} > 0$ and $\mathcal{B} = \Delta/(2c)$ for some positive integer c. Let $R(t)$ denote the interval $(L(t), U(t))$ and let T^* be the first time that $\mathcal{W}(t, \Delta) \notin R(t)$. Finally, let $\mathcal{E}$ be the event that $\mathcal{W}(T^*, \Delta) \leqslant L(T^*)$. Then,

$$\Pr\{\mathcal{E}\} = \sum_{\ell=1}^{c} (-1)^{\ell+1} \left(1 - \frac{1}{2}\mathcal{I}(\ell = c)\right) \exp\{-2\mathcal{A}\mathcal{B}(2c - \ell)\ell\}.$$

REMARK. The event $\mathcal{E}$ will correspond to an incorrect selection (incorrectly eliminating the best system from consideration by exiting the region going down).

Without loss of generality, suppose that the true steady-state means of the systems are indexed so that $\mu_k \geqslant \mu_{k-1} \geqslant \cdots \geqslant \mu_1$. Now we present our main result:

THEOREM 1. *If $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k$ are independent, each satisfies Assumption* 1 *and Equation* (3)*, and $m_0 V_{i\ell}^2$ is distributed as $v_{i\ell}^2 \chi_d^2/d$ and is asymptotically independent of $C_{i\ell}(\cdot, r)$, then $\liminf_{\delta \to 0} \Pr\{\mathcal{KN}+ \text{ selects } k\} \geqslant 1 - \alpha$ provided $\mu_k \geqslant \mu_{k-1} + \delta$.*

REMARK. The requirement that $m_0 V_{i\ell}^2$ is asymptotically independent of $C_{i\ell}(\cdot, r)$ is plausible because the variance estimator $m_0 V_{i\ell}^2$ is based on a fixed initial sample of n_0 observations from each system. Asymptotic independence implies that as $r \to \infty$ (that is, more and more observations are collected), the behavior of the standardized partial sum $C_{i\ell}(\cdot, r)$ depends less and less on the initial sample.

PROOF. We begin by considering the case of only two systems, denoted k and i, with $\mu_k \geqslant \mu_i + \delta$. Select a value of η such that $g(\eta) = 1 - (1 - \alpha)^{1/(k-1)}$ from Equation (4).

With probability 1, there exists δ_0 such that $N_{ik} \geqslant n_0$ for all $\delta \leqslant \delta_0$ because $m_0 V_{ik}^2$ is finite and positive with probability 1. For δ smaller than δ_0, let

$$T(\delta) = \min\{r\colon r \geqslant n_0 \text{ and } |\overline{X}_k(r) - \overline{X}_i(r)| \geqslant W_{ik}(r)\}.$$

Thus, $T(\delta)$ is the stage at which the procedure terminates by leaving the continuation region. Let ICS denote the event that an incorrect selection is made. Then,

$$\begin{aligned}
&\Pr\{\mathrm{ICS}\} \\
&= \Pr\{\overline{X}_k(T(\delta)) - \overline{X}_i(T(\delta)) \leqslant -W_{ik}(T(\delta))\} \\
&= \Pr\left\{\sum_{j=1}^{T(\delta)} (X_{kj} - X_{ij}) \leqslant -T(\delta) \cdot W_{ik}(T(\delta))\right\} \\
&= \Pr\left\{\sum_{j=1}^{T(\delta)} (X_{kj} - X_{ij}) \leqslant \min\left\{0, \frac{-h^2 m_0 V_{ik}^2}{2\delta c} + \frac{\delta T(\delta)}{2c}\right\}\right\} \\
&= \Pr\left\{\frac{\sum_{j=1}^{T(\delta)} (X_{kj} - X_{ij}) - (\mu_k - \mu_i)T(\delta)}{v_{ik}\sqrt{N_{ik}+1}} + \frac{(\mu_k - \mu_i)T(\delta)}{v_{ik}\sqrt{N_{ik}+1}}\right. \\
&\qquad \left. \leqslant \min\left\{0, \frac{-h^2 m_0 V_{ik}^2}{2\delta c v_{ik}\sqrt{N_{ik}+1}} + \frac{\delta T(\delta)}{2c v_{ik}\sqrt{N_{ik}+1}}\right\}\right\} \\
&\leqslant \Pr\left\{\frac{\sum_{j=1}^{T(\delta)} (X_{kj} - X_{ij}) - (\mu_k - \mu_i)T(\delta)}{v_{ik}\sqrt{N_{ik}+1}} + \frac{\delta T(\delta)}{v_{ik}\sqrt{N_{ik}+1}}\right. \\
&\qquad \left. \leqslant \min\left\{0, \frac{-h^2 m_0 V_{ik}^2}{2\delta c v_{ik}\sqrt{N_{ik}+1}} + \frac{\delta T(\delta)}{2c v_{ik}\sqrt{N_{ik}+1}}\right\}\right\} \\
&= \mathrm{E}\left[\Pr\left\{\frac{\sum_{j=1}^{T(\delta)} (X_{kj} - X_{ij}) - (\mu_k - \mu_i)T(\delta)}{v_{ik}\sqrt{N_{ik}+1}} + \frac{\delta T(\delta)}{v_{ik}\sqrt{N_{ik}+1}}\right.\right. \\
&\qquad \left.\left. \leqslant \min\left\{0, \frac{-h^2 m_0 V_{ik}^2}{2\delta c v_{ik}\sqrt{N_{ik}+1}} + \frac{\delta T(\delta)}{2c v_{ik}\sqrt{N_{ik}+1}}\right\} \middle| m_0 V_{ik}^2\right\}\right].
\end{aligned} \tag{5}$$

The inequality arises because we replace $\mu_k - \mu_i$ with δ, which is no larger. To establish the result, we will show that $\limsup_{\delta \to 0} \Pr\{\mathrm{ICS}\} \leqslant 1 - (1 - \alpha)^{1/(k-1)}$. To do so, let

$$C_{ki}(t, \delta) = \frac{\sum_{j=1}^{\lfloor (N_{ik}+1)t \rfloor} (X_{kj} - X_{ij}) - (N_{ik}+1)(\mu_k - \mu_i)t}{v_{ik}\sqrt{N_{ik}+1}}$$

for $0 \leqslant t \leqslant 1$, where we express C_{ki} as a function of δ, instead of $N_{ik} + 1$, because N_{ik} is a function of δ. Further, define

$$\begin{aligned}
\widehat{T}(\delta) = \min\Bigg\{ t \in \Bigg\{ &\frac{n_0}{N_{ik}+1}, \frac{n_0+1}{N_{ik}+1}, \ldots, 1 \Bigg\}\colon \\
&\left| C_{ki}(t, \delta) + \frac{(N_{ik}+1)\delta t}{v_{ik}\sqrt{N_{ik}+1}} \right| \\
&\qquad \geqslant \frac{h^2 m_0 V_{ik}^2}{2\delta c v_{ik}\sqrt{N_{ik}+1}} - \frac{(N_{ik}+1)\delta t}{2c v_{ik}\sqrt{N_{ik}+1}} \Bigg\}.
\end{aligned}$$

Clearly, $\widehat{T}(\delta) = T(\delta)/(N_{ik} + 1)$. Also, define the stopping time of the corresponding continuous-time process as

$$\begin{aligned}
\widetilde{T}(\delta) = \min\Bigg\{ t \geqslant \frac{n_0}{N_{ik}+1}\colon &\left| C_{ki}(t, \delta) + \frac{(N_{ik}+1)\delta t}{v_{ik}\sqrt{N_{ik}+1}} \right| \\
&\qquad \geqslant \frac{h^2 m_0 V_{ik}^2}{2\delta c v_{ik}\sqrt{N_{ik}+1}} - \frac{(N_{ik}+1)\delta t}{2c v_{ik}\sqrt{N_{ik}+1}} \Bigg\}.
\end{aligned}$$

Note that for fixed δ, $C_{ki}(\widehat{T}(\delta), \delta)$ corresponds to the right-hand limit of a point of discontinuity of $C_{ki}(\cdot, \delta)$. We can show that $\widehat{T}(\delta) \to \widetilde{T}(\delta)$ with probability 1 as $\delta \to 0$, making use of the fact that $1/(N_{ik}+1) \to 0$ with probability 1. Thus, in the limit, we can focus on $C_{ki}(\widetilde{T}(\delta), \delta)$.

Now, condition on $m_0 V_{ik}^2$. Then, Assumption 1, Lemma 1, and the CMT imply that

$$C_{ki}(t, \delta) + \frac{(N_{ik}+1)\delta t}{v_{ik}\sqrt{N_{ik}+1}} \Longrightarrow \mathscr{W}(t, \Delta)$$

as $\delta \to 0$, where

$$\Delta = \lim_{\delta \to 0} \frac{(N_{ik}+1)\delta}{v_{ik}\sqrt{N_{ik}+1}} = \frac{\sqrt{h^2 m_0 V_{ik}^2}}{v_{ik}}.$$

Still conditional on $m_0 V_{ik}^2$, let

$$\mathscr{A}(\delta) = \frac{h^2 m_0 V_{ik}^2}{2\delta c v_{ik}\sqrt{N_{ik}+1}} \xrightarrow{\delta \to 0} \frac{\sqrt{h^2 m_0 V_{ik}^2}}{2c v_{ik}} \equiv \mathscr{A},$$

$$\mathscr{B}(\delta) = \frac{(N_{ik}+1)\delta}{2c v_{ik}\sqrt{N_{ik}+1}} \xrightarrow{\delta \to 0} \frac{\sqrt{h^2 m_0 V_{ik}^2}}{2c v_{ik}} \equiv \mathscr{B}.$$

Note that the stopping time $\widetilde{T}(\delta)$ is the first time t at which the event

$$\left\{ \left| C_{ki}(t, \delta) + \frac{(N_{ik}+1)\delta t}{v_{ik}\sqrt{N_{ik}+1}} \right| - \mathscr{A}(\delta) + \mathscr{B}(\delta) t \geqslant 0 \right\}$$

occurs. Define the mapping s_δ: $D[0,1] \to \Re$ such that $s_\delta(Y) = Y(T_{Y,\delta})$, where

$$T_{Y,\delta} = \inf\{t\colon |Y(t)| - \mathscr{A}(\delta) + \mathscr{B}(\delta) t \geqslant 0\}$$

for every $Y \in D[0,1]$ and $\delta > 0$. Similarly, define $s(Y) = Y(T_Y)$, where

$$T_Y = \inf\{t\colon |Y(t)| - \mathscr{A} + \mathscr{B} t \geqslant 0\}$$

for every $Y \in D[0,1]$ and $\delta > 0$. Note that

$$s_\delta\left(C_{ki}(t, \delta) + \frac{(N_{ik}+1)\delta t}{v_{ik}\sqrt{N_{ik}+1}} \right) = C_{ki}(\widetilde{T}(\delta), \delta) + \frac{(N_{ik}+1)\delta \widetilde{T}(\delta)}{v_{ik}\sqrt{N_{ik}+1}},$$

$$s(\mathscr{W}(\cdot, \Delta)) = \mathscr{W}(T_{\mathscr{W}(\cdot,\Delta)}, \Delta).$$

We need to show that

$$s_\delta(\mathscr{G}_{ki}(\cdot, \delta)) \Longrightarrow s(\mathscr{W}(\cdot, \Delta))$$

as $\delta \to 0$, where

$$\mathscr{G}_{ki}(t, \delta) \equiv C_{ki}(t, \delta) + \frac{(N_{ik}+1)\delta t}{v_{ik}\sqrt{N_{ik}+1}}$$

for $t \in [0,1]$ and $\delta > 0$. This follows from Proposition 2 of Kim et al. (2005), which establishes that the extended CMT (Theorem 5.5 of Billingsley 1968) applies.

Now, unconditioning on $m_0 V_{ik}^2$ gives

$$\begin{aligned}
&\limsup_{\delta \to 0} \Pr\{\text{ICS}\} \\
&\leqslant \mathrm{E}[\Pr\{\mathscr{W}(t, \Delta) \text{ exits continuation region through the lower boundary} \mid m_0 V_{ik}^2\}] \\
&= \mathrm{E}\left[\sum_{\ell=1}^{c} (-1)^{\ell+1} \left(1 - \frac{1}{2} \mathcal{I}(\ell = c) \right) \cdot \exp\left\{ -2 \frac{h^2 m_0 V_{ik}^2}{(2c)^2 v_{ik}^2} (2c - \ell)\ell \right\} \right] \quad \text{(by Lemma 2)} \\
&= \mathrm{E}\left[\sum_{\ell=1}^{c} (-1)^{\ell+1} \left(1 - \frac{1}{2} \mathcal{I}(\ell = c) \right) \cdot \exp\left\{ -\frac{\eta d m_0 V_{ik}^2}{c v_{ik}^2} (2c - \ell)\ell \right\} \right] \\
&= \sum_{\ell=1}^{c} (-1)^{\ell+1} \left(1 - \frac{1}{2} \mathcal{I}(\ell = c) \right) \left(1 + \frac{2\eta(2c - \ell)\ell}{c} \right)^{-d/2} \\
&= 1 - (1 - \alpha)^{1/(k-1)},
\end{aligned}$$

where we make use of the fact that $m_0 V_{ik}^2$ and $C_{ki}(t)$ are asymptotically independent, and that the argument of expectation (5) is bounded by one for all δ, so it is uniformly integrable. The final equality follows from the way we choose η and the chi-squared property of $m_0 V_{ik}^2$.

Now consider $k \geqslant 2$ systems and let CS be the event that k is selected (correct selection) and let ICS_i be the event that an incorrect selection is made when systems k and i are considered *in isolation*. Then,

$$\begin{aligned}
1 - \Pr\{\text{CS}\} &= \Pr\left\{ \bigcup_{i=1}^{k-1} (\text{system } i \text{ eliminates } k) \right\} \\
&\leqslant \sum_{i=1}^{k-1} \Pr\{\text{ICS}_i\},
\end{aligned}$$

which implies that

$$\begin{aligned}
&\liminf_{\delta \to 0} \Pr\{\text{CS}\} \\
&\geqslant \liminf_{\delta \to 0} \left(1 - \sum_{i=1}^{k-1} \Pr\{\text{system } i \text{ eliminates system } k\} \right) \\
&\geqslant \liminf_{\delta \to 0} \Pr\left\{ \bigcap_{i=1}^{k-1} (\text{system } k \text{ eliminates system } i) \right\} \quad \text{(by DeMorgan's law)} \\
&= \mathrm{E}\left[\liminf_{\delta \to 0} \Pr\left\{ \bigcap_{i=1}^{k-1} (\text{system } k \text{ eliminates system } i) \;\middle|\; X_{k1}, \ldots, X_{k(N_k+1)}, m_0 V_{ik}^2 \right\} \right]
\end{aligned}$$

$$= \mathrm{E}\left[\liminf_{\delta\to 0}\prod_{i=1}^{k-1}\Pr\left\{\text{system } k \text{ eliminates system } i \mid X_{k1},\ldots,X_{k(N_k+1)}, m_0V_{ik}^2\right\}\right] \quad (6)$$

because the events {system k eliminates system i} are asymptotically conditionally independent given $X_{k1},\ldots,X_{k(N_k+1)}$ and $m_0V_{ik}^2$. Thus, by applying the same argument that we used to bound the probability of correct selection for the pair (i,k), and noting that the conditional probability is a bounded, continuous function of the condition, we get

$$\begin{aligned}
&\liminf_{\delta\to 0}\Pr\{\mathrm{CS}\}\\
&\quad\geqslant \mathrm{E}\left[\liminf_{\delta\to 0}\left(\Pr\left\{\text{system } k \text{ eliminates system } i \mid X_{k1},\ldots,X_{k(N_{ik}+1)}, m_0V_{ik}^2\right\}\right)^{k-1}\right]\\
&\quad\geqslant \left(\liminf_{\delta\to 0}\Pr\{\text{system } k \text{ eliminates system } i\}\right)^{k-1}\\
&\qquad\qquad\text{(by Jensen's inequality).}
\end{aligned}$$

As a result,

$$\liminf_{\delta\to 0}\Pr\{\mathrm{CS}\}\geqslant\left(1-\limsup_{\delta\to 0}\Pr\{\mathrm{ICS}_i\}\right)^{k-1}$$

and $\liminf_{\delta\to 0}\Pr\{\mathrm{CS}\}\geqslant 1-\alpha$. □

Remark. Note that we assume a chi-squared distribution for the first-stage variance estimator, rather than converging to it as the sample size increases (in $\mathcal{KN}+$ the first-stage sample size is fixed). Is it reasonable to make this assumption? There are at least two ways that it can fail. If the initial sample size is not large enough, then the distribution of $m_0V_{ik}^2$ may not be well approximated by a scaled chi-squared distribution. Even if the chi-squared distribution is appropriate, the scaling constant may not be v_{ik}^2 in finite samples, meaning that $m_0V_{ik}^2$ is biased. Goldsman et al. (2002) looked at the impact of both deviations from the chi-squared assumption. When the output data actually come from an AR(1) process, they showed that the bias in $m_0V_{ik}^2$ does not seriously degrade PCS as long as there are a small number of large batches in the first stage. They also found this to be true in an extensive empirical study. Theorem 1 shows that if we can get n_0 right, then making elimination decisions one observation at a time can be justified, in a limiting sense. In the next section, we present a procedure that overcomes the need to get n_0 "right."

5. $\mathcal{KN}++$ Procedure

The variance estimators employed in $\mathcal{KN}+$ depend only on the first-stage data. In this section, we present a refinement of $\mathcal{KN}+$ in which we update the variance estimators as more data are obtained. The asymptotic validity of $\mathcal{KN}++$ depends on having variance estimators with the strong consistency property. Thus, $\mathcal{KN}++$ requires a batching sequence (m_r, b_r) that ensures that the strong consistency property holds. There are several different batching sequences that achieve almost sure convergence or mean-square convergence. Goldsman et al. (2000, 2002) modified existing batching sequences to introduce more updates for $\mathcal{KN}++$ and performed experiments to examine the performance of these batching sequences. We use one of them in §6. It is also worth noting that weak (convergence in probability) consistency of variance estimators is not sufficient to prove the asymptotic validity of $\mathcal{KN}++$ (Theorem 2 below); see Glynn and Whitt (1992).

In this section, we highlight the differences between $\mathcal{KN}+$ and $\mathcal{KN}++$ and defer the full details of the procedure to the appendix. The primary differences are that $\mathcal{KN}++$ does not use the same critical values h^2 and η, and the batch size in $\mathcal{KN}++$ becomes a function of r rather than being a fixed constant. Therefore, whenever m_r changes value (due to an increase in r), the variance estimators $m_rV_{i\ell}^2(r)$ and all other quantities that depend on them—including $N_{i\ell}(r)$, $N_i(r)$, and $W_{i\ell}(r)$—are recomputed in a new step called **Update**. The **Initialization**, **Screening**, and **Stopping Rule** steps are essentially unchanged.

In the appendix we prove the following theorem:

Theorem 2. *If $\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_k$ are independent, each satisfies Assumption* 1 *and Equation* (3)*, and m_r is an integer-valued nondecreasing function of r such that $m_rV_{i\ell}^2(r)\longrightarrow v_{i\ell}^2$ with probability one (thus it satisfies Assumption* 3*), then, $\liminf_{\delta\to 0}\Pr\{\mathcal{KN}++ \text{ selects } k\}\geqslant 1-\alpha$ provided $\mu_k\geqslant\mu_{k-1}+\delta$.*

The continuation region that is defined by $r\cdot W_{i\ell}(r)$ is critical to achieving the desired PCS. If it is too wide, then a fully sequential procedure loses its ability to detect inferior systems. On the other hand, if it is too narrow, then the likelihood of eliminating good systems is greater than desired. After choosing a batching sequence, a variance estimator and a value of c, the continuation region is completely determined by h^2 and $m_rV_{i\ell}^2(r)$.

The strong consistency property of variance estimators leads to a proof of asymptotic validity that is similar to a known-variances case, and the value of h^2 used in $\mathcal{KN}++$ is the same as if the asymptotic variances were known. Not surprisingly, h^2 in $\mathcal{KN}++$ is smaller than h^2 in $\mathcal{KN}+$ which treats the variances as unknown. As a result, the continuation region for $\mathcal{KN}++$ is narrower than the continuation region for $\mathcal{KN}+$. Although asymptotically valid, this narrower continuation region could cause early elimination of good systems, and the performance of $\mathcal{KN}++$ for finite samples might suffer (see Goldsman et al. 2000). We can overcome this disadvantage by using a larger value of h^2 than is required to prove asymptotic validity. Goldsman et al. (2002) used h^2 computed as in $\mathcal{KN}+$ with $c=1$, and their experiments showed that $\mathcal{KN}++$ works better with

this adjustment. Procedure $\mathcal{KN}{+}{+}$ is still asymptotically valid with the larger value of h^2.

6. Example

In this section, we illustrate the performance of our new procedures using an example based on the $M/M/1$ queue, and from these experiment results we glean guidelines for choosing m_0 and the variance estimator to use *in practice*. Goldsman et al. (2002) present an extensive empirical evaluation based on a larger, but different, collection of examples, and our recommendations are also influenced by their results. As a baseline, we compare our procedures to procedure $\mathcal{R}+$, due to Goldsman and Marshall (1999), which modifies Rinott's (1978) procedure by replacing the usual sample variance of each system with an estimator of the asymptotic variance. Procedure $\mathcal{R}+$ has two stages: In the first stage, an initial sample of size n_0 is taken from each system and used to estimate the asymptotic variance and compute the second-stage sample size. In the second stage, additional observations are taken from each system and the system with the largest or smallest sample mean is selected as the best. Procedure $\mathcal{R}+$ has no elimination step.

The performance measure in our example is w_i, the expected waiting time in the queue of system i. Thus, *smaller w_i is better*. We set the parameters of the systems so that system 1 is always the best.

The number of systems in each experiment varied over $k = 2, 5, 10$. Three variance estimators were tested: the nonoverlapping batch means (BM) estimator, the overlapping batch means (OBM) estimator, and a weighted area (A) estimator. Detailed specifications may be found in Goldsman et al. (2000, 2002).

We chose the first-stage sample size n_0 such that the ratio of the variance of n_0 observations ($v^2(n_0) = n_0\mathrm{Var}[\bar{X}(n_0)]$) and the asymptotic variance is approximately equal to 1; more specifically, $|1 - v^2(n_0)/v^2| \approx 0.01$. This guarantees that enough data are available to estimate the asymptotic variance, but not so much data that it is easy to estimate v^2. After n_0 was determined empirically, all divisors of n_0 were employed as batch sizes m_0, implying n_0/m_0 batches for BM and A, and $n_0 - m_0 + 1$ batches for OBM.

The indifference-zone parameter was set to $\delta = v_1/\sqrt{n_0}$, where v_1^2 is the asymptotic variance of the best system. Thus, δ is approximately the standard deviation of the first-stage sample mean of the best system.

In all cases the service rate for system 1 was set to one, and the service rates of the other systems were set to obtain different configurations of the means, w_i. The arrival rate is the same for all systems in a particular experiment, but varied over $\lambda = 0.3, 0.6, 0.9$ in different experiments. This allowed us to obtain traffic intensities for system 1 of $\rho = 0.3, 0.6, 0.9$. However, because larger ρ implies stronger dependence, a procedure that works well for a large value of ρ will do so for smaller values. For this reason, we only present experiment results for $\rho = 0.9$ here.

For configurations of w_i, we considered the slippage configuration (SC)—which was not considered in Goldsman et al. (2002)—and monotonically decreasing means (MDM). In SC, $w_1 = \rho^2/\lambda(1-\rho)$, while $w_2 = \cdots = w_k = w_1 + \delta$. In MDM, $w_i = w_1 + (i-1)\delta$.

For $\mathcal{KN}{+}{+}$, the modified batching sequence in Goldsman et al. (2002) and the value of h^2 computed as in $\mathcal{KN}+$ were used. This batching sequence is essentially $m_r = \sqrt{r}$, but with more frequent updates of the number of batches when r is small.

For each configuration, 1,000 macroreplications of the entire experiment were performed. In all experiments, the nominal probability of correct selection was set to $1 - \alpha = 0.95$.

Tables 1–4 give results for the average number of observations and estimated PCS when there are $k = 5$ systems, traffic intensity is $\rho = 0.9$, and the initial number of observations from each system is $n_0 = 24{,}000$. Tables 1 and 3 show that the sample average of the total number of basic observations decreases as m_0, the initial batch size, decreases. Tables 2 and 4 show that the estimated PCS also decreases as m_0 decreases. Thus, a large initial batch size

Table 1. Sample average of the total number of basic observations when $M/M/1$ processes are tested under the MDM configuration, $k = 5$, $\rho = 0.9$, and $n_0 = 24{,}000$.

	$\mathcal{R}+$			$\mathcal{KN}+$			$\mathcal{KN}{+}{+}$		
m_0	BM	OBM	AREA	BM	OBM	AREA	BM	OBM	AREA
24,000			389.23			13,878.83			
12,000	391.24	158.52	150.61	1,329.52	65.49	67.85	7.89	7.99	7.44
8,000	154.15	82.62	81.59	66.38	27.27	27.74	6.91	6.98	6.61
6,000	82.40	50.97	57.49	26.68	18.03	17.60	6.48	6.51	6.19
4,800	60.04	44.70	46.48	18.04	14.10	13.24	6.04	6.05	5.68
4,000	50.18	38.81	40.12	14.22	12.00	10.82	5.85	5.85	5.36
3,000	39.58	33.27	30.90	10.44	9.73	8.36	5.44	5.43	4.80
2,400	35.32	31.31	25.79	8.94	8.48	6.78	5.15	5.12	4.25
2,000	32.39	28.69	21.59	8.00	7.70	5.76	4.92	4.93	3.99
1,600	28.86	26.20	17.07	7.07	6.86	4.60	4.70	4.72	3.51
1,000	22.76	21.48	10.51	5.48	5.40	2.82	4.11	4.11	2.49

Note. All numbers are in units of 10^5.

Table 2. Estimated PCS when $M/M/1$ processes are tested under the MDM configuration, $k = 5$, $\rho = 0.9$, and $n_0 = 24{,}000$.

	$\mathcal{R}+$			$\mathcal{KN}+$			$\mathcal{KN}++$		
m_0	BM	OBM	AREA	BM	OBM	AREA	BM	OBM	AREA
24,000			0.986			0.987			
12,000	0.984	0.994	0.971	0.979	0.997	0.979	0.975	0.982	0.959
8,000	0.976	0.982	0.957	0.992	0.990	0.979	0.970	0.987	0.967
6,000	0.968	0.967	0.962	0.985	0.984	0.974	0.961	0.979	0.967
4,800	0.970	0.959	0.952	0.978	0.981	0.972	0.958	0.965	0.955
4,000	0.957	0.956	0.951	0.974	0.975	0.956	0.966	0.971	0.959
3,000	0.956	0.950	0.949	0.965	0.969	0.962	0.967	0.972	0.957
2,400	0.953	0.948	0.940	0.961	0.970	0.959	0.971	0.971	0.960
2,000	0.956	0.945	0.934	0.965	0.965	0.941	0.963	0.970	0.943
1,600	0.949	0.941	0.934	0.959	0.958	0.942	0.948	0.946	0.915
1,000	0.936	0.938	0.911	0.946	0.950	0.896	0.959	0.958	0.909

helps to achieve the nominal PCS, but at the cost of a larger total number of observations. Note that $\mathcal{R}+$ outperforms $\mathcal{KN}+$ only when we have just one degree of freedom for the variance estimator. Note also that $\mathcal{KN}++$ is much less sensitive to the choice of initial batch size m_0 because m_r and b_r increase as more data are obtained, which helps to correct a poor initial variance estimate.

The fully sequential procedures are more efficient under MDM than under SC. The SC is a difficult configuration for procedures that eliminate systems because all inferior systems are exactly δ from being the best. Comparing Table 1 and Table 3, we find that $\mathcal{R}+$ consumes more observations under MDM than under SC; this is simply because larger means yield larger variances in our $M/M/1$ example, and the total sample size for $\mathcal{R}+$ depends only on the variances, not on the means. On the other hand, $\mathcal{KN}+$ and $\mathcal{KN}++$ consume fewer observations under MDM than under SC, because under MDM it is relatively easy to detect and eliminate inferior systems.

Tables 1 and 3 show that $\mathcal{KN}+$ and $\mathcal{KN}++$ can effectively eliminate inferior systems while maintaining the desired PCS when implemented with appropriate choices of variance estimator and batching strategy. These results (and others not reported) give us insight into how to make the choices. OBM appears to be the best variance estimator for our purposes. For good performance of all of the procedures, we need the initial batch size m_0 large enough to achieve the nominal PCS, while not being so large that it implies a huge total sample size. For $\mathcal{R}+$ and $\mathcal{KN}+$, we suggest setting m_0 to produce at least five degrees of freedom for the variance estimator. For $\mathcal{KN}++$, we recommend choosing a very large batch size, such as $n_0/4$, because updating allows the variance estimator to improve as the procedure progresses. Procedure $\mathcal{R}+$ should only be used if implementing a fully sequential procedure is infeasible. Similarly, $\mathcal{KN}+$ should be chosen only if variance updating is too difficult. Because Procedure $\mathcal{KN}++$, combined with the OBM variance estimator, is highly efficient and robust, it should be used if possible.

7. Conclusion and Future Work

In this paper, we proved the asymptotic validity of two new fully sequential selection procedures for steady-state simulation, and also provided a general framework for such

Table 3. Sample average of the total number of basic observations when $M/M/1$ processes are tested under the SC configuration, $k = 5$, $\rho = 0.9$, and $n_0 = 24{,}000$.

	$\mathcal{R}+$			$\mathcal{KN}+$			$\mathcal{KN}++$		
m_0	BM	OBM	AREA	BM	OBM	AREA	BM	OBM	AREA
24,000			371.16			1,684.48			
12,000	374.40	109.54	105.43	1,700.85	83.66	82.67	8.89	9.10	8.64
8,000	109.01	52.46	52.61	90.80	36.10	35.42	7.93	8.11	7.71
6,000	52.04	30.86	35.29	34.47	24.26	24.14	7.38	7.44	7.09
4,800	36.56	26.40	29.19	22.39	19.09	17.88	7.02	7.01	6.69
4,000	29.57	22.95	24.66	18.31	16.27	14.49	6.77	6.81	6.41
3,000	23.82	20.29	20.07	13.86	12.94	10.96	6.26	6.29	5.59
2,400	20.91	18.74	16.85	11.86	11.20	8.87	5.95	6.01	5.30
2,000	19.10	17.52	14.71	10.35	10.15	7.64	6.00	6.00	4.96
1,600	17.75	16.64	12.40	9.13	9.02	6.23	5.51	5.51	4.17
1,000	14.23	13.95	8.02	7.08	6.98	3.58	4.90	4.91	2.92

Note. All numbers are in the units of 10^5.

Table 4. Estimated PCS when $M/M/1$ processes are tested under the SC configuration, $k = 5$, $\rho = 0.9$, and $n_0 = 24{,}000$.

	$\mathcal{R}+$			$\mathcal{KN}+$			$\mathcal{KN}++$		
m_0	BM	OBM	AREA	BM	OBM	AREA	BM	OBM	AREA
24,000			0.959			0.971			
12,000	0.958	0.984	0.928	0.982	0.989	0.956	0.910	0.949	0.892
8,000	0.937	0.940	0.888	0.958	0.968	0.938	0.905	0.932	0.894
6,000	0.911	0.891	0.869	0.942	0.943	0.920	0.899	0.914	0.896
4,800	0.894	0.878	0.867	0.928	0.932	0.913	0.901	0.920	0.895
4,000	0.878	0.864	0.854	0.931	0.927	0.905	0.889	0.894	0.878
3,000	0.866	0.849	0.839	0.910	0.908	0.880	0.891	0.904	0.863
2,400	0.863	0.831	0.826	0.897	0.898	0.874	0.895	0.900	0.863
2,000	0.847	0.834	0.817	0.892	0.890	0.848	0.879	0.889	0.826
1,600	0.832	0.835	0.803	0.870	0.878	0.837	0.878	0.882	0.816
1,000	0.825	0.817	0.753	0.840	0.845	0.746	0.825	0.844	0.719

proofs. Even though asymptotic validity does not imply that PCS $\geqslant 1 - \alpha$ for finite samples, it does suggest that our procedures will work well under difficult situations, such as when our indifference level is small.

The procedures presented here are asymptotically valid under very general conditions, including variances that are unknown and unequal, output data that are nonnormal, and output data from *within* each system that are dependent in series. However, they do not account for dependence *across* the outputs from different systems due to the use of common random numbers (CRN). Although we suspect that our procedures are still valid when CRN is employed, they do not exploit it. Our experience with procedures for i.i.d. data that do incorporate CRN leads us to believe that further gains in efficiency are possible, making the development of such procedures for steady-state simulation an important open research problem.

Appendix

In this appendix, we give full details of $\mathcal{KN}++$ and prove Theorem 2.

$\mathcal{KN}++$ Procedure

Setup. Select confidence level $1/k < 1 - \alpha < 1$, indifference-zone parameter $\delta > 0$, first-stage sample size $n_0 \geqslant 2$, and initial batch size $m_0 < n_0$. Calculate η and c as described below.

Initialization. Let $I = \{1, 2, \ldots, k\}$ be the set of systems still in contention, and let $h^2 = 2c\eta$.

Obtain n_0 observations X_{ij}, $j = 1, 2, \ldots, n_0$, from each system $i = 1, 2, \ldots, k$.

Set the observation counter $r = n_0$ and $m_r = m_0$.

Update. If m_r has changed since the last update, then for all $i \neq \ell$, $i, \ell \in I$, compute estimator $m_r V_{i\ell}^2(r)$, the sample asymptotic variance of the difference between systems i and ℓ based on b_r batches of size m_r. Let

$$N_{i\ell}(r) = \left\lfloor \frac{h^2 m_r V_{i\ell}^2(r)}{\delta^2} \right\rfloor$$

and let

$$N_i(r) = \max_{\ell \neq i} N_{i\ell}(r).$$

If $r \geqslant \max_i N_i(r) + 1$, then stop and select the system in I with the largest $\bar{X}_i(r)$ as the best.

Otherwise go to **Screening**.

Screening. Set $I^{\text{old}} = I$. Let

$$I = \{i\colon i \in I^{\text{old}} \text{ and } \bar{X}_i(r) > \bar{X}_\ell(r) - W_{i\ell}(r)\ \forall \ell \in I^{\text{old}}, \ell \neq i\},$$

where

$$W_{i\ell}(r) = \max\left\{0, \frac{\delta}{2cr}\left(\frac{h^2 m_r V_{i\ell}^2(r)}{\delta^2} - r\right)\right\}.$$

Stopping Rule. If $|I| = 1$, then stop and select the system whose index is in I as the best.

Otherwise, take one additional observation $X_{i,r+1}$ from each system $i \in I$, set $r = r + 1$, and go to **Update**.

Constants. The constant c may be any nonnegative integer. The constant η is the solution to the equation

$$g(\eta) = \sum_{\ell=1}^{c} (-1)^{\ell+1}\left(1 - \frac{1}{2}\mathcal{I}(\ell = c)\right)\exp\left(-\frac{\eta}{c}(2c - \ell)\ell\right) = 1 - (1 - \alpha)^{1/(k-1)}, \tag{7}$$

where $\mathcal{I}$ is the indicator function.

For proof of the validity of $\mathcal{KN}++$, we need a lemma from Billingsley (1968, Theorem 4.4, p. 27), in addition to Lemma 2.

LEMMA 3. *If* $Y_n \Longrightarrow Y$ *and* $Z_n \xrightarrow{P} a$, *where* a *is a constant, then* $(Y_n, Z_n) \Longrightarrow (Y, a)$.

PROOF OF THEOREM 2. We begin by considering the case of only two systems, denoted k and i, with $\mu_k \geqslant \mu_i + \delta$. We also assume that v_{ik}^2 is known, so that

$$N_{ik}(r) = N_{ik} = \left\lfloor \frac{h^2 v_{ik}^2}{\delta^2} \right\rfloor$$

for all r, where we select the value of η so that $g(\eta) = 1 - (1-\alpha)^{1/(k-1)}$ from Equation (7). We relax the assumption of known asymptotic variance later.

Let

$$T(\delta) = \min\{r\colon r \geqslant n_0 \text{ and } |\overline{X}_k(r) - \overline{X}_i(r)| \geqslant W_{ik}(r)\}.$$

Thus, $T(\delta)$ is the stage at which the procedure terminates by leaving the continuation region.

Let ICS denote the event that an incorrect selection is made. Then,

$$\begin{aligned}
&\Pr\{\text{ICS}\}\\
&= \Pr\{\overline{X}_k(T(\delta)) - \overline{X}_i(T(\delta)) \leqslant -W_{ik}(T(\delta))\}\\
&= \Pr\left\{\sum_{j=1}^{T(\delta)}(X_{kj} - X_{ij}) \leqslant -T(\delta)\cdot W_{ik}(T(\delta))\right\}\\
&= \Pr\left\{\sum_{j=1}^{T(\delta)}(X_{kj} - X_{ij}) \leqslant \min\left\{0, \frac{-h^2 v_{ik}^2}{2\delta c} + \frac{\delta T(\delta)}{2c}\right\}\right\}\\
&= \Pr\left\{\frac{\sum_{j=1}^{T(\delta)}(X_{kj} - X_{ij}) - (\mu_k - \mu_i)T(\delta)}{v_{ik}\sqrt{N_{ik}+1}} + \frac{(\mu_k - \mu_i)T(\delta)}{v_{ik}\sqrt{N_{ik}+1}}\right.\\
&\qquad \left.\leqslant \min\left\{0, \frac{-h^2 v_{ik}^2}{2\delta c v_{ik}\sqrt{N_{ik}+1}} + \frac{\delta T(\delta)}{2c v_{ik}\sqrt{N_{ik}+1}}\right\}\right\}\\
&\leqslant \Pr\left\{\frac{\sum_{j=1}^{T(\delta)}(X_{kj} - X_{ij}) - (\mu_k - \mu_i)T(\delta)}{v_{ik}\sqrt{N_{ik}+1}} + \frac{\delta T(\delta)}{v_{ik}\sqrt{N_{ik}+1}}\right.\\
&\qquad \left.\leqslant \min\left\{0, \frac{-h^2 v_{ik}^2}{2\delta c v_{ik}\sqrt{N_{ik}+1}} + \frac{\delta T(\delta)}{2c v_{ik}\sqrt{N_{ik}+1}}\right\}\right\}. \qquad (8)
\end{aligned}$$

The inequality arises because we replace $\mu_k - \mu_i$ with δ, which is no larger. To establish the result, we will show that $\limsup_{\delta\to 0}\Pr\{\text{ICS}\} \leqslant 1 - (1-\alpha)^{1/(k-1)}$. To do so, let

$$C_{ki}(t,\delta) = \frac{\sum_{j=1}^{\lfloor (N_{ik}+1)t\rfloor}(X_{kj} - X_{ij}) - (N_{ik}+1)(\mu_k - \mu_i)t}{v_{ik}\sqrt{N_{ik}+1}}$$

for $0 \leqslant t \leqslant 1$, where we express C_{ki} as a function of δ, instead of $N_{ik}+1$, because N_{ik} is a function of δ. Further, define

$$\begin{aligned}
\widehat{T}(\delta) = \min\Bigg\{&t \in \left\{\frac{n_0}{N_{ik}+1}, \frac{n_0+1}{N_{ik}+1}, \ldots, 1\right\}\colon\\
&\left|C_{ki}(t,\delta) + \frac{(N_{ik}+1)\delta t}{v_{ik}\sqrt{N_{ik}+1}}\right|\\
&\qquad \geqslant \frac{h^2 v_{ik}^2}{2\delta c v_{ik}\sqrt{N_{ik}+1}} - \frac{(N_{ik}+1)\delta t}{2c v_{ik}\sqrt{N_{ik}+1}}\Bigg\}.
\end{aligned}$$

Clearly, $\widehat{T}(\delta) = T(\delta)/(N_{ik}+1)$. Also, define the stopping time of the corresponding continuous-time process as

$$\begin{aligned}
\widetilde{T}(\delta) = \min\Bigg\{&t \geqslant \frac{n_0}{N_{ik}+1}\colon \left|C_{ki}(t,\delta) + \frac{(N_{ik}+1)\delta t}{v_{ik}\sqrt{N_{ik}+1}}\right|\\
&\qquad \geqslant \frac{h^2 v_{ik}^2}{2\delta c v_{ik}\sqrt{N_{ik}+1}} - \frac{(N_{ik}+1)\delta t}{2c v_{ik}\sqrt{N_{ik}+1}}\Bigg\}.
\end{aligned}$$

Note that for fixed δ, $C_{ki}(\widehat{T}(\delta),\delta)$ corresponds to the right-hand limit of a point of discontinuity of $C_{ki}(\cdot,\delta)$. We can show that $\widehat{T}(\delta) \to \widetilde{T}(\delta)$ with probability 1 as $\delta \to 0$, making use of the fact that $1/(N_{ik}+1) \to 0$ with probability 1. Thus, in the limit, we can focus on $C_{ki}(\widetilde{T}(\delta),\delta)$.

Then, Assumption 1, Lemma 1, and the CMT imply that

$$C_{ki}(t,\delta) + \frac{(N_{ik}+1)\delta t}{v_{ik}\sqrt{N_{ik}+1}} \Longrightarrow \mathscr{W}(t,\Delta)$$

as $\delta \to 0$, where

$$\Delta = \lim_{\delta\to 0}\frac{(N_{ik}+1)\delta}{v_{ik}\sqrt{N_{ik}+1}} = h.$$

Let

$$\begin{aligned}
\mathscr{A}(\delta) &= \frac{h^2 v_{ik}^2}{2\delta c v_{ik}\sqrt{N_{ik}+1}} \xrightarrow{\delta\to 0} \frac{h}{2c} \equiv \mathscr{A},\\
\mathscr{B}(\delta) &= \frac{(N_{ik}+1)\delta}{2c v_{ik}\sqrt{N_{ik}+1}} \xrightarrow{\delta\to 0} \frac{h}{2c} \equiv \mathscr{B}. \qquad (9)
\end{aligned}$$

Note that the stopping time $\widetilde{T}(\delta)$ is the first time t at which the event

$$\left\{\left|C_{ki}(t,\delta) + \frac{(N_{ik}+1)\delta t}{v_{ik}\sqrt{N_{ik}+1}}\right| - \mathscr{A}(\delta) + \mathscr{B}(\delta)t \geqslant 0\right\}$$

occurs. Define the mapping $s_\delta\colon D[0,1] \to \Re$ such that $s_\delta(Y) = Y(T_{Y,\delta})$, where

$$T_{Y,\delta} = \inf\{t\colon |Y(t)| - \mathscr{A}(\delta) + \mathscr{B}(\delta)t \geqslant 0\}$$

for every $Y \in D[0,1]$ and $\delta > 0$. Similarly, define $s(Y) = Y(T_Y)$, where

$$T_Y = \inf\{t\colon |Y(t)| - \mathscr{A} + \mathscr{B}t \geqslant 0\}$$

for every $Y \in D[0,1]$ and $\delta > 0$. Note that

$$\begin{aligned}
&s_\delta\left(C_{ki}(t,\delta) + \frac{(N_{ik}+1)\delta t}{v_{ik}\sqrt{N_{ik}+1}}\right)\\
&\qquad = C_{ki}(\widetilde{T}(\delta),\delta) + \frac{(N_{ik}+1)\delta\widetilde{T}(\delta)}{v_{ik}\sqrt{N_{ik}+1}},\\
&s(\mathscr{W}(\cdot,\Delta)) = \mathscr{W}(T_{\mathscr{W}(\cdot,\Delta)},\Delta).
\end{aligned}$$

We need to show that

$$s_\delta(\mathscr{G}_{ki}(\cdot,\delta)) \Longrightarrow s(\mathscr{W}(\cdot,\Delta)) \qquad (10)$$

as $\delta \to 0$, where

$$\mathscr{G}_{ki}(t,\delta) \equiv C_{ki}(t,\delta) + \frac{(N_{ik}+1)\delta t}{v_{ik}\sqrt{N_{ik}+1}}$$

for $t \in [0,1]$ and $\delta > 0$. This follows from Proposition 2 of Kim et al. (2005), which establishes that the extended CMT (Theorem 5.5 of Billingsley 1968) applies.

Therefore, by Lemma 1 and (10),

$$\begin{aligned}
&\limsup_{\delta\to 0}\Pr\{\text{ICS}\}\\
&\quad\leqslant \Pr\{\mathscr{W}(t,\Delta) \text{ exits continuation region through the lower boundary}\}\\
&\quad=\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1-\frac{1}{2}\mathscr{I}(\ell=c)\right)\\
&\qquad\cdot\exp\left\{-2\frac{h^2}{(2c)^2}(2c-\ell)\ell\right\}\quad\text{(by Lemma 2)}\\
&\quad=\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1-\frac{1}{2}\mathscr{I}(\ell=c)\right)\\
&\qquad\cdot\exp\left\{-\frac{\eta}{c}(2c-\ell)\ell\right\}=1-(1-\alpha)^{1/(k-1)},
\end{aligned}$$

where the equality follows from the way we choose η.

Now consider $k\geqslant 2$ systems and let CS be the event that k is selected (correct selection) and let ICS_i be the event that an incorrect selection is made when systems k and i are considered *in isolation*. Thus, by applying the same argument that we used to bound the probability of correct selection for the pair (i,k), we get

$$\liminf_{\delta\to 0}\Pr\{\text{CS}\}\geqslant\left(1-\limsup_{\delta\to 0}\Pr\{\text{ICS}_i\}\right)^{k-1}.$$

Thus,

$$\liminf_{\delta\to 0}\Pr\{\text{CS}\}\geqslant 1-\alpha.$$

This argument establishes the asymptotic validity of a special case of $\mathscr{KN}{+}{+}$ in which v_{ik}^2 is known. To prove the validity of $\mathscr{KN}{+}{+}$ in general, we replace v_{ik}^2 by a strongly consistent estimator of it. Therefore—considering again the case $k=2$—at termination we have

$$N_{ik}(T(\delta))=\left\lfloor\frac{h^2 m_{T(\delta)}V_{ik}^2(T(\delta))}{\delta^2}\right\rfloor.$$

To establish the result, we note the following:

(1) With probability 1, $T(\delta)$ goes to infinity as $\delta\to 0$. This is because the continuation region implied by δ' contains the continuation region implied by δ if $\delta'<\delta$.

(2) As a consequence of 1, the number of sampling stages goes to infinity as $\delta\to 0$, insuring that $m_{T(\delta)}V_{ik}^2(T(\delta))$ converges to v_{ik}^2 with probability 1. Because strong consistency implies convergence in probability, Lemma 3 can be applied to $(C_{ik}(T(\delta),\delta), m_{T(\delta)}V_{ik}^2(T(\delta)))$.

(3) As a consequence of 2, $N_{ik}(T(\delta))$ goes to infinity with probability 1 as $\delta\to 0$.

(4) As a consequence of 3, and an application of the random-change-of-time theorem (Billingsley 1968, Theorem 17.1), the standardized difference still converges in distribution to $\mathscr{W}(t,\Delta)$ with $\Delta=h$. The terms in (9) also converge as in the known variance case.

A subtle point in the derivation is that in a finite sample we may have $T(\delta)>N_{ik}(T(\delta))+1$ if a variance update occurs at time $T(\delta)$. However, in the limit $\tilde{T}(\delta)\leqslant 1$ because it is not possible for $\mathscr{W}(t,\Delta)$ to exit the continuation region for the first time beyond the end of the region. □

Remark. That item 1 is correct might not be immediately obvious. To simplify the exposition, let $v\equiv v_{ik}^2=1$, and represent the output process $X_{kj}-X_{ij}$ under the slippage configuration by $Z_j+\delta$, where Z_j is a mean-zero, stationary process satisfying our assumptions. Then, the upper boundary of the continuation region is defined by the intercept $h^2/(2c\delta)$ and slope $-\delta/(2c)$. Let $T(\delta)$ be the first time $\sum_{j=1}^{r}Z_j+\delta r$ leaves the region. We consider paths that leave the region going up; an entirely analogous argument applies for paths that leave going down.

Consider a sample path (going up) for which $\limsup_{\delta\to 0}T(\delta)=T^*<\infty$. For this event to occur, there must exist $\delta^*>0$ (that may depend on the sample path) such that for all $\delta\leqslant\delta^*$,

$$\sum_{j=1}^{T^*}Z_j+T^*\delta\geqslant\frac{h^2}{2c\delta}-\frac{\delta T^*}{2c},\tag{11}$$

which can be rewritten as

$$\sum_{j=1}^{T^*}Z_j+\left(\frac{2c+1}{2c}\right)T^*\delta\geqslant\frac{h^2}{2c\delta}.\tag{12}$$

Note that for T^* finite, $\sum_{j=1}^{T^*}Z_j$ is finite with probability 1, but $T^*\delta$ goes to 0 while $h^2/(2c\delta)$ goes to infinity as $\delta\to 0$. Thus, (12) has probability 0.

Acknowledgments

This project was supported by NSF grants DMI-9622065 and DMI-0217690, Rockwell Software, and Symix/Pritsker. The authors thank the referees and the editors for their extensive advice and patience; Dave Goldsman for his cheerful willingness to endure detailed discussions of the technical details of the paper; and James R. Wilson for proving a difficult result that was needed in the proofs of Theorems 1 and 2.

References

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

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

Chick, S., K. Inoue. 2001. New two-stage and sequential procedures for selecting the best simulated system. *Oper. Res.* **49** 1609–1624.

Chien, C., D. Goldsman, B. Melamed. 1997. Large-sample results for batch means. *Management Sci.* **43** 1288–1295.

Dalal, S. R., W. J. Hall. 1977. Most economical robust selection procedures for location parameters. *Ann. Statist.* **7** 1321–1328.

Damerdji, H. 1994. Strong consistency of the variance estimator in steady-state simulation output analysis. *Math. Oper. Res.* **19** 494–512.

Damerdji, H. 1995. Mean-square consistency of the variance estimator in steady-state simulation output analysis. *Oper. Res.* **43** 282–291.

Damerdji, H., D. Goldsman. 1995. Consistency of several variants of the standardized time series area variance estimator. *Naval Res. Logist.* **42** 1161–1176.

Damerdji, H., M. K. Nakayama. 1999. Two-stage multiple-comparison procedures for steady-state simulations. *ACM TOMACS* **9** 1–30.

Fabian, V. 1974. Note on Anderson's sequential procedures with triangular boundary. *Ann. Statist.* **2** 170–176.

Glynn, P. W., D. L. Iglehart. 1990. Simulation output analysis using standardized time series. *Math. Oper. Res.* **15** 1–16.

Glynn, P. W., W. Whitt. 1992. The asymptotic validity of sequential stopping rules for stochastic simulations. *Ann. Appl. Probab.* **2** 180–198.

Goldsman, D., W. S. Marshall. 1999. Selection procedures with standardized time series variance estimators. P. A. Farrington, H. B. Nembhard, D. T. Sturrock, G. W. Evans, eds. *Proc. 1999 Winter Simulation Conf.*, Institute of Electrical and Electronics Engineers, Piscataway, NJ, 382–388.

Goldsman, D., B. L. Nelson. 1998. Comparing systems via simulation. J. Banks, ed. *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*. John Wiley and Sons, New York, 273–306.

Goldsman, D., S.-H. Kim, W. S. Marshall, B. L. Nelson. 2000. Ranking and selection for steady-state simulation. J. Joines, R. R. Barton, P. Fishwick, K. Kang, eds. *Proc. 2000 Winter Simulation Conf.*, Institute of Electrical and Electronics Engineers, Piscataway, NJ, 544–553.

Goldsman, D., S.-H. Kim, W. Marshall, B. L. Nelson. 2002. Ranking and selection procedures for steady-state simulation: Perspectives and procedures. *INFORMS J. Comput.* **14** 2–19.

Kim, S.-H., B. L. Nelson. 2001. A fully sequential procedure for indifference-zone selection in simulation. *ACM TOMACS* **11** 251–273.

Kim, S.-H., B. L. Nelson, J. R. Wilson. 2005. Some almost-sure convergence properties useful in sequential analysis. *Sequential Anal.* **24**(4) 411–419.

Law, A. M., D. Kelton. 2000. *Simulation Modeling and Analysis*, 3rd ed. McGraw-Hill, New York.

Lehmann, E. L. 1999. *Elements of Large-Sample Theory*. Springer, New York.

Meketon, M. S., B. Schmeiser. 1984. Overlapping batch means: Something for nothing? S. Sheppard, U. Pooch, D. Pedgen, eds. *Proc. 1984 Winter Simulation Conf.*, Institute of Electrical and Electronics Engineers, Piscataway, NJ, 227–230.

Mukhopadhyay, N., T. K. S. Solanky. 1994. *Multistage Selection and Ranking Procedures: Second-Order Asymptotics*. Marcel Dekker, New York.

Nakayama, M. K. 1997. Multiple-comparison procedures for steady-state simulations. *Ann. Statist.* **25** 2433–2450.

Rinott, Y. 1978. On two-stage selection procedures and related probability-inequalities. *Comm. Statist.—Theory Methods* **A7** 799–811.

Sargent, R. G., K. Kang, K. Goldsman. 1992. An investigation of finite-sample behavior of confidence interval estimators. *Oper. Res.* **40** 898–913.

Schruben, L. 1983. Confidence interval estimation using standardized time series. *Oper. Res.* **31** 1090–1108.

Siegmund, D. 1985. *Sequential Analysis: Tests and Confidence Intervals*. Springer-Verlag, New York.