Proceedings of the 1995 Winter Simulation Conference
ed. C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman

SELECTING THE BEST SYSTEM IN STEADY-STATE SIMULATIONS USING BATCH MEANS

Marvin K. Nakayama

Department of Computer and Information Science
New Jersey Institute of Technology
Newark, NJ 07102, U.S.A.

ABSTRACT

Suppose that we want to compare k different systems, where μ_i denotes the steady-state mean performance of system i. Our goal is to use simulation to pick the "best" system (i.e., the one with the largest or smallest steady-state mean). To do this, we present some two-stage procedures based on the method of batch means. Our procedures also construct multiple-comparisons-with-the-best (MCB) confidence intervals for $\mu_i - \max_{j\neq i}\mu_j$, $i = 1, \ldots, k$. Under the assumption of an indifference zone of (absolute or relative) width δ, we can show that asymptotically (as $\delta \to 0$ with the size of the batches proportional to $1/\delta^2$), the joint probability of correctly selecting the best system and of the MCB confidence intervals simultaneously containing $\mu_i - \max_{j\neq i}\mu_j$, $i = 1, \ldots, k$, is at least $1 - \alpha$, where α is prespecified by the user.

1 INTRODUCTION

Suppose that we want to compare k different systems (i.e., stochastic processes), where system i has (unknown) steady-state mean μ_i and (unknown) asymptotic variance σ_i^2. We allow for the variances to be unequal. Our goal is to run independent simulations of the various systems to determine which has the largest (or smallest) steady-state mean. For example, the different systems may represent various service disciplines in a queueing system, and we want to select the one that will result in the largest steady-state throughput.

Since the steady-state means are not known and have to be estimated, we can never be certain that the system we eventually choose is actually the best one. Thus, we desire a procedure that (under certain assumptions) will correctly select the best system with some (prespecified) high probability. Also, we may be indifferent between two systems if their steady-state means are very close in value. This leads us to define an *indifference zone* (i.e., an interval whose upper endpoint is given by the largest steady-state mean) of width $\delta > 0$ (which the user prespecifies), and we assume that we are equally satisfied with any system having a mean lying in the indifference zone. This type of problem formulation has its origins in the work of Bechhofer (1954). Moreover, we want to construct simultaneous confidence intervals for $\mu_i - \max_{j\neq i}\mu_j$, for $i = 1, 2, \ldots, k$. The confidence intervals are known as *multiple comparisons with the best* (MCB); see Hsu (1984).

In this paper, we propose some two-stage procedures based on the method of batch means for simultaneous indifference-zone selection and MCB. It can be shown (see Nakayama 1995) that asymptotically (if the batch sizes are proportional to $1/\delta^2$ and the indifference-zone width $\delta \to 0$), the joint probability that our procedure makes a correct selection (i.e., it selects the best system or a system whose mean lies in the indifference zone) and that the true differences $\mu_i - \max_{j\neq i}\mu_j$, $i = 1, 2, \ldots, k$, simultaneously lie in the MCB confidence intervals is at least $1 - \alpha$ (which the user prespecifies). We present procedures for both absolute- and relative-width indifference zones.

Our two-stage procedure for absolute-width indifference zones generalizes some previous results established for i.i.d. normal random variables. Specifically, Rinott (1978) developed a procedure for absolute-width indifference-zone selection for i.i.d. normals, and he proved that the probability of correct selection is at least $1-\alpha$. Matejcik and Nelson (1992) modified Rinott's method to also construct MCB confidence intervals for i.i.d. normals with an absolute-width indifference zone, and they proved that the joint probability of correct selection and simultaneous MCB coverage is at least $1 - \alpha$ (also see Hsu 1984). For a review of these and other procedures, see Goldsman and Nelson (1994). None of these papers covers relative-width indifference zones, as we do here.

The rest of the paper has the following organization. In Section 2 we define the notation used and state an assumption on the processes being simulated.

We present our procedures in Section 3. Section 4 contains a brief discussion on how to specify values for the parameters needed to run our procedures.

2 NOTATION AND ASSUMPTIONS

Suppose that there are k systems, labeled $1, 2, \ldots, k$, that we want to compare. For system $i = 1, 2, \ldots, k$, let $\mathbf{Y}_i = \{Y_i(t) : t \geq 0\} \in D[0, \infty)$ be a real-valued stochastic process representing the simulation output of system i, where $D[0, \infty)$ is the space of right continous functions on $[0, \infty)$ having left limits (see Ethier and Kurtz 1986 or Glynn 1990 for more details on the space $D[0, \infty)$). Essentially all stochastic processes arising in practice have sample paths lying in $D[0, \infty)$.

Let μ_i and $\sigma_i^2 > 0$ be the (unknown) steady-state mean and (unknown) asymptotic variance, respectively, of $\mathbf{Y}_i$. We assume that $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_k$ are mutually independent. (In practice, this means that for all i and j with $j \neq i$, the simulations of systems i and j are generated using non-overlapping streams of uniform random numbers.) Also, let $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_k)$ and $Y(t) = (Y_1(t), Y_2(t), \ldots, Y_k(t))$. Let $(1), (2), \ldots, (k)$ be defined such that $\mu_{(1)} \leq \mu_{(2)} \leq \cdots \leq \mu_{(k)}$, and the exact values of $(1), (2), \ldots, (k)$ are unknown to us. In other words, system (j) has the jth smallest steady-state mean, and our goal is to determine the value of (k).

To establish our results, we need to assume that our process $\mathbf{Y}$ satisfies a functional central limit theorem (FCLT). More specifically, let "$\Rightarrow$" denote weak convergence (see Billingsley 1968), and then assume the following:

A1 *There exist a nonsingular $k \times k$ matrix Σ and a constant $\mu = (\mu_1, \mu_2, \ldots, \mu_k) \in \Re^k$ such that*

$$X_\delta \Rightarrow \Sigma B$$

as $\delta \to 0$, where B is a k-dimensional standard Brownian motion, $X_\delta = (X_{1,\delta}, X_{2,\delta}, \ldots, X_{k,\delta})$, and

$$X_{i,\delta}(t) = \frac{1}{\delta}\left(\frac{\int_0^{t/\delta^2} Y_i(s)ds}{1/\delta^2} - \mu_i t\right), \quad t \geq 0,$$

for $i = 1, 2, \ldots, k$.

The constants μ_i appearing in A1 are precisely the steady-state means of the process $\mathbf{Y}$. Also, the elements of $\Sigma = (\sigma_{i,j} : i, j = 1, \ldots, k)$ satisfy $\sigma_{i,i} = \sigma_i^2$ for all $i = 1, \ldots, k$, and $\sigma_{i,j} = 0$ for $i \neq j$. Thus, Σ is the asymptotic covariance matrix of the process $\mathbf{Y}$.

Virtually all stochastic systems arising in the "real world" and having a steady state satisfy the FCLT in Assumption A1. For example, Assumption A1 holds if the process $\mathbf{Y}$ satisfies any of the following: $\mathbf{Y}$ is regenerative and satisfies suitable moment conditions (see Glynn and Whitt 1987); $\mathbf{Y}$ is a martingale process (see Chapter 7 of Ethier and Kurtz 1986); $\mathbf{Y}$ satisfies appropriate mixing conditions (see Chapter 7 of Ethier and Kurtz 1986); or the $\mathbf{Y}(t)$ are associated (see Newman and Wright 1981).

3 OUR PROCEDURES

First we consider an absolute-width indifference zone; i.e., the indifference zone is defined to be the interval $(\mu_{(k)} - \delta,\ \mu_{(k)}]$. Our goal is to select a system having a steady-state mean lying in the indifference zone and to specify simultaneous confidence intervals for $\mu_i - \max_{i \neq j} \mu_j$, $i = 1, 2, \ldots, k$.

The basic idea of our two-stage procedures is as follows. In the first stage we run independent simulations of the different systems. We apply the method of batch means with m batches to the output of each system, thus yielding an estimate of the variance of the first-stage sample mean. This is used to compute how many total batches we need to simulate for each system. In the second stage, we collect the additional batches for each system. Finally, we select the system with the largest overall sample mean as our choice for the best system (or a system having a mean lying in the indifference zone) and construct simultaneous MCB confidence intervals.

More precisely, our two-stage batch means algorithm for absolute-width indifference-zone selection and MCB is as follows:

Procedure 1

1. Specify the absolute width of the indifference zone δ (where δ is small); the desired probability, $1 - \alpha$, of simultaneous correct selection and MCB coverage; and the number of initial batches m. Let a_α solve Rinott's integral for m, k, and α. (Wilcox 1984 presents tables for a_α. Note that our notation differs from that used in Wilcox's tables. In particular, our m and a_α correspond to his n_0 and h, respectively.)

2. Independently simulate systems $i = 1, 2, \ldots, k$, with run lengths $T_i = T_i(\delta)$, which are proportional to $1/\delta^2$. For each system i, group the output into m (non-overlapping) batches, each of size T_i/m, and compute

   $$Z_{i,j} = \frac{1}{T_i/m} \int_{(j-1)T_i/m}^{jT_i/m} Y_i(s)\, ds, \quad j \geq 1,$$

   which is the sample mean of the jth batch.

3. For each system $i = 1, 2, \ldots, k$, compute the sample variance of the m batch means from the first stage as

$$S_i^2 = \frac{1}{m-1} \sum_{j=1}^{m} \left(Z_{i,j} - \frac{1}{m} \sum_{k=1}^{m} Z_{i,k} \right)^2 .$$

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

$$N_{a,i}(\delta) = \max \left\{ m, \left\lceil \left(\frac{a_\alpha S_i}{\delta} \right)^2 \right\rceil \right\}.$$

5. For the second stage, independently simulate systems $i = 1, 2, \ldots, k$, to collect additional (non-overlapping) batches $m+1, m+2, \ldots, m+N_{a,i}(\delta)$, each of size T_i/m, and compute the batch means $Z_{i,m+1}, Z_{i,m+2}, \ldots, Z_{i,N_{a,i}(\delta)}$.

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

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

7. Select the system with the largest $\hat{\mu}_{a,i}$.

8. Simultaneously construct the absolute-precision MCB confidence intervals

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

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

In order to study the asymptotic properties of Procedure 1, we define $k_a(\delta)$, $1 \leq k_a(\delta) \leq k$, such that

$$\begin{aligned} \mu_{(k)} - \mu_{(l)} &< \delta \quad \text{for all} \quad l \geq k_a(\delta); \\ \mu_{(k)} - \mu_{(i)} &\geq \delta \quad \text{for all} \quad i < k_a(\delta). \end{aligned}$$

Thus, we are equally satisfied with selecting any of the systems $(k_a(\delta)), (k_a(\delta)+1), \ldots, (k)$ since each of their means lies in the indifference zone. On the other hand, systems $(1), (2), \ldots, (k_a(\delta)-1)$ have means lying outside of the indifference zone, and so we do not want to select one of these. Now define the events

$$\begin{aligned} CS_a(\delta) &= \left\{ \max_{l \geq k_a(\delta)} \hat{\mu}_{a,(l)}(\delta) > \max_{i < k_a(\delta)} \hat{\mu}_{a,(i)}(\delta) \right\}, \\ JC_a(\delta) &= \left\{ \mu_i - \max_{j \neq i} \mu_j \in I_{a,i}(\delta),\ i = 1, 2, \ldots, k \right\}. \end{aligned}$$

Note that $CS_a(\delta)$ is the event of a *correct selection*, which we define as choosing a system with a mean lying in the indifference zone. Also, $JC_a(\delta)$ is the event that all of the true differences $\mu_i - \max_{j \neq i} \mu_j$, $i = 1, 2, \ldots, k$, are jointly covered by their MCB joint confidence intervals. Nakayama (1995) establishes the validity of the following result:

Theorem 1 *Assume that Assumption A1 holds and that the (absolute-width) indifference zone is defined as the interval* $(\mu_{(k)} - \delta,\ \mu_{(k)}]$. *Then if Procedure 1 is used,*

$$\lim_{\delta \to 0} P\{CS_a(\delta) \cap JC_a(\delta)\} \geq 1 - \alpha.$$

Theorem 1 establishes that asymptotically (as the width of the indifference zone $\delta \to 0$ with the batch sizes proportional to $1/\delta^2$), the joint probability that we make a correct selection and that the true differences simultaneously lie within their MCB confidence intervals is at least $1 - \alpha$.

Procedure 1 above is for the case when the indifference zone has an absolute width. However, in certain settings, we may be equally satisfied with either of two systems if the difference in their means is less than, say, 5%. Thus, we now consider a relative-width indifference zone defined as the interval $(\mu_{(k)} - \delta|\mu_{(k)}|,\ \mu_{(k)}]$.

Our two-stage batch means algorithm for relative-width indifference-zone selection and MCB is as follows:

Procedure 2

1. Specify the relative width of the indifference zone δ (where δ is small); the desired probability, $1 - \alpha$, of simultaneous correct selection and MCB coverage; and the number of initial batches m. Let a_α solve Rinott's integral for m, k, and α.

2. Independently simulate systems $i = 1, 2, \ldots, k$, with run lengths $T_i = T_i(\delta)$, which are proportional to $1/\delta^2$. For each system i, group the output into m (non-overlapping) batches, each of size T_i/m, and compute the first m batch means $Z_{i,1}, Z_{i,2}, \ldots, Z_{i,m}$.

3. For each system $i = 1, 2, \ldots, k$, compute the sample mean and the sample variance of the m batch means from the first stage as

$$\tilde{\mu}_i = \frac{1}{m} \sum_{j=1}^{m} Z_{i,k}$$

and

$$S_i^2 = \frac{1}{m-1} \sum_{j=1}^{m} \left(Z_{i,j} - \tilde{\mu}_i \right)^2 ,$$

respectively.

4. Define $\tilde{\mu} = \max(\tilde{\mu}_1, \tilde{\mu}_2, \ldots, \tilde{\mu}_k)$, and for each system $i = 1, 2, \ldots, k$, compute the total number of batches to collect as
$$N_{r,i}(\delta) = \max\left\{ m, \left\lceil \left(\frac{a_\alpha S_i}{\delta \tilde{\mu}} \right)^2 \right\rceil \right\}.$$
5. For the second stage, independently simulate systems $i = 1, 2, \ldots, k$, to collect additional (non-overlapping) batches $m+1, m+2, \ldots, m+N_{r,i}(\delta)$, each of size T_i/m, and compute the batch means $Z_{i,m+1}, Z_{i,m+2}, \ldots, Z_{i,N_{r,i}(\delta)}$.
6. For each system $i = 1, 2, \ldots, k$, compute the overall sample mean as
$$\hat{\mu}_{r,i} = \frac{1}{N_{r,i}(\delta)} \sum_{j=1}^{N_{r,i}(\delta)} Z_{i,j}.$$
7. Select the system with the largest $\hat{\mu}_{r,i}$.
8. Simultaneously construct the relative-precision MCB confidence intervals
$$I_{r,i}(\delta) = \left[\left(\hat{\mu}_{r,i} - \max_{j \neq i} \hat{\mu}_{r,j} - \delta \max_{j \neq i} |\hat{\mu}_{r,j}| \right)^-, \right. \left. \left(\hat{\mu}_{r,i} - \max_{j \neq i} \hat{\mu}_{r,j} + \delta\, |\hat{\mu}_{r,i}| \right)^+ \right]$$
for $\mu_i - \max_{j \neq i} \mu_j$, $i = 1, 2, \ldots, k$.

In order to study the asymptotic properties of Procedure 2, we define $k_r(\delta)$, $1 \leq k_r(\delta) \leq k$, such that
$$\begin{aligned} \mu_{(k)} - \mu_{(l)} &< \delta |\mu_{(k)}| \quad \text{for all} \quad l \geq k_r(\delta); \\ \mu_{(k)} - \mu_{(i)} &\geq \delta |\mu_{(k)}| \quad \text{for all} \quad i < k_r(\delta). \end{aligned}$$

Thus, we are equally satisfied with selecting any of the systems $(k_r(\delta)), (k_r(\delta)+1), \ldots, (k)$ since each of their means lies in the (relative-width) indifference zone. On the other hand, we do not want to select any of the systems $(1), (2), \ldots, (k_r(\delta)-1)$ since their means lie outside of the indifference zone. Now define the events
$$\begin{aligned} CS_r(\delta) &= \left\{ \max_{l \geq k_r(\delta)} \hat{\mu}_{r,(l)}(\delta) > \max_{i < k_r(\delta)} \hat{\mu}_{r,(i)}(\delta) \right\}, \\ JC_r(\delta) &= \left\{ \mu_i - \max_{j \neq i} \mu_j \in I_{r,i}(\delta),\ i = 1, 2, \ldots, k \right\}. \end{aligned}$$

Nakayama (1995) establishes the validity of the following theorem:

Theorem 2 *Assume that Assumption A1 holds and that the (relative-width) indifference zone is defined as the interval* $(\mu_{(k)} - \delta|\mu_{(k)}|,\ \mu_{(k)}]$. *Also, assume that* $\mu_{(k)} \neq 0$. *Then if Procedure 2 is used,*
$$\lim_{\delta \to 0} P\{CS_r(\delta) \cap JC_r(\delta)\} \geq 1 - \alpha.$$

From a theoretical standpoint, the formulation of the asymptotic results in Theorems 1 and 2 may not be completely appropriate. In particular, we assumed that the steady-state means $\mu_1, \mu_2, \ldots, \mu_k$ are fixed (and do not change with δ), and so taken by itself, the probability of correct selection (PCS) converges to 1 as $\delta \to 0$ by the strong law of large numbers. Perhaps a more theoretically interesting result would allow the steady-state means to vary with δ, thereby leading to a PCS which converges to something strictly less than 1 (and at least $1-\alpha$). (For theorems of this type, see Damerdji et al. 1995.) However, our Theorems 1 and 2 still have theoretical value since the limiting coverage probability of our MCB confidence intervals is strictly less than 1 (and at least $1-\alpha$). (From a practical standpoint, though, these issues are not a concern.)

4 SPECIFYING VALUES FOR PARAMETERS

To use Procedures 1 and 2 in practice, the user must specify values for several parameters. These include the (absolute or relative) width of the indifference zone δ, the run length of the first stage T_i (which must be proportional to $1/\delta^2$) for each system, and the number of initial batches m.

Because of the similarity of Procedures 1 and 2 and the two-stage stopping procedures developed by Nakayama (1994), it is probably reasonable to assume that appropriate values for the parameters of Nakayama's (1994) algorithm are also valid for our new procedures. Nakayama (1994) suggests that we should choose $5 \leq m \leq 15$ and $\delta < 0.025$. However, as Nakayama (1994) notes, selecting a reasonable value for T_i given δ is a delicate matter. In the case when simulating queueing systems, though, Nakayama (1994) proposes using some of the results of Whitt (1989a,1989b); for more details, see Sections 5 and 6 of Nakayama (1994).

ACKNOWLEDGMENTS

The author would like to thank Halim Damerdji and Peter Glynn for some enlightening discussions. Also, part of this work was supported by NJIT SBR Grant 421180.

REFERENCES

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.

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

Damerdji, H., P. W. Glynn, M. K. Nakayama, and J. R. Wilson. 1995. Selecting the Best System in Steady-State Simulations. Forthcoming technical report.

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

Glynn, P. W. 1990. Diffusion approximations. Chapter 4 of *Handbooks in Operations Research and Management Science, Vol. 2, Stochastic Models*, ed. D. Heyman and M. Sobel. Elsevier Science Publishers B. V. (North-Holland).

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

Goldsman, D. and B. L. Nelson. 1994. Ranking, selection and multiple comparisons in computer simulation. In *Proceedings of the 1994 Winter Simulation Conference*, ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila, 192–199. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

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

Matejcik, F. J. and B. L. Nelson. 1992. Two-stage multiple comparisons with the best for computer simulation. *Operations Research*, forthcoming.

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

Nakayama, M. K. 1995. Procedures for simultaneous selection and multiple comparisons in steady-state simulations. Unpublished manuscript.

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

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

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

Whitt, W. 1989b. Simulation run length planning. In *Proceedings of the 1989 Winter Simulation Conference*, ed. E. A. MacNair, K. J. Musselman, and P. Heidelberger, 106–112. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

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

AUTHOR'S BIOGRAPHY

MARVIN K. NAKAYAMA is an assistant professor in the Department of Computer and Information Science at the New Jersey Institute of Technology. Previously, he was a visiting assistant professor in the Management Science/Computer Information Systems Department of the Graduate School of Management at Rutgers University, and before that, a post-doctoral fellow at the IBM Thomas J. Watson Research Center in Yorktown Heights, New York. He received a B.A. in Mathematics/Computer Science from University of California, San Diego and an M.S. and Ph.D. in Operations Research from Stanford University. His research interests include applied probability, simulation output analysis, gradient estimation, rare-event simulation, and random-variate generation.