TWO-STAGE MULTIPLE COMPARISONS WITH THE BEST FOR COMPUTER SIMULATION

FRANK J. MATEJCIK

South Dakota School of Mines and Technology, Rapid City, South Dakota

BARRY L. NELSON

Northwestern University, Evanston, Illinois

(Received October 1992; revisions received February, July 1994; accepted September 1994)

We consider the problem of comparing a small number of stochastic systems via computer simulation when the basis for comparison is the expected value of some system performance measure. To solve this problem we develop two-stage sampling procedures that provide confidence intervals for the difference between the expected performance of each system and the best of the others. These confidence intervals are valid under mild conditions, and the procedures allow the experimenter to specify the desired precision in advance. Special cases of our results include standard indifference-zone selection procedures. The paper includes guidelines for experiment design and an illustrative example.

We consider the problem of comparing a small number of stochastic systems via simulation. Although our work involves one-way experiment designs used in ANOVA, our goal is to find the system with the largest expected performance, where μ_i for $i = 1, 2, \ldots, k$ denotes the expected performance of system i. We achieve this goal by using a class of multiple-comparisons procedures known as multiple comparisons with the best (MCB). We define MCB procedures as those that provide joint confidence intervals for $\mu_i - max_{j \neq i} \mu_j$ for $i = 1, 2, \ldots, k$. We will derive and illustrate two-stage MCB procedures that are valid under mild assumptions, thereby allowing MCB to be used more readily in computer simulations. Although our research focuses on computer simulation experiments, our results are applicable in other settings.

Computer Simulation and Two-Stage Methods

We restrict our discussion to one popular output-analysis method: *batch means*. To preview our procedures we present a general description of the combined use of two-stage methods and batch means in typical computer simulation experiments.

Schmeiser (1982) and others have considered the use of batch means to transform stationary, but dependent, simulation data into nearly independent and nearly normally distributed data. Batch means may also be used with independent and identically distributed (i.i.d.) data, because i.i.d. data comprise a special case of stationary, dependent data. To use (nonoverlapping) batch means in a two-stage, steady-state simulation, we first generate long sequences of performance observations from each system, $i = 1, 2, \ldots, k$. The sequences are then partitioned into $b_{0,i}$ batches of m_i observations. The mean of each batch is computed. Under mild conditions the batch means will be nearly independent and normally distributed, so we may use normal-theory methods. In particular, we may compute the sample variances of the first-stage batch means and design a second-stage sample to allow inferences with a prespecified precision. We will derive MCB procedures of this form. Our procedures extend a rich body of literature on two-stage procedures for inference when variances are unknown and unequal. Stein (1945) is the foundation paper; see Bishop (1976) for an extensive survey.

Consider also that computer simulations often begin with a *pilot study*, followed by a *production run* (see e.g., Goldsman, Nelson and Schmeiser 1991). Classical design-of-experiment techniques are awkward to use in this situation because the experimenter must select a particular variance for their design, and the design does not guarantee inference at a prespecified precision. Moreover, classical designs typically require strong variance assumptions, but simulators often know less about the variance than they do about the expected performance. The two-stage procedures in this paper work directly with pilot results, require mild assumptions, and guarantee inference at a prespecified precision.

Notation and Organization of the Paper

A detailed presentation of our procedures requires some notation. Many of these terms are vectors, so we display vectors in boldfaced type.

- Y is a single observation of a performance measure;
- k is the number of systems to be ranked;
- $\boldsymbol{\mu}$ is a $k \times 1$ vector of mean performance measures;
- $\hat{\boldsymbol{\mu}}$ is an estimator of $\boldsymbol{\mu}$;

Subject classifications: Simulation: statistical analysis; design of experiments.
Area of review: SIMULATION.

0030-364X/95/4304-0633 $01.25

$\bar{\mathbf{Y}}$ is the overall sample mean, the estimator of $\boldsymbol{\mu}$ when batch means are used;

i is a subscript corresponding to a system, $i = 1, 2, \ldots, k$;

$n_{0,i}$ is the first-stage sample size, in terms of basic observations;

$b_{0,i}$ is the first-stage sample size, in terms of the number of batches;

N_i is the estimated number of basic observations for the sum of the first-stage and second-stage samples;

B_i is the estimated number of batches for the sum of the first-stage and second-stage samples;

m_i is the number of observations in a batch, so $n_{0,i} = b_{0,i}m_i$;

$S_i^2(b_{0,i})$ is the estimator of the variance of the batch means based on $b_{0,i}$ batches; we write S_i^2 when the number of batches is obvious.

The paper is organized into six sections: MCB preliminaries, procedures, selecting batch-size in two-stage MCB, example, and conclusions.

1. MCB PRELIMINARIES

Before we present our procedures for two-stage MCB we present some general MCB results. MCB provides simultaneous confidence intervals for the difference between the expected performance of each system and the best of the other systems, $\mu_i - \max_{j\neq i}\mu_j$ for $i = 1, 2, \ldots, k$. If $\mu_i - \max_{j\neq i}\mu_j > 0$, then system i is the best, for it is better than the best of the other systems. If $\mu_i - \max_{j\neq i} \mu_j < 0$, then system i is not the best, because there is another better system. However, even if $\mu_i - \max_{j\neq i} \mu_j < 0$, if $\mu_i - \max_{j\neq i} \mu_j > -\epsilon$, where ϵ is a positive number, then system i is within ϵ of the best. MCB confidence intervals for these parameters may therefore conclusively identify the system with the largest expected performance, conclusively eliminate some systems from contention for being the best, or bound how far each system is from being the best. Procedures for finding and using these confidence intervals were first published in Hsu (1981).

Our presentation is structured around two results, which we call Hsu's lemmas. Although Hsu did not present "Hsu's lemmas" in the form that we will present them, his work provided the proofs of the lemmas and the foundations of MCB. MCB intervals are constrained, so it is convenient to use the notation $x^+ = \max(x, 0)$ and $-x^- = \min(0, x)$.

Lemma 1. (Hsu's single-bound lemma) *Let* $\mu_{(1)} \leq \mu_{(2)} \leq \cdots \leq \mu_{(k)}$ *be the (unknown) ordered performance parameters of k systems, and let* $\hat{\mu}_{(1)}, \hat{\mu}_{(2)}, \ldots, \hat{\mu}_{(k)}$ *be any estimators of the parameters. If*

$$\Pr\{\hat{\mu}_{(k)} - \hat{\mu}_{(i)} - (\mu_{(k)} - \mu_{(i)}) > -w, \quad i = 1, 2, \ldots, k-1\} = 1 - \alpha, \tag{1}$$

then

$$1 - \alpha \leq \Pr\{\mu_i - \max_{j\neq i} \mu_j \in [-(\hat{\mu}_i - \max_{j\neq i} \hat{\mu}_j - w)^-, (\hat{\mu}_i - \max_{j\neq i} \hat{\mu}_j + w)^+], \text{ for all } i\}. \tag{2}$$

If we replace the $=$ *in* (1) *with* $\geq$, *then* (2) *still holds.*

To simplify notation we define the event

$$\{\mathrm{CS}(\hat{\boldsymbol{\mu}}, w)\} = \{\mu_i - \max_{j\neq i} \mu_j \in [-(\hat{\mu}_i - \max_{j\neq i} \hat{\mu}_j - w)^-, (\hat{\mu}_i - \max_{j\neq i} \hat{\mu}_j + w)^+] \text{ for all } i\},$$

where CS is short for "correct selection," a term used in ranking and selection to mean correctly choosing the best system.

Hsu's single-bound lemma has essentially the same proof as the balanced case of MCB in Hochberg and Tamhane (1987, p. 151). The letter w is chosen because the bound in (1) is a whisker length in a whisker plot of the MCB intervals. Informally, this lemma says that if we can find a single bound w for all the differences with the best, then we can do MCB. We will see that this implies that solving a selection of the best problem in ranking and selection will allow us to do MCB.

Lemma 2. (Hsu's multiple-bound lemma) *Let* $\mu_{(1)} \leq \mu_{(2)} \leq \cdots \leq \mu_{(k)}$ *be the (unknown) ordered performance parameters of k systems. Let* T_{ij} *be a point estimator of the parameter* $\mu_i - \mu_j$. *If for each i individually*

$$\Pr\{T_{ij} - (\mu_i - \mu_j) > -w_{ij}, \text{ for all } j\neq i\} = 1 - \alpha, \tag{3}$$

then we can make the joint probability statement

$$1 - \alpha \leq \Pr\{\mu_i - \max_{j\neq i} \mu_j \in [D_i^-, D_i^+], \text{ for all } i\}, \tag{4}$$

where

$$D_i^+ = (\min_{j\neq i} [T_{ij} + w_{ij}])^+,$$

$$\mathcal{G} = \{\ell : D_\ell^+ > 0\},$$

and

$$D_i^- = \begin{cases} 0 & \text{if } \mathcal{G} = \{i\} \\ -(\min_{j\in\mathcal{G}, j\neq i}[-T_{ji} - w_{ji}])^- & \text{otherwise.} \end{cases}$$

If we replace the $=$ *in* (3) *with* $\geq$, *then* (4) *still holds.*[1]

To simplify notation we define the event

$$\{\mathrm{CS}(\mathbf{T}, \mathbf{w})\} = \{\mu_i - \max_{j\neq i} \mu_j \in [D_i^-, D_i^+], \text{ for all } i\},$$

where $\mathbf{T} = \{T_{ij}; i, j = 1, 2, \ldots, k, i \neq j\}$ and $\mathbf{w} = \{w_{ij}; i, j = 1, 2, \ldots, k, i \neq j\}$.

Hsu's multiple-bound lemma is proved in Chang and Hsu (1992). Informally, this lemma says that if we can do multiple comparisons with a control (MCC) with each

system as the control, then we can do MCB. More precisely, (3) is the probability statement required for one-sided MCC with system i as the control ($\mu_i - \mu_j$, for all $j \neq i$). The "bounds" w and w_{ij} have a helpful graphical interpretation. MCB intervals are often presented as whisker diagrams, so w and w_{ij} are interpreted as "whisker lengths" in these diagrams. Accordingly, w and w_{ij} are measures of the precision, which help define the lengths of the MCB intervals. Procedures based on the multiple-bound lemma can be sharper than those based on the single-bound lemma because the length of each whisker w_{ij} can be specified individually, rather than requiring a single whisker w to apply to all MCC differences.

2. PROCEDURES

We begin by introducing a means MCB procedure in a general form. Next, we prove its validity. We also present a corollary that allows even more flexibility in the procedure. Following this introduction we develop two useful special cases: balanced first-stage sample with a constant whisker length w, and a general single-stage-sample form.

Procedure 1. Two-Stage Means MCB

STEP 1. Specify the target whisker length w^*, and error rate α. (Comment: The target whisker length w^* is analogous to the indifference zone in indifference-zone selection.)

STEP 2. Take an initial stationary sample $Y_{i,1}, Y_{i,2}, \ldots, Y_{i,n_{0,i}}$ of size $n_{0,i}$ from system i, $i = 1, 2, \ldots, k$.

STEP 3. Compute the batch means $\bar{Y}_{i,j}$ so that they are approximately independent and normally distributed:

$$\bar{Y}_{i,j} = \frac{1}{m_i} \sum_{p=(j-1)m_i+1}^{jm_i} Y_{i,p} \quad i = 1, 2, \ldots, k$$

$$j = 1, 2, \ldots, b_{0,i}.$$

(Comment: We may skip Step 3 for system i if $Y_{i,1}, \ldots, Y_{i,n_{0,i}}$ are approximately independent and normally distributed. When we skip this step for system i, we assign $m_i = 1$, $b_{0,i} = n_{0,i}$, and $\bar{Y}_{i,j} = Y_{i,j}$ for $j = 1, 2, \ldots, b_{0,i}$.)

STEP 4. Compute these intermediate statistics for each sample:

$$\bar{Y}_i(b_{0,i}) = \frac{1}{b_{0,i}} \sum_{j=1}^{b_{0,i}} \bar{Y}_{i,j}$$

$$S_i^2 = \frac{1}{b_{0,i}-1} \sum_{j=1}^{b_{0,i}} (\bar{Y}_{i,j} - \bar{Y}_i(b_{0,i}))^2 \quad i = 1, 2, \ldots, k.$$

STEP 5. Compute the first-stage whisker lengths w'_{ij}

$$w'_{ij} = h_i \max\left\{\frac{S_i}{\sqrt{b_{0,i}}}, \frac{S_j}{\sqrt{b_{0,j}}}\right\}.$$

(Comment: Here h_i is the solution to Rinott's (1978) integral equation, which is also (6) in the Appendix.)

STEP 6. If $\{w'_{ij};\ i, j = 1, 2, \ldots, k,\ i \neq j\}$ are sufficiently narrow in the judgment of the experimenter, no second-stage sample is taken. If no second-stage sample is taken, then assign $w_{ij} = w'_{ij}$, $\bar{Y}_i = \bar{Y}_i(b_{0,i})$, and go to Step 11. If a second-stage sample is to be taken then go to Step 7.

STEP 7. Compute the final whisker lengths, $w_{ij} = \min(w^*, w'_{ij})$.

STEP 8. Compute the total sample size, $\mathbf{N} = (N_1, N_2, \ldots, N_k)$, where $N_i = m_i B_i$ and

$$B_i = \max\{b_{0,i}, \lceil S_i^2 \max_{j \neq i} (h_i/w_{ij})^2 \rceil, \lceil S_i^2 \max_{j \neq i} (h_j/w_{ji})^2 \rceil\}.$$

STEP 9. Take $N_i - n_{0,i}$ additional observations, $Y_{i,n_{0,i}+1}, \ldots, Y_{i,N_i}$, from system i, $i = 1, 2, \ldots, k$.

STEP 10. Compute the overall means,

$$\bar{Y}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} Y_{i,j} = \frac{1}{N_i}\left(n_{0,i}\bar{Y}_i(b_{0,i}) + \sum_{j=n_{0,i}+1}^{N_i} Y_{i,j}\right).$$

STEP 11. Compute comparison estimates, $T_{i,j} = \bar{Y}_i - \bar{Y}_j$.

STEP 12. Construct MCB intervals using $\{CS(\mathbf{T}, \mathbf{w})\}$.

Theorem 1. *If the batch means $\bar{Y}_{i,j}$ of Procedure 1 are independent and normally distributed with finite mean and variance, then $Pr\{CS(\mathbf{T}, \mathbf{w})\} \geq 1 - \alpha$.*

Proof. The proof is in the Appendix.

We may generalize Procedure 1 by replacing Steps 6 and 7 with Steps 6′ and 7′:

STEP 6′. Independent of the first-stage means, $\bar{Y}_i(b_{0,i})$, select the set of the systems II, for which a second-stage sample will be taken.

STEP 7′. Compute the final whisker lengths w_{ij} as

$$w_{ij} = \begin{cases} \min(w^*, w'_{ij}) & \text{if } i, j \in \prod \\ w'_{ij} & \text{otherwise.} \end{cases}$$

The following corollary is a consequence of the proof of Theorem 1.

Corollary 1. (Partial Stopping Rules) *Under the assumptions of Theorem 1, if we replace Steps 6 and 7 of Procedure 1 with Steps 6′ and 7′ then for this modified procedure $Pr\{CS(\mathbf{T}, \mathbf{w})\} \geq 1 - \alpha$.*

Step 6′ allows many useful and practical experimenter interventions. However, modifications not stochastically independent of first-stage means are not justified. Our proof of Theorem 1 uses the property of independence across systems, so if we choose Π using a method that depends on first-stage means, then we no longer have

independence across systems and our proof fails. However, we may choose not to include system i in Π (and thereby not sample it in the second stage) for the following practical reasons:

- System i is no longer available for testing.
- The cost of using system i is too high.
- Administrative decisions prevent further consideration of system i.
- The w^* was vaguely specified, and although $w^* < w'_{ij}$ or $w^* < w'_{ji}$ for some values of j, all w'_{ij} and w'_{ji} are sufficiently small to avoid further sampling of system i. Notice that the whisker lengths w'_{ij} must be evaluated based in their absolute size. Comparing these whisker lengths across systems would violate our independence assumptions.
- A system parameter unrelated to the objective observation Y discourages use of system i.
- Objectionable transient behaviors in a steady-state experiment discourage use of system i.

We now list two useful special cases of Procedure 1. The first one may also be proved directly using Hsu's single-bound lemma and Rinott. The second may be proved by letting Π in Corollary 1 be the empty set.

Procedure 2. Two-Stage Means MCB With Common w and b_0

STEP 1. Specify w and α.

STEP 2. Take an initial stationary sample $Y_{i,1}, Y_{i,2}, \ldots, Y_{i,n_{0,i}}$ of size $n_{0,i}$ from system i, such that $b_0 = n_{0,i}/m_i$ and $b_0 \geq 2$ for $i = 1, 2, \ldots, k$.

STEP 3. Compute the batch means and intermediate statistics for each sample as in Procedure 1, Steps 3 and 4.

STEP 4. Compute the total sample size $\mathbf{N} = (N_1, N_2, \ldots, N_k)$, where $N_i = m_i B_i$ and

$$B_i = \max \{b_0, \lceil (h/w)^2 S_i^2 \rceil\}.$$

(Comment: Here h is the solution to a special case of the integral in Procedure 1. A table for h is available in Wilcox (1984).)

STEP 5. Take $N_i - n_{0,i}$ additional observations, $Y_{i,n_{0,i}+1}, \ldots, Y_{i,N_i}$, from system i, $i = 1, 2, \ldots, k$.

STEP 6. Compute the overall means $\bar{Y}_i$ as in Procedure 1, Step 10.

STEP 7. Construct MCB intervals using $\{CS(\bar{\mathbf{Y}}, w)\}$.

Procedure 3. Two-Stage Means MCB Stopping After First Stage

STEP 1. Specify the error rate α.

STEP 2. Take an initial stationary sample, compute the batch means, intermediate statistics, and first-stage whisker lengths as in Procedure 1, Steps 3, 4, and 5.

STEP 3. Let $w_{ij} = w'_{ij}$, and $\bar{Y}_i = \bar{Y}_i(b_{0,i})$.

STEP 4. Compute $\mathbf{T}$ as in Procedure 1, Step 11.

STEP 5. Construct MCB intervals using $\{CS(\mathbf{T}, \mathbf{w})\}$.

We believe that Procedures 2 and 3 can be used frequently by simulators. Procedure 2 is convenient for simulators because h is readily available and because Procedure 2 involves no complex decision making regarding w'_{ij}. Procedure 3 can be used as a one-stage procedure that produces conservative MCB intervals when variances are unknown and unequal. Thus, our results should appeal to experiments who prefer one-stage procedures. However, the additional features provided by Procedure 1 and its corollary may prove useful in some situations.

3. SELECTING BATCH SIZE IN TWO-STAGE MCB

As stated in the Introduction, to use batch means we first generate long sequences of performance observations $Y_{i,j}$, from each system $i = 1, \ldots, k$. The first-stage sample $n_{0,i}$ is typically determined by judgments regarding normality and independence. However, we can make recommendations for selecting other constants.

To use the two-stage MCB procedures of Section 2 we must specify three parameters: the confidence level $1 - \alpha$, the target whisker length w^*, and the first-stage number of batches $b_{0,i}$. The target whisker length w^* is analogous to the indifference zone in indifference-zone selection. Simulators can specify w^* from practical considerations. The confidence level α is typically selected from the traditional values 0.10, 0.05, and 0.01. The choice of the first-stage number of batches $b_{0,i}$ affects the performance of the procedure itself, so this section will develop recommendations for $b_{0,i}$.

Simulation experiments often require $n_{0,i} \gg b_{0,i}$, so the selection of $b_{0,i}$ is a compromise between a small number of batches which improves the approximations of normality and independence, and a large number of batches which may result in a smaller total sample, N_i. We recommend using $b_{0,i} \in [10, 40]$. Our recommendation is based on an examination of Procedure 2, but it carries over naturally to Procedure 1.

For ease of exposition, assume that $Y_{i,1}, Y_{i,2}, \ldots, Y_{i,N_i}$ are i.i.d. $N(\mu_i, \sigma_i^2)$, so that the batch means $\bar{Y}_{i,j}$ are i.i.d. $N(\mu_i, \sigma_i^2/m_i)$. Analysis under these assumptions is worst case in the sense that we could set $b_0 = n_0$ (that is, not batch at all). Our results show the penalty for batching when it is not needed. Alternatively, we could make the detailed assumptions assuring the appropriateness of our b_0 value as was done by Schmeiser.

With these assumptions we may develop recommendations for b_0 in Procedure 2. For Procedure 2 the following formula defines the total number of batches B_i,

$$B_i = \max \left(b_0, \left\lceil \left(\frac{h}{w}\right)^2 S_i^2(b_0) \right\rceil\right).$$

Therefore,

$$N_i = m_i B_i \geq \left(\frac{h}{w}\right)^2 m_i S_i^2(b_0).$$

If we assume that second-stage sampling is likely, then $E(N_i) \approx (h/w)^2 \sigma_i^2$. Since the $E(N_i)$ decreases as b_0 increases, the relative savings from rebatching (RSR) the first-stage sample to $b_0' > b_0$ batches is

$$\mathrm{RSR}(b_0, b_0') = \frac{\mathrm{E}(N_{i,b_0}) - \mathrm{E}(N_{i,b_0'})}{\mathrm{E}(N_{i,b_0})} \times 100\%,$$

where $E(N_{i,b_0})$ is the expected total sample size when the initial number of batches is b_0.

From our approximation for $E(N_i)$ and the h values in Wilcox, we can plot approximate $\mathrm{RSR}(b_0, b_0')$ surfaces. However, we present instead an upper bound for RSR, which we call the limiting relative savings from rebatching (LRSR). The LRSR is

$$\mathrm{LRSR}(b_0) = \lim_{b_0' \to \infty} \mathrm{RSR}(b_0, b_0') = \frac{h_{b_0}^2 - h_\infty^2}{h_\infty^2} \times 100\%,$$

where h_{b_0} is the solution to Rinott's integral equation when the first-stage sample size is b_0, and h_∞ is $\lim_{b_0 \to \infty} h_{b_0}$. It can be shown that $h_\infty = \sqrt{2}\Phi^{-1}[(1 - \alpha)1/k - 1]$, where Φ is the cdf of the standard normal distribution. Therefore, the $\mathrm{LRSR}(b_0)$ may also be calculated with the aid of the tabled values in Wilcox.

Figure 1 is a plot of $\mathrm{LRSR}(b_0)$ for different numbers of systems, $k = 2, 6, 10$, and confidence level $\alpha = 0.05$. We judge from the graph that selecting $b_0 < 10$ leads to a large total sample size, and selecting $b_0 > 40$ gains little with regard to total sample size. Other analysts may draw somewhat different conclusions from the figure. However, since selecting a large b_0 for fixed n_0 may make normality and independence assumptions inappropriate, we see little benefit and potential harm from venturing beyond 40 batches.

Our analysis assumes that the conditions required for the method of batch means to be valid in small samples are exactly satisfied. Nakayama (1994) established conditions under which a two-stage, batch-means procedure for the expected performance of a single system is asymptotically valid for more general output processes. Asymptotic validity obtains as the absolute error (indifference zone w in our terminology) decreases to zero. In brief, his procedure requires that the first-stage batch size must grow proportional to $1/w^2$ as $w \to 0$.

We are now prepared to apply two-stage MCB to an example.

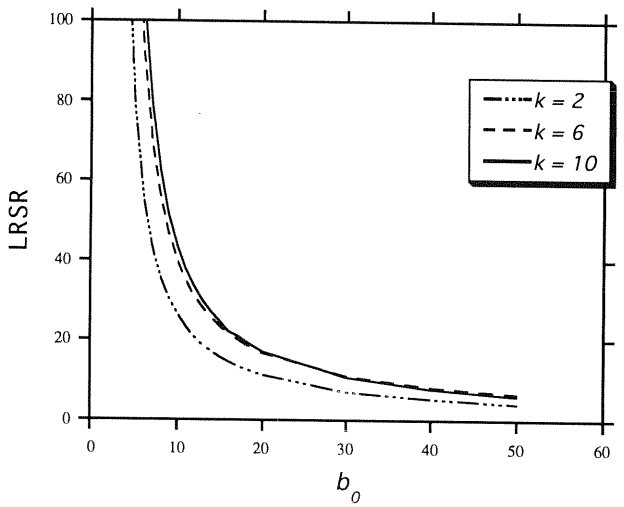

Figure 1. The function $\mathrm{LRSR}(b_0)$ for $\alpha = 0.05$.

4. EXAMPLE

Goldsman, Nelson and Schmeiser propose the following problem:

> We consider $k = 4$ different airline-reservation systems. The single measure of performance is the expected time to failure, E[TTF]—the larger the better. The system works if either of two computers work. Computer failures are rare, repair times are fast, and the resulting E[TTF]s are roughly 100,000 minutes (about 70 days) for all four systems. We are indifferent to expected differences of less than 3,000 minutes (about two days).

Goldsman, Nelson and Schmeiser used Rinott's indifference-zone selection procedure, which is a two-stage procedure similar to our Procedure 2, to solve this problem. They chose a probability of correct selection $P^* = 0.90$, and an indifference-zone $\delta = 3{,}000$ minutes to start the experiment. They choose an initial sample size of $n_{0,i} = 400$, and the number of batches $b_0 = 20$. They verified normality following the first-stage sample. In this example observations were initially independent, so no test of the independence of the batch means was required. Table I shows Goldsman, Nelson and Schmeiser's results from their first-stage sample.

Goldsman, Nelson and Schmeiser obtained the additional samples required for the second stage, as prescribed by Rinott's procedure, and computed $\bar{Y}_1 = 110816.5$, $\bar{Y}_2 = 106411.8$, $\bar{Y}_3 = 99093.1$, and $\bar{Y}_4 = 86568.9$. Accordingly, Goldsman, Nelson and Schmeiser selected system 1 as the best. At this point they appropriately claimed:

> We can make the formal statement that we are at least 90% sure that we have made the correct selection (with the proviso that the true difference between the best and the second best E[TTF]'s is at least $\delta = 3{,}000$ minutes).

The formal statement is correct, and it is the statement desired by an indifference-zone selection. However, it says nothing about the possibility that another system could actually be best if the difference between the two

Table I
Goldsman, Nelson and Schmeiser's
First-Stage Results ($b_0 = 20$)

i	1	2	3	4
$\bar{Y}_i(b_0)$	108286.0	107686.0	96167.7	89747.9
S_i	29157.3	24289.9	25319.5	20810.8
$\bar{B}_i$	699	485	527	356

best systems is less than δ. Furthermore, it says nothing about eliminating any of the systems from being considered the best. Fortunately, as we see from our solution, MCB allows inferences beyond the indifference-zone selection procedure with just a simple additional calculation.

4.1. Our Solution

A comparison of our Procedure 2 and Rinott's selection procedure finds differences only in Steps 1, 4, and 7. If we allow α from both procedures to be the same and set $w = \delta$, then the procedures differ only in the final step. In Procedure 2 we set $w = 3{,}000$, $\alpha = 0.10$, and used Goldsman, Nelson and Schmeiser's results. After performing the calculations in Step 7 we obtained the MCB intervals in Table II. With a 90% confidence level the MCB intervals presented in Table II exclude systems 2–4 from being the best and clearly identify system 1 as being the best. This a much stronger statement than was given by Goldsman, Nelson and Schmeiser using the same data and a selection procedure. These stronger statements are a characteristic of MCB and demonstrate an advantage of MCB over indifference-zone selection procedures. Furthermore, indifference-zone results usually can be supplemented by MCB intervals as we have done in this example. Later, we formalize this statement into a theorem.

Additionally, we could have made a useful inference following Step 3, if, for some reason, obtaining a second-stage sample was not possible. Procedure 3 allows us to find MCB intervals for an experiment terminated at Step 3. Applying Procedure 3 to Goldsman, Nelson and Schmeiser's first-stage data we obtain Table III. With a 90% confidence level the MCB intervals presented in the table exclude system 4 from being the best. They fail to identify any of the other systems as the best, but they do provide bounds on the potential difference between each one and the best. This ability to make reports if the experiment must be stopped at the first stage can be useful in some circumstances. Notice that forming Table III following Step 3, but then eliminating system 4 from second-stage sampling based on the first-stage results, is not an appropriate use of Corollary 1.

4.2. An Observation from the Example

We commented previously that we can typically supplement an indifference-zone selection procedure with MCB intervals. We state this result as a theorem to be used to supplement such procedures.

Table II
Two-Stage Means MCB Solution for the Airline Reservation Problem

i	$\bar{Y}_i$	Lower Limit	$\bar{Y}_i - \max_{j \neq i} \bar{Y}_j$	Upper Limit
1	110816.5	0	4,405	7,405
2	106411.8	−7,405	−4,405	0
3	99093.1	−14,724	−11,723	0
4	86568.9	−27,248	−24,248	0

Table III
MCB for Goldsman, Nelson and Schmeiser's First-Stage Data

i	Lower Limit	$\bar{Y}_i - \max_{j \neq i} \bar{Y}_j$	Upper Limit
1	−17,134	600	18,334
2	−18,234	−600	17,134
3	−29,852	−12,118	5,616
4	−36,272	−18,538	0

Theorem 2. (MCB intervals for indifference-zone selection procedures) *Using the data produced by the indifference-zone selection procedures presented in Dudewicz and Dalal* (1975), *Rinott* (1978), *and Clark and Yang* (1986), *we may supplement the analysis with MCB intervals as defined by*

$$P^* = 1 - \alpha \leq \Pr\{\mu_i - \max_{j \neq i} \mu_j \in [-(\bar{Y}_i - \max_{j \neq i} \bar{Y}_j - \delta)^-, (\bar{Y}_i - \max_{j \neq i} \bar{Y}_j + \delta)^+], \text{ for all } i\},$$

where δ is the indifference zone.

Proof. See Nelson and Matejcik (1994).

It is worth noting that the selection procedure of Clark and Yang permits the use of common random numbers by employing the Bonferroni inequality. The details of our two-stage, Bonferroni MCB procedure are in Matejcik (1992). We hope that this theorem will cause selection procedures to be more useful in practice because of the additional inference that it allows.

5. CONCLUSIONS

We have developed two-stage means MCB procedures. We have shown that they are versatile and easy to apply. Furthermore, we have argued that two-stage procedures are appropriate for computer simulation because computer simulations allow design from a pilot study. Also, we have suggested guidelines for selecting the first-stage number of batches, b_0. Additionally, we have shown by example and analysis that MCB can supplement many common selection procedures, thus allowing many helpful inferences.

APPENDIX

Proof of Theorem 1

Hsu's multiple bound lemma tells us that we need only show that **T** is obtained so that we may make the appropriate MCC claims in (3). To do this we modify the proof in Rinott.

Fix i. Define $Pr\{Coverage\}$ as

$$\Pr\{Coverage\}$$

$$\equiv \Pr\{\bar{Y}_i - \bar{Y}_j - (\mu_i - \mu_j) > -w_{ij}, \text{ for all } j \neq i\}$$

$$= \Pr\left\{\frac{(\bar{Y}_j - \mu_j) - (\bar{Y}_i - \mu_i)}{[(\sigma_i^2/B_i) + (\sigma_j^2/B_j)]^{1/2}} < \frac{w_{ij}}{[(\sigma_i^2/B_i) + (\sigma_j^2/B_j)]^{1/2}}, \text{ for all } j \neq i\right\},$$

where $\sigma_i^2 \equiv E(S_i^2)$.

Let

$$Z_j \equiv \frac{(\bar{Y}_j - \mu_j) - (\bar{Y}_i - \mu_i)}{[(\sigma_i^2/B_i) + (\sigma_j^2/B_j)]^{1/2}}, \quad \text{for all } j \neq i$$

and

$$Q_j \equiv \frac{h_i}{[(\sigma_i^2/S_i^2) + (\sigma_j^2/S_j^2)]^{1/2}}, \quad \text{for all } j \neq i.$$

Since $B_i \geq (h_i/w_{ij})^2 S_i^2$ and $B_j \geq (h_i/w_{ij})^2 S_j^2$ it follows that

$$\frac{w_{ij}}{[(\sigma_i^2/B_i) + (\sigma_j^2/B_j)]^{1/2}} \geq Q_j$$

so that

$$\Pr\{Coverage\} \geq \Pr\{Z_j < Q_j, \text{ for all } j \neq i\}.$$

By standard arguments it follows that the conditional joint distribution of the Z_j's given $S_1^2, \ldots, S_k^2$ is multivariate normal with mean vector equal to zero and covariance matrix determined by $\text{Var}(Z_j) = 1$, for all $j \neq i$, and for $j \neq \ell$ and $\ell, j \neq i$:

$$\text{Cov}(Z_j, Z_\ell) = \frac{\sigma_i^2/B_i}{[(\sigma_j^2/B_j) + (\sigma_i^2/B_i)]^{1/2}[(\sigma_\ell^2/B_\ell) + (\sigma_i^2/B_i)]^{1/2}} > 0.$$

Since the variances are constant, the covariances are positive, and the Q_j are functions of $S_1^2, \ldots, S_k^2$ and may therefore be regarded as constants when we condition on $S_1^2, \ldots, S_k^2$, we have by Slepian's inequality

$$\Pr\{Z_j < Q_j, \text{ for all } j \neq i | S_1^2, \ldots, S_k^2\} > \prod_{j \neq i} \Pr\{Z_j < Q_j | S_1^2, \ldots, S_k^2\}.$$

Let Φ denote the standard normal distribution function. Notice that the marginal conditional distribution of each Z_j given $S_1^2, \ldots, S_k^2$ is Φ. Thus

$$\Pr\{Z_j < Q_j | S_1^2, \ldots, S_k^2\} = \Phi(Q_j).$$

Combining the previous equation and the previous inequality we obtain

$$\Pr\{Z_j < Q_j, \text{ for all } j \neq i\}$$

$$= E\{\Pr(Z_j < Q_j, \text{ for all } j \neq i | S_1^2, \ldots, S_k^2)\}$$

$$> E\left\{\prod_{j \neq i} \Phi(Q_j)\right\}$$

$$= E\left\{\prod_{j \neq i} \Phi\left(\frac{h_i}{[(\sigma_i^2/B_i) + (\sigma_j^2/B_j)]^{1/2}}\right)\right\}.$$

Let $X_j \equiv (b_{0,j} - 1)S_j^2/\sigma_j^2$, $j = 1, \ldots, k$. By our assumptions the X_j are independent χ^2 random variables with $b_{0,j} - 1$ degrees of freedom. By substitution we have

$$\Pr\{Coverage\} \geq E\left\{\prod_{j \neq i} \Phi\left(\frac{h_i}{\{([b_{0,i}-1]/X_i) + ([b_{0,j}-1]/X_j)\}^{1/2}}\right)\right\}. \quad (5)$$

Notice that the expression on the right-hand side of (5) is independent of $\sigma_1^2, \ldots, \sigma_k^2$. Let f_m denote the density of the χ^2 distribution with m degrees of freedom. Exploiting the independence of the X_j, $j = 1, \ldots, k$, we can simplify (5) to

$$E\left\{\prod_{j \neq i} \Phi\left(\frac{h_i}{\{([b_{0,i} - 1]/X_i) + ([b_{0,j} - 1]/X_j)\}^{1/2}}\right)\right\}$$

$$= E\left\{E\left[\prod_{j \neq i} \Phi\left(\frac{h_i}{\{([b_{0,i} - 1]/X_i) + ([b_{0,j} - 1]/X_j)\}^{1/2}}\right) | X_i\right]\right\}$$

$$= \int_0^\infty \prod_{j \neq i} \left[\int_0^\infty \Phi\left(\frac{h_i}{\{([b_{0,j} - 1]/x) + ([b_{0,i} - 1]/y)\}^{1/2}}\right) \cdot f_{b_{0,j}-1}(x)\, dx\right] f_{b_{0,i}-1}(y)\, dy.$$

Thus, we have shown that $Pr\{Coverage\} \geq 1 - \alpha$ if h_i is determined by the integral equation

$$\int_0^\infty \prod_{j \neq i} \left[\int_0^\infty \Phi\left(\frac{h_i}{\{([b_{0,j} - 1]/x) + ([b_{0,i} - 1]/y)\}^{1/2}}\right) \cdot f_{b_{0,j}-1}(x)\, dx\right] f_{b_{0,i}-1}(y)\, dy = 1 - \alpha. \quad (6)$$

So, by the definition of $Pr\{Coverage\}$,

$$\Pr\{T_{i,j} - (\mu_i - \mu_j) > -w_{ij}, \text{ for all } j \neq i\} \geq 1 - \alpha.$$

Notice that the above inequality holds for any i, so applying Hsu's multiple-bound lemma proves the theorem.

NOTES

[1]Notice that $\mathcal{G}$ is the set of systems i whose MCB upper bound D_i^+ for $\mu_i - \max_{j \neq i} \mu_j$ is positive. Thus, $\mathcal{G}$ is the set of all systems that may be inferred to be the best. However, the most that MCB can infer is that a system is no worse than the best. Therefore, when $\mathcal{G}$ is a singleton $\{i\}$ the lower bound D_i^- must be 0.

ACKNOWLEDGMENT

This work was partially supported by NSF grant DDM-8922721. The authors thank the two anonymous referees for their valuable insights and suggestions.

REFERENCES

BISHOP, T. A. 1976. Heteroscedastic Anova, Manova, and Multiple-Comparisons. Ph.D. Dissertation, Department of Statistics, The Ohio State University, Columbus, Ohio.

CHANG, J. Y., AND J. C. HSU. 1992. Optimal Designs for Multiple Comparisons with the Best. *J. Statist. Plan. and Infer.* **30,** 45–62.

CLARK, G. M., AND W. N. YANG. 1986. A Bonferroni Selection Procedure When Using Common Random Numbers with Unknown Variances. In *Proceedings of the 1986 Winter Simulation Conference*, J. R. Wilson, J. O. Henriksen and S. D. Roberts (eds.), 313–315.

DUDEWICZ, E. J., AND S. R. DALAL. 1975. Allocation of Observations in Ranking and Selection with Unequal Variances. *Sankhya: The Indian J. Statist. Series B* **37,** 28–78.

GOLDSMAN, D., B. L. NELSON AND B. SCHMEISER. 1991. Methods for Selecting the Best System. In *Proceedings of the 1991 Winter Simulation Conference*, B. L. Nelson, G. M. Clark and W. D. Kelton (eds.), 177–186.

HOCHBERG, Y., AND A. C. TAMHANE. 1987. *Multiple Comparison Procedures.* John Wiley, New York.

HSU, J. C. 1981. Simultaneous Confidence Intervals for all Distances from the Best. *Anns. Statist.* **9,** 1026–1034.

MATEJCIK, F. J. 1992. Heteroscedastic Multiple Comparison Procedures for Computer Simulation. Ph.D. Dissertation, Department of Industrial and Systems Engineering, The Ohio State University, Columbus, Ohio.

NAKAYAMA, M. K. 1994. Two-Stage Stopping Procedures Based on Standardized Time Series. *Mgmt. Sci.* **40,** 1189–1206.

NELSON, B. L., AND F. J. MATEJCIK. 1994. Using Common Random Numbers for Indifference-Zone Selection and Multiple Comparisons in Simulation. *Mgmt. Sci.* (to appear).

RINOTT, Y. 1978. On Two-Stage Selection Procedures and Related Probability-Inequalities. *Commun. in Statist. Series A* **7,** 799–811.

SCHMEISER, B. 1982. Batch Size Effects in the Analysis of Simulation Output. *Opns. Res.* **30,** 556–568.

STEIN, C. 1945. A Two-Sample Test for a Linear Hypothesis Whose Power is Independent of the Variance. *Anns. Math. Statist.* **17,** 243–258.

WILCOX, R. R. 1984. A Table For Rinott's Selection Procedure. *J. Qual. Tech.* **16,** 97–100.