Proceedings of the 2001 Winter Simulation Conference
B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, eds.

STATISTICAL SELECTION OF THE BEST SYSTEM

David Goldsman

School of Industrial & Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332, U.S.A.

Barry L. Nelson

Department of Industrial Engineering
and Management Sciences
Northwestern University
Evanston, IL 60208-3119, U.S.A.

ABSTRACT

This tutorial discusses some statistical procedures for selecting the best of a number of competing systems. The term "best" may refer to that simulated system having, say, the largest expected value or the greatest likelihood of yielding a large observation. We describe six procedures for finding the best, three of which assume that the underlying observations arise from competing normal distributions, and three of which are essentially nonparametric in nature. In each case, we comment on how to apply the above procedures for use in simulations.

1 INTRODUCTION

Experiments are often performed to compare two or more system designs in order to determine which scenario is the best. The statistical methods of *screening, selection*, and *multiple comparisons* are applicable when we are interested in making comparisons among a finite, possible large, number of scenarios. The particular method that is appropriate depends on the type of comparison desired and properties of the data under study. For instance, are we interested in comparing means or quantiles? Are the available data independent or correlated within and/or among systems?

In this review, the term "best" may refer to that simulated system having, say, the largest expected value or the greatest likelihood of yielding a large observation. We will typically, but not always, regard the best population as the one having the largest expected value.

We describe six procedures for finding the best, three of which assume that the underlying observations arise from competing normal distributions, and three of which are essentially nonparametric in nature. In each case, we comment on how to apply the above procedures for use in simulations.

Of the normal procedures, the first is a screen-and-select procedure for finding the population with the largest expected value; in this procedure, inferior competitors are screened out after an initial stage of sampling. The second is a sequential procedure that can eliminate inferior choices at any stage and uses the common random numbers variance reduction technique in which we intentionally induce positive correlation between scenarios. The third is an efficient two-stage procedure that also uses common random numbers.

Of the "nonparametric" procedures, the first is a single-stage procedure for finding the most probable multinomial cell, the second is sequential, and the third is a clever augmentation that makes more efficient use of the underlying observations.

The remainder of this article is organized as follows. In the next section, we will give some additional background on screening, selection, and multiple comparisons procedures. Section 3 establishes relevant notation and ground rules. Section 4 discusses the three normal means procedures for selecting the best (or nearly the best) scenario, while Section 5 deals with the three nonparametric procedures. We give conclusions in Section 6.

2 BACKGROUND

We will usually assume that the observations coming from a particular scenario are independent and identically distributed (i.i.d.). Since this is never the case when dealing with simulation output (which is, for instance, almost always serially correlated), we will make appropriate comments to show how to apply the above procedures for use in simulations.

What are screening, selection, and multiple comparisons procedures? Screening and selection procedures (SSPs) are statistical methods designed to find the "best" (or "nearly the best") system from among a collection of competing alternatives. For example, such procedures could be efficacious in any of the following practical situations:

- Find the normal population with the largest mean: A manufacturer would like to know which of three potential plant layouts will maximize expected revenues.

- Find the most probable multinomial cell: A polling service wishes to determine the most popular candidate before a certain election.
- Find the Bernoulli population having the largest success parameter: A medical research team conducts a clinical study comparing the success rates of five different drug regimens for a particular disease.

Informally speaking, SSPs are used to

- *screen* the competitors in order to find a small subset of those systems that contains the best system (or at least a "good" one).
- *select* outright the best system.

In practice, we could invoke a screening procedure to pare down a large number of alternatives into a palatable number; at that point, we might use a selection procedure to make the more fine-tuned choice of the best. Provided that certain assumptions are met, SSPs usually guarantee a user-specified probability of advertised performance—i.e., with high probability, a screening procedure will choose a subset containing the best (or a good) alternative, and a selection procedure will pick the best.

Multiple-comparison procedures (MCPs) treat the comparison problem as an inference problem on the performance parameters of interest. MCPs account for the error that arises when making simultaneous inferences about differences in performance among the systems. Usually, MCPs report to the user simultaneous confidence intervals for the differences. Recent research has shown that MCPs can be combined with SSPs for a variety of problems—including the manufacturing, medical, and polling examples outlined above. In fact, the field has progressed steadily since our most recent tutorials, Goldsman and Nelson (1994, 1998ab).

What is particularly nice about SSPs and MCPs is that they are relevant, easily adaptable, and statistically valid in the context of computer simulation because the assumptions behind the procedures can frequently be satisfied: For example, these procedures sometimes require normality of the observations, an assumption that can often be secured by batching large numbers of (cheaply generated) outputs. Independence can be obtained by controlling random-number assignments. And multiple-stage sampling—which is required by some methods—is feasible in computer simulation because a subsequent stage can be initialized simply by retaining the final random-number seeds from the preceding stage. As a bonus, it is possible to enhance (in a theoretically rigorous way) the performance of some of the procedures through the use of common random numbers, a popular variance reduction technique sometimes used in simulation.

3 SOME NOTATION

To facilitate what follows we define some notation: Let Y_{ij} represent the jth simulation output from system design i, for $i = 1, 2, \ldots, k$ alternatives and $j = 1, 2, \ldots$. For fixed i, we will always assume that the outputs from system i, $Y_{i1}, Y_{i2}, \ldots$, are i.i.d. These assumptions are plausible if $Y_{i1}, Y_{i2}, \ldots$ are outputs across independent replications, or if they are appropriately defined batch means from a single replication after accounting for initialization effects. Let $\mu_i = \mathrm{E}[Y_{ij}]$ denote the expected value of an output from the ith system, and let $\sigma_i^2 = \mathrm{Var}[Y_{ij}]$ denote its variance. Further, let

$$p_i \;=\; \Pr\left\{ Y_{ij} > \max_{\ell \neq i} Y_{\ell j} \right\}$$

be the probability that Y_{ij} is the largest of the jth outputs across all systems *when* $Y_{1j}, Y_{2j}, \ldots, Y_{kj}$ *are mutually independent*.

The methods we describe make comparisons based on either μ_i or p_i. Although not a restriction on either SSPs or MCPs, we will only consider situations in which there is no known functional relationship among the μ_i or p_i (other than $\sum_{i=1}^{k} p_i = 1$). Therefore, there is no potential information to be gained about one system from simulating the others—such as might occur if the μ_i were a function of some explanatory variables—and no potential efficiency to be gained from fractional-factorial experiment designs, group screening designs, etc.

4 NORMAL MEANS PROCEDURES

In this section, we begin with a motivational example, then go over the approaches that are relevant to the goal of selecting the best, and then outline three basic procedures for finding the best.

Example 1 *Simulation models of 25 different inventory policies have been developed for potential implementation at a large distribution/warehouse center. The single measure of system performance is the expected profit achieved while a particular policy is in effect. Differences between different policies' expected profits of less than about \$10,000 are considered practically equivalent.*

4.1 Three Approaches

We discuss briefly the three approaches employed here—subset selection (screening), indifference-zone selection (choosing the single best), and multiple comparisons (inference).

4.1.1 Subset Selection

The subset selection approach is a screening device that attempts to select a (random-size) *subset* of the $k = 25$ competing designs of Example 1 that contains the design with the greatest expected profit. Gupta (1956, 1965) proposed a single-stage procedure for this problem that is applicable in cases when the data from the competing designs are balanced (i.e., having the same number of observations from each contender) and are normal with common (unknown) variance σ^2. Nelson, et al. (2001) handle more general cases—in particular, that in which the unknown variances σ_i^2, $i = 1, 2, \ldots, k$, are not necessarily equal.

4.1.2 Indifference-Zone Selection

If expected profit is taken as the performance measure of interest, then the goal in this example is to select the system with the largest expected profit. In a stochastic simulation such a "correct selection" can never be guaranteed with certainty. A compromise solution offered by *indifference-zone selection* is to guarantee to select the best system with high probability whenever it is at least a user-specified amount better than the others; this "practically significant" difference is called the indifference parameter. In the example the indifference parameter is $\delta = \$10000$. Law and Kelton (2000) describe a number of indifference-zone procedures that have proven useful in simulation, while Bechhofer, Santner, and Goldsman (BSG) (1995) provide a comprehensive review of SSPs.

4.1.3 Multiple Comparisons

MCPs approach the problem of determining the best system by forming simultaneous confidence intervals on the parameters $\mu_i - \max_{j \neq i} \mu_j$ for $i = 1, 2, \ldots, k$, where μ_i denotes the expected profit for the ith inventory policy. These confidence intervals are known as *multiple comparisons with the best (MCB)*, and they bound the difference between the expected performance of each system and the best of the others. The first MCB procedures were developed by Hsu (1984); a thorough review is found in Hochberg and Tamhane (1987).

Matejcik and Nelson (1995) and Nelson and Matejcik (1995) established a fundamental connection between indifference-zone selection and MCB by showing that *most indifference-zone procedures can simultaneously provide MCB confidence intervals with the width of the intervals corresponding to the indifference zone*. The procedure we display below in Section 4.2.1 is a combined subset, indifference-zone selection, and MCB procedure. The advantage of a combined procedure is that we not only select a system as best, we also gain information about how close each of the inferior systems is to being the best. This information is useful if secondary criteria that are not reflected in the performance measure (such as ease of installation, cost to maintain, etc.) may tempt us to choose an inferior system if it is not deficient by much.

4.2 Putting It All Together

We now present three procedures for finding the best (largest mean) normal distribution. The first is a combined screen-select-infer procedure for finding the population with the largest expected value; in this procedure, which uses all three of the approaches outlined in Sections 4.1.1–4.1.3, inferior competitors are screened out after an initial stage of sampling. The second is a sequential procedure that can eliminate inferior choices at any stage and uses the common random numbers variance reduction technique in order to make more precise (and therefore efficient) comparisons among the competing populations. The third is an efficient two-stage procedure that also uses common random numbers.

4.2.1 Subset + Rinott + MCB Procedure

The combined procedure that follows uses a sampling strategy in which the normal observations between scenarios are independent, i.e., Y_{ij} is independent of $Y_{i',j}$ for all $i \neq i'$ and all j. Nelson, et al. (2001) show how to combine a simple subset (screening) procedure with a two-stage indifference-zone selection procedure due to Rinott (1978). After the fact, MCB confidence intervals are then provided for free. The procedure simultaneously guarantees a probability of correct selection and confidence-interval coverage probability of at least $1 - \alpha$ under the stated assumptions. This combined procedure is of great utility when the experimenter is initially faced with a large number of alternatives—the idea is for the subset procedure to pare out non-contending systems, after which Rinott selects the best from the survivors.

Procedure Subset + Rinott + MCB

1. Specify the overall desired probability of correct selection $1 - \alpha$, the indifference-zone parameter δ, a common initial sample size from each scenario $n_0 \geq 2$, and the initial number of competing systems k.
   Further, set

   $$t = t_{1-(1-\alpha/2)^{\frac{1}{k-1}}, n_0-1},$$

   where $t_{\gamma,\nu}$ is the upper-γ quantile of a t-distribution with ν degrees of freedom, and let h solve the following integral.

$$1-\frac{\alpha}{2} = \int_0^\infty \left[\int_0^\infty \Phi\left(\frac{h}{\sqrt{\nu(\frac{1}{x}+\frac{1}{y})}} \right) f_\nu(x)\,dx \right]^{k-1} f_\nu(y)\,dy,$$

where $\Phi(\cdot)$ is the standard normal c.d.f. and $f_\nu(\cdot)$ is the p.d.f. of the χ^2-distribution with $\nu = n_0 - 1$ degrees of freedom. The FORTRAN program `rinott` in BSG (1995) calculates values of h, or one can use the tables in Wilcox (1984) or BSG (1995).

2. Take an i.i.d. sample $Y_{i1}, Y_{i2}, \ldots, Y_{in_0}$ from each of the k normal populations, obtained independently.
3. Calculate the first-stage sample means $\bar{Y}_i^{(1)} = \sum_{j=1}^{n_0} Y_{ij}/n_0$, and marginal sample variances

$$S_i^2 = \frac{\sum_{j=1}^{n_0}(Y_{ij} - \bar{Y}_i^{(1)})^2}{n_0 - 1},$$

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

4. Calculate the quantity

$$W_{ij} = t\left(\frac{S_i^2 + S_j^2}{n_0}\right)^{1/2}$$

for all $i \neq j$. Form the screening subset I, containing every alternative i such that $1 \le i \le k$ and

$$\bar{Y}_i^{(1)} \ge \bar{Y}_j^{(1)} - (W_{ij} - \delta)^+ \text{ for all } j \neq i.$$

5. If I contains a single index, then stop and return that system as the best. Otherwise, for all $i \in I$, compute the second-stage sample sizes

$$N_i = \max\left\{n_0, \left\lceil (hS_i/\delta)^2 \right\rceil\right\},$$

where $\lceil\cdot\rceil$ is the ceiling (integer round-up) function.

6. Take $N_i - n_0$ additional i.i.d. observations from all systems $i \in I$, independently of the first-stage sample and the other systems.
7. Compute the overall sample means $\bar{\bar{Y}}_i = \sum_{j=1}^{N_i} Y_{ij}/N_i$ for $i \in I$.
8. Select the system with the largest $\bar{\bar{Y}}_i$ as best.
9. With probability at least $1-\alpha$, we can claim that

* For all $i \in I^c$, we have $\mu_i < \max_{j\neq i}\mu_j$ (i.e., the systems excluded by the screening are not the best), and

* If we define $J_i = \{j : j \in I \text{ and } j \neq i\}$, then for all $i \in I$,

$$\mu_i - \max_{j\in J_i}\mu_j \in \left[-\left(\bar{\bar{Y}}_i - \max_{j\in J_i}\bar{\bar{Y}}_j - \delta\right)^-, \left(\bar{\bar{Y}}_i - \max_{j\in J_i}\bar{\bar{Y}}_j + \delta\right)^+ \right].$$

(Thus, these confidence intervals bound the difference between each alternative and the best of the others in I.)

4.2.2 Common Random Numbers

A fundamental assumption of the Subset+Rinott+MCB procedure is that the k systems are simulated independently (see Step 2 in that procedure). In practice this means that different streams of pseudo-random numbers are assigned to the simulation of each system. However, under fairly general conditions, assigning common random numbers (CRN) to the simulation of each system decreases the variances of estimates of the pairwise differences in performance. Unfortunately, CRN also complicates the statistical analysis when $k > 2$ systems are involved. The following procedures from Kim and Nelson (2001a) and Nelson and Matejcik (1995) provide (almost) the same guarantees as procedure Subset+Rinott+MCB under a more complex set of conditions, but have been shown to be quite robust to departures from those conditions. And unlike Subset+Rinott+MCB, they are designed to exploit the use of CRN to reduce the total number of observations required to make a correct selection.

4.2.3 Procedures that Allow Common Random Numbers

We next examine a sequential procedure due to Kim and Nelson (2001a) that can eliminate inferior choices at any stage. This procedure uses a sampling strategy in which the normal observations may be dependent due to the use of common random numbers.

The KN procedure is a bit more complicated to implement than the vanilla Rinott (1978) procedure, but it has several distinct advantages. First, once an initial set of n_0 observations is collected from each treatment, KN is parsimonious in taking additional observations in that they are added one-at-a-time and the data are examined to determine if sufficient information has been collected to stop. In contrast, Rinott and its enhancements take potentially large groups of observations. Second, KN allows treatments to be discarded before the final decision; those treatments that appear inferior can legitimately be dropped from further consideration. See Kim and Nelson (2001ab) and Goldsman, et al. (2001) for more details.

Procedure KN

1. Specify the overall desired probability of correct selection $1-\alpha$, the indifference-zone parameter δ, a common initial sample size from each scenario $n_0 \geq 2$, and the initial number of competing systems k. Calculate the constant
$$\eta \;=\; \frac{1}{2}\left[\left(\frac{2\alpha}{k-1}\right)^{-2/(n_0-1)} - 1\right].$$
Further, set $I = \{1, 2, \ldots, k\}$ and let $h^2 = 2\eta(n_0 - 1)$.
2. Take a random sample of n_0 observations Y_{ij} ($1 \leq j \leq n_0$) from population i ($1 \leq i \leq k$). For treatment i compute the sample mean based on the n_0 observations, $\bar{Y}_i(n_0) = \sum_{j=1}^{n_0} Y_{ij}/n_0$ ($1 \leq i \leq k$). For all $i \neq \ell$, compute the sample variance of the difference between treatments i and ℓ,
$$S_{i\ell}^2 \;=\; \frac{1}{n_0-1}\sum_{j=1}^{n_0}\left(Y_{ij} - Y_{\ell j} - [\bar{Y}_i(n_0) - \bar{Y}_\ell(n_0)]\right)^2$$
and set
$$N_{i\ell} \;=\; \left\lfloor h^2 S_{i\ell}^2/\delta^2 \right\rfloor,$$
where $\lfloor\cdot\rfloor$ is the floor (integer round-down) function. Finally, for all i set
$$N_i \;=\; \max_{\ell \neq i} N_{i\ell}.$$
If $n_0 > \max_i N_i$, then stop and select the population with the largest sample mean $\bar{Y}_i(n_0)$ as one having the largest mean. Otherwise, set the sequential counter $r = n_0$ and go to the Screening phase of the procedure.
3. *Screening*: Set $I^{\text{old}} = I$ and re-set
$$I \;=\; \{i : i \in I^{\text{old}} \text{ and } \bar{Y}_i(r) \geq \bar{Y}_\ell(r) - W_{i\ell}(r), \text{ for all } \ell \in I^{\text{old}}, \ell \neq i\},$$
where
$$W_{i\ell}(r) \;=\; \max\left\{0, \frac{\delta}{2r}\left(\frac{h^2 S_{i\ell}^2}{\delta^2} - r\right)\right\}.$$
4. *Stopping Rule*: If $|I| = 1$, then stop and select the treatment with index in I as having the largest mean. If $|I| > 1$, take one additional observation $Y_{i,r+1}$ from each treatment $i \in I$. Increment $r = r + 1$ and go to the screening stage if $r < \max_i N_i + 1$. If $r = \max_i N_i + 1$, then stop and select the treatment associated with the largest $\bar{Y}_i(r)$ having index $i \in I$.

The following procedure due to Nelson and Matejcik (1995) is a two-stage procedure, like Subset+Rinott+MCB, but different in that it exploits the use of common random numbers.

Procedure NM + MCB

1. Specify the constants δ, α, and $n_0 \geq 2$. Let $g = T^{(\alpha)}_{k-1,(k-1)(n_0-1),0.5}$, an equicoordinate critical point of the equicorrelated multivariate central t-distribution; this constant can be found in Hochberg and Tamhane (1987), Appendix 3, Table 4; BSG (1995); or by using the FORTRAN program `AS251` of Dunnett (1989).
2. Take an i.i.d. sample $Y_{i1}, Y_{i2}, \ldots, Y_{in_0}$ from each of the k systems *using CRN across systems*.
3. Compute the approximate sample variance of the difference of the sample means
$$S^2 = \frac{2\sum_{i=1}^{k}\sum_{j=1}^{n_0}\left(Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot}\right)^2}{(k-1)(n_0-1)},$$
where $\bar{Y}_{i\cdot} = \sum_{j=1}^{n_0} Y_{ij}/n_0$, $\bar{Y}_{\cdot j} = \sum_{i=1}^{k} Y_{ij}/k$, and $\bar{Y}_{\cdot\cdot} = \sum_{i=1}^{k}\sum_{j=1}^{n_0} Y_{ij}/kn_0$.
4. Compute the final sample size
$$N = \max\left\{n_0, \left\lceil (gS/\delta)^2 \right\rceil\right\}.$$
5. Take $N - n_0$ additional i.i.d. observations from each system, using CRN across systems.
6. Compute the overall sample means $\bar{\bar{Y}}_i = \sum_{j=1}^{N} Y_{ij}/N$ for $i = 1, 2, \ldots, k$.
7. Select the system with the largest $\bar{\bar{Y}}_i$ as best.
8. Simultaneously form the MCB confidence intervals
$$\mu_i - \max_{j\neq i}\mu_j \in$$
$$\left[-\left(\bar{\bar{Y}}_i - \max_{j\neq i}\bar{\bar{Y}}_j - \delta\right)^-, \left(\bar{\bar{Y}}_i - \max_{j\neq i}\bar{\bar{Y}}_j + \delta\right)^+\right]$$
for $i = 1, 2, \ldots, k$.

4.3 Simulation Considerations

As we have already mentioned, the methods described in this section all rely on our ability to generate i.i.d. normal observations within each scenario. (We may or may not want the observations to be independent between systems, especially if we are thinking of using common random numbers.) Of course, data from a simulation are rarely i.i.d. normal—but we can achieve approximate notmality by taking sample averages of contiguous observations (batching); and we can achieve independence by running independent replications. Moderate departures from normality do not really pose a problem, since all of the procedures appear to be robust in that sense (see BSG 1995 or Goldsman, et al. 2001). But one really ought to make sure that the Y_{ij}'s are indeed independent within each scenario, for procedure performance seriously deteriorates when that assumption fails.

5 MULTINOMIAL PROCEDURES

This section begins with a version of the motivational example from Section 4, but with a slightly different criterion for describing the best alternative. We then describe three procedures to achieve the new goal of finding the best.

Example 2 *Simulation models of 25 different inventory policies have been developed for potential implementation at a large distribution/warehouse center. The goal now is to select the system that is most likely to have the largest actual profit (instead of the largest expected profit).*

5.1 Setup

We define p_i as the probability that design i will produce the largest profit from a given vector-observation $\boldsymbol{Y}_j = (Y_{1j}, Y_{2j}, \ldots, Y_{kj})$. The goal now is to select the design associated with the largest p_i-value. This goal is equivalent to that of finding the multinomial category having the largest probability of occurrence; and there is a rich body of literature concerning such problems. In fact, we make almost no assumptions on the underlying distributions of the competing populations—thus, the procedures to be discussed below are, in a sense, nonparametric.

More specifically, suppose that we want to select the correct category with probability $1-\alpha$ whenever the ratio of the largest to second largest p_i is greater than some user-specified constant, say $\theta > 1$. The indifference constant θ can be regarded as the smallest ratio "worth detecting."

5.2 The Procedures

This subsection describes three "nonparametric" procedures. The first is a single-stage procedure for finding the most probable multinomial cell, the second is a sequential procedure, and the third is a clever augmentation of the first that makes more efficient use of the underlying observations.

5.2.1 Single-Stage Procedure

The following *single-stage* procedure was proposed by Bechhofer, Elmaghraby, and Morse (BEM) (1959) to guarantee the above probability requirement.

Procedure BEM

1. For the given k, and (α, θ) specified prior to the start of sampling, find n from the tables in BEM (1959), Gibbons, Olkin, and Sobel (1977) or BSG (1995).
2. Take a random sample of n observations $Y_{i1}, Y_{i2}, \ldots, Y_{in}$ from each alternative i, $i = 1, 2, \ldots, k$. Turn these into n independent multinomial observations, $\boldsymbol{X}_j = (X_{1j}, X_{2j}, \ldots, X_{kj})$, $j = 1, 2, \ldots, n$, by setting

$$X_{ij} = \begin{cases} 1, & \text{if } Y_{ij} > \max_{\ell \neq i}\{Y_{\ell j}\} \\ 0, & \text{otherwise,} \end{cases}$$

   where we assume (for notational convenience) that there are never ties for the maximum observation within a particular vector $\boldsymbol{Y}_j$.
3. Let $W_i = \sum_{j=1}^{n} X_{ij}$ for $i = 1, 2, \ldots, k$. Select the design that yielded the largest W_i as the one associated with the largest p_i (randomize in the case of ties).

5.2.2 Sequential Procedure

A more efficient procedure, due to Bechhofer and Goldsman (1986), uses *closed, sequential* sampling; that is, the procedure stops when one design is "sufficiently ahead" of the others.

Procedure BG

1. For the given k, and (α, θ) specified prior to the start of sampling, find the *truncation number* (i.e., an upper bound on the number of vector-observations) n_0 from the tables in Bechhofer and Goldsman (1986) or BSG (1995).
2. At the mth stage of experimentation ($m \geq 1$), take the random multinomial observation $\boldsymbol{X}_m = (X_{1m}, X_{2m}, \ldots, X_{km})$ (defined above) and calculate the *ordered* category totals $W_{[1]m} \leq W_{[2]m} \leq \cdots \leq W_{[k]m}$; also calculate

$$Z_m = \sum_{i=1}^{k-1} (1/\theta)^{(W_{[k]m} - W_{[i]m})}.$$

3. Stop sampling at the first stage when *either*

$$Z_m \le \alpha/(1-\alpha) \quad or \quad m = n_0$$

$$or \quad W_{[k]m} - W_{[k-1]m} \ge n_0 - m,$$

whichever occurs first.
4. Let N (a random variable) denote the stage at which the procedure terminates. Select the design that yielded the largest W_{iN} as the one associated with the largest p_i (randomize in the case of ties).

5.2.3 Augmentation of BEM

Miller, Nelson, and Reilly (1998) present a remarkably efficient procedure that directly uses the original Y_{ij} observations (instead of the 0-1 X_{ij}, which lose information). Their procedure AVC, based on all possible vector comparisons of the observations, always results in an increased probability of correct selection when compared to the analogous implementation of the BEM procedure.

Procedure AVC

1. For the given k, and (α, θ) specified prior to the start of sampling, use the same n as in BEM.
2. Take a random sample of n observations $Y_{i1}, Y_{i2}, \ldots, Y_{in}$ from each alternative i, $i = 1, 2, \ldots, k$. Consider all n^k vectors of the form $\boldsymbol{Y}'_j = (Y'_{1j}, Y'_{2j}, \ldots, Y'_{kj})$, $j = 1, 2, \ldots, n^k$, where Y'_{ij} is one of the n observations from alternative i. Turn these into n^k (non-independent) multinomial observations, $\boldsymbol{X}'_j = (X'_{1j}, X'_{2j}, \ldots, X'_{kj})$, $j = 1, 2, \ldots, n^k$, by setting

$$X'_{ij} = \begin{cases} 1, & \text{if } Y'_{ij} > \max_{\ell \ne i}\{Y'_{\ell j}\} \\ 0, & \text{otherwise,} \end{cases}$$

where we again assume that there are never ties for the maximum observation within a particular vector $\boldsymbol{Y}'_j$.
3. Let $W'_i = \sum_{j=1}^{n^k} X'_{ij}$ for $i = 1, 2, \ldots, k$. Select the design that yielded the largest W'_i as the one associated with the largest p_i (randomize in the case of ties).

6 FINAL THOUGHTS

Space limitations preclude detailed discussion, but we also mention the interesting technical results to be found in Damerdji, et al. (1997ab), Damerdji and Nakayama (1996), and Nakayama (1997), in which the authors rigorously show that certain selection-of-the-best procedures satisfy probability requirements similar to those in Section 4. Chen, et al. (1997) propose a completely different approach in their discussion on optimal budget strategies. Chick (1997) takes a decision-theoretic view on the selection-of-the best problem. And Boesel and Nelson (1998) use the techniques discussed in the current paper to present a methodology for optimization of stochastic systems.

ACKNOWLEDGMENTS

Portions of this article were published previously in Goldsman and Nelson (1994, 1998a). Our work was supported by National Science Foundation Grants DMI-9622065 and DMI-9622269.

REFERENCES

Bechhofer, R. E., S. Elmaghraby, and N. Morse. 1959. A single-sample multiple decision procedure for selecting the multinomial event which has the highest probability. *Ann. Math. Stat.* 30:102–119.

Bechhofer, R. E., and D. M. Goldsman. 1986. Truncation of the Bechhofer-Kiefer-Sobel sequential procedure for selecting the multinomial event which has the largest probability (II): Extended tables and an improved procedure. *Comm. Stat.—Simul. and Comp.* B15:829–851.

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

Boesel, J., and B. L. Nelson. 1998. Accounting for randomness in heuristic simulation optimization. *IIE Transactions*, forthcoming.

Chen, H.-C., C.-H. Chen, L. Dai, and E. Yücesan. 1997. New development of optimal computing budget allocation for discrete event simulation. In *Proc. 1997 Winter Simulation Conference*, ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson, 334–341. Piscataway, New Jersey: IEEE.

Chick, S. E. 1997. Selecting the best system: A decision-theoretic approach. In *Proc. 1997 Winter Simulation Conference*, ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson, 326–333. Piscataway, New Jersey: IEEE.

Damerdji, H., P. W. Glynn, M. K. Nakayama, and J. R. Wilson. 1997a. Selection of the best system in steady-state simulations. Technical Report 97–5, Dept. of Industrial Engineering, North Carolina State Univ., Raleigh, North Carolina.

Damerdji, H., P. W. Glynn, M. K. Nakayama, and J. R. Wilson. 1997b. Selection of the best system in transient simulations. Technical Report 97–6, Dept. of Industrial

Engineering, North Carolina State Univ., Raleigh, North Carolina.

Damerdji, H. and M. K. Nakayama. 1996. Two-stage multiple-comparison procedures for steady-state simulations. Technical Report, Dept. of Industrial Engineering, North Carolina State Univ., Raleigh, North Carolina.

Dunnett, C. W. 1989. Multivariate normal probability integrals with product correlation structure. *Applied Statistics* 38:564–579. Correction: 42:709.

Gibbons, J. D., I. Olkin, and M. Sobel. 1977. *Selecting and Ordering Populations: A New Statistical Methodology*. New York: John Wiley and Sons.

Goldsman, D., S.-H. Kim, W. S. Marshall, and B. L. Nelson. 2001. Ranking and selection for steady-state simulation: Procedures and perspectives. Working Paper, Department of Industrial Engineering & Management Sciences, Northwestern University, Evanston, Illinois.

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. Tew, M. S. Manivannan, D. A. Sadowski, and A. F. Seila, 192–199. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Goldsman, D., and B. L. Nelson. 1998a. Statistical screening, selection, and multiple comparison procedures in computer simulation. In *Proceedings of the 1998 Winter Simulation Conference*, ed. D. J. Medeiros, E. F. Watson, M. Manivannan, and J. S. Carson, 159–166. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Goldsman, D., and B. L. Nelson. 1998b. Comparing systems via simulation. *Handbook of Simulation*, ed. J. Banks, Chapter 8. New York: John Wiley and Sons.

Gupta, S. S. 1956. *On a Decision Rule for a Problem in Ranking Means*. Ph.D. Dissertation (Mimeo. Ser. No. 150). Inst. of Statist., Univ. of North Carolina, Chapel Hill, North Carolina.

Gupta, S. S. 1965. On some multiple decision (selection and ranking) rules. *Technometrics* 7:225–245.

Hochberg, Y., and A. C. Tamhane. 1987. *Multiple Comparison Procedures*. New York: John Wiley and Sons.

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

Kim, S-H. and B. L. Nelson. 2001a. A fully sequential procedure for indifference-zone selection in simulation. *ACM TOMACS*, in press.

Kim, S-H. and B. L. Nelson. 2001b. On the asymptotic validity of fully sequential selection procedures for steady-state simulation. Working Paper, Department of Industrial Engineering & Management Sciences, Northwestern University, Evanston, Illinois.

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

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

Miller, J. O., B. L. Nelson, and C. H. Reilly. Efficient multinomial selection in simulation. 1998. *Naval Research Logistics* 45:459–482.

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

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

Nelson, B. L., J. Swann, D. Goldsman, and W.-M. Song. 2001. Simple procedures for selecting the best system when the number of alternatives is large. To appear in *Operations Research*.

Rinott, Y. 1978. On two-stage selection procedures and related probability-inequalities. *Comm. Stat.—Thy. and Meth.* A7:799–811.

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

AUTHOR BIOGRAPHIES

DAVID GOLDSMAN is a Professor in the School of Industrial and Systems Engineering at the Georgia Institute of Technology. His research interests include simulation output analysis and ranking and selection. He is the Simulation Department Editor for *IIE Transactions*, and an Associate Editor for *Operations Research Letters*. He is an active participant in the Winter Simulation Conference, having been Program Chair in 1995. His e-mail address is `<sman@isye.gatech.edu>`, and his web page is `<www.isye.gatech.edu/~sman>`.

BARRY L. NELSON is a Professor in the Department of Industrial Engineering and Management Sciences at Northwestern University, and is Director of the Master of Engineering Management program there. He is interested in the design and analysis of computer simulation experiments, particularly statistical efficiency, multivariate output analysis, and input modeling. He was Program Chair of the 1997 Winter Simulation Conference, and is currently a member of the Board of Directors. His e-mail address is `<nelsonb@northwestern.edu>`, and his web page is `<www.iems.northwestern.edu/~nelsonb/>`