

SIMULTANEOUS RANKING, SELECTION AND MULTIPLE COMPARISONS FOR SIMULATION

Frank J. Matejick

Industrial Engineering
South Dakota School of Mines & Technology
Rapid City, South Dakota 57701-3995

Barry L. Nelson

Department of Industrial & Systems Engineering
The Ohio State University
Columbus, Ohio 43210-1271

ABSTRACT

We present three experiment design and analysis procedures that simultaneously provide indifference-zone selection and multiple-comparison inference for choosing the best among k systems. One procedure is appropriate when the systems are simulated independently; the other two are appropriate in conjunction with common random numbers. All are easy to apply. We illustrate all three procedures with a numerical example.

1 INTRODUCTION

In this paper we consider the problem of comparing a small number of systems, say 2 to 20, in terms of the expected value of some stochastic performance measure. We assume that expected performance will be estimated via a simulation experiment. At a gross level we are interested in which system is best, where "best" is defined to be maximum or minimum expected performance. At a more refined level we may also be interested in how much better the best is relative to each alternative, since secondary criteria not reflected in the performance parameter (such as ease of installation, cost to maintain, etc.) may tempt us to choose an inferior system if it is not deficient by much.

Since we are estimating expected performance we can neither select the best system nor bound the differences between systems with certainty. Instead, we present procedures that simultaneously control the error in selecting the best and bounding the differences. These procedures unify standard ranking and selection procedures—that control the error when choosing the best—and standard multiple-comparison procedures—that control the error in making simultaneous comparisons. The procedures depend upon having normally distributed data, but they do not require known or equal variances across

systems, and some of the procedures allow the use of common random numbers to reduce the computational effort. When the simulation outputs are sample averages, the normality assumption is typically not a serious restriction.

The paper is organized as follows: The next section presents a motivating example that will be used throughout the paper to illustrate the new procedures. We then provide background necessary to understand the inference that the new procedures provide. Section 4 contains the procedures themselves along with a numerical example. We close with some conclusions in Section 5.

2 A MOTIVATING EXAMPLE

Consider an (s, S) inventory system in which some discrete item is periodically reviewed. If the inventory level is found to be below s units, then an order is issued to bring the inventory level up to S units; otherwise no additional items are ordered. Different (s, S) inventory policies result in different inventory systems. Koenig and Law (1985) used this example to illustrate a subset selection procedure; see their paper for a detailed description of the model. The only stochastic input process in the simulation is the demand for inventory in each period.

Suppose that five (s, S) inventory policies have been identified for study, and we are interested in which policy has the minimum expected cost per period for 30 periods, where cost is measured in thousands of dollars. Differences of less than 1 (thousand dollars) are considered practically insignificant, so while we want to choose the best system we also want to know which policies are practically equivalent to the best policy.

3 BACKGROUND

Let Y_{ij} represent the output from the j th replication (or batch mean) of system i , for $i = 1, 2, \dots, k$, so that $\mathbf{Y}_j = (Y_{1j}, Y_{2j}, \dots, Y_{kj})'$ is the $k \times 1$ vector of outputs across all systems on replication j . We assume throughout that $\mathbf{Y}_1, \mathbf{Y}_2, \dots$ are i.i.d., and that $\mathbf{Y}_j \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the multivariate normal distribution with unknown mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_k)'$ and unknown variance-covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ & & \ddots & \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{pmatrix}.$$

We are interested in comparing the k systems in terms of their expected performance, μ_i . In the inventory example there are $k = 5$ policies, Y_{ij} is the average cost for 30 periods observed on the j th replication of the i th inventory policy, and μ_i is the expected cost per period of the i th inventory policy

If we simulate the systems independently—meaning we use different random numbers to drive the simulation of each system—then

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \sigma_{kk} \end{pmatrix}.$$

However, since using common random numbers (CRN) across systems is known to reduce the variance of comparisons, we are also interested in the case when CRN forces the covariances $\sigma_{ij} > 0$, for $i \neq j$. In the inventory example we used CRN to force each inventory policy to be subjected to the same sequence of demands, providing a fairer comparison of policy performance.

Output-analysis methods that exploit CRN and furnish appropriate statistical inference have long been of interest to the simulation community. Yang and Nelson (1991) provide additional references and some solutions for multiple comparisons in conjunction with CRN. This paper provides methods for simultaneously selecting the best system and providing confidence-intervals for the differences under CRN, as well as for independently simulated systems. Moreover, all of our methods allow for unequal and unknown variances across the systems.

The following sections present the confidence-interval procedure, multiple comparisons with the best, and the decision-theory procedure, indifference-zone selection. We show that both types of inference can be attained simultaneously from a single experiment.

3.1 Multiple Comparisons with the Best

Suppose that larger μ_i is better. *Multiple Comparisons with the Best (MCB)* provides simultaneous confidence intervals for the parameters

$$\mu_i - \max_{j \neq i} \mu_j$$

for $i = 1, 2, \dots, k$. These confidence intervals bound the difference between the performance of each system and the best of the others with prespecified confidence level. For minimization problems, such as the inventory example, we consider $\mu_i - \min_{j \neq i} \mu_j$, for $i = 1, 2, \dots, k$.

Most MCB procedures assume the variances across systems are equal. See Hochberg and Tamhane (1987) for a general discussion of MCB procedures.

3.2 Ranking and Selection

Let

$$\mu_{(1)} \leq \mu_{(2)} \leq \cdots \leq \mu_{(k)}$$

be the (unknown) ordered means. Two-stage, indifference-zone selection procedures yield estimators $\hat{\mu}_i, i = 1, 2, \dots, k$, that guarantee

$$\Pr \{ \hat{\mu}_{(k)} > \hat{\mu}_{(i)}, \forall i \neq k \mid \mu_{(k)} - \mu_{(i)} \geq \delta \} \geq 1 - \alpha$$

independent of the system variances, where $\hat{\mu}_{(i)}$ is the estimate associated with the (unknown) system having the i th smallest expectation. The user-specified value δ is called the *indifference zone*. The implication is that if we use the procedure and then select the system with the largest performance estimate $\hat{\mu}_i$ as the best system we will be correct with probability $\geq 1 - \alpha$ when the best is at least δ better than the others. In the inventory example we could set $\delta = 1$ thousand dollars since we are indifferent to policies with expected costs that differ by less than 1 thousand dollars.

Indifference-zone-selection procedures typically do not exploit CRN, and do not provide inference about systems other than the best. Goldsman (1983) provides an exposition of indifference-zone selection and related topics.

3.3 Simultaneous Ranking, Selection and Multiple Comparisons

In this section we establish that MCB intervals and indifference-zone selection can be derived simultaneously from the same experiment. We begin with a result, due to Hsu (1984), that establishes sufficient conditions under which MCB intervals can be formed.

Result (Hsu): If

$$\Pr \{ \hat{\mu}_{(k)} - \hat{\mu}_{(i)} - (\mu_{(k)} - \mu_{(i)}) > -w, \forall i \neq k \} \geq 1 - \alpha$$

then with probability $\geq 1 - \alpha$

$$\mu_i - \max_{j \neq i} \mu_j \in$$

$$\left[- \left(\hat{\mu}_i - \max_{j \neq i} \hat{\mu}_j - w \right)^-, \left(\hat{\mu}_i - \max_{j \neq i} \hat{\mu}_j + w \right)^+ \right]$$

for $i = 1, 2, \dots, k$, where $-x^- = \min\{0, x\}$ and $x^+ = \max\{0, x\}$.

The quantity w is the *whisker length* of the MCB intervals, and it is analogous to the half width of symmetric confidence intervals. In standard MCB procedures w is a random variable, but the next theorem establishes that the estimators resulting from indifference-zone-selection procedures with indifference zone δ satisfy the condition in Hsu's result with $w = \delta$. Therefore, both types of inference can be derived simultaneously from the same experiment, and the whisker length of the MCB intervals can be specified in advance. This theorem and its proof appear in Matejcek and Nelson (1992).

Theorem (Matejcek & Nelson): For standard ranking and selection estimators $\hat{\mu}_i$ (e.g., sample means and generalized sample means), the statement

$$\Pr \{ \hat{\mu}_{(k)} > \hat{\mu}_{(i)}, \forall i \neq k \mid \mu_{(k)} - \mu_{(i)} \geq w \} \geq 1 - \alpha$$

implies that

$$\Pr \{ \hat{\mu}_{(k)} - \hat{\mu}_{(i)} - (\mu_{(k)} - \mu_{(i)}) > -w, \forall i \neq k \} \geq 1 - \alpha$$

for *any* values of the true means.

The primary consequence of this result is that we can use the outcome of a two-stage, indifference-zone-selection procedure to form MCB intervals with whisker length equal to the indifference zone, and simultaneously guarantee both the probability of correct selection and coverage of the MCB differences with overall confidence level $1 - \alpha$. In the following section we present procedures that exploit this theorem.

4 PROCEDURES

We now present three combined indifference-zone selection and MCB procedures, and illustrate them using the inventory example described in Section 2. As

is traditional, the procedures are stated in a maximization context, but the inventory example used to illustrate the procedures is a minimization problem.

The first procedure assumes the systems are simulated independently (different random numbers). The second and third procedures allow CRN. All of the procedures assume that the simulation output data are normally distributed. We present versions of each procedure that are easy to apply in practice; there are more complex versions that have advantages in certain situations.

In the procedures we use the convention that a “.” subscript indicates averaging with respect to that subscript. For example, \bar{Y}_i is the sample average of $Y_{i1}, Y_{i2}, \dots, Y_{in}$.

4.1 Rinott's Procedure

The first procedure, which requires independently simulated systems, is an extension of Rinott's (1978) indifference-zone-selection procedure.

Procedure \mathcal{R}

1. Specify w , α , and n_0 . Let h solve Rinott's integral for n_0, k and α (see tables in Wilcox 1984).
2. Take i.i.d. sample $Y_{i1}, Y_{i2}, \dots, Y_{in_0}$ from each of the k systems *simulated independently*.
3. Compute the sample variances

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

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

4. Compute the final sample sizes

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

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

5. Take $N_i - n_0$ additional i.i.d. observations from system i , independently of the first-stage sample and the other systems.
6. Compute the overall sample means

$$\bar{\bar{Y}}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} Y_{ij}$$

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

7. Select the system with the largest $\bar{\bar{Y}}_i$ as best and simultaneously form the MCB confidence intervals

$$\mu_i - \max_{j \neq i} \mu_j \in \left[- \left(\bar{Y}_{i.} - \max_{j \neq i} \bar{Y}_{j.} - w \right)^-, \left(\bar{Y}_{i.} - \max_{j \neq i} \bar{Y}_{j.} + w \right)^+ \right]$$

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

To illustrate the procedure we simulated the $k = 5$ inventory systems independently with first-stage number of replications $n_0 = 10$, indifference-zone $w = 1$ (thousand dollars), and confidence level $1 - \alpha = 0.95$. From the tables in Wilcox (1984) we obtain $h = 3.692$. The procedure selected inventory policy 2 as the best (that is, policy 2 had the smallest estimated cost per period), and provided the MCB intervals in Table 1. Recall that these are confidence intervals on $\mu_i - \min_{j \neq i} \mu_j$, where μ_i is the unknown expected cost per period of inventory policy i .

The point estimate for $\mu_2 - \min_{j \neq 2} \mu_j$ is -1.4 , indicating that policy 2 appears to be 1.4 thousand dollars less expensive than the best of the other policies. The intervals tell us that, with confidence level 0.95, policy 2 is no worse than any of the others (the upper endpoint of the confidence interval is 0), and it may be as much as 2.4 thousand dollars less expensive (the lower endpoint is -2.4). Notice that the whisker length is $w = -1.4 - (-2.4) = 1$, precisely as specified.

The intervals also indicate that the other four policies are inferior to policy 2 (the lower endpoints of their intervals are all 0), and may be as much as 2.4, 18.1, 18.6 and 35.3 thousand dollars more expensive for policies 1, 3, 4 and 5, respectively. These constrained intervals—which either contain 0 or have 0 as one endpoint—are a feature of MCB. Technically, the most MCB can declare is that a system is *no worse* than the best, it cannot declare that the system is better.

Also presented in Table 1 are the total sample sizes (number of replications) for each policy, N_i . They range from 130 replications for policy 5 to 270 for policy 3, for a total of 1072 replications. The differences are a function of the variances of the systems; the larger the variance the greater the number of replications. Procedures that exploit CRN should reduce the number of replications required to attain the same confidence level and whisker length.

4.2 Clark and Yang’s Procedure

The second procedure, which allows the systems to be simulated under CRN, is an extension of Clark

Table 1: MCB Results for Procedure \mathcal{R}

i	N_i	lower	$\bar{Y}_{i.} - \min_{j \neq i} \bar{Y}_{j.}$	upper
1	220	0	1.4	2.4
2	211	-2.4	-1.4	0
3	270	0	17.1	18.1
4	241	0	17.6	18.6
5	130	0	34.3	35.3

and Yang’s (1986) indifference-zone-selection procedure. This procedure exploits the Bonferroni inequality to account for the dependence induced by CRN. Thus, it is a conservative procedure that typically prescribes more replications than actually needed to make a correct selection under CRN. In the procedure, $t = t_{1 - \frac{\alpha}{k-1}, n_0 - 1}$ is the $(1 - \frac{\alpha}{k-1})$ -quantile of the t distribution with $n_0 - 1$ degrees of freedom.

Procedure \mathcal{CY}

1. Specify w , α and n_0 . Let $t = t_{1 - \frac{\alpha}{k-1}, n_0 - 1}$.
2. Take i.i.d. sample $Y_{i1}, Y_{i2}, \dots, Y_{in_0}$ from each of the k systems *using CRN across systems*.
3. Compute the sample variances of the differences

$$S_{ij}^2 = \frac{1}{n_0 - 1} \sum_{t=1}^{n_0} (Y_{it} - Y_{jt} - (\bar{Y}_{i.} - \bar{Y}_{j.}))^2$$

for all $i \neq j$.

4. Compute the second-stage sample size

$$N = \max \left\{ n_0, \left[\max_{j \neq i} (t S_{ij} / w)^2 \right] \right\}.$$

5. Take $N - n_0$ additional i.i.d. observations from each system, using CRN across systems.
6. Compute the overall sample means

$$\bar{Y}_{i.} = \frac{1}{N} \sum_{j=1}^N Y_{ij}$$

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

7. Select the system with the largest $\bar{Y}_{i.}$ as best and simultaneously form the MCB confidence intervals

$$\mu_i - \max_{j \neq i} \mu_j \in$$

$$\left[- \left(\bar{Y}_{i.} - \max_{j \neq i} \bar{Y}_{j.} - w \right)^-, \left(\bar{Y}_{i.} - \max_{j \neq i} \bar{Y}_{j.} + w \right)^+ \right]$$

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

We performed the same experiment for the inventory example, but this time using CRN across systems. The value of $t = t_{1-\frac{0.05}{4}, 9} = 2.685$. This procedure also selected inventory policy 2 as the best, but did so with many fewer total replications (875 to 1072). The MCB results are displayed in Table 2.

Procedure \mathcal{CY} is clearly superior to \mathcal{R} in this example. Unfortunately, procedures based on the Bonferroni inequality become more conservative as the number of systems, k , increase. At some point this conservatism overwhelms the benefit from CRN; avoiding this problem is the motivation for the procedure presented in the next section.

Table 2: MCB Results for Procedure \mathcal{CY}

i	N	lower	$\bar{Y}_{i.} - \min_{j \neq i} \bar{Y}_{j.}$	upper
1	175	0	1.4	2.4
2	175	-2.4	-1.4	0
3	175	0	17.9	18.9
4	175	0	18.1	19.1
5	175	0	34.4	35.4

4.3 Nelson and Matejcek's Procedure

The third procedure, which also allows the systems to be simulated under CRN, is an extension of Nelson's (1993) robust MCB procedure. This procedure assumes that Σ has a particular structure known as *sphericity*, specifically

$$\Sigma = \begin{pmatrix} 2\psi_1 + \tau^2 & \psi_1 + \psi_2 & \cdots & \psi_1 + \psi_r \\ \psi_2 + \psi_1 & 2\psi_2 + \tau^2 & \cdots & \psi_2 + \psi_r \\ & & \ddots & \\ \psi_r + \psi_1 & \psi_r + \psi_2 & \cdots & 2\psi_r + \tau^2 \end{pmatrix}.$$

Sphericity implies that

$$\text{Var}[Y_{ij} - Y_{\ell j}] = 2\tau^2$$

for all $i \neq \ell$. In other words, the variances of all pairwise differences across systems are equal, even though the marginal variances and covariance may be

unequal. Sphericity generalizes *compound symmetry*, which is

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ & & \ddots & \\ \rho & \rho & \cdots & 1 \end{pmatrix}.$$

Compound symmetry has been assumed by many researchers to account for the effect of CRN.

The procedure below is exact when Σ satisfies sphericity. Nelson (1993) showed that an MCB procedure based on the assumption of sphericity is remarkably robust to departures from sphericity provided that the covariances $\sigma_{ij} \geq 0$, the assumed effect of CRN. The combined indifference-zone-selection and MCB procedure will be similarly robust.

In the procedure, $g = T_{k-1, (k-1)(n_0-1), \frac{1}{2}}^{(1-\alpha)}$ is the $(1-\alpha)$ -quantile of the maximum of a multivariate t random variable of dimension $k-1$ with $(k-1)(n_0-1)$ degrees of freedom and common correlation $1/2$; see, for instance, Table 4 in Hochberg and Tamhane (1987).

Procedure \mathcal{NM}

1. Specify w , α and n_0 . Let $g = T_{k-1, (k-1)(n_0-1), \frac{1}{2}}^{(1-\alpha)}$.
2. Take i.i.d. sample $Y_{i1}, Y_{i2}, \dots, Y_{in_0}$ from each of the k systems using CRN across systems.
3. Compute the approximate sample variance of the difference

$$S^2 = \frac{2 \sum_{i=1}^k \sum_{j=1}^{n_0} (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2}{(k-1)(n_0-1)}.$$

4. Compute the second-stage sample size

$$N = \max \{ n_0, \lceil (gS/w)^2 \rceil \}.$$

5. Take $N - n_0$ additional i.i.d. observations from each system, using CRN across systems.
6. Compute the overall sample means

$$\bar{Y}_{i.} = \frac{1}{N} \sum_{j=1}^N Y_{ij}$$

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

7. Select the system with the largest $\bar{Y}_{i.}$ as best and simultaneously form the MCB confidence intervals

$$\mu_i - \max_{j \neq i} \mu_j \in \left[- \left(\bar{Y}_i - \max_{j \neq i} \bar{Y}_j - w \right)^-, \left(\bar{Y}_i - \max_{j \neq i} \bar{Y}_j + w \right)^+ \right]$$

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

We performed the same experiment for the inventory example using CRN across systems. The value of $g = T_{4,36,\frac{1}{2}}^{(0.95)} = 2.24$. This procedure also selected inventory policy 2 as the best, but did so with significantly fewer total replications (330 to 875 to 1072). The MCB results are displayed in Table 3, and are nearly identical to the results obtained by the other procedures.

Table 3: MCB Results for Procedure \mathcal{NM}

i	N	lower	$\bar{Y}_i - \min_{j \neq i} \bar{Y}_j$	upper
1	66	0	1.5	2.5
2	66	-2.5	-1.5	0
3	66	0	18.2	19.2
4	66	0	18.2	19.2
5	66	0	34.7	35.7

5 SUMMARY AND CONCLUSIONS

In this paper we have presented three new procedures that simultaneously control the error in selecting the best of k systems, and comparing the best system to each of the other competitors. These procedures are based on a theorem that allows MCB confidence intervals to be appended to indifference-zone-selection procedures, thereby unifying an inference approach and a decision-theory approach. This inference can be displayed in an intuitive graphical representation, and facilitates making selections based on secondary criteria that are not reflected in the performance parameter.

Two of the procedures allow CRN to be used to reduce the sample size required to attain a fixed precision. This is in contrast to most applications of variance reduction that increase the precision of an estimator for a fixed sample size.

Each of the procedures employs a two-stage-sampling approach, which is more natural in simulation than in most other sampling experiments. The procedures that incorporate CRN have equal second-stage sample sizes for all systems, which is convenient for programming purposes.

REFERENCES

- 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*, ed. J. R. Wilson, J. O. Henriksen, and S. D. Roberts, 313-315. Institute of Electrical and Electronics Engineers, Washington, D.C.
- Goldman, D. 1983. Ranking and selection in simulation. In *Proceedings of the 1983 Winter Simulation Conference*, ed. S. D. Roberts, J. Banks and B. Schmeiser, 387-393. Institute of Electrical and Electronics Engineers, Washington, D.C.
- 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. *The Annals of Statistics* 12: 1136-1144.
- Koenig, L. W., and A. M. Law. 1985. A procedure for selecting a subset of size m containing the ℓ best of k independent normal populations, with applications to simulation. *Communications in Statistics* B14: 719-734.
- Matejckik, F. J., and B. L. Nelson. 1992. Two-stage multiple comparisons with the best for computer simulation. Working Papers Series No. 1992-004, Department of Industrial and Systems Engineering, The Ohio State University, Columbus, Ohio.
- Nelson, B. L. 1993. Robust multiple comparisons under common random numbers. *ACM Transactions on Modeling and Computer Simulation*, forthcoming.
- Rinott, Y. 1978. On two-stage selection procedures and related probability-inequalities. *Communications in Statistics* A7: 799-811.
- Wilcox, R. R. 1984. A table for Rinott's selection procedure. *Journal of Quality Technology* 16: 97-100.
- Yang, W., and B. L. Nelson. 1991. Using common random numbers and control variates in multiple-comparison procedures. *Operations Research* 39: 583-591.

AUTHOR BIOGRAPHIES

FRANK J. MATEJCIK is an Assistant Professor of Industrial Engineering at the South Dakota School

of Mines and Technology. He received his Ph.D. in Industrial and Systems Engineering from The Ohio State University in 1992. His research interests include simulation experimental design and sequential methods. He is a member of ORSA, ASA, and ASQC.

BARRY L. NELSON is an Associate Professor in the Department of Industrial and Systems Engineering at The Ohio State University. His research interests are experiment design and analysis of stochastic simulations. He is President of the TIMS College on Simulation and an Associate Editor for *Operations Research* and *TOMACS*.