Proceedings of the 1996 Winter Simulation Conference
ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain

SELECTING THE BEST SYSTEM IN TRANSIENT SIMULATIONS WITH VARIANCES KNOWN

Halim Damerdji
Department of Industrial Engineering
North Carolina State University
Raleigh, NC 27695–7906, U.S.A.

Peter W. Glynn
Department of Operations Research
Stanford University
Stanford, CA 94305–4022, U.S.A.

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

James R. Wilson
Department of Industrial Engineering
North Carolina State University
Raleigh, NC 27695–7906, U.S.A.

ABSTRACT

Selection of the best system among k different systems is investigated. This selection is based upon the results of finite-horizon simulations. Since the distribution of the output of a transient simulation is typically unknown, it follows that this problem is that of selection of the best population (best according to some measure) among k different populations, where observations within each population are independent, and identically distributed according to some general (unknown) distribution. In this work in progress, it is assumed that the population variances are known. A natural single-stage sampling procedure is presented. Under Bechhofer's indifference zone approach, this procedure is asymptotically valid.

1 INTRODUCTION

Simulation is often used in order to select the best system among a set of k, say, different systems. For example, in the design of an (s, S)-inventory system one may want to try k different settings for the parameters (s, S), the objective being to find the system with the smallest "mean inventory level at end of day." The decision of selecting the best of these systems might be based upon the results of finite-horizon simulations.

In statistical terminology the setting is the following. There are k distinct populations, and a sample of independent and identically distributed variables $X_{i,1}$, $X_{i,2}$, ... can be collected from each population (i =1,...,k). (When simulating the ith system, observation $X_{i,j}$ is the output of the jth replicate.) Let F_i be the cumulative distribution function (cdf) of $X_{i,1}$, whose exact form is unknown, and

$$\mu(F_i) \equiv \mathsf{E}_{F_i} X_{i,1}$$

its mean. (The expectation is indexed to stress that it is with respect to F_i.) The goal is to find an efficient procedure so that the system with the largest mean is selected with probability at least P^*, which is a prespecified parameter (e.g., $P^* = 0.9$). Under the indifference-zone approach, one will want to select the best population with a high likelihood when the second largest mean is at least δ units away from the largest. The procedure's tolerance parameter δ is prespecified.

The parametric Gaussian procedures (e.g., the Dudewicz-Dalal and Rinott procedures) seem to be used in practice, even when the distributions of the samples depart from normality. The small robustness study of Sullivan and Wilson (1989) shows clearly that in populations with even moderate skewness, there can be substantial discrepancy between nominal and actual probabilities of correct selection that are obtained with classical two-stage Dudewicz-Dalal-type procedures. Selection of the best system in the nonparametric setting has both practical and theoretical importance.

It will be assumed here that population variances are known. In practice, it is unlikely that cdf's are not known, yet variances are. However, the problem is difficult. The proofs we have used relied heavily on Berry-Esseen bounds and local limit theorems. We believe that some of these techniques will also be fruitful for tackling the general unknown-variances case. We also believe that solving the general i.i.d. nonparametric case is a step towards solving the problem of selection of the best system in the (stationary) dependent case (which arises in steady-state simulation).

The rest of the paper is organized as follows. Section 2 contains a brief literature review. Some preliminary notations are given in Section 3. The procedure is presented in Section 4 along with the paper's main result, whose proof is sketched in Section 5. A two-stage sampling procedure for the case of unknown variances is presented in Section 6. Section 7 is the conclusion.

2 LITERATURE REVIEW

Selection of the best system, under the indifference-zone approach, is briefly reviewed. Related topics such as subset selection, elimination, and sequential procedures are not discussed. The problem of selection of the best system has been extensively studied in the i.i.d. Gaussian context. The setting considered by Bechhofer (1954), in his seminal paper, was that of known variances. The case of unknown but common variances was investigated by Bechhofer, Dunnett, and Sobel (1954). Dudewicz and Dalal (1975) proposed two two-stage sampling procedures for the case of most interest in practice, where variances are unknown and possibly unequal. Dudewicz and Dalal (1975) then showed the validity of one of the methods. Rinott (1978) considered a modification of the other procedure and showed its validity.

Selection of the best system in the i.i.d. non-Gaussian context has been investigated for certain probability distributions, such as the exponential, gamma, Weibull, Bernoulli (selection of the population with largest probability of success), and multinomial (selection of the cell with largest probability). Nonparametric models have been considered, but mostly in the context of location parameter distributions. (These are distributions with cdf's $F(x;\theta) \equiv F(x-\theta)$.) There are also results on scale distributions. (These are distributions with cdf's $F(x;\theta) \equiv F(x/\theta)$.) Gupta and McDonald (1980) discuss the case where the populations have stochastically increasing (in the parameter) cumulative distribution functions. For relevant discussions on selection of the best system in the nonparametric setting, see Chapter 8 of Gupta and Panchapakesan (1979), Chapter 7 of Gibbons, Olkin, and Sobel (1977), Section 8.6 of Mukhopadhyay and Solanky (1994), and Bechhofer, Santner, and Goldsman (1995, p. 68).

3 PRELIMINARY NOTATIONS

Let Λ be the set of possible distribution functions the various populations can be coming from. In the inventory example, Λ is the set of cdf's that can model the end-of-day inventory level. Let $\tilde{F} = (F_1, \ldots, F_k) \in \Lambda^k$ and $\mu(\tilde{F}) = (\mu(F_1), \ldots, \mu(F_k))$. Denote $\mu(\tilde{F},[i])$ as the ith largest mean in $\tilde{F}$ (and so $\mu(\tilde{F},[k])$ is the mean of the best population in $\tilde{F}$). The preference region is given by

$$\Theta(\delta) \equiv \{\tilde{F} \in \Lambda^k : \mu(\tilde{F},[k]) \geq \mu(\tilde{F},[k-1]) + \delta\}.$$

The preference region is the set of joint cdf's for which the best and second best means are at least δ units apart from one another. When $\tilde{F} \in \Theta(\delta)$, the procedure should select the best system, with probability at least P^*. The least favorable configuration is the set

$$\bar{\Theta}(\delta) \equiv \{\tilde{F} \in \Lambda^k : \mu(\tilde{F},[1]) = \cdots = \mu(\tilde{F},[k-1]) = \mu(\tilde{F},[k]) - \delta\},$$

which is assumed to be nonempty. The procedure will have most difficulty in selecting the best system when $\tilde{F} \in \bar{\Theta}(\delta)$.

Once the systems are chosen, the joint experiment depends upon $\tilde{F}$, which is simply the product of the k marginal cdf's $F_1, \ldots, F_k$. The probability measure associated with this particular experiment will be denoted by $P_{\tilde{F}}$. It is assumed that the variance $\sigma^2(i,\tilde{F}) = \mathsf{E}_{F_i}[(X_{i,1} - \mu(F_i))^2]$ is known for each population ($i = 1, \ldots, k$).

Another parameter used in the selection procedure is the solution h to the equation

$$\int_{-\infty}^{\infty} \Phi(z+h)^{k-1} \Phi(dz) = P^*,$$

where $\Phi(.)$ is the standard-normal cdf. The solution to the equation exists and is unique from classical arguments. Bechhofer, Santner, and Goldsman (1995, pp. 61–63 and pp. 293–294) contains a table of values, and a computer program used to determine h when this parameter cannot be deduced from the table.

4 THE PROCEDURE

Since variances are assumed to be known, the following single-stage procedure for selecting the best system is proposed.

1. Choose P^* ($1/k < P^* < 1$) and $\delta > 0$. Compute h.

2. For $i = 1, \ldots, k$, compute
$$N(i,\tilde{F},\delta) \equiv \left\lceil \frac{h^2\sigma^2(i,\tilde{F})}{\delta^2} \right\rceil,$$
where $\lceil x \rceil$ is the smallest integer greater or equal to x.

3. For every population i, sample $X_{i,1}, \ldots, X_{i,N(i,\tilde{F},\delta)}$. Let
$$\bar{X}_i(\delta) \equiv \frac{1}{N(i,\tilde{F},\delta)} \sum_{j=1}^{N(i,\tilde{F},\delta)} X_{i,j}$$
be the sample mean over the available observations.

4. Select the population with the largest sample mean.

Note that $N(i, \tilde{F}, \delta)$ is not a random variable but a deterministic integer. The procedure uses no distributional information beyond the second moments of the underlying distributions. Since the forms of the cdf's are unknown, it is unrealistic to expect the procedure to be exact for a finite sample; asymptotics have then to be considered in order to show its (asymptotic) validity.

The asymptotics we consider are as the tolerance parameter δ shrinks to zero, which will result in the sample size for each population tending to infinity. The second effect of letting $\delta \to 0$ will be that the preference region gets closer to the whole set Λ^k, making the correct selection increasingly more difficult (because when $\tilde{F}$ belongs to the least favorable configuration, the best mean will be closer and closer to the remaining $(k-1)$ means); nontrivial results are thereby obtained in the limit. Consider the following two assumptions.

Assumption 1 *For all $F \in \Lambda$, F admits a density f, the distribution mean is finite, its variance is finite and uniformly bounded away from zero, and its absolute centered third moment is uniformly bounded (from above).*

Assumption 2 *There exists a finite constant K such that, for all $F \in \Lambda$, $\sup_{\{-\infty<x<\infty\}} f(x) \le K$.*

Our goal is to show that the above procedure is asymptotically valid. One must show then that the probability of correct selection is at least P^* for any joint distribution $\tilde{F}$ in the preference region. (Correct selection will be denoted by $\mathrm{CS}(\delta)$.) Since the procedure must be efficient (from a sampling standpoint), it is actually desired that

$$\lim_{\delta \to 0} \inf_{\{\tilde{F}=(F_1,\ldots,F_k)\in\Theta(\delta)\}} P_{\tilde{F}}\Big[\mathrm{CS}(\delta)\Big] = P^*. \qquad (1)$$

Consider the following results.

Proposition 3 *Under Assumptions 1 and 2, we have that for every $\epsilon > 0$, there exists $\delta_0 \equiv \delta_0(\epsilon)$ such that when $0 < \delta \le \delta_0$,*

$$P_{\tilde{F}}\Big[\mathrm{CS}(\delta)\Big] \;\ge\; P^* - \epsilon \qquad \textit{for all } \tilde{F} \in \Theta(\delta).$$

Proposition 4 *Under Assumptions 1 and 2, we have that for all $\epsilon > 0$, there exists $\delta_0 \equiv \delta_0(\epsilon)$ such that when $0 < \delta \le \delta_0$,*

$$P_{\tilde{F}}\Big[\mathrm{CS}(\delta)\Big] \;\le\; P^* + \epsilon \qquad \textit{for all } \tilde{F} \in \bar{\Theta}(\delta).$$

The main result of the paper follows as a corollary.

Corollary 5 *Under Assumptions 1 and 2, (1) holds, i.e., the procedure is asymptotically valid (as the tolerance parameter $\delta \to 0$).*

The proofs of the propositions are sketched in the next section. The complete proofs are given in Damerdji et al. (1996).

5 SKETCH OF THE PROOFS

As previously mentioned, Berry-Esseen bounds on the rate of convergence in the central limit theorem for i.i.d. variables as well as local limit theorems are pivotal.

Lemma 6 (Petrov, 1975) *Let $Y_1,\ldots,Y_n$ be independent random variables having a common distribution. Suppose $\mathsf{E}Y_1 = \mu$, $\mathsf{E}[(Y_1-\mu)^2] = \sigma^2$, and $\mathsf{E}[|Y_1-\mu|^3] < \infty$. Then*

$$\sup_x \left| P\Big(\frac{1}{\sigma\sqrt{n}}\sum_{j=1}^{n}(Y_j-\mu) \;<\; x\Big) - \Phi(x)\right| \;\le\; A\frac{\mathsf{E}[|Y_1-\mu|^3]}{\sigma^3}\frac{1}{\sqrt{n}},$$

where A is a universal constant.

Lemma 7 (Petrov, 1975) *Let $Y_1,\ldots,Y_n$ be independent random variables having a common distribution, with density $p(x)$, such that $\mathsf{E}Y_1 = \mu$, $\mathsf{E}[(Y_1-\mu)^2] = \sigma^2$, $\mathsf{E}[|Y_1-\mu|^3] < \infty$, and $\sup_{-\infty<x<\infty} p(x) \le K$ for some constant K. Let $p_n(x)$ be the density of the random variable $(1/(\sigma\sqrt{n}))\sum_{j=1}^{n}(Y_j-\mu)$. Then*

$$\sup_{-\infty<x<\infty}\left|p_n(x) - \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\right| \;\le\; A'\frac{\mathsf{E}[|Y_1-\mu|^3]}{\sigma^3\sqrt{n}}\max(1,K^3),$$

and for all x

$$\left|p_n(x) - \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\right| \;\le\; A'\Big(\frac{\mathsf{E}[|Y_1-\mu|^3}{\sigma^3]}\Big)^{2m-1}\frac{\max(1,K^{2m+1})}{\sqrt{n}(1+|x|^m)}$$

for $m = 2$ and $m = 3$, where A' is a universal constant.

Denote $\bar{X}_{(i)}(\delta)$ as the sample mean of the population with ith largest mean. The probability of correct selection is

$$P_{\tilde{F}}\Big[\mathrm{CS}(\delta)\Big] \;=\; P_{\tilde{F}}\Big[\bar{X}_{(i)}(\delta) \;<\; \bar{X}_{(k)}(\delta) \quad \forall i \ne k\Big].$$

Let

$$a(t;\tilde{F},(i),\delta) \equiv P_{\tilde{F}}\Big[\frac{\bar{X}_{(i)}(\delta)-\mu(\tilde{F},[i])}{\delta/h} < t\Big].$$

We now sketch the proof of Proposition 3. For all $\tilde{F} \in \Theta(\delta)$, we have that

$$P_{\tilde{F}}\Big[\mathrm{CS}(\delta)\Big] \geq P_{\tilde{F}}\Big[\frac{\bar{X}_{(i)}(\delta)-\mu(\tilde{F},[i])}{\delta/h} < \frac{\bar{X}_{(k)}(\delta)-\mu(\tilde{F},[k])}{\delta/h}+h \ \ \forall i \neq k\Big]$$
$$= \int_{-\infty}^{\infty}\Big(\prod_{i\neq k} a(z+h;\tilde{F},(i),\delta)\Big)a(dz;\tilde{F},(k),\delta). \quad (2)$$

Some more notation is needed. Let

$$b(t;\tilde{F},(i),\delta) \equiv P_{\tilde{F}}\Big[\frac{\bar{X}_{(i)}(\delta)-\mu(\tilde{F},[i])}{\sqrt{\sigma^2((i),\tilde{F})/N((i),\tilde{F},\delta)}} < t\Big],$$

$$c(t;\tilde{F},(i),\delta) = a(t;\tilde{F},(i),\delta) - b(t;\tilde{F},(i),\delta),$$

and

$$d(t;\tilde{F},(i),\delta) = b(t;\tilde{F},(i),\delta) - \Phi(t).$$

We naturally have that

$$a(z+h;\tilde{F},(i),\delta) = \Phi(z+h) + c(z+h;\tilde{F},(i),\delta) + d(z+h;\tilde{F},(i),\delta).$$

The expression $\prod_{i\neq k} a(z+h;\tilde{F},(i),\delta)$ in (2) can then be expanded into a sum of terms involving $c(z+h;\tilde{F},(i),\delta)$, $d(z+h;\tilde{F},(i),\delta)$, and $\Phi(z+h)$. Lemma 6 is used to show that for all $\tilde{F} \in \Theta(\delta)$,

$$P_{\tilde{F}}\Big[\mathrm{CS}(\delta)\Big] \geq \int_{-\infty}^{\infty}\Big(\Phi(z+h)\Big)^{k-1} a(dz;\tilde{F},(k),\delta) - H(\tilde{F},(k),\delta),$$

where the exact form of $H(\tilde{F},(k),\delta)$ is given in Damerdji et al. (1996). Under Assumptions 1 and 2, this term can be made arbitrarily small uniformly in $\tilde{F}$.

We now rewrite

$$a(dz;\tilde{F},(k),\delta) = \Phi(dz) + \Big(a(dz;\tilde{F},(k),\delta) - \Phi(dz)\Big),$$

and use Lemma 7 to bound $|a(dz;\tilde{F},(k),\delta) - \Phi(dz)|$.

As for Proposition 4, we start by considering $\tilde{F}$ in the least favorable configuration. For such $\tilde{F}$, we get that

$$P_{\tilde{F}}\Big[\mathrm{CS}(\delta)\Big] \leq \int_{-\infty}^{\infty}\Big(\Phi(z+h)\Big)^{k-1} a(dz;\tilde{F},(k),\delta) + H(\tilde{F},(k),\delta).$$

Using the bound on $|H(\tilde{F},(k),\delta)|$ developed earlier, we get that $P_{\tilde{F}}\Big[\mathrm{CS}(\delta)\Big] \leq P^* + \epsilon$, where ϵ depends only upon δ. The detailed proof is given in Damerdji, Glynn, Nakayama, and Wilson (1996).

6 THE TWO-STAGE SAMPLING PROCEDURE

The following is a two-stage procedure for selection of the best system for i.i.d. observations with unknown distributions (and variances). The (asymptotic) validity of this procedure is yet unavailable.

1. Choose P^* $(1/k < P^* < 1)$ and $\delta > 0$. Compute h.
2. Let $m(\delta) \equiv \lceil \delta^{-2} \rceil$ be the size of the first subsample for each population.
3. For every population $i = 1, \ldots, k$, let $X_{i,1}$, $X_{i,2}$, $\ldots$, $X_{i,m(\delta)}$ be the corresponding sample. Let
$$\bar{X}_{i,1}(\delta) \equiv \frac{1}{m(\delta)} \sum_{j=1}^{m(\delta)} X_{i,j}$$
and
$$S_i^2(\delta) \equiv \frac{1}{m(\delta)-1} \sum_{j=1}^{m(\delta)} \Big(X_{i,j} - \bar{X}_{i,1}(\delta)\Big)^2.$$
4. For every population i, compute the total sample size:
$$N(i,\tilde{F},\delta) = \max\Big\{m(\delta), \Big\lceil \frac{h^2 S_i^2(\delta)}{\delta^2} \Big\rceil\Big\}.$$
5. If $N(i,\tilde{F},\delta) \geq m(\delta)+1$, sample $X_{i,m(\delta)+1}, \ldots, X_{i,N(i,\tilde{F},\delta)}$.
6. Let
$$\bar{X}_i(\delta) \equiv \frac{1}{N(i,\tilde{F},\delta)} \sum_{j=1}^{N(i,\tilde{F},\delta)} X_{i,j}.$$
7. Select the population with the largest sample mean.

Note that in the unknown-variances case, the total sample sizes are not deterministic quantities but, rather, random variables. This is not a major issue in the Gaussian case; in that setting the mean and sample variances are independent (the total sample is function of the sample variance), which allows passage to the t-distribution. See Remark 4.1 of Dudewicz and Dalal (1975). Another great difficulty in showing the validity of this procedure is that we want equality in (1) so the procedure is computationally efficient. This work is still in progress.

7 CONCLUSION

Suppose a number of systems are to be investigated via simulation, where the nature of the problem is such that the simulations are of the finite-horizon type. Simulation of each system will then involve running independent replicates, each replicate providing a single "observation," whose distribution is unknown.

Selection of the best system is undertaken here. In this work in progress, it is assumed that the population variances are known. A single-stage procedure was presented, and its asymptotic validity discussed. The detailed proofs are given in Damerdji, Glynn, Nakayama, and Wilson (1996).

The assumption of known variances is unrealistic in practice. For the unknown-variances case, a two-stage procedure was presented. Asymptotic validity of this procedure, i.e., Equation (1), is still an open problem for the two-stage case. Some of the tools described here might be useful. Another problem under investigation is that of selection of the best system in steady-state simulation, i.e., when the problems are such that the simulations are of the steady-state type.

REFERENCES

Bechhofer, R. E. 1954. A single-sample multiple decision procedure for ranking means of normal populations with known variances. *The Annals of Mathematical Statistics,* **25**, 16–39.

Bechhofer, R. E., Dunnett, C. W., and M. Sobel. 1954. A two-sample multiple-decision procedure for ranking means of normal populations with a common unknown variance. *Biometrika,* **41**, 170–176.

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

Damerdji, H., P. W. Glynn, M. K. Nakayama, and J. R. Wilson. 1996. Selection of the best system in a finite-horizon simulation context. Working paper.

Dudewicz, E. J., and S. R. Dalal. 1975. Allocation of observations in ranking and selection with unequal variances. *Sankhyā,* **B37**, 28–78.

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

Gupta, S. S., and G. C. McDonald. 1980. Nonparametric procedures in multiple decisions (ranking and selection procedures). *Colloquia Mathematica Societatis János Bolyai,* **32**, 361–389.

Gupta, S. S., and S. Panchapakesan. 1979. *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations.* Wiley, New York.

Mukhopadhyay, N., and T. K. S. Solanky. 1994. *Multistage Selection and Ranking Procedures.* Marcel Dekker, New York.

Petrov, V. V. 1975. *Sums of Independent Random Variables.* Springer-Verlag, New York.

Rinott, Y. 1978. On two-stage selection procedures and related probability-inequalities. *Communication in Statistics: Theory and Methods,* **A7(8)**, 799–811.

Sullivan, D. W., and J. R. Wilson. 1989. Restricted subset selection procedures for simulation. *Operations Research,* **37**, 52–71.

AUTHOR BIOGRAPHIES

HALIM DAMERDJI is an assistant professor in the Department of Industrial Engineering at North Carolina State University. He received a Ph.D. degree in industrial engineering from the University of Wisconsin-Madison. He has held positions at the Ecole Nationale Polytechnique of Algiers and Purdue University. His research interests are in simulation and applied stochastic processes. He is a member of INFORMS.

PETER W. GLYNN received his Ph.D. from Stanford University, after which he joined the faculty of the Department of Industrial Engineering at the University of Wisconsin-Madison. In 1987, he returned to Stanford, where he currently holds the Thomas Ford Faculty Scholar Chair in the Department of Operations Research. He was a co-winner of the 1993 Outstanding Simulation Publication Award sponsored by the TIMS College on Simulation. His research interests include discrete-event simulation, computational probability, queueing, and general theory for stochastic systems.

MARVIN K. NAKAYAMA is an assistant professor in the Department of Computer and Information Science at the New Jersey Institute of Technology. Previously, he has held positions in the Management Science/Computer Information Systems Department of the Graduate School of Management at Rutgers University, and at the IBM Thomas J. Watson Research Center in Yorktown Heights, New York. He received Ph.D. in Operations Research from Stanford University. His research interests include applied probability, simulation output analysis, gradient estimation, rare-event simulation, and random-variate generation.

JAMES R. WILSON is Professor and Director of Graduate Programs in the Department of Industrial Engineering at North Carolina State University. He received a B.A. degree in mathematics from Rice University, and he received M.S. and Ph.D. degrees in industrial engineering from Purdue University. His current research interests are focused on the design and analysis of simulation experiments. He also has an active interest in applications of operations research techniques to all areas of industrial engineering. From 1988 to 1992, he served as Departmental Editor of *Management Science* for Simulation. He was *Proceedings* Editor for WSC '86, Associate Program Chair for WSC '91, and Program Chair for WSC '92. He has also held various offices in TIMS (now INFORMS) College on Simulation. He is a member of ASA, ACM/SIGSIM, IIE, and INFORMS.