Proceedings of the 1980 Winter Simulation Conference
T.I. Ören, C.M. Shub, P.F. Roth (eds.)

SELECTION PROCEDURES BASED ON PROPORTION PARAMETERS

Paul D. Amer
Department of Computer and Information Sciences
University of Delaware
Newark, Delaware 19711

The data analysis component of a simulation study comparing a set of alternatives should include a confidence statement regarding the accuracy of the selection results. The usefulness of statistical ranking and selection procedures in providing such confidence statements in simulation experiments has become increasingly clear. Five general selection goals are described. Procedures are given when the alternatives are compared on the basis of proportion parameters. Statistical derivations and tables underlying these procedures are not included, but instead are appropriately referenced.

1. INTRODUCTION

The comparison of alternatives is the purpose of many simulation studies. These alternatives may be resource scheduling algorithms, transportation systems, computer hardware configurations, communication strategies, etc. Although the complexity of the simulations vary, a central objective often is either to determine which alternative is best (according to some predefined criteria) or to determine those alternatives which perform at least as well as a control or standard alternative. Because of the variability inherent in simulation studies due to probabilistic activities (those activities dependent on random number generation), unless the designs for the studies are well-conceived and statistically sound, the results will lack significance. Ideally, the data analysis component of a simulation study should provide an experimenter with a statistical confidence regarding the accuracy of the results.

Based on the program of recent Winter/Summer Simulation Conferences (e.g., see Highland, Sargent & Schmidt (1977)), the usefulness of statistical ranking and selection procedures in providing accurate simulation results has become increasingly clear. Given that a comparison of alternatives is to be based on some set of simulated performance measurements, ranking and selection procedures can explicitly specify the number of observations and/or the appropriate selection rule necessary to guarantee a predetermined probability that the alternative(s) selected actually satisfy the experimental objective.

In this paper, five experimental designs (procedures) are described for selecting amongst alternatives on the basis of a proportion parameter. For increased readability, statistical derivations and tables underlying these procedures are not included, but instead are appropriately referenced.

2. BACKGROUND

When a comparison is based on the quantitative estimates of the merits of a set of alternatives (perhaps derived from a simulation study), the selection activity is termed a statistical selection procedure. The alternatives can be considered populations, $\pi_1,\pi_2,\ldots,\pi_k$ with associated probability distribution functions $f_1,f_2,\ldots,f_k$. Given a set of parameters $\theta_1,\theta_2,\ldots,\theta_k$ where θ_i summarizes f_i (i.e., θ_i is the mean, percentile, etc.), statistical ranking and selection procedures provide information regarding the true ordering of the θ_i, denoted $\theta_{[1]}\le\theta_{[2]}\le\ldots\le\theta_{[k]}$, under the assumption that the pairing of π_i (and θ_i) with the $\theta_{[j]}$ $(1\le i,j\le k)$ is completely unknown.

A variety of selection goals as well as distributional assumptions, including the underlying form of f_i (e.g. normal, gamma, etc.) and the dependence (or lack of) between collected observations, can be made. Different procedures are required for different assumptions. Two summaries of the numerous ranking and selection procedures already developed and their respective assumptions are Gibbons, Olkin and Sobel (1977) and Kleijnen (1975). The reader is referred to Dudewicz and Koo (1978) for a categorized bibliography of ranking and selection procedures.

2.1 Experimental Goals

The definition of a correct decision (or correct selection) and the calculation of its probabilistic chances of occurring are dependent on the objective of the simulation study. In the present paper, five goals are considered.

Goal 1: Select π_i associated with $\theta_{[k]}$ (or for some applications $\theta_{[1]}$).

Goal 2: Select a subset of minimal size that includes the π_i associated with $\theta_{[k]}$.

Goal 3: Select a subset of minimal size that includes all π_i with $\theta_i \geq \theta_s$ (where θ_s is a control/standard constant).

Goal 4: Select a subset of minimal size that both includes all π_i with $\theta_i \geq \theta_{s_u}$ and excludes all π_i with $\theta_i \leq \theta_{s_\ell}$ (where θ_{s_u}, θ_{s_ℓ} are upper, lower control/standard constants).

In the case where each f_i is characterized by two parameters θ_{i_1} and θ_{i_2}, the following goal is considered.

Goal 5: Select a subset of minimal size that includes all π_i with $\theta_{i_1} \geq \theta_{s_1}$ and $\theta_{i_2} \geq \theta_{s_2}$ (where θ_{s_1}, θ_{s_2} are dual control/standard constants).

Ranking and selection techniques have been developed for different combinations of Goals 1-5 and specific parameter (e.g. mean, alpha quantile). Procedures are presented in Section 3.0 when the parameter(s) of interest is a proportion parameter. Section 2.2 first discusses such a parameter.

2.2 Proportion Parameter

It is often the case that an experimenter has a performance index and natural "threshold" value in mind (denoted C_{thld}), and wishes to make a comparison/selection based on each alternative's proportion of values less than (or greater than) this threshold. For example,response time at a computer terminal can be segregated into those times which promote human-machine interaction (< 15 seconds, (Carbonell, Elkind and Nickerson (1968)) and those which do not. An experimenter selecting from a set of computer services might then only be interested in the service with the largest proportion of response times < 15 seconds, or those services for which at least 80% of the responses were less than 15 seconds. Similarly, a purchaser of light bulbs might only be interested in the brands for which 80% of the bulbs remain operational at least 1000 hours.

3. PROCEDURES

Substituting a proportion parameter, p_i, for the general parameter θ_i in Goals 1-5, the following procedures have been derived. No statistical justification for the procedures is presented. The interested reader is referred to the original sources for the appropriate statistical theory. All procedures are derived such that the probability of a correct decision (or correct selection), defined by the experimental goal, is at least a user specified probability P*, (0<P*<1).

These procedures were applied in an experiment to evaluate four remote access interactive computer services. Details of the experiment may be found in Amer (1980), and Mamrak and Amer (1979). Following a definition of each procedure, a figure indicating an example application is provided to further clarify the experimental goal.

3.1 Selection of the Best

Sobel and Huyett (1957) provide the following procedure for determining the "best" alternative (Goal 1). The exact mathematical solution employs an "indifference zone" approach originally described in the pioneering ranking and selection treatise by Bechhofer (1954). An indifference zone is a minimum distance, d*>0, between the best proportion, $p_{[k]}$, and the remaining others, which an analyst feels it is significant to be able to detect. Any loss incurred by incorrectly selecting a population whose p_i is less than $p_{[k]}$, but not more than d* less, is considered insignificant. For practical purposes, all such populations are equivalent to the best and satisfy the selection goal. A description of this goal in terms of a computer system selection can be found in Amer and Mamrak (1978).

Procedure Best: Choose appropriate C_{thld}, P^* and d^* values according to nonstatistical considerations. Collect n independent observations from each π_i where n is determined by table look-up based on the parameters k, P^* and d^*. (Tables are available in Sobel and Huyett (1957).) Compute X_i=no. of observations from π_i which are < C_{thld}. Select as best alternative the π_i associated with $\max_i\{X_i\}$. (See Figure 1)

Example: Select the computer service which provides the largest proportion of response times less than 5 seconds.

C_{thld} = 5 seconds

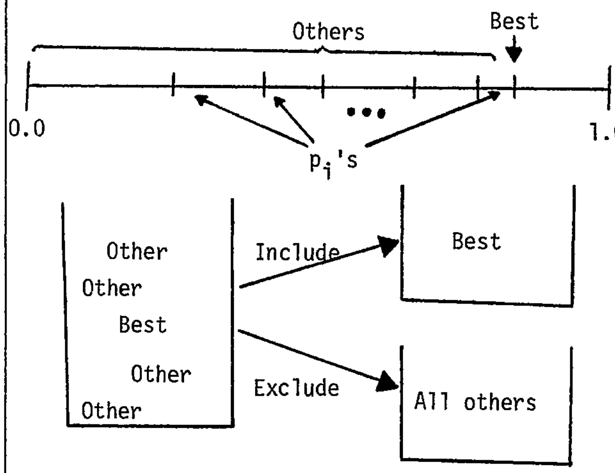

Figure 1 Selection of the Best

3.2 Subset Selection of the Best

Goal 2 describes a "subset" approach which often is employed as a first stage preliminary screening of inferior populations. The objective is to select a subset which contains the π_i associated with $p_{[k]}$. Procedure Best may then be applied to the reduced set as a second stage thereby defining a two step process for selection of the best (see Amer (1979), Chapter 5). Gupta and Sobel (1960) provide the following procedure for determining a subset which contains the best (Goal 2).

Procedure Best-Subset: Choose appropriate C_{thld} and P^* values according to nonstatistical considerations. Collect n independent observations from each π_i where n is determined by practical constraints. Compute X_i=no. of observations from π_i which are < C_{thld}. Place π_i in the selected subset if $X_i \geq \max_i\{X_i\}-d$, where d is determined by table look-up based on the parameters k, P^* and n. (Tables are available in Gupta and Sobel (1960).) (See Figure 2)

Example: Select a subset which includes the computer service that provides the largest proportion of response times less than 5 seconds.

C_{thld} = 5 seconds

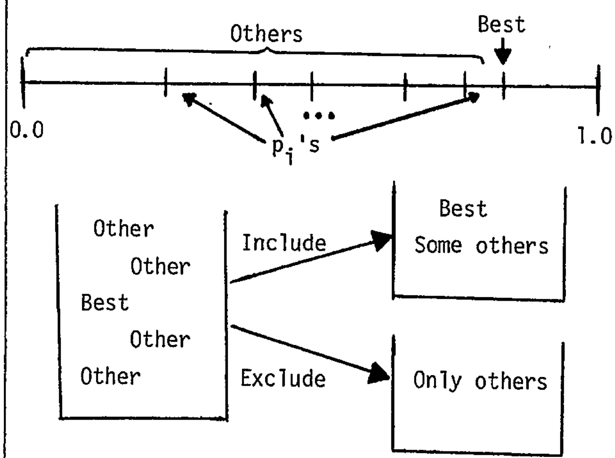

Figure 2 Subset Selection of the Best

3.3 Selection Better Than a Control/Standard

Goals 3-5 describe subset approaches for determining those alternatives which are better than a specified "control" or "standard" alternative. These goals are applicable in instances where an experimenter wishes to separate "good" alternatives from "bad" alternatives. For example, a farmer might wish to determine which hens are profitable egg layers (> x eggs per week at least 70% of the time) and which hens are not.

When applying control/standard procedures, an experimenter would like to perfectly separate all better than standard alternatives from all worse than standard alternatives. Due to the inherent variation in an empirical study, perfect separation is not possible. Goals 3 and 4 assume it is better to minimize the probability of rejecting alternatives that are in actuality better than standard at the expense of occasionally selecting a worse than standard alternative.

For a subset selection approach, there is the implicit assumption of minimizing the size of the selected subset subject to a P^* constraint on the probability of error. A correct subset selection can be attained with probability of error equal to zero by including all alternatives in the selected subset, yet this contradicts the implicit desire of eliminating inferior (or substandard) alternatives.

3.3.1 Selection Based on a Single Control/Standard

Gupta and Sobel (1958) provide the following procedure for selecting a subset of $\pi_1,\ldots,\pi_k$ such that all π_i with $p_i \geq p_s$ (p_s is a control/standard proportion) are included therein with probability $\geq P^*$ (Goal 3).

<u>Procedure Single Standard</u>: Choose C_{thld}, P^* and p_s values according to nonstatistical considerations. Collect n independent observations from each π_i where n is determined by practical constraints. Compute X_i=no. of observations from π_i which are $< C_{thld}$. Place π_i in the selected subset if $X_i \geq M$, where M is determined by table look-up based on the parameters k, n, P^* and p_s. (Limited tables are available in Gupta, Huyett and Sobel (1957); more extensive tables appear in Mamrak and Amer (1979).) (See Figure 3)

Example: Select all services which have at least 80% turnaround times less than 10 minutes.

C_{thld} = 10 minutes $\qquad p_s$ = .80

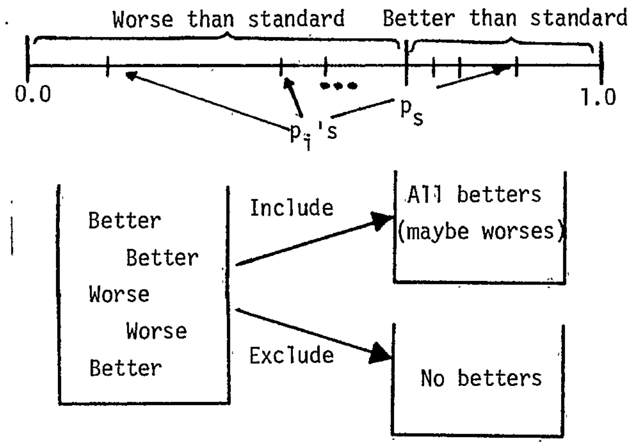

Figure 3 Selection Based on a Single Control/Standard

Goals 4 and 5 were developed in reaction to certain computer selection problems (see Amer (1979)). They are extensions of Goal 3 and provide experimenters with additional flexibility and control in applying control standards.

3.3.2 Selection Based on an Upper-Lower Control/Standard

Goal 4 is applicable when an experimenter wishes to control how far below standard a population may perform and still be selected due to variability in sampling. Values from π_i lie either $<$ or $\geq C_{thld}$ with probabilities p_i, $1-p_i$, respectively. An "upper-lower" proportion standard (Goal 4) is satisfied by a selected subset S if all π_i with $p_i \geq p_{s_u}$ are included in S (p_{s_u} is an upper control/standard proportion) and all π_i with $p_i \leq p_{s_\ell}$ are not included in S (p_{s_ℓ} a lower control/standard proportion).

For example, in a selection effort to choose a computerized reservation service, reasons of economics might restrict attention to those services for which at least 70% of the reservation interactions were processed in less than 3 seconds. Instead of an experiment designed to select a subset which contained all reservation services whose proportion of responses less than 3 seconds was at least .70 (Goal 3), an analyst could design the same experiment also to guarantee (with probability P^*) that the subset contained no services whose proportion was less than .50.

Amer and Dudewicz (1980) provide the following procedure for determining a subset selection satisfying Goal 4 with probability $\geq P^*$.

<u>Procedure Upper-Lower Standard</u>: Choose appropriate C_{thld}, p_{s_u}, p_{s_ℓ} and P^* values based on nonstatistical considerations. Collect n independent observations from each π_i where n is constrained by practical considerations, but must be large enough for a selection rule to exist (determined by table look-up). Compute X_i=no. of observations from π_i which are $< C_{thld}$. Place π_i in the selected subset if $X_i \geq M$, where M is determined by table

Example: Select all services which have at least 70% response times less than 3 seconds also guaranteeing to exclude all services with at most 50% response times less than 3 seconds.

C_{thld} = 3 seconds $\qquad p_{s_u}$ = .70 $\qquad p_{s_\ell}$ = .50

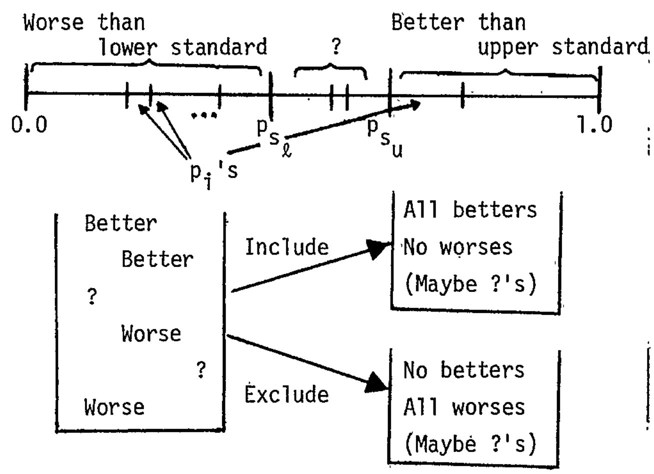

Figure 4 Selection Based on an Upper-Lower Control/Standard

look-up based on the parameters n, P*, k, p_{s_ℓ}, p_{s_u}. (Tables are available in Amer (1980).) (See Figure 4)

3.3.3 Selection Based on a Dual Control/Standard
Goal 5 is applicable when a set of populations each have two related parameters of interest and the experimenter wishes to determine which π_i's have both parameters better than a pair of respective control/standard parameters. In the case of proportions, values from π_i lie in one of three intervals, $I_1=[-\infty, C_{thld1})$, $I_3=[C_{thld1}, C_{thld2})$ or $I_2=[C_{thld2}, \infty]$ with probabilities p_{i_1}, $1-p_{i_1}-p_{i_2}$, p_{i_2}, respectively. A "dual" proportion standard (Goas 5) is satisfied by a selected subset S if all π_i with $p_{i_1} \geq p_{s_1}$ and $1-p_{i_2} \geq p_{s_2}$ are included in S (p_{s_1}, p_{s_2} are dual control/standard proportions).

For example, in a life testing simulation of light bulbs, an experimenter could define a brand of bulbs to be better than standard if the proportion of its bulbs lasting 1250 hours was ≥ .50 and the proportion of its bulbs lasting 750 hours was ≥ .80.

Amer and Dudewicz (1980) provide the following procedure for determining a subset selection satisfying Goal 5 with probability ≥ P*.

Procedure Dual Standard: Choose appropriate C_{thld1}, C_{thld2}, p_{s_1}, p_{s_2}, and P* values according to nonstatistical considerations. Collect n independent observations from π_i where n is as large as practical constraints permit. Compute X_{i_1}, X_{i_2}=no. of observations from π_i which are < C_{thld1}, < C_{thld2}, respectively. Place π_i in the selected subset if $X_{i_1} \geq M1$ and $X_{i_2} \geq M2$ where M1, M2 are determined by table look-up based on the parameters k, n, P*, p_{s_1} and p_{s_2}. (Tables are available in Amer (1980).) (See Figure 5)

Example: Select all services which have at least 60% response times less than 30 seconds and at least 90% response times less than 60 seconds.

C_{thld1} = 30 seconds $\quad p_{s_1}$ = .60

C_{thld2} = 60 seconds $\quad p_{s_2}$ = .90

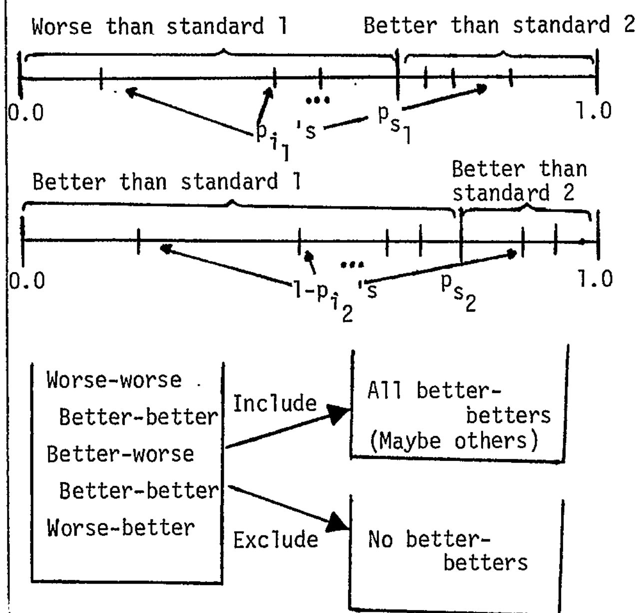

Figure 5 Selection Based on a Dual Control/Standard

4. CONCLUSIONS

It is important that the data analysis component of a simulation study comparing alternatives provide some sort of confidence statement regarding the accuracy of the results. Five general selection goals are described and procedures are presented when the parameter of interest is a proportion parameter. The decision as to which selection goal (and procedure) is most appropriate for a simulation study must be based on the study's objectives. This paper has summarized five procedures to provide the reader with an understanding of some experimental designs that are available.

In terms of a comparison of remote access interactive computer services, the kind of statements that can be made upon accurate application of the procedures presented herein are (it is suggested that the reader think of "population" and "proportion(s)" in terms of an environment with which he/she is familiar): "With P* confidence (e.g. 90%), it can be stated that:

Goal 1 - computer service #3 provides the largest proportion of response times less than x seconds,

Goal 2 - the set of services {#1,#4,#7} contains the computer service which provides the largest proportion of response times less than x seconds,

Goal 3 - the set of services {#2,#3,#9,#10} contains all computer services whose proportion of response times less than x seconds is at least .80,

Goal 4 - the set of services {#1,#5,#6} contains all computer services whose proportion of response times less than x seconds is at least .80, and excludes all services whose proportion is at most .60,

Goal 5 - the set of services {#1,#2,#6,#8} contains all computer services whose proportions of response times less than x and 2x seconds is at least .60 and .90, respectively."

REFERENCES

Amer, P. D. & Mamrak, S. A. (1978), "Statistical Methods in Computer Performance Evaluation: A Binomial Approach to the Comparison Problem", Proc. Computer Science and Statistics: Eleventh Annual Symposium on the Interface, March, 314-319.

Amer, P. D. (1979), Experimental Design for Computer Comparison and Selection, Ph.D. dissertation, The Ohio State Univ., Columbus, March.

Amer, P. D. (1980), Application of Measurement Criteria in the Selection of Interactive Computer Services, NBS Special Publication 500-58, National Bureau of Standards, Wash., D.C., April.

Amer, P. D. & Dudewicz, E. J. (1980), "Two New Goals for Selection Based on Proportions", Comm. in Statistics, A9, 14, 1461-1472.

Bechhofer, R. E. (1954), "A Single-Sample Multiple-Decision Procedure for Ranking Means of Normal Populations with Known Variances", Annals of Mathematical Statistics 25, 1, 16-39.

Carbonell, J. R., Elkind, J., & Nickerson, R. (1968), "On the Psychological Importance of Time in a Time Sharing System", Human Factors 10, 2, 135-142.

Dudewicz, E. J. & Koo, J. O. (1978), "A Categorized Bibliography on Ranking and Selection Procedures", Tech. Report 163, Dept. of Statistics, Ohio State Univ., June.

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

Gupta, S. S., Huyett, M. J. & Sobel, M. (1957), "Selection and Ranking Problems with Binomial Populations", Transactions American Society for Quality Control, Baltimore, 635-643.

Gupta, S. S. & Sobel, M. (1958), "On Selecting a Subset Containing the Best of Several Binomial Populations", Annals of Mathematical Statistics 29, 235-244.

______ (1960), "Selecting a Subset Containing the Best of Several Binomial Populations", Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, (ed. Olkin et.al.), 224-248.

Highland, H. J., Sargent, R. G. & Schmidt, J. W. (eds.) (1977), Proc. 1977 Winter Simulation Conference, National Bureau of Standards, Gaithersburg, MD, December.

Kleijnen, J. P. C. (1975), Statistical Techniques in Simulation: Part II, Marcel Dekker, New York.

Mamrak, S. A. & Amer, P. D. (1979), A Methodology for the Selection of Interactive Computer Services, NBS Special Publication 500-44, National Bureau of Standards, Wash., D.C., January.

Sobel, M. & Huyett, M. J. (1957), "Selecting the Best One of Several Binomial Populations", Bell Systems Tech. Journal 36, March, 537-576.