NEW PROCEDURES FOR SELECTION AMONG (SIMULATED) ALTERNATIVES

Edward J. Dudewicz
Department of Statistics
The Ohio State University

ABSTRACT

"Ranking and selection" procedures were developed for use in situations where the experimenter's goal is to "select the best" (selection) or to "rank competing alternatives" (ranking). These are typical goals when a simulation study is performed, often in order to select that one of several procedures (for running a real-world system) which is best (with regard to a specified criterion of goodness). Therefore these procedures have become a standard analysis method in simulation work in recent years (e.g., see the references in (8)).

In this paper we: review some of the selection procedures most often found useful in design and analysis of simulation experiments; discuss some other important procedures and related problems; and state some unsolved statistical problems of importance in simulation work.

I. USED SELECTION PROCEDURES

One of the most common situations in simulation is that where one has: $k \geq 2$ populations (sources of observations) $\pi_1,\ldots,\pi_k$ with respective unknown means $\mu_1,\ldots,\mu_k$ for their observations, where observations from π_i follow a normal probability distribution with variance σ_i^2 about their respective means; a goal of selecting the population associated with $\mu_{[k]} = \max(\mu_1,\ldots,\mu_k)$; a probability requirement that the probability of correct selection Prob{CS} be $\geq P^*$ $(1/k < P^* < 1)$ if $\mu_{[k]} - \mu_{[k-1]} \geq \delta^*$, where the experimenter specifies P^* and δ^*, and where $\mu_{[k-1]}$ denotes the second-best mean; and a procedure of selecting the population yielding the largest sample mean $\bar{X}_{MAX} = \max(\bar{X}_1, \bar{X}_2,\ldots,\bar{X}_k)$.

This problem was first solved, by telling how many observations N to take from each population (in simulations, how to set the run length) by Bechhofer (2), who assumed that $\sigma_1^2 = \ldots = \sigma_k^2 = \sigma^2$ with σ^2 known and that all observations were independent. The case of $\sigma_1^2,\ldots,\sigma_k^2$ unknown and unequal was solved more recently by Dudewicz and Dalal (9), and requires a two-stage procedure (which is no problem in simulation). (For more details of these two procedures, see (7) and (8).) For the more common case where the observations are also correlated within each population (but not across populations), a heuristic procedure was recently given (11). The recommended Dudewicz-Zaino Procedure $A(\hat{\rho}_i, s_i^2)$ is as follows. Take an initial sample of $N_0 = 30$ observations from each population. Calculate

$$M_i = \max\left(N_0, \left\lceil \frac{s_i^2 h^2}{(\delta^*)^2} \right\rceil\right) \tag{1}$$

(which is the number of observations which would be needed if we had all correlations $\rho_i = 0$ (zero correlation), where h depends on k and P^* and is given in Table 1 below, extracted from (10)). Calculate

$$\hat{\rho}_i = \frac{\sum_{n=2}^{N_0}(X_{in}-\bar{X}_i)(X_{i,n-1}-\bar{X}_i)}{\sum_{n=1}^{N_0}(X_{in}-\bar{X}_i)^2} \tag{2}$$

and form the $100(1-\alpha)\%$ confidence interval for ρ_i from

$$(\rho_i-\hat{\rho}_i)^2 \leq \frac{N_0-1}{N_0(N_0-3)}\,(1-\hat{\rho}_i^2)t^2_{N_0-3}(1-\alpha/2) \tag{3}$$

with $\alpha = .05$. (Here $t_r(q)$ is the 100q percent point of Student's-t distribution with r degrees of freedom.) If this 95% confidence interval contains $\rho_i = 0$, judge the sample size M_i as being adequate for population i. Otherwise calculate

$$N_{2i} = \left\lceil M_i \frac{1+\hat{\rho}_i}{1-\hat{\rho}_i} \right\rceil \tag{4}$$

and continue the run until we have N_{2i} observations from π_i. Finally calculate $\bar{X}_1,\ldots,\bar{X}_k$ based on all available observations and select (as being best) that population which produced the largest of $\bar{X}_1,\ldots,\bar{X}_k$.

While Procedure $A(\hat{\rho}_i, s_i^2)$ is heuristic (unlike the procedure of Dudewicz and Dalal for $\rho_i = 0$, which is entirely rigorously derived), studies show it should be sufficient to preclude gross errors due to significant correlations. (Results of Bishop (4) will soon yield theorems for this heuristic procedure.)

TABLE 1

Quantity h Needed in Equation (1)

	$P^* = .95$	$P^* = .99$
k = 2	2.41	3.45
k = 3	2.81	3.81
k = 4	3.03	4.01
k = 5	3.18	4.14
k = 6	3.30	4.25
k = 7	3.39	4.33
k = 8	3.46	4.40
k = 9	3.53	4.46
k = 10	3.58	4.51
k = 15	3.79	4.71
k = 20	3.92	4.84
k = 25	4.03	4.94

Another common situation, for which a procedure is available, is that where one desires to select the population with the smallest median $\xi_{.5}$ or some other quantile, e.g. the smallest p^{th} quantile ξ_p. (For example, in computer performance evaluation such is a reasonable goal for the 90^{th} percentile of the processing time distribution, as discussed by Mamrak and DeRuyter (16).) Here a procedure developed by Sobel (18) is used, with new tables of Dhariyal and Dudewicz (6). This procedure has recently been exposited and applied in an excellent and accessible paper (16), and a book (12), and so will not be discussed in further detail here.

II. NEW PROCEDURES AND PROBLEMS

While Section I noted the most widely used selection procedures for simulation applications, in fact there are a number of others (and related considerations) which should be borne in mind, since for design goals other than selecting the population with the "largest mean" or "smallest quantile" these will be the appropriate procedures and considerations. These may be classified into categories of factorial experiments, estimation, subset selection, and nonparametric selection.

When one's simulation is being run to select the optimum levels of two or more factors, one is dealing with a factorial experiment. (For example, one could be considering the best combination of the two factors "number of vehicles" and "routing algorithm" for a transportation system where loads are generated at random.) In the setting where there is no interaction (i.e., where the effects due to each factor simply add to produce the effect of the combination) the problem was discussed by Bechhofer (2), (3), and by Bawa (1). When the factors do interact, Lun (15) showed that one might want to select based on cell means of factor combinations (rather than based on marginal means for each factor), and that to do otherwise could lead to arbitrarily low Prob{CS}.

To be specific, suppose that we have a two-factor experiment with k_1 levels of factor 1 (e.g., k_1 possibilities for "number of vehicles" in the example of the above paragraph) and k_2 levels of factor 2 (k_2 would then be the number of routing algorithms under consideration). We assume our observations are normally distributed with known common variance σ^2 and that if Y_{ij} is an observation taken at level i of factor 1 and level j of factor 2 then

$$E(Y_{ij}) = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

$(i = 1,\ldots,k_1;\ j = 1,\ldots,k_2)$. The "best" population is that one of the $K = k_1k_2$ combinations of factor levels which maximizes the system's mean yield $E(Y_{ij})$. If $\gamma_{ij} = 0$ for all i and j, we have no "interaction" (in which case the same algorithm yields the best results regardless of the number of vehicles available to the system).

One can formulate many procedures for dealing with the above problem, especially if all $\gamma_{ij} = 0$. In the literature we find SP1 and SP2 (see below), while SP3 is new with (15) and this paper.

Procedure SP1 takes N independent observations from each of the K populations, and selects the level associated with the largest marginal sample mean of each of the two factors; the combination of these two levels is asserted to be the best factor combination.

Procedure SP2 takes N_1 independent observations at each level of factor 1, with factor 2 held fixed at one of its levels. The level of factor 1 yielding the largest sample mean is selected. That selected level is then used in experimentation taking N_2 observations at each level of factor 2, after which the level of factor 2 yielding the largest sample mean is selected. The combination of the levels of factor 1 and factor 2 so selected is asserted to be best.

Procedure SP3 takes M independent observations from each of the $K = k_1k_2$ populations and selects the factor combination (population) yielding the largest sample mean as best.

Procedures SP1, SP2, SP3 are respectively the "Factorial", "One-at-a-time", and "Interaction" methods. Bawa (1) compared SP1 and SP2 when there is no interaction. Lun (15) showed that if there is interaction, then the inf (over $\mu_{[K]}-\mu_{[K-1]}\geq\delta^*$) of the P(CS) for SP1 is $\leq 1/k_1$ and $\leq 1/k_2$, hence $\leq 1/\max(k_1,k_2)$. We believe it can also be shown that this inf of the P(CS) is, for SP1, $\leq 1/(k_1k_2)$, which shows SP1 to be unreasonable for situations where interaction may be present in unknown magnitude (since one can achieve P(CS) of $1/(k_1 k_2)$ by a totally random selection). If one estimates the interactions after the experiment, a reasonable SP4 should be able to be developed which acts as SP1 does if the interactions are "small", and otherwise acts as does SP3. The above considerations generalize to any fixed number of factors.

Recently this problem has been considered further by Bechhofer (3).

In the related estimation problem, one asks "How good is the best alternative?". (E.g., if one establishes it is not very good compared to the presently used alternative, continuing the simulation until the "best" is found may not make sense. While if one establishes it is very superior, then a good deal of simulation may be economically justified.) Some of the procedures useful here are considered by Chen (5), who also gives new procedures for two-factor experiments with no interaction.

In subset selection one's interest is in selecting a subset of the populations $\pi_1, \ldots, \pi_k$ in which one can be fairly sure that the best population lies. This is a particularly appropriate goal when one cannot afford the computer time needed to select the best, or when (with a modest outlay) one wishes to eliminate the worst alternatives from contention. This has been discussed well by Gupta and Hsu (13).

Some nonparametric aspects, particularly applicable in cases where multiple criteria are used in evaluation or where some observations may be missing, were given by Lee (14). Other nonparametric procedures for a "complete data" setting where the interest is in selecting a subset are given by McDonald (17). Finally, Sobel (19) discusses cases where the dispersion is important and one does not know the underlying distributions (e.g., this is important when several alternatives have the same median yield, and one wishes to choose one to minimize variability about that yield).

III. UNSOLVED PROBLEMS

While the results discussed in Sections I and II and the references are directly and appropriately applicable today, they do not cover all important situations. For the purpose of stimulating statistical research in these directions, we wish to briefly note some important situations omitted.

If one has multivariate observations, how does one select (or even define) the best population (without some artificial definition of a linear or other simple univariate function of the multivariate responses)? Will work of Bishop (4) and the Heteroscedastic Method yield a solution here? How can multiple-criteria problems be handled in the computer (without resorting to rankings by "managers" as described by Lee (14))?

How can costs and gains/losses be taken into account in decision-theoretic selection (as can be appropriate when these can be quantified)?

In how far can the Central Limit Theorem justify use of existing procedures for non-normal cases?

IV. CONCLUSIONS

Selection procedures are available and used for a number of important simulation problems (Section I). Besides these major procedures, other important work is available for use now; using the procedure which best fits one's goals makes for the most efficient (and least costly) experimentation (Section II). New results and procedures for factorial settings have been given (Section II). Nevertheless, important problems in the area are still available for theoretical statisticians (Section III).

BIBLIOGRAPHY

1. Bawa, V. S. "Asymptotic Efficiency of One R-Factor Experiment Relative to R One-Factor Experiments for Selecting the Best Normal Population," in Journal of the American Statistical Association, Vol 67, No. 339 (September 1972), pp. 660-661.

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

3. Bechhofer, R. E. "Selection in Factorial Experiments," in Proceedings of the 1977 Winter Simulation Conference, 1977.

4. Bishop, Thomas A. Heteroscedastic ANOVA, MANOVA, and Multiple-Comparisons. Unpublished Ph. D. Dissertation, Department of Statistics, The Ohio State University, Columbus, Ohio, 1976.

5. Chen, H. J. "Estimation of the Best Alternative," in Proceedings of the 1977 Winter Simulation Conference, 1977.

6. Dhariyal, I. D. and Dudewicz, Edward J. "Sobel's Nonparametric Selection Procedure for Quantiles: Extensions and Tables," Technical Report, Department of Statistics, The Ohio State University, Columbus, Ohio, 1977, to appear.

7. Dudewicz, Edward J. Introduction to Statistics and Probability. Holt, Rinehart and Winston, New York, 1976.

8. Dudewicz, Edward J. "Statistics in Simulation: How to Design for Selecting the Best Alternative," in Proceedings of the 1976 Bicentennial Winter Simulation Conference, edited by H. J. Highland, T. J. Schriber, and R. G. Sargent, 1976, pp. 67-71.

9. Dudewicz, Edward J. and Dalal, Siddhartha R. "Allocation of Observations in Ranking and Selection with Unequal Variances," in Sankhyā, Series B, Vol. 37, Part 1 (1975), pp. 28-78.

10. Dudewicz, E. J., Ramberg, J. S., and Chen, H. J. "New Tables for Multiple Comparisons With a Control (Unknown Variances)," in Biometrische Zeitschrift, Vol. 17, Part 1 (1975), pp. 13-26.

11. Dudewicz, Edward J. and Zaino, Nicholas A., Jr. "Allowance for Correlation in Setting Simulation Run-length via Ranking-and-Selection Procedures," Technical Report, Department of Statistics, Stanford University, Stanford, California, 1976. To appear in Management Science, 1977.

12. Gibbons, Jean Dickinson, Olkin, Ingram, and Sobel, Milton Selecting and Ordering Populations: A New Statistical Methodology. John Wiley & Sons, Inc., New York, 1977.

13. Gupta, Shanti S. and Hsu, Jason C. "Subset Selection Procedures with Speical Reference to the Analysis of Two-Way Layout: Application to Motor-Vehicle Fatality Data," in Proceedings of the 1977 Winter Simulation Conference, 1977.

14. Lee, Young Jack "Winner Selection," in Proceedings of the 1977 Winter Simulation Conference, 1977.

15. Lun, Francis W. "Selection in Factorial Experiments," Project Report, Statistics 828 (Ranking, Selection, and Multiple-Decision), offered Spring Quarter 1977, Department of Statistics, The Ohio State University, Columbus, Ohio.

16. Mamrak, Sandra A. and DeRuyter, Patricia A. "Statistical Methods for Comparison of Computer Services," in Computer, Vol. 10, No. 11 (November 1977), to appear.

17. McDonald, Gary C. "An Application of Nonparametric Selection Procedures to an Analysis of Motor-Vehicle Traffic Fatality Rates," in Proceedings of the 1977 Winter Simulation Conference, 1977.

18. Sobel, M. "Nonparametric Techniques for Selecting the t Populations with the Largest α-Quantiles," in Annals of Mathematical Statistics, Vol. 38, 1967, pp. 1804-1816.

19. Sobel, Milton "Selecting the Population with the Smallest Dispersion in a Nonparametric Setting," in Proceedings of the 1977 Winter Simulation Conference, 1977.