WINNER SELECTION

Young Jack Lee

University of Maryland, College Park, and
National Cancer Institute, Bethesda, Maryland

ABSTRACT

When k alternatives are simulated (e.g. in a study to find which one optimizes the performance of some real or planned system) it is not rare for multiple criteria to be employed (i.e., several system responses are of interest and no simple combination of them, called a measure of effectiveness, is available). In such settings, n persons (who are responsible for the system, e.g. administrators, investigators, etc.) may be asked to rank the performances of the k from best to worst. An optimal method for performing this ranking will be discussed. A similar problem occurs when each of n referees ranks k contestants to select the winner of a prize (the best contestant), and it is common that not all the referees complete the full ranking. Similar situations arise when selecting the most efficient simulation algorithm. Our methods for selecting the winner also apply to data of this type.

I. INTRODUCTION

In the k-population-n-block model, McDonald (7) selects the best (worst) population among k given populations based on rank scores within each of n blocks. We consider a similar model with one difference. In McDonald's model no observation is missing and thus full ranking within each block is available. The full ranking, however, sometimes is not practical or can be expensive.

When k alternatives are simulated (e.g. in a study to find which one optimizes the performance of some real or planned system) it is not rare for multiple criteria to be employed (i.e., several system responses are of interest and no simple combination of them, called a measure of effectiveness, is available). In such settings, n persons (who are responsible for the system, e.g. administrators, investigators, etc.) may be asked to rank the performances of the k from best to worst. In this case, it is quite likely that not all the n persons complete the ranking particularly when k is not small.

Suppose k given computing algorithms with similar capabilities for a certain optimization problem. The k algorithms are tried on n test problems to select an algorithm which is most efficient (in computer time). Some test problems are relatively simple so that time to completion may be measured for all the k packages. But other test problems may be complicated and require longer computer time. In the latter case not all the k algorithms may be observed to completion, and thus full ranking of the k algorithms may not be available for all the n test problems. For illustration suppose we have 5 algorithms and test them sequentially for a test problem, yielding the following result:

package 1	40 minutes	
package 2	40+ minutes	70 minutes
package 3	40+ minutes	60 minutes
package 4	40+ minutes	65 minutes
package 5	35 minutes,	

where algorithms 2, 3, and 4 would have taken 70, 60, and 65 minutes, had they not been censored at 40 minutes. Then since algorithms 2, 3, and 4 were censored at 40 minutes, the observed ranking of the 5 algorithms is

$$(4,\_,\_,\_,5)$$

where the higher the rank score, the better the algorithm is rated. Note that in this illustration had the algorithms been tested in the order of 2, 4, 3, 1, and 5, the observed ranking would have been

$$(4,1,3,2,5).$$

On the other hand, if algorithm 5 had been tested first, then all others would have been censored at 35 minutes, yielding the ranking

$$(\_,\_,\_,\_,5).$$

Note that the censoring time for each algorithm is random, since the order of testing the 5 algorithm is random and the time to completion for an algorithm may also be random.

Another example in which the full ranking may not be available is when each of n referees ranks k finalists to select the best contestant. For example, the Chemical Division of the American Society for Quality Control awards the Frank Wilcoxon Prize and the Jack Youden Prize each year. These prizes are awarded for outstanding articles in Technometrics, the Wilcoxon being for the best practical application paper and the Youden for the best expository paper. Table 1 gives the rank scores of seven finalists for the 1976 Youden award. Referees were not mandated to complete the ranking, score 6 is given to a contestant judged best, score 5 to one judged second best, etc. Twelve referees out of twenty-four completed the ranking.

TABLE 1

Rank Scores of 7 Finalists for the 1976 Youden Prize*

Referee	C_1	C_2	C_3	C_4	C_5	C_6	C_7
1	4	2	1	5	6	0	3
2	2	0	5	1	4	3	6
3	1	5	2	3	6	0	4
4	2	5	6	0	3	1	4
5	1	2	5	3	4	0	6
6	6	2	3	4	5	0	1
7	5	4	1	3	2	0	6
8	2	1	0	6	5	3	4
9	4	5	2	1	3	6	0
10	6	2	3	1	4	0	5
11	4	5	3	2	6	1	0
12	2	3	4	0	1	5	6
13	2	6	-	-	4	3	5
14	2	-	3	6	4	-	5
15	5	4	-	6	2	-	3
16	4	5	-	-	-	3	6
17	6	4	-	5	3	-	-
18	4	5	3	-	-	-	6
19	4	5	-	-	3	-	6
20	3	6	5	-	4	-	-
21	-	6	5	-	-	4	-
22	5	-	-	-	4	-	6
23	-	5	-	-	6	-	-
24	-	-	-	-	-	6	-

*The referee numbers are arranged for convenience of preparing the table.

In making the final recommendation for the 1976 Youden Prize by the Awards Committee (this author served as a member of the Committee), rating a first place vote as a 2, a second place vote as a 1, and all others as 0, contestant C_7 was selected (Table 2).

A question arises on the way contestant C_7 was recommended. One may suggest to select a subset of 7 contestants based on Table 3 and select one contestant through a runoff if the subset size is greater than 1: Subset selection procedure R_1 by McDonald (7) selects C_1, C_2, C_5, and C_7, and the probability that one of them is best is approximately 0.95. Assume that referees vote consistently. Table 4 is derived from Table 1 using the relative rank among C_1, C_2, C_5, and C_7, 2 points for the highest score, 1 point for the second highest score, and 0 for others; sums of scores of Table 4 indicate that C_7 be recommended, a result consistent with that of Table 2.

TABLE 2

Sum of Modified Scores of the First Twenty-three Referees*

C_1	C_2	C_3	C_4	C_5	C_6	C_7
9	14	6	8	10	3	19

* 2 points for a first place vote, 1 point for a second place vote, and 0 point for all others.

TABLE 3

Sum of all Scores in Table 1 with Missing Scores Replaced by Average Scores

C_1	C_2	C_3	C_4	C_5	C_6	C_7
80	86.5	61	58	85	43.5	90

In preparing Table 2, one referee-report was not considered. If two referee-reports (referees 23 and 24) are not taken into consideration, then Table 5 (3 points for a first place vote, 2 points for second, 1 point for third, and 0 for others) is obtained and a contestant may be selected based on it. Or shouldn't one select according to Table 6 or Table 7? What is the theoretical background for selecting the contestant based on Table 2? The answer to the first question is "No" and we will justify this answer in Section 3.

TABLE 4

Runoff Scores of C_1, C_2, C_5, and C_7

Obtained from Table 1

Referee	C_1	C_2	C_5	C_7
1	1	0	2	0
2	0	0	1	2
3	0	1	2	0
4	0	2	0	1
5	0	0	1	2
6	2	0	1	0
7	1	0	0	2
8	0	0	2	1
9	1	2	0	0
10	2	0	0	1
11	0	1	2	0
12	0	1	0	2
13	0	2	0	1
14	0	0	1	2
15	2	1	0	0
16	0	1	0	2
17	2	1	0	0
18	0	1	0	2
19	0	1	0	2
20	0	2	1	0
21	1/3	2	1/3	1/3
22	1	0	0	2
23	0	1	2	0

TABLE 5

Sum of Modified Scores of the First Twenty-two Referees*

C_1	C_2	C_3	C_4	C_5	C_6	C_7
21	26	12	14	20	6	33

* 3 points for a first place vote, 2 points for a second place vote, ..., and 0 point for a fourth or fifth or sixth or seventh place vote.

TABLE 6

Sum of Scores by the First Twelve Referees

C_1	C_2	C_3	C_4	C_5	C_6	C_7
39	36	35	29	49	19	45

TABLE 7

Sum of Modified Scores of the First Fifteen Referees*

C_1	C_2	C_3	C_4	C_5	C_6	C_7
33	33	26	29	44	15	45

* 5 points for a first place vote, 4 points for a second place vote, ..., and 0 point for a sixth or seventh place vote.

TABLE 8

Frequency of the First Place

C_1	C_2	C_3	C_4	C_5	C_6	C_7
3	3	1	3	4	2	8

In Section 2, we present a definition of the "best contestant", selection procedures conforming to the definition, and their properties. In Section 3, the selection procedures are compared via relative efficiency, showing that the procedure based on Table 8 is most efficient for the Jack Youden Prize problem.

II. DEFINITION, PROCEDURES, AND PROPERTIES

Let C_i denote the i^{th} contestant.

Definition: Let $\phi_{i\ell}$ denote the probability that a referee ranks C_i as the $(k-\ell+1)^{st}$ best among $(C_1,\ldots,C_k)$. For a fixed t define $\mu_i(t) = \sum_{\ell=k-t+1}^{k}(\ell+t-k)\phi_{i\ell}$. Then the contestant associated with $\max(\mu_1(t),\ldots,\mu_k(t))$ is called the best contestant by ranking t out of k and is denoted by C - best(t).

This definition of the best contestant is a generalization of definitions given in Lee (5) where only C - best(1) and C - best(k) are considered. It is possible that different values of t may define the best contestant differently. For example if we select the best contestant based on the first 12 rows of Table 1, C_5 will be selected according to C - best(k) while C - best(1) is C_7. But if the same referees select between C_5 and C_7 and if they vote consistently, C_7 will be selected as the best contestant. We will assume below that whichever t may be given to define the best contestant, the best contestant is the same. It is our belief that when referees judge contestants with consistency with regard to their relative ranks the assumption would be well received. Otherwise the definition of the best contestant should conform to the situations surrounding the problem. From now on we often take the liberty of calling a contestant the best without specific reference.

Conforming to the definition of the best contestant for given t, the selection procedure R(t) is: Ask each referee to assign the scores t,t-1,...,2,1 and (k-t) 0's to the k contestants starting with the contestant he judges is strongest and ending with the weakest contestant. Denote by $R_{ji}(t)$ the rank score assigned to the C_i by the j^{th} referee. Let $V_i(t) = \sum_{j=1}^{n} R_{ji}(t)$, and select the contestant yielding $\max(V_1(t),\ldots,V_k(t))$ as the best, breaking ties for the maximum by randomization. (Tables 2, 3, 5, 6, 7, and 8 are bases for procedures R(2), R(7), R(3), R(7), R(5), and R(1) respectively.)

Procedure R(1) is a procedure selecting the most probable multinomial event (Bechhofer et al (1) and Lee (4)) and procedure R(k) has been studied by Dudewicz and Fan (2) and Lee and Dudewicz (6) among others.

In the problem of selecting the best contestant (or the like), we assume we do not have any information on relative ranks of the k contestants. For convenience of notation and without any loss of generality, however, we assume that the k^{th} contestant is the best. Let CS (Correct Selection) denote the event of

$$V_k(t) = \max(V_1(t),\ldots,V_k(t))$$

breaking ties for the maximum by randomization.

Assume that

$$\phi_{kk} \ge \phi_{k\ell} \text{ and } \phi_{kk} \ge \phi_{\ell' k}, \quad 1 \le \ell,\ell' \le k-1. \qquad (1)$$

Namely we assume the probability that the best contestant is rated as the best is greater than the probability that he is rated as the $(k-\ell+1)^{st}$ best $(1 \le \ell \le k-1)$ and is greater than the probability that another contestant is rated as the best.

Let n, k, t, and λ^*, $1 < \lambda^* < \infty$, be fixed and suppose (1) satisfies

$$\phi_{kk} \ge \lambda^* \phi_{k\ell} \text{ and } \phi_{kk} \ge \lambda^* \phi_{\ell' k}, \quad 1 \le \ell,\ell' \le k-1. \qquad (2)$$

Then for $t \ge 2$

$$\inf_{\phi_{ij}} P[CS|\phi_{ij},R(t)] \le P[CS|R(t),\ \phi_{kk} = \lambda^*\phi_{k\ell} = \lambda^*\phi_{\ell' k},\ \phi_{\ell,\ell'} = \phi_{\ell'\ell},\ 1 \le \ell,\ell' \le k-1] \qquad (3)$$

and for $t = 1$

$$\inf_{\phi_{ij}} P[CS|\phi_{ij},R(1)] = P[CS|R(1),\ \phi_{kk} = \lambda^*\phi_{\ell k},\ 1 \le \ell \le k-1]. \qquad (4)$$

In fact $P[CS|\phi_{ij},R(1)]$ is a function of only ϕ_{kk}, $\phi_{(k-1)k}$, ..., ϕ_{2k}, and ϕ_{1k}, and a proof of (4) is given by Kesten and Morse (3).

For large n an approximation to the right hand side of (3) and (4) is given by

$$\int_{-\infty}^{\infty} \Phi^{k-1}\left(\frac{z}{a(\lambda^*)} + \frac{n^{\frac{1}{2}}t(\lambda^*-1)(2k-t-1)b(\lambda^*)}{2(k+\lambda^*-1)(k-1)\tau(\lambda^*)a(\lambda^*)}\right)d\Phi(z), \qquad (5)$$

where

$$\Phi(x) = \int_{-\infty}^{x} (2\pi)^{-\frac{1}{2}}\exp(-x^2/2)dx,$$

$$a(\lambda^*) = \left\{\frac{\tau^2(\lambda^*) - \gamma(\lambda^*)}{\gamma(\lambda^*)}\right\}^{\frac{1}{2}},$$

$$b(\lambda^*) = \{1 + a^2(\lambda^*)\}^{\frac{1}{2}},$$

$$\tau^2(\lambda^*) = Var\{n^{-\frac{1}{2}}(V_k(t) - V_1(t))\},$$

and

$$\gamma(\lambda^*) = Cov\{n^{-\frac{1}{2}}(V_k(t) - V_1(t)),\ n^{-\frac{1}{2}}(V_k(t) - V_2(t))\}.$$

Equation (5) can be computed numerically using Gaussian quadrature. Approximation by (5) is fairly reliable, and was off by less than 0.01 for the cases studied.

III. RELATIVE EFFICIENCY

We can compare selection procedures R(t) by computing (5) for each t with fixed n, k, and λ^*. For example, see Table 9.

TABLE 9

P[CS] Comparisons

n	k	λ^*	$P[CS\|R(k)]$	$P[CS\|R(1)]$
30	5	2.0	0.731	0.787
30	6	2.0	0.559	0.728
30	7	2.0	0.470	0.673

Instead of computing $P[CS|R(t)]$ to compare $R(t)$'s for given n, k, and λ^*, however, we equate (5) to a given P^* $(1/k < P^* < 1)$ and solve the smallest n needed to satisfy the equation. Denote that n by $n_{k,t}(\lambda^*,P^*)$. The ratio $n_{k,t}(\lambda^*,P^*)/n_{k,t'}(\lambda^*,P^*)$, $1 \le t \ne t' \le k$, is called the relative efficiency of $R(t')$ with respect to $R(t)$, denoted by $Eff[R(t'),R(t)]$. If $Eff[R(t'),R(t)] \ge 1$, then procedure $R(t')$ is at least as efficient as $R(t)$. Of particular interest is $Eff[R(1),R(t)]$. Since $Eff[R(1),R(t)]$ requires a computation for each combination of (k,λ^*,P^*), we instead compute and obtain

$$\lim_{\lambda^*\to 1} Eff[R(1),R(t)] = \frac{t+1}{3}\left[\frac{(k-1)(4tk+2k-3t^2-3t)}{t(2k-t-1)^2}\right]. \quad (6)$$

If $t = k$, $\lim_{\lambda^*\to 1} Eff[R(1),R(t)] = (k+1)/3$. In our Jack Youden Prize example with $k = 7$, $\lim_{\lambda^*\to 1} Eff[R(1),R(t)]$ as a function of t is:

TABLE 10

Relative Efficiency

t	2	3	4	5	6	7
$\lim_{\lambda^*\to 1} Eff[R(1),R(t)]$	1.29	1.65	1.91	2.4	2.67	2.67

which shows that R(1) is the most efficient procedure.

The relative efficiency result demonstrates that when the best contestant is selected based on the data like that in Table 1, R(1) is the selection procedure to use. By the way, we recommended contestant C_7, using the R(2) procedure, for the 1976 Jack Youden Prize; however, note that this result is consistent with the recommendation that would have been made according to R(1). (What a relief!)

ACKNOWLEDGMENT

The author wishes to acknowledge comments and suggestions by Professor Edward J. Dudewicz. The author also appreciates his permission to use the 1976 Jack Youden Prize data.

BIBLIOGRAPHY

1. Bechhofer, R. E., Elmaghraby, S. and Morse, N. (1959), "A Single Sample Multiple Decision Procedure for Selecting the Multinomial Event which Has the Highest Probability," Annals of Mathematical Statistics, 30, pp. 102-19.

2. Dudewicz, E. J. and Fan, C-ℓ. (1973), "Further Light on Non-parametric Selection Efficiency," Naval Research Logistics Quarterly, 20, pp. 737-44.

3. Kesten, H. and Morse, N. (1959), "A Property of the Multinomial Distribution," Annals of Mathematical Statistics, 30, pp. 120-7.

4. Lee, Y. J. (1976), "On Selecting the t Most Probable Multinomial Events: Single-Sample Procedures," tentatively accepted by the Annals of Statistics.

5. Lee, Y. J. (1977), "On Selecting the Best Contestant," Technical Report TR-77-24, Department of Mathematics, University of Maryland, College Park, Maryland.

6. Lee, Y. J. and Dudewicz, E. J. (1974), "Nonparametric Ranking and Selection Procedures," Technical Report #105, Department of Statistics, The Ohio State University, Columbus, Ohio.

7. McDonald, G. C. (1977), "An Application of Nonparametric Selection Procedures to an Analysis of Motor-Vehicle Traffic Fatality Rates," in these Proceedings.