BEST-CHOICE PROBLEMS WITH DEPENDENT CRITERIA

Thomas S. Ferguson

UCLA

Abstract. The multiple criteria secretary problem in which the criteria are independent and the objective is to choose an applicant that is best in at least one criterion was solved by Gnedin (1981). We generalize this result to the case of two criteria that may be dependent. Two models, one assuming independence of the vectors of ranks and the other assuming independence of the variables on which the rankings are based, although equivalent in the case of independent criteria, are seen to be quite distinct for dependent criteria.

§**1. Introduction.** We consider a variation of the secretary problem in which the objects observed are judged and ranked on several criteria, or traits. There are n objects that are put in random order and shown one at a time to a decision maker who must select one of the objects. As each object is shown, it must be either selected or rejected, and once rejected it cannot be later selected. It is assumed that the only relevant data available to the decision maker when he must decide on the j^{th} object are the relative ranks of the first j objects on each of the traits, and it is assumed that the rankings of the different traits are independent. The decision maker wins if the object selected is best in at least one of the traits used to judge the objects.

Multiple criteria optimal selection problems were introduced in a more general form, with observations in a partially ordered set and with an arbitrary payoff utility by Berezovskii, Geninson and Rubchinskii (1980) and Stadje (1980). The above form of the problem was solved by Gnedin (1981). Such a problem may be considered as a generalization of the standard one-criterion best-choice problem, as found, for example, in Gilbert and Mosteller (1966). Problems in which the decision maker wins if the object selected is optimal with respect to a social choice function, for example Pareto optimal, were treated by Berezovskii and Gnedin (1981), Gnedin (1983), Baryshnikov, Berezovskiy and

AMS 1980 subject classification: 60G40, 62L15.

Key words and phrases: Secretary problem, Optimal stopping.

Gnedin (1984), and Berezovskiy, Baryshnikov and Gnedin (1986). In Samuels and Chotlos (1987) the goal of the decision maker is to minimize the expectation of the sum of the ranks of the object selected, rank one being best. Thus, Samuels and Chotlos generalize the one-criteria expected rank problem solved by Chow, Moriguti, Robbins and Samuels (1964).

In this paper, we generalize the paper of Gnedin (1981) by allowing the criteria to be dependent, but we restrict attention to the bivariate setting only. We assume that the observations are two-dimensional vectors $(X_1, Y_1), \ldots, (X_n, Y_n)$ where X_j (resp. Y_j) is the indicator function of the event that the j^{th} object is best among the first j objects when ranked according to the first (resp. second) trait. The distributional assumptions are:

(1.1) (a) The vectors $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent,

(b) $P(X_j = 1) = P(Y_j = 1) = 1/j$ for all j.

With this assumption, the joint distribution of the observations is completely specified by knowledge of the probabilities that object j is relatively best in both traits, denoted by θ_j:

$$\theta_j = P\left(X_j = 1 \text{ and } Y_j = 1\right), \text{ for } j = 1, \ldots, n. \tag{1.2}$$

Note that

$$\theta_1 = 1 \text{ and } 0 \le \theta_j \le 1/j, \text{ for all } j\,. \tag{1.3}$$

Three special cases should be noted.

(1.4) (a) $\theta_j = 1/j$ for all j.

(b) $\theta_j = 1/j^2$ for all j.

(c) $\theta_j = 0$ for all $j > 1$.

In the first case, there is perfect positive dependence between the traits, and the problem becomes essentially identical to the standard univariate best-choice problem. In the second case, the problem solved by Gnedin (1981), the traits are independent. Some aspects of this problem are treated in section 4. In the third case, there is perfect negative dependence between the traits, and the problem becomes equivalent to a univariate problem in which we win if we choose the best or the worst object. This problem has a straightforward solution presented in section 2.

The general case is treated in section 3. The optimal rule is seen to have the simple form for some integers $1 \le r \le s \le n$:

(1.5) $\quad R(r,s):$ for $j = 1$ to $r-1$, reject object j;

for $j = r$ to $s-1$, select object j if $X_j = Y_j = 1$;

for $j = s$ to n, select object j if $X_j = 1$ or $Y_j = 1$.

The probability of win using $R(r,s)$ is found and a simple computational scheme for finding the optimal values of r and s is given in Theorem 3.1. Asymptotic properties as $n \to \infty$ of the optimal r and s and of the optimal probability of win are given in Theorems 3.2 and 3.3, under the condition $\sum_{j=1}^{\infty} \theta_j < \infty$. The main result in this case is that the optimal s/n tends to 1/2 and the optimal probability of win tends to 1/2 also. These surprising results cover both the independent case, (1.4(b)), and the perfect negative dependence case, (1.4(c)). Moreover, the asymptotic optimal probability of win is unaffected by the choice of r provided $r \to \infty$. In Theorem 3.4, it is seen that the asymptotic value of the optimal r/n as $n \to \infty$ is $1 - 1/\sqrt{2} = 0.293\cdots$.

In the independent case, treated in section 4, it is seen that for all finite n, the optimal s is equal to the least integer greater than $n/2$. This is the same optimal value of s that is found for the perfect negative dependence case, treated in section 2.

It is natural to wonder what happens if the rankings of the objects are derived from a bivariate normal distribution, for example, as the correlation coefficient goes from -1 to $+1$. This leads to the comparison of the ranking model, given by (1.1), and the sampling model discussed in section 5. In the classical secretary problem as well as in the three cases (1.4) of the ranking model, these approaches are equivalent. In general though, there is a big difference.

Finally, in section 6, we look at the problem in which the decision maker wins if and only if he selects the object, if any, that is best in all of the traits.

§**2. The Best-or-Worst-Choice Problem.** In this section, we find the optimal rule for the 2-criteria problem under assumption (1.4(c)). This case arises, for example, when one criterion is the negation of the other. It may be considered as a problem with just one criterion and the decision maker wins if the object selected is best or worst according to this criterion among all n objects.

We say that object j is a candidate if it is best in one of the criteria among the first j objects. The standard simple argument of Gilbert and Mosteller (1966) may be used to show that there is an optimal rule, call it $R(s)$, of the form : Reject the first $s-1$ objects and then select the next candidate. (The argument is given in Lemma 3.1 in a more general setting and so is not repeated here.)

If object j is a candidate and is selected, then the probability it remains best in the criterion in which it is relatively best at time j is $y_j = j/n$, as in the classical problem. The probability of win using rule $R(s)$ is, for $s \geq 2$,

$$
\begin{aligned}
Q_n(s) &= \sum_{j=s}^{n} P(\text{select } j|s) y_j \\
&= \sum_{j=s}^{n} (1-\frac{2}{s})(1-\frac{2}{s+1})\cdots(1-\frac{2}{j-1})\frac{2}{j}\cdot\frac{j}{n} \qquad (2.1)\\
&= \frac{2(s-2)(s-1)}{n}\sum_{j=s}^{n}\frac{1}{(j-1)(j-2)} = \frac{2(s-1)(n-s+1)}{n(n-1)}\,.
\end{aligned}
$$

To find the integer s at which Q_n achieves its maximum, we look at the differences,

$$n(n-1)(Q_n(s+1)-Q_n(s)) = 2(n-2s+1)\,.$$

This is decreasing in s, so $Q_n(s)$ is unimodal and the optimal value of s is the first s such that $Q_n(s+1)-Q_n(s) \leq 0$, namely, the smallest integer bigger than or equal to $(n+1)/2$.

Theorem 2.1. *$Q_n(s)$ is maximized at $s = s(n) = \lceil (n+1)/2 \rceil$, and the optimal probability of win $Q_n(s(n)) \to 1/2$ as $n \to \infty$.*

§**3. The Bivariate Secretary.** In this section, we assume that the observations are the indicator random variables $(X_1, Y_1), \ldots, (X_n, Y_n)$ satisfying (1.1), (1.2) and (1.3).

Let $\alpha(j,n)$ denote the probability that an object, relatively best in both traits at stage j, will remain relatively best in both traits after the n^{th} object has been observed, $1 \le j \le n$,

$$(3.1) \qquad \alpha(j,n) = \prod_{i=j+1}^{n} (1 - \frac{2}{i} + \theta_i) \qquad (\alpha(n,n) = 1)\,.$$

Suppose we select object j. If object j is a single maximum, that is, if it is relatively best in exactly one of the two traits, then the probability we win is the probability that it is best overall in that trait, namely j/n, the same as for the 1-dimensional secretary problem. If it is a double maximum, that is, if it is relatively best in both traits, then by the addition rule, the probability we win is $2j/n - \alpha(j,n)$. Letting π_j denote the probability of win if we select the j^{th} object, we have

$$(3.2) \qquad \begin{array}{ll} \pi_j = j/n & \text{if object } j \text{ is a single maximum} \\ \pi_j = 2j/n - \alpha(j,n) & \text{if object } j \text{ is a double maximum} \\ \pi_j = 0 & \text{otherwise.} \end{array}$$

Lemma 3.1 *There is an optimal rule of the form* $R(r,s)$ *of* (1.5) *for some integers* r *and* s*, with* $1 \le r \le s \le n$*.*

Proof. Let V_j denote the probability of win under an optimal strategy among those rules that do not stop before stage j. Since this is a finite horizon problem, the argument of backward induction may be used to show that it is optimal to stop with object j if

$$\pi_j \ge V_{j+1}\,.$$

The V_j are nonincreasing in j, since any strategy available at stage $j+1$ is also available at stage j. Since $2j/n - \alpha(j,n)$ and (j/n) are both increasing in j, the optimal values of r and s are

$$r = \min\{j \ge 1 : 2j/n - \alpha(j,n) \ge V_j{+}1\}$$

$$s = \min\{j \ge 1 : j/n \ge V_{j+1}\}\,,$$

and since $j/n \le 2j/n - \alpha(j,n)$, we have $r \le s$. ∎

Let $\beta(r,j)$, for $r \le j \le s$, denote the probability of reaching stage j using $R(r,s)$,

$$(3.3) \qquad \beta(r,j) = \prod_{i=r}^{j-1} (1 - \theta_i) \qquad (\beta(r,r) = 1)\,.$$

Finally, let $Q_n(r,s)$ denote the probability of win using $R(r,s)$. This probability is computed in the following lemma.

Lemma 3.2. *For* $1 \le r \le s \le n$,

$$Q_n(r,s) = \sum_{j=r}^{s-1} \beta(r,j)\theta_j(\frac{2j}{n} - \alpha(j,n)) + \beta(r,s)\sum_{j=s}^{n} \alpha(s-1,j-1)(\frac{2}{n} - \theta_j\alpha(j,n)) \,. \tag{3.4}$$

Proof. For $r=1$, $Q_n(1,s) = 2/n - \alpha(1,n)$. This satisfies (3.4) since $\beta(1,1) = 1$ by convention, and $\beta(1,j) = 0$ for $j > 1$. Assume $r > 1$. The first sum on the right side of (3.4) is the sum, from stages r to $s-1$, of the probability of reaching stage j, $\beta(r,j)$, of stopping there, θ_j, and of subsequently winning, π_j with a double maximum. The second sum is the sum, from stages s to n, of reaching stage s, $\beta(r,s)$, then reaching stage j, $\alpha(s-1,j-1)$, and then stopping with a double maximum which wins, $\theta_j(2j/n - \alpha(j,n))$, or stopping with a single maximum which wins, $2(1/j - \theta_j)(j/n)$. Noting that

$$\theta_j\bigl(2j/n - \alpha(j,n)\bigr) + 2(1/j - \theta_j)(j/n) = 2/n - \theta_j\alpha(j,n) \,.$$

completes the proof. ∎

Formula (3.4) may be used to find the optimal values of (r,s). Since this is a 2-dimensional discrete maximization, one would expect it to be difficult to verify that a local maximum is a global one. However, two properties hold that simplify the computation greatly. First, $Q_n(r,s)$ is unimodal in s with r held fixed, and unimodal in r with s held fixed. The mode may be assumed at more than one point. Second, the maximum of $Q_n(r,s)$ over s occurs at a point that is independent of r, in the sense that there is a number $s(n)$ such that the maximum of $Q_n(r,s)$ for fixed $r > 1$ occurs at $s(r,n) = \max(r, s(n))$. This will allow us to find the global maximum in two 1-dimensional searches; put $r = 2$ and find $s(n)$, and then find the value of r that maximizes $Q_n(r, s(n))$.

Lemma 3.3. *For each* $r, Q_n(r,s)$ *is unimodal in* s, $r \le s \le n$. *In fact, for* $r > 1$, $Q_n(r,s+1) \le Q_n(r,s)$ *if, and only if,* $\theta_s = 1/s$, *or* $s \ge s(n)$, *where*

$$s(n) = \min\{s \ge 1 : \frac{s}{n} \ge Q_n(s+1,s+1)\} \,. \tag{3.5}$$

Proof. Computing the difference $\Delta_s Q_n(r,s) = Q_n(r,s+1) - Q_n(r,s)$, we find

$$\Delta_s Q_n(r,s) = \frac{2}{s}\beta(r,s)(1-s\theta_s)[Q_n(s+1,s+1) - \frac{s}{n}]\,.$$

Therefore, for $r > 1$, $Q_n(r,s+1) \le Q_n(r,s)$ if and only if $s\theta_s = 1$ or the term in square brackets is nonpositive. The $s(n)$ of (3.5) is the smallest s for which this term is nonpositive. Assume that the inequality in (3.5) is satisfied; we are to show that it is also satisfied if s is replaced by $s+1$. Inequality (3.5) gives

$$\sum_{s+1}^{n} \alpha(s,j-1)(2-n\theta_j\alpha(j,n)) = (2-n\theta_{s+1}\alpha(s+1,n))$$
$$+ (1 - \frac{2}{s+1} + \theta_{s+1}) \sum_{s+2}^{n} \alpha(s+1,j-1)(2-n\theta_j\alpha(j,n)) \le s\,,$$

or, equivalently,

$$\sum_{s+2}^{n} \alpha(s+1,j-1)(2-n\theta_j\alpha(j,n)) \le (1 - \frac{2}{s+1} + \theta_{s+1})^{-1}(s-2+n\theta_{s+1}\alpha(s+1,n))\,.$$

It is sufficient to show that the last term is $\le (s+1)$, or equivalently,

$$\theta_{s+1}(\alpha(s+1,n) - (s+1)/n) \le 1/n\,. \tag{3.6}$$

This holds because $\alpha(s+1,n)$ is the probability that both traits maximum at stage $s+1$ remain maximum overall, and this is less than or equal to the probability that a given trait maximum at stage $s+1$ is maximum overall, which is $(s+1)/n$. ∎

Lemma 3.4. *For fixed s, $Q_n(r,s)$ is unimodal in r for $1 \le r \le s$. In fact, $Q_n(r+1,s) \ge Q_n(r,s)$ if $2r/n - \alpha(r,n) \le Q_n(r+1,s)$ and $Q_n(r+1,s) \le Q_n(r,s)$ otherwise.*

The proof follows that of Lemma 3.3 using $Q_n(r+1,s) - Q_n(r,s) = \theta_r[Q_n(r+1,s) + \alpha(r,n) - (2r/n)]$. Details are omitted. The maximizing value of r is $r(n,s)$ where

$$\begin{aligned} r(s,n) &= \min\{\, r < s : 2r/n - \alpha(r,n) \ge Q_n(r+1,s)\,\} \quad \text{if nonempty} \\ &= s \qquad \text{otherwise.} \end{aligned} \tag{3.7}$$

From Lemma 3.3, the maximum of $Q_n(r,s)$ over values of (r,s) such that $1 \le r \le s \le n$ occurs on the line $s = s(n)$ or on the line $r = s \ge s(n)$. In the following theorem, we rule out the second case.

Theorem 3.1. *The maximum of $Q_n(r,s)$ over values of (r,s) such that $1 \le r \le s \le n$ occurs at $s = s(n)$ and $r = r(n, s(n))$.*

Proof. Since

$$Q_n(s,s) = \sum_{j=s}^{n} \alpha(s-1, j-1)(\frac{2}{n} - \theta_j \alpha(j,n)),$$

we find when $s \ge s(n)$, using the inequality in (3.5),

$$\begin{aligned} &Q_n(s+1, s+1) - Q_n(s,s) = \\ &\quad = (\frac{2}{s} - \theta_s) \sum_{s+1}^{n} \alpha(s, j-1)(\frac{2}{n} - \theta_j, \alpha(j,n)) - (\frac{2}{n} - \theta_s \alpha(s,n)) \\ &\quad \le (\frac{2}{s} - \theta_s)\frac{s}{n} - (\frac{2}{n} - \theta_s \alpha(s,n)) = \theta_s(\alpha(s,n) - \frac{s}{n}) \le 0, \end{aligned}$$

as in (3.6). Hence $Q_n(s,s)$ is decreasing in s for $s \ge s(n)$, so that $Q_n(r,s)$ takes its maximum on the line $s = s(n)$. ∎

In order to find the asymptotic behavior of $s(n)/n$ as $n \to \infty$, we assume that $\sum_1^\infty \theta_j < \infty$. A referee points out that this condition is essentially that with probability one there are only a finite number of double maxima. First, we find uniform bounds for the two terms of the sum in (3.5). This is done in the following two lemmas.

Lemma 3.5. *If $\sum_1^\infty \theta_j < \infty$, then, as $n \to \infty$,*

$$\max_{1 \le s \le n} \sum_{s+1}^{n} \alpha(s, j-1)\theta_j \alpha(j,n) \to 0.$$

Proof. Note that since $\theta_j \le 1/j$,

$$\alpha(j,n) \le (1 - \frac{1}{j+1}) \cdots (1 - \frac{1}{n}) = \frac{j}{n},$$

and we have

$$\sum_{s+1}^{n} \alpha(s, j-1)\theta_j \alpha(j,n) \le \sum_{s+1}^{n} \frac{s}{j-1}\theta_j \frac{j}{n} \le \frac{2s}{n} \sum_{s+1}^{\infty} \theta_j.$$

The maximum of this over $1 \le s \le n$ is bounded by the larger of the two maxima, over $1 \le s \le \sqrt{n}$, and over $\sqrt{n} \le s \le n$, both of which tend to zero:

$$\max_{1 \le s \le n} 2\frac{s}{n} \sum_{s+1}^{n} \theta_j \le \max\Big\{ \frac{2}{\sqrt{n}} \sum_{1}^{\infty} \theta_j, 2 \sum_{\sqrt{n}+1}^{\infty} \theta_j \Big\} \to 0. \; ∎$$

Lemma 3.6. *If* $\sum_1^\infty \theta_j < \infty$, *then, as* $n \to \infty$,

$$\max_{1 \le s \le n} \Big\{ \frac{2}{n} \sum_{s+1}^{n} \alpha(s, j-1) - \frac{2s(n-s)}{n^2} \Big\} \to 0. \tag{3.8}$$

Proof. The value of $\alpha(s, j-1)$ when $\theta_j \equiv 0$ is

$$\alpha_0(s, j-1) = (1 - \frac{2}{s+1}) \cdots (1 - \frac{2}{j-1}) = \frac{s(s-1)}{(j-1)(j-2)} \ ,$$

so that

$$\frac{2}{n} \sum_{s+1}^{n} \alpha_0(s, j-1) = \frac{2s(n-s)}{n(n-1)} \ .$$

Hence,

$$\begin{aligned}
0 &\le \frac{2}{n} \sum_{s+1}^{n} \alpha(s, j-1) - \frac{2s(n-s)}{n(n-1)} \\
&= \frac{2}{n} \sum_{s+1}^{n} (\alpha(s, j-1) - \alpha_0(s, j-1)) \\
&= \frac{2}{n} \sum_{s+1}^{n} \alpha_0(s, j-1)\, (\prod_{s+1}^{j-1} (1 + \frac{\theta_i}{(i-2)/i}) - 1) \\
&\le \frac{2}{n} \sum_{s+1}^{n} \alpha_0(s, j-1)\, (\exp\{ \sum_{s+1}^{j-1} \frac{\theta_i}{(i-2)/i} \} - 1) \\
&\le \frac{2}{n} \sum_{s+1}^{n} \alpha_0(s, j-1)\, (\exp\{ \frac{s+1}{s-1} \sum_{s+1}^{\infty} \theta_i \} - 1) \\
&= \frac{2s(n-s)}{n(n-1)} (\exp\{ \frac{s+1}{s-1} \sum_{s+1}^{\infty} \theta_i \} - 1) \ .
\end{aligned}$$

As in the proof of Lemma 3.5, the maximum over $1 \le s \le n$ goes to zero because the maximum over $0 \le s \le \sqrt{n}$ and over $\sqrt{n} \le s \le n$ both go to zero. From this, (3.8) follows without difficulty. ■

Theorem 3.2. *If* $\sum_1^\infty \theta_j < \infty$, *then* $s(n)/n \to 1/2$ *as* $n \to \infty$.

Proof. From Lemmas 3.5 and 3.6, we can find for every $\epsilon > 0$ an N sufficiently large that for all $n > N$ and all $s \le n$,

$$\left| \sum_{s+1}^{n} \alpha(s, j-1)\, (\frac{2}{n} - \theta_j \alpha(j, n)) - \frac{2s(n-s)}{n^2} \right| < \epsilon \ .$$

Hence for $n > N$, $s(n)$ is between the two bounds,

$$s_\pm(n) = \min\{\, s \geq 1 \,:\, 2\frac{s}{n}(1 - \frac{s}{n}) \leq \frac{s}{n} \pm \epsilon \,\}$$

from which we may deduce

$$\left| \frac{s(n)}{n} - \frac{1}{2} \right| \;\leq\; 2\epsilon + \frac{1}{n}\,.$$

This implies that $s(n)/n \to 1/2$ as $n \to \infty$. ∎

Next, we find the limiting optimal probability of win for the case $\sum_1^\infty \theta_j < \infty$, and show that this value does not depend otherwise on the θ_j, and that the limit of $Q_n(r,s)$ is independent of the choice of r provided $r \to \infty$ and $s/n \to 1/2$ as $n \to \infty$.

Theorem 3.3. *If* $\sum_1^\infty \theta_j < \infty$, *then* $\max_{r,s} Q_n(r,s) \to 1/2$ *as* $n \to \infty$*; moreover,* $Q_n(r,s) \to 1/2$ *as* $n \to \infty$, $r \to \infty$, *and* $s/n \to 1/2$.

Proof. We upper bound the first term of $Q_n(r,s)$ by deleting the $-\alpha(j,n)$ and replacing $\beta(r,j)$ by 1:

$$\begin{aligned} 0 \leq \; 1^{\text{st}} \text{ term of } Q_n(r,s) &\leq \frac{2}{n}\sum_{j=1}^{n} j\theta_j \\ &\leq \frac{2}{n}\sum_{k=1}^{n}\sum_{j=k}^{\infty}\theta_j \to 0\,, \quad \text{as} \quad n \to \infty\,. \end{aligned}$$

For the second term of $Q_n(r,s)$, we bound $\beta(r,s)$ by 1 and use Lemma 3.6 as in the proof of Theorem 3.2 to show that $\lim_n Q_n(r,s) \leq 1/2$. Furthermore, by letting $s/n \to 1/2$ and $r \to \infty$, we get $Q_n(r,s) \to 1/2$, since $1 \geq \beta(r,s) \geq 1 - \sum_r^\infty \theta_j \to 1$ as $n \to \infty$. ∎

Finally, we show that the limiting value of $r(n,s(n))/n$ also does not depend upon the detailed behavior of the sequence θ_j, provided that $\sum_1^\infty \theta_j < \infty$.

Theorem 3.4. *If* $\sum_1^\infty \theta_j < \infty$, *then* $r(n,s(n))/n \to 1 - 1/\sqrt{2} = .293\cdots$ *as* $n \to \infty$.

Proof. From Theorem 3.3, $Q_n(r,s(n)) \to 1/2$ so that from (3.7), we see that $r(n,s(n))/n$ is at least 1/4 in the limit. Thus, $r(n,s(n))$ tends to ∞. Using the inequalities,

$$\frac{r(r-1)}{n(n-1)} \leq \alpha(r,n) \leq \frac{r(r-1)}{n(n-1)}(1 + 2\sum_{r+1}^{\infty}\theta_i)\,,$$

as in the proof of Lemma 3.6, we may deduce from (3.7) that $r(n,s(n))/n \to x \in (0,1)$, where x satisfies $2x = 1/2 + x^2$. The solution to this is $x = 1 - 1/\sqrt{2}$. ∎

§**4. The Independent Bivariate Secretary.** In this section, we assume that the two rankings of the n objects are independent. This is equivalent to assuming that $\theta_j = (1/j)^2$. From (2.1) and (2.3), we have

$$\alpha(j,n) = \frac{j^2}{n^2} \qquad \text{and} \qquad \beta(r,j) = \frac{(r-1)j}{r(j-1)} \,. \tag{4.1}$$

Thus, the probability of win if we select the j^{th} object becomes

$$\begin{aligned} \pi_j &= j/n && \text{if object } j \text{ is a single maximum} \\ \pi_j &= (j/n)(2-j/n) && \text{if object } j \text{ is a double maximum} \\ \pi_j &= 0 && \text{otherwise.} \end{aligned} \tag{4.2}$$

From Lemma 3.1, there is an optimal rule of the form $R(r,s)$ for some integers r and s, with $1 \le r \le s \le n$. From Lemma 3.2, we find that the probability of win using $R(r,s)$ for $2 \le r \le s \le n$ simplifies to

$$Q_n(r,s) = \frac{r-1}{rn^2}\left(\sum_{j=r}^{s-1}\frac{2n-j}{j-1} + s(s-1)(2n-1)\sum_{j=s}^{n}\frac{1}{(j-1)^2}\right). \tag{4.3}$$

For $r=1$, we have $Q_n(1,s) = (2n-1)/n^2$.

Finally, we find that the optimal value of s, from Lemma 3.3 and Theorem 3.1, simplifies to $s = s(n)$, where

$$s(n) = \min\left\{ s \ge 1 : \ (2n-1)s\sum_{j=s}^{n-1}\frac{1}{j^2} \le n \right\}. \tag{4.4}$$

Using the fairly precise bounds for the sum in (4.4) given in the following lemma, we can find the exact value of $s(n)$.

Lemma 4.1. *For* $1 \le s \le n$,

$$\sum_{j=s}^{n-1}\frac{1}{j^2} = \frac{(n-s)}{(s-.5)(n-.5)} - \delta, \quad \text{where} \quad 0 \le \delta < \frac{1}{4s^3} \,. \tag{4.5}$$

Proof. Since

$$\frac{1}{j^2} = \frac{1}{j-.5} - \frac{1}{j+.5} - \frac{1}{4j^2(j^2-.25)} \,,$$

we have (4.5) with

$$\delta = \sum_{j=s}^{n-1} \frac{1}{4j^2(j^2 - .25)} \, .$$

Clearly $\delta \geq 0$. The other inequality follows from

$$\begin{aligned} \delta &\leq \sum_{j=s}^{\infty} \frac{1}{4j^2(j^2 - .25)} \\ &\leq \frac{1}{4s^2} \sum_{j=s}^{\infty} \frac{1}{(j-.5)(j+.5)} = \frac{1}{4s^2(s+.5)} \leq \frac{1}{4s^3} \, . \ \blacksquare \end{aligned}$$

Theorem 4.1. $s(n) = \lceil (n+1)/2 \rceil$ *(the smallest integer* $\geq (n+1)/2)$.

Proof. From (4.5)

$$(2n-1)s \sum_{j=s}^{n-1} \frac{1}{j^2} = (2n-1)s \left[\frac{n-s}{(s-.5)(n-.5)} - \delta \right] .$$

Since $\delta > 0$, the inequality in (4.4) is satisfied if the weaker inequality, in which δ is ignored, is satisfied:

$$2s(n-s) \quad \leq \quad n(s-.5) \, .$$

Upon solving for n, this inequality becomes

$$\frac{n+1}{2} \quad \leq \quad s + \frac{1}{4s+2} \, .$$

In particular, if $s \geq (n+1)/2$, this inequality is satisfied. This shows that $s(n) \leq (n+2)/2$.

The reverse of the inequality in (4.4) is satisfied if the weaker inequality, in which δ is replaced by $1/(4s^3)$, is satisfied. In particular, if $s \leq n/2$, then

$$\begin{aligned} (2n-1)s \left[\frac{n-s}{(s-.5)(n-.5)} - \frac{1}{4s^3} \right] \quad &\geq \quad (2n-1)s \left[\frac{n/2}{(s-.5)(n-.5)} - \frac{1}{4s^3} \right] \\ &= n + \frac{n}{2s-1} - \frac{2n-1}{4s^2} \quad > \quad n + \frac{n}{2s} - \frac{n}{2s^2} \quad \geq \quad n. \end{aligned}$$

This implies $s(n) \geq (n+1)/2$, completing the proof. $\blacksquare$

One can find upper and lower bounds for the optimal $r(n) = r(n, s(n))$ in a similar manner. Instead, we use an asymptotic analysis. Using (4.1) and (4.3), we may write (3.7) as

$$r(n,s) = \min \left\{ r : \ (r+1)(2n-r-1) \quad \geq \quad z(s) - s + \sum_{r+1}^{s-1} \frac{1}{j-1} \right\} \tag{4.6}$$

where

$$z(s) = s(s-1)(2n-1)\sum_{s}^{n} \frac{1}{(j-1)^2} \ .$$

Then, using $s(n) = n/2+O(1)$ as $n \to \infty$, we find $z(s(n))/n^2 = (1/2)(1+3/n)+O(1/n^2)$. Moreover, since $r(n)/n \to \phi = 1 - 1/\sqrt{2}$, we find

$$\sum_{r+1}^{s} \frac{1}{j-1} = -\log(2\phi) + O(1/n) \ .$$

Combining these into (4.6) gives

$$r(n, s(n)) = \min\left\{r : \ \frac{(r+1)(2n-(r+1))}{n^2} \ \geq \ \frac{1}{2}(1 + \frac{2}{n} - \frac{1}{n}\log(2\phi)\,) + O(\frac{1}{n^2})\right\} .$$

Solving the quadratic inequality for $(r+1)$, we find

$$r(n, s(n)) = \lfloor n\phi + \lambda + O(1/n) \rfloor \tag{4.7}$$

where $\lambda = (1 - 2\log(2\phi))/\sqrt{2} = 1.463428\cdots$. Surprisingly, numerical calculations show that the approximation, $r(n, s(n)) = \lfloor n\phi + \lambda \rfloor$, is valid for all values of $n < 10000$.

Table 4.1. Optimal values of r, s, $Q_n(r,s)$ and $Q_n(s)$.

n	r	s	$Q_n(r,s)$	$Q_n(s)$	n	r	s	$Q_n(r,s)$	$Q_n(s)$
1	1	-	1.0000	1.0000	10	4	6	0.5647	0.5556
2	1	-	0.7500	1.0000	20	7	11	0.5325	0.5263
3	2	2	0.6944	0.6667	30	10	16	0.5217	0.5172
4	2	3	0.6615	0.6667	40	13	21	0.5163	0.5128
5	2	3	0.6175	0.6000	50	16	26	0.5130	0.5102
6	3	4	0.6055	0.6000	60	19	31	0.5109	0.5085
7	3	4	0.5872	0.5714	70	21	36	0.5093	0.5072
8	3	5	0.5803	0.5714	80	24	41	0.5082	0.5063
9	4	5	0.5668	0.5556	90	27	46	0.5073	0.5056

In Table 4.1, the values of the optimal r and s and probability of win in the independent case (4.3), and in the perfect negative dependent case (2.1), are given for various values of n. Since the win probability in the independent case is greater than the win probability in the perfect positive dependent case, one might suspect that the optimal probability of win in the perfect negative dependent case would be greater yet. That this is true for only two values of n is seen in the table.

§5. Ranking Models and Sampling Models. The perfect positive dependence case, (1.4(a)), is equivalent to the standard best-choice problem in which the optimal s/n and the optimal probability of win both tend to $e^{-1} = .3679\cdots$ as $n \to \infty$. For both the independent case (1.4(b)) and the perfect negative dependence case (1.4(c)), the optimal s/n and the optimal probability of win tend to $1/2$ as $n \to \infty$. It is natural to wonder what happens when the two traits have intrinsic values given, say, by a bivariate normal distribution with zero means, unit variances and a correlation coefficient ρ. This leads us to sampling models, ranking models, and to the reasons why the ranking model in section 1 may be unrealistic.

In the univariate case, the sampling model and the ranking model are equivalent. In the sampling model, the relative rankings of the objects are given as the actual rankings of their intrinsic worths, $Z_1, \ldots, Z_n$, with the assumption,

$$\begin{aligned} &Z_1, \ldots, Z_n \quad \text{are i.i.d. } F(z), \text{ continuous,} \\ &R_j = \text{rank of } Z_j \text{ among } Z_1, \ldots, Z_j. \end{aligned} \tag{5.1}$$

All the decision maker sees are the relative ranks, $R_1, \ldots, R_n$, of the Z's. In the ranking model, the decision maker observes the relative ranks, $R_1, \ldots, R_n$, directly, where it is assumed that

$$\begin{aligned} &(a) \quad R_1, \ldots, R_n \quad \text{are independent.} \\ &(b) \quad \text{For all j, } P(R_j = r) = 1/j \text{ for } r = 1, \ldots, j. \end{aligned} \tag{5.2}$$

It is not necessarily assumed that the objects have intrinsic worths. These models are equivalent in the sense that (5.1) implies (5.2). This equivalence does not depend on F, and we may take F to be the uniform distribution on the interval (0,1) by applying F^{-1} to each observation.

For the independent m-variate case, a similar equivalence holds. But when the traits are dependent, there are immediate differences. The sampling model is given by

$$
\begin{aligned}
(5.3) \qquad & \mathbf{Z}_1, \ldots, \mathbf{Z}_n \quad \text{are i.i.d. } F(\mathbf{z}) \text{ with continuous marginals.} \\
& R_{ij} = \text{rank of } Z_{ij} \text{ among } Z_{i1}, \ldots, Z_{ij}.
\end{aligned}
$$

First of all, the vectors $\mathbf{R}_1, \ldots, \mathbf{R}_n$ of relative ranks are not necessarily independent, and second, although the marginal distribution of R_{ij} is uniform on $\{1, \ldots, j\}$, the joint distribution of $R_{1j}, \ldots, R_{mj}$ depends on F. There are some general lower bounds on the optimal value given in Theorem 2 of Berezovskii et al (1986) that are valid for both ranking models and sampling models and are found using threshold rules. For the m-dimensional nonsingular multivariate normal distribution, the lower bound is asymptotic to $(1/m)^{1/(m-1)}$ for $m > 1$, giving 1/2 for the bivariate case.

The differences in the two models may be illustrated by examples in the bivariate case, even when we restrict the observations to the occurences of relative maxima. The results of Samuel and Chotlos (1985) require the entire relative ranks and are even more difficult to generalize to dependent sampling models. We take $m = 2$ and let $(Z_{11}, Z_{21}), \ldots, (Z_{1n}, Z_{2n})$ denote the vectors of the intrinsic worths of the objects, assumed to be i.i.d. with distribution $F(\mathbf{z})$. As in the independent case, we may assume without loss of generality that the marginal distributions of Z_{11} and Z_{21} are uniform on the interval (0,1), denoted by $\mathcal{U}(0,1)$.

Example. Let Z_{11} be $\mathcal{U}(0,1)$ and for some $0 < \delta < 1$, let

$$Z_{21} = \delta - Z_{11} \quad \text{if} \quad Z_{11} < \delta, \qquad \text{and} \qquad Z_{21} = Z_{11} \quad \text{if} \quad Z_{11} > \delta\,.$$

Then Z_{21} is also $\mathcal{U}(0,1)$ and the correlation coefficient, $\rho = 1 - 2\delta^3$, goes from -1 to $+1$ as δ goes from 1 to 0. As in section 1, we let X_j (resp. Y_j) be the indicator function of the event that Z_{1j} (resp. Z_{2j}) is a relative maximum for the first (resp. second) coordinate, then we can see that (1.1(a)) is not satisfied for any $0 < \delta < 1$:

$$
\begin{aligned}
P_\delta(X_3 = 1, Y_3 = 1 \mid X_2 = 1, Y_2 = 0) &= P_\delta(X_3 = 1, Y_3 = 1 \mid X_2 = 0, Y_2 = 1) = 1 - \delta, \\
P_\delta(X_3 = 1, Y_3 = 1 \mid X_2 = 0, Y_2 = 0) &= P_\delta(X_3 = 1, Y_3 = 1 \mid X_2 = 1, Y_2 = 1) \\
&= (1 - \delta)(1 + 2\delta)/(3(1 + \delta)).
\end{aligned}
$$

This implies that when making a decision at stage j, data gathered before stage j cannot be ignored.

Suppose at stage two we observe $X_2 = Y_2 = 1$. For any fixed n, if δ is sufficiently close to 1, this object is very likely to be the best in both traits and we should select it. In other words, the fact that a joint maximum occurred gave us a good idea as to the intrinsic worth of the object.

Another difficulty arises because sufficiency does not reduce the data. At any stage j, we must remember not only how many double maxima and single maxima occurred, but when they occurred. For example, let d denote a double maximum, s a single maximum and n neither maximum. Then, observing the sequence $d, s, d, n, \ldots, n, n, n$ is different from observing $d, s, n, n, \ldots, n, n, d$ because in the former the n's have a chance to be associated with Z_{1i}'s $> \delta$; thus, $P_\delta(d \text{ next} \mid d, s, d, n, \ldots, n) < P_\delta(d \text{ next} \mid d, s, n, \ldots, n, d)$.

Thus, solving any given dependent sampling model for general n looks very difficult. It is unknown whether or not there exists a sampling model that gives rise to (1.1(a)), except for the three cases of (1.4).

§**6. The (Best-Choice)^m Problem.** We look briefly at the problem with $m \geq 2$ traits and in which to win you must select the object, if any, which is best in all traits. This problem was treated for $m = 2$ in Stadje (1980). Since we assume the traits are independent, the probability that there is an object which is best simultaneously in all traits is only $1/n^{(m-1)}$; so the probability of a win is at most $1/n^{(m-1)}$. We consider the questions: How much smaller than $1/n^{(m-1)}$ is the optimal probability of win, and what does the optimal strategy look like?

By the argument of Lemma 3.1, it follows that there is an optimal rule, call it $R(s)$, of the form: Reject the first $s-1$ objects and then select the next object which is relatively best in all traits. The probability of win, $Q_n(s)$, using such a rule may be written as:

$$Q_n(s) = n^{-m} \sum_{j=s}^{n} \prod_{k=s}^{j-1} (1 - k^{-m}).$$

Theorem 6.1. *The optimal value of s is $s(n)$, where*

(6.1) $$s(n) = \min\{\, s \geq 1 \,:\, Q_n(s+1) \,\leq\, s^m/n^m \,\}.$$

As $n \to \infty$, $s(n)^m/n \to 1$ *and* $n^{m-1}Q_n(s(n)) \to 1$.

Proof. To find the optimal value of s, we look at the differences,

$$n^m\,[\,Q_n(s+1) - Q_n(s)\,] = s^{-m}\sum_{j=s+1}^{n}\prod_{k=s+1}^{j-1}(1-k^{-m})\,-1\,.$$

This is decreasing in s, so the optimal value of s is given by (6.1). Using this, we find lower and upper bounds for $s = s(n)$:

$$(6.2)\qquad \begin{aligned} &(a)\quad \sum_{j=s+1}^{n}\prod_{k=s+1}^{j-1}(1-k^{-m})\,\le\, s^m\\ &(b)\quad \sum_{j=s}^{n}\prod_{k=s}^{j-1}(1-k^{-m})\,>\,(s-1)^m\,. \end{aligned}$$

From (b), $(s-1)^m < n-s+1$, so limsup $s^m/n \le 1$. From (a),

$$s^m \,\ge\, \sum_{j=s+1}^{n}(1-\sum_{k=s+1}^{j-1}k^{-m})\,\ge\, n-s-n\sum_{s+1}^{n-1}1/k^m\,\ge\, n-s-n/s^{m-1},$$

so $\liminf\, s^m/n \ge 1$. Thus, $\lim\, s^m/n = 1$, and from (6.1) we also have $\lim\, n^{m-1}\,Q(s(n)) = 1$. ■

We see that the optimal rule waits until about $n^{1/m}$ of the observations have been made, and then selects the next object, if any, that is best in all traits. Using more precise bounds for the terms in (6.1), we can find some very precise estimates of $s(n)$. We treat separately the cases $m = 2$ and $m > 2$.

First assume $m > 2$ and $s = O(n^{1/m})$. Then, using the elementary expansions,

$$(6.3)\qquad \prod_{k=s+1}^{n}(1-k^{-m}) = 1 - \sum_{k=s+1}^{n}k^{-m} + O(s^{-2m-2})$$

and

$$(6.4)\qquad \sum_{k=s+1}^{\infty}k^{-m} = \frac{s^{-(m-1)}}{m-1} - \frac{s^{-m}}{2} + O(s^{-(m+1)})\,,$$

we find,

$$\sum_{j=s+1}^{n} \prod_{k=s+1}^{j-1} (1-k^{-m}) = \sum_{j=s+1}^{n} (1 - \sum_{k=s+1}^{j-1} k^{-m}) + O(ns^{-2m-2})$$

$$= n - s - n\sum_{s+1}^{n} k^{-m} + \sum_{s}^{n} k^{-m+1} + O(s^{-m-2})$$

$$= n - s - \frac{ns^{-(m-1)}}{m-1} + \frac{ns^{-m}}{2} + O(s^{-1})\,,$$

from which we can obtain a very accurate approximation for $s(n)$ as the smallest integer greater than the solution of

$$s^m = n - \frac{ms}{m-1} + \frac{1}{2} \tag{6.5}$$

A simple approximation for $s(n)$ is found by replacing s in the left side of (6.5) by $n^{1/m}$ and solving for s. However, it is more accurate to use one iteration of Newton's method with $s = n^{1/m}$ as the initial value. This leads to the approximation

$$s(n) \sim \left\lceil \frac{n + 1/2m}{n^{(m-1)/m} - 1/(m-1)} \right\rceil .$$

For $m = 3$, I could find no values of n for which this did not give the correct answer.

For $m = 2$, the formula for Q_n simplifies to

$$Q_n(s) = n^{-2} \sum_{j=s}^{n} \frac{s-1}{s} \frac{j}{j-1}$$

so that $s(n)$ reduces to

$$s(n) = \min \left\{ s \geq 1 \,:\, n + 1 + \sum_{j=s}^{n-1} \frac{1}{j} \;\leq\; (s+1)^2 \right\} .$$

From this, we can derive a very good approximation to $s(n)$ when $m = 2$:

$$s(n) \sim \left\lfloor \left(n + 1 + \frac{log(n)}{2} \right)^{1/2} \right\rfloor .$$

This gives the optimal value of s when $n \leq 10000$ in all but five cases ($n =$ 7, 46, 357, 396, 2911) .

References.

Baryshnikov, Yu. M., Berezovskii, B. A. and Gnedin, A. V. (1984) "On the probability of stopping on a nondominated option" *Automat. & Remote Contro* **45** No. **10** 1354-1359.

Berezovskiy, B. A., Baryshnikov, Yu. M. and Gnedin, A. V. (1986) "On a class of best-choice problems" *Information Sciences* **39** 111-127.

Berezovskii, B. A., Geninson, B. A. and Rubchinskii, A. A. (1980) "Optimal stopping on partially ordered objects" *Automat. & Remote Control* **41** No. **11** 1537-1542.

Berezovskii, B. A. and Gnedin, A. V. (1981) "Theory of choice and the problem of optimal stopping at a best entity" *Automat. & Remote Control* **42** No. **9** 1221-1225.

Chow, Y. S., Moriguti, S.,Robbins, H., and Samuels, S. M. (1964) "Optimal selection based on relative rank (The "secretary" problem)" *Israel J. Math.* **2** 81-90.

Gilbert, J. and Mosteller, F. (1966) "Recognizing the maximum of a sequence" *J. Amer. Statist. Assoc.* **61** 35-73.

Gnedin, A. V. (1981) "A multicriteria problem of optimal stopping of a selection process" *Automat. & Remote Control* **42** No. **7** 981- 986.

Gnedin, A. V. (1983) "Efficient stopping on a Pareto optimal option" *Avtomat. i Telemekh.* **1983** No.**3** 166-170 (in Russian).

Samuels, S. M. and Chotlos, B. (1987) "A multiple criteria optimal selection problem" *Adaptive Statistical Procedures and Related Topics*, J. Van Ryzin, Ed. I.M.S. Lecture Note - Monograph Series, Vol. **8**.

Stadje, W. (1980) "Efficient stopping of a random series of partially ordered points" *Multiple Criteria Decision Making Theory and Applications* Lecture notes in Econ. and Math. Syst. vol. **177**, Springer-Verlag, 430-447.

Thomas S. Ferguson

Mathematics Department

UCLA

Los Angeles, CA 90024