OPTIMAL SELECTION BASED ON RELATIVE RANK*
(the "Secretary Problem")

BY

Y. S. CHOW, S. MORIGUTI, H. ROBBINS AND S. M. SAMUELS

ABSTRACT

n rankable persons appear sequentially in random order. At the ith stage we observe the relative ranks of the first i persons to appear, and must either select the ith person, in which case the process stops, or pass on to the next stage. For that stopping rule which minimizes the expectation of the absolute rank of the person selected, it is shown that as $n \to \infty$ this tends to the value

$$\prod_{j=1}^{\infty} \left(\frac{j+2}{j} \right)^{1/j+1} \cong 3.8695.$$

1. **Introduction.** n girls apply for a certain position. If we could observe them all we could rank them absolutely with no ties, from best (rank 1) to worst (rank n). However, the girls present themselves one by one, in random order, and when the ith girl appears we can observe only her rank relative to her $i-1$ predecessors, that is, 1 + the number of her predecessors who are better than she. We may either select the ith girl to appear, in which case the process ends, or reject her and go on to the $(i+1)$th girl; in the latter case the ith girl cannot be recalled. We must select one of the n girls. Let X denote the absolute rank of the girl selected. The values of X are $1, \cdots, n$, with probabilities determined by our selection strategy. What selection strategy (*i.e.* stopping rule) will minimize the expectation $EX =$ expected absolute rank of the girl selected?

To formulate the problem mathematically, let $x_1, \cdots, x_n$ denote a random permutation of the integers $1, \cdots, n$, all $n!$ permutations being equally likely. The integer 1 corresponds to the best girl, $\cdots, n$ to the worst. For any $i = 1, \cdots, n$ let $y_i = 1 +$ number of terms $x_1, \cdots, x_{i-1}$ which are $< x_i$ (y_i = relative rank of ith girl to appear). It is easy to see that the random variables $y_1, \cdots, y_n$ are independent, with the distribution

$$P(y_i = j) = \frac{1}{i} \qquad (j = 1, \cdots, i), \tag{1}$$

and that

$$P(x_i = k \mid y_1 = j_1, \cdots, y_{i-1} = j_{i-1}, y_i = j) = P(x_i = k \mid y_i = j) \tag{2}$$
$$= \binom{k-1}{j-1} \binom{n-k}{i-j} \Big/ \binom{n}{i},$$

Received August 22, 1964

*Research supported by Office of Naval Research and Aerospace Research Laboratories.

so that

$$E(x_i \mid y_i = j) = \sum_{k=1}^{n} kP(x_i = k \mid y_i = j) = \frac{n+1}{i+1} j. \tag{3}$$

For any stopping rule τ the expected absolute rank of the girl selected is therefore $EX = E\left(\frac{n+1}{\tau+1} y_\tau\right)$. We wish to minimize this value by optimal choice of τ.

To find an optimal τ by the usual method of backward induction we define for $i = 0, 1, \cdots, n-1$, $c_i = c_i(n) =$ minimal possible expected absolute rank of girl selected if we must confine ourselves to stopping rules τ such that $\tau \geqq i+1$. We are trying to find the value c_0. Now

$$c_{n-1} = E\left(\frac{n+1}{n+1} y_n\right) = \frac{1}{n} \sum_{1}^{n} j = \frac{n+1}{2}, \tag{4}$$

and for $i = n-1, n-2, \cdots, 1$,

$$c_{i-1} = E\left(\min\left(\frac{n+1}{i+1} y_i, c_i\right)\right) = \frac{1}{i} \sum_{j=1}^{i} \min\left(\frac{n+1}{i+1} j, c_i\right). \tag{5}$$

These equations allow us to compute successively the values $c_{n-1}, c_{n-2}, \cdots, c_1, c_0$ and contain the implicit definition of an optimal stopping rule. Equation (5) can be rewritten more simply if we denote by $[x]$ the greatest integer $\leqq x$ and set

$$s_i = \left[\frac{i+1}{n+1} c_i\right] \qquad (i = n-1, \cdots, 1); \tag{6}$$

then (5) becomes

$$\begin{aligned} c_{i-1} &= \frac{1}{i}\left\{\frac{n+1}{i+1}(1 + 2 + \cdots + s_i) + (i - s_i) c_i\right\} \\ &= \frac{1}{i}\left\{\frac{n+1}{i+1} \cdot \frac{s_i(s_i+1)}{2} + (i - s_i) c_i\right\}. \end{aligned} \tag{7}$$

Defining $s_n = n$, an optimal stopping rule is, stop with the first $i \geqq 1$ such that $y_i \leqq s_i$; the expected absolute rank of the girl selected using this rule is c_0.

We observe from (4) and (5) that

$$c_0 \leqq c_1 \leqq \cdots \leqq c_{n-1} = \frac{n+1}{2}, \tag{8}$$

and from (6) and (8) that $s_i \leqq i$ and

$$s_1 \leqq s_2 \leqq \cdots \leqq s_{n-1} = \left[\frac{n}{2}\right]. \tag{9}$$

For example, let $n = 4$. Then from (4), (6),and(7),

$$c_3 = \frac{5}{2}, s_3 = 2, c_2 = \frac{1}{3}\left(\frac{5}{4}\cdot\frac{2.3}{2} + \frac{5}{2}\right) = \frac{25}{12}, s_2 = 1,$$

$$c_1 = \frac{1}{2}\left(\frac{5}{3}\cdot\frac{1.2}{2} + \frac{25}{12}\right) = \frac{15}{8}, \quad s_1 = 0, c_0 = c_1 = \frac{15}{8},$$

and an optimal stopping rule is given by the vector $(s_1, \cdots, s_4) = (0,1,2,4)$. The values of c_0 for $n = 10$, 100, 1000 are found by similar computation to be respectively 2.56, 3.60, 3.83.

D. V. Lindley [1] has treated this problem heuristically for large n by replacing (7) by a single differential equation. His results indicate that for $n \to \infty$, c_0 should approach a finite limit, but his method is too rough to give the value of this limit. A more adequate but still heuristic approach involves replacing (7) by an infinite sequence of differential equations, one for each value $0, 1, \cdots$ of s_i. This method indicates that $\lim c_0$ (all limits as $n \to \infty$) has the value

$$\prod_{j=1}^{\infty}\left(\frac{j+2}{j}\right)^{1/j+1} \cong 3.8695. \tag{10}$$

It is not clear how to make this heuristic argument rigorous by appealing to known theorems on the approximation of difference equations by differential equations. Instead, we shall give a direct proof that c_0 tends to the value (10).

2. **The basic inequalities.** We shall derive rather crude upper and lower bounds for the constants c_i which will permit the evaluation of $\lim c_0$. To this end we define the constants

$$t_i = \frac{i+1}{n+1}c_i \qquad (i = 0, 1, \cdots, n-1). \tag{1}$$

From (1.8) it follows that

$$t_0 < t_i < \cdots < t_{n-1} = \frac{n}{2} \tag{2}$$

and (1.7) becomes

$$t_{i-1} = \frac{s_i(s_i+1) + 2(i-s_i)t_i}{2(i+1)}, \quad s_i = [t_i] \, (i = 1, \cdots, n-1). \tag{3}$$

For fixed i set

$$t_i = s_i + \alpha, \qquad 0 \leqq \alpha < 1; \tag{4}$$

then (3) becomes

$$t_{i-1} = \frac{t_i(1+2i-t_i)}{2(i+1)} - \frac{\alpha(1-\alpha)}{2(i+1)}. \tag{5}$$

Put

$$T(x) = \frac{x(1+2i-x)}{2(i+1)}; \tag{6}$$

then for $x \leqq i + \frac{1}{2}$,

$$T'(x) = \frac{1+2i-2x}{2(i+1)} \geqq 0 \tag{7}$$

so $T(x)$ is increasing for $x \leqq i + \frac{1}{2}$, and by (5),

$$t_{i-1} \leqq T(t_i). \tag{8}$$

We now prove the first basic inequality.

LEMMA 1.

$$t_i \leqq \frac{2n}{n-i+3} \qquad (i = 0, \cdots, n-1). \tag{9}$$

Proof. (9) is true for $i = n-1$ since by (2)

$$t_{n-1} = \frac{n}{2} = \frac{2n}{4}.$$

Assume (9) holds for some $1 \leqq i \leqq n-1$; we shall prove that it holds for $i-1$. We know by (1.8) and (1) that

$$t_{i-1} = \frac{i}{n+1} c_{i-1} \leqq \frac{i}{n+1} \cdot \frac{n+1}{2} = \frac{i}{2}. \tag{10}$$

If $\frac{i}{2} \leqq \frac{2n}{n-i+4}$ then (9) holds for $i-1$. If $\frac{i}{2} > \frac{2n}{n-i+4}$ then $i + \frac{1}{2} \geqq \frac{2n}{n-i+3}$, so by (8), since $T(x)$ is increasing for $x \leqq i + \frac{1}{2}$,

$$t_{i-1} \leqq T(t_i) \leqq T\left(\frac{2n}{n-i+3}\right) = \frac{n\{(1+2i)(n-i+3)-2n\}}{(i+1)(n-i+3)^2} \leqq \frac{2n}{n-i+4}, \tag{11}$$

since the last inequality is equivalent to

$$(n-i+3)^2 + (2n-2i-1)(n-i+3) + 2n \geqq 0,$$

which is true for $n - i \geqq 1$. Hence (9) holds for $i-1$ in this case also, and the lemma is proved.

COROLLARY 1.

$$c_0 < 8 \qquad (n = 1, 2, \cdots). \tag{12}$$

Proof. In (9) set $i = \left[\frac{n}{2}\right]$. Then

$$c_0 \leqq c_i = \frac{n+1}{i+1} t_i \leqq \frac{n+1}{i+1} \cdot \frac{2n}{n-i+3} \leqq \frac{2n(n+1)}{\frac{n}{2}\left(\frac{n}{2}+3\right)} < 8. \tag{13}$$

We next observe that

$$t_{i-1} \geqq \frac{i}{i+1}\left\{1 - \frac{t_i}{2(i+1)}\right\} \qquad (i = 1, \cdots, n-1). \tag{14}$$

This inequality follows from (5) if we show that

$$\frac{t_i(1+2i-t_i)}{2(i+1)} - \frac{\alpha(1-\alpha)}{2(i+1)} \geqq \frac{i}{i+1} t_i \left\{1 - \frac{t_i}{2(i+1)}\right\},$$

which reduces to

$$t_i\left(1 - \frac{t_i}{i+1}\right) \geqq \alpha(1-\alpha) \qquad (\alpha = t_i - [t_i]), \tag{15}$$

and this is true if $t_i \geqq 1$, since $t_i \leqq \frac{i+1}{2}$ by (10), and if $t_i < 1$, since then $t_i = \alpha$.

We now establish the second basic inequality.

LEMMA 2.

$$t_i \geqq \frac{3(i+1)}{2(n-i+2)} \qquad (i = 0, \cdots, n-1). \tag{16}$$

Proof. (16) is true for $i = n-1$; suppose it is true for some $1 \leqq i \leqq n-1$. Define

$$T(x) = x\left\{1 - \frac{x}{2(i+1)}\right\},$$

which is increasing for $x \leqq i+1$. Since by (10) $t_i \leqq \frac{i+1}{2}$, we have by (14),

$$t_{i-1} \geqq \frac{i}{i+1} T(t_i) \geqq \frac{i}{i+1} T\left(\frac{3(i+1)}{2(n-i+2)}\right)$$
$$= \frac{3i(4n-4i+5)}{8(n-i+2)^2} \geqq \frac{3i}{2(n-i+3)}.$$

which is equivalent to $i \leqq n-1$. This proves the lemma.

We have seen (1.9) that for any positive integer k, if $n \geqq 2k$ then $s_{n-1} \geqq k$. We now define for each $k = 1, 2, \cdots$ and each $n \geqq 2k$,

(17) $\qquad i_k =$ smallest integer $j \geqq 1$ such that $s_j \geqq k$.

We note that $s_{i_1 - 1} = 0$ and hence from (1.7),

$$c_0 = c_1 = \cdots = c_{i_1 - 1}. \tag{18}$$

COROLLARY 2.

$$\varliminf \frac{i_1}{n} \geqq \frac{1}{8}. \tag{19}$$

Proof. If $i_1 > [n/2]$ then $i_1 \geqq [n/2]+1 > n/2$, $(i_1/n) > \frac{1}{2}$. If $i_1 \leqq [n/2]$ then by (13),

$$1 \leqq s_{i_1} \leqq t_{i_1} = \frac{i_1+1}{n+1} c_{i_1} \leqq \frac{i_1+1}{n+1} c_{[n/2]} < \frac{i_1+1}{n+1} \cdot 8.$$

(We remark that $i_1 > 1$ for $n > 2$, since if $i_1 = 1$ then $s_1 = 1$ and $c_0 = \dfrac{n+1}{2}$, which only holds for $n \leqq 2$.)

COROLLARY 3. *On every set*

$$\left\{\alpha \leqq \frac{i}{n} \leqq \beta;\ 0 < \alpha < \beta < 1\right\}, \tag{20}$$

$$\lim (t_i - t_{-1}) = 0 \tag{21}$$

uniformly.

Proof. From (14) and (9),

$$0 \leqq t_i - t_{i-1} \leqq t_i - \frac{i}{i+1} t_i \left\{1 - \frac{t_i}{2(i+1)}\right\} = \frac{t_i}{i+1} + \frac{i t_i^2}{2(i+1)^2}$$

$$\leqq \frac{(1+t_i)^2}{2(i+1)} \leqq \frac{\left(1+\dfrac{2n}{n-i}\right)^2}{2(i+1)} \leqq \frac{\left(1+\dfrac{2}{1-\beta}\right)^2}{2\alpha} \cdot \frac{1}{n} \to 0.$$

COROLLARY 4. *For* $k = 1, 2, \cdots$ *and* $n \geqq 12k$,

$$\frac{i_k}{n} \geqq 1 - \frac{2}{k}, \tag{22}$$

$$\frac{i_k}{n} \leqq 1 - \frac{1}{2k}. \tag{23}$$

Proof. By (9),

$$s_{i_k} \geqq k \Rightarrow t_{i_k} \geqq k \Rightarrow \frac{2n}{n-i_k} \geqq k \Rightarrow \frac{i_k}{n} \geqq 1 - \frac{2}{k},$$

which proves (22). (23) holds if $i_k \leqq \left[\dfrac{n}{2}\right]$, for then

$$\frac{i_k}{n} \leqq \frac{1}{2} \leqq 1 - \frac{1}{2k},$$

and if $i_k > \left[\frac{n}{2}\right]$ then by (16),

$$s_{i_k-1} < k \Rightarrow t_{i_k-1} < k \Rightarrow \frac{3i_k}{2(n-i_k+3)} < k \Rightarrow \frac{3i_k}{2k} < n - i_k + 3 \Rightarrow \frac{3n}{4k} < n - i_k + 3$$

$$\Rightarrow \frac{3}{4k} < 1 - \frac{i_k}{n} + \frac{3}{n} \Rightarrow \frac{i_k}{n} < 1 - \frac{3}{4k} + \frac{3}{n} \leqq 1 - \frac{1}{2k} \text{ for } n \geqq 12k.$$

COROLLARY 5.

$$\lim t_{i_k} = \lim t_{i_k-\gamma} = k \qquad (k,\gamma = 1,2,\cdots). \tag{24}$$

Proof. $t_{i_k-\gamma} < k \leqq t_{i_k}$.

Choose α, β so that $0 < \alpha < \frac{1}{8}$, $1 - \frac{1}{2k} < \beta < 1$. Then by (19) and (23),

$$\alpha < \frac{i_k - \gamma}{n} < \frac{i_k}{n} < \beta \tag{25}$$

for sufficiently large n. Hence by Corollary 3,

$$\lim (t_{i_k} - t_{i_k-\gamma}) = 0.$$

COROLLARY 6. *For* $k = 1,2,\cdots$

$$s_{i_k} = k \textit{ for sufficiently large } n, \tag{26}$$

$$\lim (i_{k+1} - i_k) = \infty. \tag{27}$$

Proof. $k \leqq s_{i_k} \leqq t_{i_k}$ together with (24) proves (26).

$$\lim (t_{i_{k+1}} - t_{i_k}) = 1 \text{ and } \lim (t_{i_{k+1}} - t_{i_{k+1}-\gamma}) = 0$$

by (24), and these relations imply (27).

3. **Proof of the Theorem.** Choose and fix a positive integer k and let n by (2.26) be so large that $s_{i_k} = k$, $s_{i_{k+1}} = k+1$. For $i_k \leqq i < i_{k+1}$ define

$$v_i = t_i - \frac{k}{2}. \tag{1}$$

Substituting in (2.3) we find that

$$v_{i-1} + \frac{k}{2} = \frac{k(k+1) + 2(i-k)\left(v_i + \frac{k}{2}\right)}{2(i+1)} = \frac{k}{2} + \frac{i-k}{i+1} v_i,$$

$$v_i = \frac{i+1}{i-k} v_{i-1}. \tag{2}$$

Hence for $i_k < i < i_{k+1}$,

$$v_i = \frac{i+1}{i-k}v_{i-1} = \frac{i+1}{i-k}\,\frac{i}{i-k-1}\cdots\frac{i_k+2}{i_k-k+1}v_{i_k} \tag{3}$$

$$= \frac{i+1}{i_k+1}\cdot\frac{i}{i_k}\cdots\frac{i+1-k}{i_k+1-k}\cdot v_{i_k} = v_{i_k}\prod_{j=1}^{k+1}\left(\frac{i+j-k}{i_k+j-k}\right),$$

and hence

$$t_i = \frac{k}{2} + \left(t_{i_k} - \frac{k}{2}\right)\prod_{j=1}^{k+1}\left(\frac{i+j-k}{i_k+j-k}\right). \tag{4}$$

Set $i = i_{k+1} - 1$; then

$$t_{i_{k+1}-1} = \frac{k}{2} + \left(t_{i_k} - \frac{k}{2}\right)\prod_{j=1}^{k+1}\left(\frac{i_{k+1}+j-k-1}{i_k+j-k}\right). \tag{5}$$

From (2.19) and (2.24) it follows that

$$k+1 = \frac{k}{2} + \frac{k}{2}\prod_{j=1}^{k+1}\lim\left(\frac{i_{k+1}+j-k-1}{i_k+j-k}\right) = \frac{k}{2} + \frac{k}{2}\lim\left(\frac{i_{k+1}}{i_k}\right)^{k+1}$$

and hence

$$\lim \frac{i_{k+1}}{i_k} = \left(\frac{k+2}{k}\right)^{1/k+1}, \lim\frac{i_1}{i_k} = \lim\frac{i_1/n}{i_k/n} = \prod_{j=1}^{k-1}\left(\frac{j}{j+2}\right)^{1/j+1}. \tag{6}$$

From (2.22)

$$\left(1-\frac{2}{k}\right)\prod_{j=1}^{k-1}\left(\frac{j}{j+2}\right)^{1/j+1} \leqq \varliminf \frac{i_1}{n} \leqq \varlimsup \frac{i_1}{n} \leqq \prod_{j=1}^{k-1}\left(\frac{j}{j+2}\right)^{1/j+1}. \tag{7}$$

Letting $k \to \infty$,

$$\lim \frac{i_1}{n} = \prod_{j=1}^{\infty}\left(\frac{j}{j+2}\right)^{1/j+1}. \tag{8}$$

Now by (2.24) and (2.18),

$$1 = \lim t_{i_1-1} = \lim\left(\frac{i_1}{n+1}c_{i_1-1}\right) = \lim\left(\frac{i_1}{n}c_0\right)$$

$$= \lim c_0 \cdot \prod_{j=1}^{\infty}\left(\frac{j}{j+2}\right)^{1/j+1}. \tag{9}$$

Thus we have proved the

THEOREM. $$\lim c_0 = \prod_{j=1}^{\infty}\left(\frac{j+2}{j}\right)^{1/j+1}.$$

4. Remarks.

1. It is interesting to note that $c_0 = c_0(n)$ is a strictly increasing function of n; thus in view of the Theorem,

$$c_0(n) < 3.87 \qquad (n = 1, 2, \cdots). \tag{1}$$

A direct proof that $c_0(n)$ is strictly increasing based on the formulas of Section 1 is difficult, since there is no obvious relation between the c_i for different values of n. However, a direct probabilistic proof can be given which involves no use of the recursion formulas. Let $(s_1, \cdots, s_n, n+1)$ be any stopping rule for the case of $n+1$ girls such that $s_i < i$ for $i = 1, \cdots, n$ (any optimal stopping rule has this property). Define for $i = 1, \cdots, n$

$$t_j(i) = \begin{cases} s_j & \text{for } j = 1, \cdots, i-1, \\ s_{j+1} & \text{for } j = i, \cdots, n, \end{cases} \tag{2}$$

and

$$t_j(n+1) = \begin{cases} s_j & \text{for } j = 1, \cdots, n-1, \\ n & \text{for } j = n. \end{cases} \tag{3}$$

It is easy to see that at least one of the stopping rules defined by $(t_1(i), \cdots, t_n(i))$, $i = 1, \cdots, n+1$ must give a value of c_0 for the case n which is less than that given by $(s_1, \cdots, s_n, n+1)$ for the case $n+1$. Hence $c_0(n) < c_0(n+1)$.

2. We assumed that the n girls appear in random order, all $n!$ permutations being equally likely. The minimal expected absolute rank of the girl chosen is then $c_0 < 3.87$ for all n. Suppose now that the order in which the girls are to appear is determined by an opponent who wishes to maximize the expected absolute rank of the girl we choose. No matter what he does, by choosing at random the first, second, $\cdots$, last girl to appear we can achieve the value

$$EX = (1 + 2 + \cdots + n)/n = (n+1)/2.$$

And in fact *there exists an opponent strategy such that, no matter what stopping rule we use,* $EX = (n+1)/2$. Let $x_1 = 1$ or n, each with probability $1/2$; let x_{i+1} = either the largest or the smallest of the integers remaining after $x_1, \cdots, x_i$ have been chosen, each with probability $1/2$. If we define for $i = 1, \cdots, n$

$$z_i = E(x_i \mid y_1, \cdots, y_i), \quad \mathscr{B}_i = \mathscr{B}(y_1, \cdots, y_i), \tag{4}$$

then it is easy to see that $\{z_i, \mathscr{B}_i (i = 1, \cdots, n)\}$ is a martingale, so that for any stopping rule τ,

$$E(X) = E(z_\tau) = E(z_1) = (n+1)/2. \tag{5}$$

Many extensions and generalizations of the problem considered in this paper suggest themselves at once. Some further results will be presented elsewhere.

Reference

1. D. V. Lindley, *Dynamic programming and decision theory*, Applied Statistics **10** (1961), 39–51.

Purdue University
University of Tokyo
Columbia University
Purdue University