

SHARP INEQUALITIES FOR OPTIMAL STOPPING WITH REWARDS BASED ON RANKS

BY T. P. HILL¹ AND D. P. KENNEDY

Georgia Institute of Technology and University of Cambridge

A universal bound for the maximal expected reward is obtained for stopping a sequence of independent random variables where the reward is a nonincreasing function of the rank of the variable selected. This bound is shown to be sharp in three classical cases: (i) when maximizing the probability of choosing one of the k best; (ii) when minimizing the expected rank; and (iii) for an exponential function of the rank.

1. Introduction. For every finite sequence of independent random variables there is a stopping time which stops at the maximum value with probability at least $1/e$, one which stops with one of the two largest values with probability at least $e^{-\sqrt{2}}(1 + \sqrt{2})$ and, in general, one which stops with one of the k largest values with probability at least $p(k) = \exp(-\binom{k!}{k}) \sum_{r=0}^{k-1} \{k!\}^r / k^r / r!$. These bounds are best possible, and follow from the main result of this paper (Theorem 1), which gives a universal bound for the maximal expected reward for stopping a sequence of random variables when the reward is a nonincreasing function of the rank of the variable selected.

Let X_1, X_2, \dots be a sequence of independent random variables and denote by \mathcal{S} the set of positive integer-valued stopping times relative to the natural filtration $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots$, where $\mathcal{F}_r = \sigma(X_1, X_2, \dots, X_r)$. Let \mathcal{S}_n be the subset of stopping times taking values in $\{1, 2, \dots, n\}$. For each $n \geq 1$ and $r = 1, 2, \dots, n$, take M_r^n to be the r th largest order statistic among X_1, X_2, \dots, X_n , so that

$$M_r^n = \bigvee_{1 \leq i_1 < \dots < i_r \leq n} \bigwedge_{j=1}^r X_{i_j}.$$

The rank of X_k among X_1, X_2, \dots, X_n is defined to be

$$R_k^n = \min\{r: X_k = M_r^n\}, \quad \text{for } k = 1, 2, \dots, n.$$

When considering stopping times in \mathcal{S} , to simplify the notation, set $R_k^n = 0$ for $k > n$. Notice that if two (or more) values tie, then the rank is taken as the smaller for each; for example, if two random variables are largest, both have rank 1 (see Remark 1.5). The object of the present paper is to obtain bounds on the optimal reward for the problem of choosing a stopping time $T \in \mathcal{S}$ or

Received May 1990; revised January 1991.

¹Partially supported by NSF Grants DMS-87-01691 and 89-01267, a Fulbright Research Grant and a travel grant from the U.S.–U.K. Educational Commission.

AMS 1980 subject classification. Primary 60G40.

Key words and phrases. Optimal stopping, rank, sharp inequalities, best choice problem, secretary problem, prophet inequality, Schur convexity.

$T \in \mathcal{S}_n$ so as to maximize $Ef_n(R_T^n)$ for functions $f_n(r)$ which are nonincreasing in r , $1 \leq r \leq n$. One may think of a gambler with a horizon n wishing to select one of the random variables so as to maximize the expected value of a nonincreasing function of the rank of the random variable chosen from among the first n . Here, for each $n \geq 1$, take $f_n: \{0, 1, 2, \dots, n\} \rightarrow \mathbb{R}_+$, with $f_n(1) \geq f_n(2) \geq \dots \geq f_n(n) = 0$. The assumption that $f_n(n) = 0$ is only for convenience; clearly any nonincreasing function \hat{f}_n may be reduced to this case by taking $f_n(r) = \hat{f}_n(r) - \hat{f}_n(n)$. The inclusion of $f_n(0)$ is to allow for the possibility of not selecting one of the n random variables; this corresponds to choosing a $T \in \mathcal{S}$ with $P(T > n) > 0$. For $T \in \mathcal{S}$, if $f_n(0) \geq f_n(1)$, clearly it is optimal if the horizon is n to take $T \equiv n + 1$. Similarly, if $f_n(0) \leq f_n(n)$, it is optimal to take $P(T \leq n) = 1$ and so $T \in \mathcal{S}_n$; that is, if $f_n(0) \leq f_n(n)$,

$$\sup_{T \in \mathcal{S}} Ef_n(R_T^n) = \sup_{T \in \mathcal{S}_n} Ef_n(R_T^n).$$

Hence the interesting case is when $f_n(1) \geq f_n(0) \geq f_n(n)$, which will be assumed throughout this paper.

The present setup may be seen to be related to the classical secretary problem [cf. Freeman (1983) and Ferguson (1989)] as follows. The gambler may choose one of the n items which appear in random order, each order being equally likely. When item r is viewed, only its rank among the first r items is observed. Let Y_1, \dots, Y_k be a random permutation of $1, \dots, n$ and denote by Z_k the number of Y_1, \dots, Y_k not exceeding Y_k . For $k = 1, \dots, n$, define $X_k = k$ if $Z_k = 1$ and $X_k = -Z_k$ if $Z_k \geq 2$. Then, since Z_1, \dots, Z_n are independent random variables, X_1, \dots, X_n are independent. Furthermore, $R_k^k = 1$ if and only if $Z_k = 1$, and $R_k^n = 1$ if and only if $Y_k = 1$. The secretary problem corresponds to choosing a stopping time $T \leq n$ relative to the sequence Z_1, \dots, Z_n so as to maximize $P(Y_T = 1)$, but this is clearly equivalent to choosing a stopping time T relative to X_1, \dots, X_n so as to maximize $Ef_n(R_T^n)$ where $f_n(1) = 1$, $f_n(r) = 0$ for $r \neq 1$. Thus the secretary problem is a particular case of the class of optimal stopping problems considered in this paper. However, in general, extensions of the secretary problem involving ranks greater than 1 (e.g., choosing one of the best k items, $k > 1$) cannot be reduced to the present context (see Remark 1.1).

The principal result of this paper is the following theorem.

THEOREM 1.1. *Suppose that $f_n(1) \geq f_n(2) \geq \dots \geq f_n(n) = 0$ and that $f_n(1) \geq f_n(0) \geq f_n(n)$. For all independent random variables X_1, \dots, X_n , there is a stopping time $T \in \mathcal{S}$ satisfying*

$$(1) \quad Ef_n(R_T^n) \geq \sup_{0 \leq p \leq 1} \left[\sum_{r=0}^{n-1} f_n(r) \binom{n}{r} p^r (1-p)^{n-r} \right].$$

When the random variables are continuous it will be shown that the stopping time T in the statement of Theorem 1.1 may be taken to be a threshold stopping time; that is, stop at the first random variable that exceeds a fixed level. If the random variables are not continuous the proof of the

existence of such a T is nonconstructive. By taking the derivative with respect to p of the summation in the right-hand side of (1), it follows that if $f_n(1) > 0$, then the supremum is attained by the unique p , $0 < p < 1$, satisfying

$$(2) \quad f_n(1) - f_n(0) = \sum_{r=1}^{n-1} \{f_n(r) - f_n(r+1)\} \binom{n-1}{r} \left(\frac{p}{1-p}\right)^r.$$

The uniqueness of p satisfying (2) follows by observing that the right-hand side is strictly increasing in p , $0 < p < 1$. Appropriate choices of f_n lead to the inequalities in the following three theorems; the sharpness claims, however, require separate proofs.

THEOREM 1.2 (Best k of n). *For all independent random variables X_1, \dots, X_n , there is a stopping time $T \in \mathcal{S}_n$ satisfying*

$$(3) \quad P(X_T \geq M_k^n) \geq \frac{\sum_{r=1}^k \binom{n}{r} \binom{n-1}{k}^{-r/k}}{\left[1 + \left\{\binom{n-1}{k}\right\}^{-1/k}\right]^n},$$

and this bound is sharp.

Taking the limit of the bound in (3) as $n \rightarrow \infty$ yields the inequality stated in the first paragraph, since the sharpness of the bounds implies that they are monotone decreasing.

THEOREM 1.3 (Expected rank). *For all independent random variables X_1, \dots, X_n there is a stopping time $T \in \mathcal{S}_n$ satisfying*

$$(4) \quad E(R_T^n) \leq (n-1)[1 - n^{-1/(n-1)}] + 1,$$

and this bound is sharp.

Observe that if C_n denotes the bound on the right-hand side of (4), then $(C_n - \log n) \rightarrow 1$ as $n \rightarrow \infty$.

THEOREM 1.4 (Exponential rank function). *For all independent random variables X_1, \dots, X_n and $0 < z < 1$ there is a stopping time $T \in \mathcal{S}_n$ satisfying*

$$(5) \quad E[z^{R_T^n}] \geq \frac{z(1-z^n)}{1-z} \left[1 + \frac{1}{z} \left\{\left(\frac{1-z^n}{1-z}\right)^{1/(n-1)} - 1\right\}\right]^{-(n-1)},$$

and this bound is sharp.

As $n \rightarrow \infty$, the bound in (5) approaches $z(1-z)^{(1-z)/z}$. To illustrate this bound, consider the particular case where the reward structure is such that stopping on the best random variable yields 1, on the second best yields 1/2, and the third best yields 1/4 and so on. Taking $z = 1/2$, for large n the last inequality shows that the optimal expected reward is bounded below by 1/2.

TABLE 1

k	1	2	3	4	5	6	...	12
$p(k)$	0.368	0.587	0.726	0.817	0.877	0.917	...	0.992
$q(k)$	0.368	0.574	0.708	0.799	0.860	0.903	...	0.988

The proof of Theorem 1.1 will be given in Section 2 and the proofs of Theorems 1.2–1.4 will be given in Section 3. It is not difficult to show that the bound in (1) is sharp for all $n \leq 4$ for all f_n . An example and proof are given in Section 4 to show that the bound in (1) is not sharp for $n = 5$ and the rank function $f_5(1) = 2, f_5(2) = f_5(3) = f_5(4) = 1$ and $f_5(0) = f_5(5) = 0$. Surprisingly enough, such an example to demonstrate that the bound in (1) is not sharp in general seems to be difficult to construct, and even in the example mentioned previously, the true bound is very close to the general bound. This suggests that the bound in (1) may be fairly sharp for a large class of f_n .

REMARK 1.1. The bound $p(1) = e^{-1}$ obtained from (3) in the limit as $n \rightarrow \infty$ in the case $k = 1$ is familiar from the classical secretary problem as the limiting probability of choosing the best item using an optimal policy as the number of items n tends to infinity. This observation shows that the secretary problem behaves asymptotically like the worst case when $f_n(1) = 1, f_n(r) = 0$ for $r > 1$. This is not true for $k > 1$, the case of choosing one of the best k . Here the limiting bound $p(k)$ differs from the limit of the optimal probability $q(k)$ [cf. Frank and Samuels (1980)] of choosing one of the k best in the classical problem. It should also be noted that $p(k) \rightarrow 1$ very quickly as k increases, as Table 1 illustrates. As Table 1 also suggests, since $p(k) > q(k)$ for $k > 1$, there is some slight advantage in knowing the *distributions* of the X_i , as opposed to knowing only the ranks.

REMARK 1.2. Sakaguchi (1984) considers the following variation of the secretary problem. If the gambler selects the best item from $n (\geq 2)$, he receives 1 unit; if he selects any but the best, he pays 1 unit, and if he opts not to select an item, he receives or pays nothing. The last option corresponds to stopping at a time that exceeds n . A corresponding generalization of the bound in Theorem 1.2 may be derived from Theorem 1.1. For any $\delta, 0 \leq \delta \leq 1$, it follows that for all independent random variables X_1, \dots, X_n , there is a stopping time $T \in \mathcal{L}$ satisfying

$$\begin{aligned}
 & P(X_T \geq M_k^n, T \leq n) + (1 - \delta)P(T > n) \\
 (6) \quad & \geq \frac{1 - \delta + \sum_{r=1}^k \binom{n}{r} \left\{ \delta / \binom{n-1}{k} \right\}^{r/k}}{\left[1 + \left\{ \delta / \binom{n-1}{k} \right\}^{1/k} \right]^n},
 \end{aligned}$$

and this bound is sharp. Notice that the limit of the bound in (6) as $n \rightarrow \infty$ is

$$\exp(-\{k!\delta\}^{1/k}) \sum_{r=0}^{k-1} \frac{\{k!\delta\}^{r/k}}{r!}.$$

This is seen to be an extension of Sakaguchi's case by taking $\delta = 1/2$ (and considering the reward/loss function $2f_n(r) - 1$); then the gambler wins 1 if he selects one of the k best of the random variables, he loses 1 if he selects one of the $n - k$ worst and he receives or pays nothing if he does not select. From (6), in the case $k = 1$, it follows that

$$\limsup_{n \rightarrow \infty} \sup_{T \in \mathcal{J}} [P(R_T^n = 1) - P(R_T^n \geq 2)] \geq \frac{2}{\sqrt{e}} - 1.$$

Comparing this last bound with the result in Sakaguchi (1984) shows that again in this situation the secretary problem behaves asymptotically like the worst case.

REMARK 1.3. The case of (3) when $k = 1$ provides an interesting bound related to the prophet inequality [Krengel and Sucheston (1978)], which establishes that for independent nonnegative random variables X_1, \dots, X_n ,

$$\sup_{T \in \mathcal{J}_n} EX_T \geq \frac{1}{2} E\left(\max_{1 \leq r \leq n} X_r\right),$$

and that $1/2$ is the best possible bound for each $n \geq 2$. It follows from (3) that for nonnegative, independent random variables

$$(7) \quad \sup_{T \in \mathcal{J}_n} E\left(\frac{X_T}{\max_{1 \leq r \leq n} X_r}\right) \geq \sup_{T \in \mathcal{J}_n} P\left(X_T = \max_{1 \leq r \leq n} X_r\right) \\ \geq \left(1 - \frac{1}{n}\right)^{n-1} \quad \text{for each } n \geq 1.$$

Here, interpret $0/0$ as 1. It will be seen that the bound in (7) is sharp for each n (where the bound is taken to be 1 for $n = 1$). Notice that the limit of the bound in (7) as $n \rightarrow \infty$ is $e^{-1} < 1/2$, which contrasts with the standard prophet inequality.

REMARK 1.4. Note that the proof of Theorem 1.1 shows the existence of a *threshold* stopping time T satisfying (1) for continuous random variables and that the threshold c may be determined explicitly from the distributions of the random variables. In the case where the random variables are i.i.d. (and continuous), there is even a threshold stopping time which stops with the maximum observation with probability at least 0.517 as was shown by Gilbert and Mosteller [(1966), page 57].

REMARK 1.5. The definition of the rank given above is a “generous” one, in that ties always move up, and the result (1) depends heavily on this. If ties move down, for example, then there is no nontrivial analog of (1), since the case $X_1 = X_2 = \dots = X_n \equiv 1$ would yield $E[f_n(R_T^n)] = f_n(n) = 0$ for all $T \in \mathcal{S}_n$. Similarly, if an averaging definition of rank is used [e.g., the reward is $(f_n(1) + f_n(2))/2$ if the value selected is tied with one other value for the maximum], then the best lower bound is easily seen to be $(f_n(1) + \dots + f_n(n))/n$ (which is attained if $X_1 = X_2 = \dots = X_n \equiv 1$), and this is also the best lower bound for arbitrary dependent random variables (via the randomized stopping time “stop with probability $1/n$ at time i , independently of the X process”). The best lower bound for the arbitrarily dependent case under the definition of relative rank used in this paper is not known for general objective functions f_n , although for the best-choice problem $f_n(1) = 1, f_n(i) = 0$ for $i \neq 1$, it is easily seen to be $1/n$.

2. Proof of Theorem 1.1.

LEMMA 2.1. For fixed real numbers $a \geq a_2 \geq \dots \geq a_n \geq 0$, the function

$$\sum_{r=1}^n a_r \left[\sum_{1 \leq i_1 < \dots < i_r \leq n} \prod_{j=1}^r (x_{i_j} - 1) \right]$$

is minimized over $x_r \geq 1, 1 \leq r \leq n$, with $\prod_1^n x_r = b > 1$ when $x_1 = \dots = x_n = b^{1/n}$.

PROOF. First, it will be shown that the function

$$\Phi(\mathbf{y}) = \sum_{r=1}^n a_r \left[\sum_{1 \leq i_1 < \dots < i_r \leq n} \prod_{j=1}^r (e^{y_{i_j}} - 1) \right]$$

is Schur convex for $\mathbf{y} = (y_1, \dots, y_n) \in \mathbb{R}_+^N$. [For the definition and properties of Schur convexity, see Marshall and Olkin (1979), in particular, page 54 and Theorem A4 of page 57.] Let

$$\phi_r^n(\mathbf{y}) = \sum_{1 \leq i_1 < \dots < i_r \leq n} \prod_{j=1}^r (e^{y_{i_j}} - 1), \quad r = 1, \dots, n,$$

with $\phi_0^n \equiv 1, \phi_{-1}^n \equiv 0$ and $\phi_r^n \equiv 0$ for $r > n$. For the n -vector $\mathbf{y} = (y_1, \dots, y_n)$ let

$$\hat{\mathbf{y}}_i = (y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n)$$

be the $(n - 1)$ -vector obtained by dropping the i th component and let $\hat{\mathbf{y}}_{ij}$ be the $(n - 2)$ -vector obtained by dropping the i th and j th components, $i \neq j$.

Observe that

$$\frac{\partial \phi_r^n}{\partial y_i} = e^{y_i} \phi_{r-1}^{n-1}(\hat{y}_i)$$

and

$$\phi_{r-1}^{n-1}(\hat{y}_i) = \phi_{r-1}^{n-2}(\hat{y}_{ij}) + \phi_{r-2}^{n-2}(\hat{y}_{ij})(e^{y_j} - 1), \text{ for } 1 \leq r \leq n.$$

This gives

$$\frac{\partial \phi_r^n}{\partial y_i} - \frac{\partial \phi_r^n}{\partial y_j} = (e^{y_i} - e^{y_j}) [\phi_{r-1}^{n-2}(\hat{y}_{ij}) - \phi_{r-2}^{n-2}(\hat{y}_{ij})],$$

implying that

$$\begin{aligned} \frac{\partial \Phi}{\partial y_i} - \frac{\partial \Phi}{\partial y_j} &= (e^{y_i} - e^{y_j}) \sum_{r=1}^n a_r [\phi_{r-1}^{n-2}(\hat{y}_{ij}) - \phi_{r-2}^{n-2}(\hat{y}_{ij})] \\ &= (e^{y_i} - e^{y_j}) \sum_{r=1}^{n-1} (a_r - a_{r+1}) \phi_{r-1}^{n-2}(\hat{y}_{ij}). \end{aligned}$$

This, in turn, implies that

$$(y_i - y_j) \left(\frac{\partial \Phi}{\partial y_i} - \frac{\partial \Phi}{\partial y_j} \right) \geq 0, \text{ for } i \neq j,$$

which by Theorem A4 of Marshall and Olkin (1979) proves that Φ is Schur convex. Schur convexity, on the other hand, implies that

$$\Phi((\bar{y}, \dots, \bar{y})) \leq \Phi((y_1, \dots, y_n))$$

where $\bar{y} = (y_1 + \dots + y_n)/n$, which completes the proof. \square

PROOF OF THEOREM 1.1. Without loss of generality assume that $f_n(1) > 0$. Let F_i denote the distribution of X_i , and suppose that p is the unique root of (2), $0 \leq p < 1$. First consider the case where there exists c with $\prod_{i=1}^n F_i(c) = (1 - p)^n > 0$. Define the pure-threshold stopping time $T(c) = \min\{k \leq n: X_k > c\}$, with $T(c) = n + 1$ if there is no such k . If exactly $r \geq 1$ of the random variables X_1, \dots, X_n exceed c , then $1 \leq R_{T(c)}^n \leq r$, from which it follows that

$$P(1 \leq R_{T(c)}^n \leq k) \geq \left[\prod_{i=1}^n F_i(c) \right] \sum_{r=1}^k \left\{ \sum_{1 \leq i_1 < \dots < i_r \leq n} \prod_{j=1}^r \left(\frac{1}{F_{i_j}(c)} - 1 \right) \right\}$$

and

$$P(R_{T(c)}^n = 0) = \prod_{i=1}^n F_i(c).$$

Using these relations,

$$\begin{aligned}
 Ef_n(R_{T(c)}^n) &= f_n(0)P(R_{T(c)}^n = 0) + \sum_{k=1}^{n-1} [f_n(k) - f_n(k+1)]P(1 \leq R_{T(c)}^n \leq k) \\
 &\geq \left[\prod_{i=1}^n F_i(c) \right] \left[f_n(0) + \sum_{r=1}^{n-1} f_n(r) \right. \\
 &\qquad \qquad \qquad \times \left. \left\{ \sum_{1 \leq i_1 < \dots < i_r \leq n} \prod_{j=1}^r \left(\frac{1}{F_{i_j}(c)} - 1 \right) \right\} \right] \\
 &\geq \sum_{r=0}^{n-1} f_n(r) \binom{n}{r} p^r (1-p)^{n-r},
 \end{aligned}$$

the last inequality following by Lemma 2.1, and establishing (1).

Now suppose that there does not exist c with $\prod_{i=1}^n F_i(c) = (1-p)^n$, but take c satisfying $\prod_{i=1}^n F_i(c-) \leq (1-p)^n < \prod_{i=1}^n F_i(c)$. Let $\{U_r, 1 \leq r \leq n\}$ be independent random variables each with the uniform distribution on $[0, 1]$ and independent of $\{X_r, 1 \leq r \leq n\}$ (defined on an enlarged probability space, if necessary). Define independent random variables $\hat{X}_r, r = 1, \dots, n$, by $\hat{X}_r = X_r$ if $X_r < c$, $\hat{X}_r = X_r + 1$ if $X_r > c$ and $\hat{X}_r = X_r + U_r$ if $X_r = c$. Then, if \hat{F}_r is the distribution function of \hat{X}_r , there exists $d, c \leq d < c + 1$ with $\prod_{i=1}^n \hat{F}_i(d) = (1-p)^n$. Let \hat{R}_k^n denote the relative rank of \hat{X}_k among $\hat{X}_1, \dots, \hat{X}_n$; it is immediate that $\hat{R}_k^n \geq R_k^n, k = 1, \dots, n$. Let $\hat{T}(d) = \min\{k \leq n: \hat{X}_k > d\}$, with $\hat{T}(d) = n + 1$ if there is no such n . Observe that $\hat{T}(d) \in \mathcal{S}$, where $\mathcal{S} \supseteq \mathcal{S}$ is the class of randomized stopping times for X_1, X_2, \dots . By the previous argument applied to $\hat{X}_1, \dots, \hat{X}_n$,

$$Ef_n(R_{\hat{T}(d)}^n) \geq Ef_n(\hat{R}_{\hat{T}(d)}^n) \geq \sum_{r=0}^{n-1} f_n(r) \binom{n}{r} p^r (1-p)^{n-r}.$$

Observing that $\sup_{T \in \mathcal{S}} Ef_n(R_T^n) = \sup_{T \in \mathcal{S}} f_n(R_T^n)$, and that for finite-horizon optimal stopping an optimal stopping time always exists, completes the proof of the theorem. \square

3. Proofs of Theorems 1.2-1.4. In establishing the sharpness of the inequalities derived in the theorems of Section 1 it will be seen that the extremal distributions for $\{X_1, \dots, X_n\}$ take the same form in each case. Say that (X_1, \dots, X_n) form a *Bernoulli pyramid* with parameter $p, 0 \leq p \leq 1$, if $X_1 \equiv 1$ and $P(X_r = r) = p = 1 - P(X_r = 1/r)$ for $r = 2, \dots, n$. For each choice of f_n define

$$V_r(x) = \sup_{T \geq r} E[f_n(R_T^n) | X_r = x],$$

so that $V_r(x)$ is the optimal expected reward if stopping takes place at time r , or later, conditional on the observed value $X_r = x$. In each of the three cases it will be seen that an optimal stopping time is always to stop at time 1; and in

each case the gambler is indifferent between stopping at time 1 and continuing.

PROOF OF THEOREM 1.2. The proof of the more general sharp inequality (6) will be established, which implies the result for (3) by taking $\delta = 1$. Here, take $f_n(1) = \dots = f_n(k) = 1$, $f_n(k + 1) = \dots = f_n(n) = 0$ and $f_n(0) = 1 - \delta$. Solving (2) for p yields

$$p = \left\{ 1 + \left[\binom{n-1}{k} / \delta \right]^{1/k} \right\}^{-1}.$$

Substituting into (1) gives (6). For the sharpness, suppose that (X_1, \dots, X_n) is a Bernoulli pyramid with parameter p where p is as before, so that $\delta = \binom{n-1}{k}(p/(1-p))^k$. Let a_r be the conditional probability that X_r is among the k best values given that $X_r = r$; likewise, let c_r be the conditional probability that X_r is among the k best given that $X_r = 1/r$. It is clear that

$$a_r = \sum_{i=0}^{(n-r) \wedge (k-1)} \binom{n-r}{i} p^i (1-p)^{n-r-i} \quad \text{for } 1 \leq r \leq n$$

and

$$c_r = \sum_{i=0}^{k-r} \binom{n-r}{i} p^i (1-p)^{n-r-i} \quad \text{for } 1 \leq r \leq k,$$

with $c_r = 0$ for $r > k$. Note that $a_1 = c_1$. For $1 \leq r < n$,

$$V_r(r) = \max\{a_r, pV_{r+1}(r+1) + (1-p)V_{r+1}(1/(r+1))\}$$

and

$$V_r(1/r) = \max\{c_r, pV_{r+1}(r+1) + (1-p)V_{r+1}(1/(r+1))\},$$

with $V_n(n) = \max\{1, 1 - \delta\} = 1$, $V_n(1/n) = \max\{0, 1 - \delta\} = 1 - \delta$. Last, define b_r , $1 \leq r \leq n$, setting $b_n = 1 - \delta$, and then letting $b_r = pa_{r+1} + (1-p)b_{r+1}$ for $1 \leq r < n$, from which it follows that for $r < n$,

$$b_r = \sum_{i=1}^{(n-r) \wedge k} \binom{n-r}{i} p^i (1-p)^{n-r-i} + (1-\delta)(1-p)^{n-r}.$$

It will be established that $a_r \geq b_r \geq c_r$, for each r , which implies by backward induction on $r = n, n - 1, \dots, 1$ that $V_r(r) = a_r$ and $V_r(1/r) = b_r$. This gives $V_1(1) = a_1$, which in turn demonstrates the sharpness of the bound (6) when it is observed that the right-hand side of (6) is a_1 , since

$$a_1 = \sum_{i=0}^{k-1} \binom{n-1}{i} p^i (1-p)^{n-i-1} = (1-\delta)(1-p)^n + \sum_{i=1}^k \binom{n}{i} p^i (1-p)^{n-i}.$$

The last relation follows using the definition of p and the identity

$$\binom{n}{i} = \binom{n-1}{i-1} + \binom{n-1}{i}.$$

First, to check that $b_r \geq c_r$, it is only necessary to consider $1 \leq r \leq k$. Since $k - r < (n - r) \wedge k$, it is then sufficient to show that

$$(1 - p)^{n-r} \delta \leq \binom{n-r}{k-r+1} p^{k-r+1} (1-p)^{n-k-1}.$$

This is trivial for $\delta = 0$, and for $1 \geq \delta > 0$, after substituting the value of p , it reduces to showing that

$$\left[\frac{1}{\delta} \binom{n-1}{n-k-1} \right]^{1/k} \leq \left[\frac{1}{\delta} \binom{n-r}{n-k-1} \right]^{1/(k-r+1)};$$

this inequality is true because $\binom{n}{k}^{1/(n-k)}$ is decreasing in $n \geq k + 1$ and $\delta^{-1/k}$ is nonincreasing in $k \geq 1$. To check $a_r \geq b_r$, note that this inequality is immediate for $n - r \leq k - 1$, while for $n - r \geq k$, using the expression for p ,

$$\begin{aligned} a_r - b_r &= \delta(1 - p)^{n-r} - \binom{n-r}{k} p^k (1-p)^{n-r-k} \\ &= p^k (1-p)^{n-r-k} \left[\binom{n-1}{k} - \binom{n-r}{k} \right] \geq 0. \end{aligned}$$

To verify the sharpness of the inequality in (7), consider the slight variation of the preceding example, where $X_1 \equiv 1$, and $P(X_r = \rho^{r-1}) = 1/n = 1 - P(X_r = 1/\rho^{r-1})$, $r = 2, \dots, n$, where $\rho > 1$. Then it is immediate that

$$\left(1 - \frac{1}{n}\right)^{n-1} \leq \sup_{T \in \mathcal{S}_n} E \left(\frac{X_T}{\max_{1 \leq r \leq n} X_r} \right) \leq \sup_{T \in \mathcal{S}_n} P(X_T = \max_{1 \leq r \leq n} X_r) + \frac{1}{\rho}.$$

But the same argument as previously given shows that

$$\sup_{T \in \mathcal{S}_n} P(X_T = \max_{1 \leq r \leq n} X_r) = \left(1 - \frac{1}{n}\right)^{n-1},$$

which proves that (7) is sharp when $\rho \rightarrow \infty$. \square

PROOF OF THEOREM 1.3. Taking $f_n(r) = n - r$, for $1 \leq r \leq n$, and $= 0$ otherwise, solution of (2) yields $p = 1 - n^{-1/(n-1)}$. Substituting into (1) and converting $E(R_T^n) = n - E[f_n(R_T^n)]$ gives (4). To establish the sharpness of (4), take a Bernoulli pyramid (X_1, \dots, X_n) with parameter p satisfying $(1 - p)^{-(n-1)} = n$. For $1 \leq r < n$, since $(n - r)p$ is the expected number of indices j , $r < j \leq n$, with $X_j = j$, it follows that

$$V_r(r) = \max\{n - (n - r)p - 1, pV_{r+1}(r + 1) + (1 - p)V_{r+1}(1/(r + 1))\}$$

and

$$V_r(1/r) = \max\{n - (n - r)p - r, pV_{r+1}(r + 1) + (1 - p)V_{r+1}(1/(r + 1))\},$$

with $V_n(n) = n - 1$ and $V_n(1/n) = 0$. First observe that

$$1 \leq n(1 - p)^{n-r} \leq r, \quad \text{for } 1 \leq r \leq n.$$

For the left-hand inequality $n(1-p)^{n-r} = (1-p)^{-(r-1)} \geq 1$ and for the right-hand side $n(1-p)^{n-r} = n^{(r-1)/(n-1)}$, but $n^{1/(n-1)} \leq r^{1/(r-1)}$, since $(1+x)^{1/x}$ is decreasing in $x > 0$, and the inequality follows. Now letting

$$b_r = pV_{r+1}(r+1) + (1-p)V_{r+1}(1/(r+1)) \quad \text{for } 1 \leq r < n,$$

it follows by backward induction on r that

$$b_r = n - (n-r)p - n(1-p)^{n-r}$$

and $V_r(r) = n - (n-r)p - 1$ and $V_r(1/r) = b_r$. Again, notice that $b_1 = (n-1)(1-p)$. It follows that $V_1(1) = (n-1)/n^{1/(n-1)}$, showing that (4) is sharp. \square

PROOF OF THEOREM 1.4. Fix $z \in (0, 1)$ and define $f_n(r) = z^r - z^n$ for $1 \leq r \leq n$, and $= 0$ otherwise. Solving (2) yields

$$p = \left(z \left[\left(\frac{1-z^n}{1-z} \right)^{1/(n-1)} - 1 \right]^{-1} + 1 \right)^{-1}$$

and substitution into (1) (recalling that z^n was subtracted off) gives (5). To show that (5) is sharp, again take a Bernoulli pyramid with p as before, so that

$$\frac{1-z^n}{1-z} = \left\{ 1 + z \left(\frac{p}{1-p} \right) \right\}^{n-1}.$$

First note the inequalities, for $1 \leq r \leq n$,

$$\begin{aligned} z\{1 - (1-z)p\}^{n-r} &\geq \{1 - (1-z)p\}^{n-r} - (1-p)^{n-r}(1-z^n) \\ &\geq z^r\{1 - (1-z)p\}^{n-r}. \end{aligned}$$

By rearrangement, the first of these inequalities may be demonstrated by showing that

$$\frac{1-z^n}{1-z} \geq \left\{ 1 + z \left(\frac{p}{1-p} \right) \right\}^{n-r},$$

but this is true, using the equation for p , since $((1-z^n)/(1-z)) \geq 1$. For $r = 1$ the second inequality is immediate, while for $r > 1$ it reduces to showing that

$$\left(\frac{1-z^r}{1-z} \right)^{1/(r-1)} \geq \left(\frac{1-z^n}{1-z} \right)^{1/(n-1)}, \quad \text{for } r = 2, \dots, n,$$

which holds since the left-hand side is nonincreasing in $r = 2, \dots, n$. Now, as in the previous two examples, set $b_r = pV_{r+1}(r+1) + (1-p)V_{r+1}(1/(r+1))$ for $r < n$, and since $\binom{n-r}{i} p^i (1-p)^{n-r-i}$ represents the probability that for exactly i of the indices $j = r+1, \dots, n$, the random variable $X_j = j$, it follows

that

$$V_r(r) = \max \left\{ \sum_{i=0}^{n-r} z^{i+1} \binom{n-r}{i} p^i (1-p)^{n-r-i} - z^n, b_r \right\}$$

and

$$V_r(1/r) = \max \left\{ \sum_{i=0}^{n-r} z^{i+r} \binom{n-r}{i} p^i (1-p)^{n-r-i} - z^n, b_r \right\},$$

with $V_n(n) = z - z^n$, $V_n(1/n) = 0$. Backward induction now gives

$$b_r = \{1 - (1 - z)p\}^{n-r} - (1 - p)^{n-r} (1 - z^n) - z^n \text{ and } V_r(1/r) = b_r.$$

This shows that $V_r(r) = z\{1 - (1 - z)p\}^{n-r} - z^n$ and $V_r(1/r) = b_r$. Again note from the relation giving p that $b_1 = z\{1 - (1 - z)p\}^{n-1} - z^n$ and that $V_1(1) + z^n$ is the bound on the right-hand side of (5), establishing the sharpness. \square

4. An example where the inequality (1) is not sharp. Take $n = 5$ and define $f_5(1) = 2$, $f_5(2) = f_5(3) = f_5(4) = 1$ and $f_5(5) = f_5(0) = 0$ (so that $\hat{f}_5 = f_5 - 1$ is $+1$ for rank 1, -1 for rank 5 and 0 otherwise). Let B denote the bound on the right-hand side of (1), so that

$$\begin{aligned} B &= \sum_{r=0}^4 f_5(r) \binom{5}{r} p^r (1-p)^{5-r} \\ &= 5p(1-p)^4 + 1 - p^5 - (1-p)^5, \end{aligned}$$

where p is the unique solution in $[0, 1)$ of the equation

$$(8) \quad 4 \left(\frac{p}{1-p} \right) + \left(\frac{p}{1-p} \right)^4 = 2.$$

Observe that $p < 1/3$ since the left-hand side of (8) is increasing in p , $0 \leq p < 1$, and exceeds 2 when $p = 1/3$.

PROPOSITION 4.1.

$$\inf \left\{ \sup_{T \in \mathcal{S}_5} E[f_5(R_T^5)]: X_1, \dots, X_5 \text{ independent} \right\} > B.$$

PROOF. Define functions $\psi, \phi: [0, 1]^5 \rightarrow \mathbb{R}$, by

$$\psi(q_1, \dots, q_5) = 1 + \sum_{j=1}^5 \left((1 - q_j) \prod_{\substack{i \neq j \\ i=1}}^5 q_i \right) - \prod_{i=1}^5 (1 - q_i) - \prod_{i=1}^5 q_i,$$

$$\phi(q_1, \dots, q_5) = 1 + \sum_{j=1}^4 \left((1 - q_j) \prod_{\substack{i \neq j \\ i=1}}^5 q_i \right) - \prod_{i=1}^5 (1 - q_i) - (1 - q_5) \prod_{i=1}^4 q_i.$$

For q_1, \dots, q_5 satisfying the constraint $\prod_1^5 q_i = (1 - p)^5$, observe that $\prod_1^5(1 - q_i)$ and $\sum_1^5 1/q_i$ are uniquely maximized and minimized, respectively, when all the q_i are equal, so that under the constraint,

$$(9) \quad \psi(q_1, \dots, q_5) = 1 + \left(\sum_{j=1}^5 \left(\frac{1}{q_j} - 1 \right) \right) \prod_{i=1}^5 q_i - \prod_{i=1}^5 (1 - q_i) - \prod_{i=1}^5 q_i$$

$$\geq \psi(1 - p, \dots, 1 - p) = B.$$

Also, setting $B^* = \phi(1 - p, \dots, 1 - p)$, note that

$$B^* = \psi(1 - p, \dots, 1 - p) + (1 - p)^4(1 - 3p) > B,$$

using the observation that $p < 1/3$. Now, if F_i denotes the distribution function of X_i , assume that there exists c with $\prod_1^5 F_i(c) = (1 - p)^5$. (For the general case use randomization exactly as in the proof of Theorem 1.1.) As before, define the pure-threshold stopping time $T(c) = \min\{1 \leq k \leq 5: X_k > c\}$ or $T(c) = 6$ if no such k exists. Then set $T^* = T(c) \wedge 5$, $T^{**} = T(c) \wedge 4$, and define the event $A = \{X_4 < X_5 \leq c\}$.

CLAIM 1.

$$E[f_5(R_{T^*}^5)] \geq \psi(F_1(c), \dots, F_5(c)) + F_1(c)F_2(c)F_3(c)P(A).$$

Claim 1 follows by noting that

$$P(R_{T^*}^5 = 5) \leq \prod_{i=1}^5 (1 - F_i(c)) + P\left(\bigvee_{i=1}^4 X_i \leq c, X_5 < \bigwedge_{i=1}^4 X_i\right)$$

$$\leq \prod_{i=1}^5 (1 - F_i(c)) + F_1(c)F_2(c)F_3(c)P(X_5 \leq X_4 \leq c)$$

$$= \prod_{i=1}^5 (1 - F_i(c)) + \prod_{i=1}^5 F_i(c) - F_1(c)F_2(c)F_3(c)P(A)$$

and that

$$P(R_{T^*}^5 = 1) \geq \sum_{j=1}^5 \left((1 - F_j(c)) \prod_{\substack{1 \neq i \\ i=1}}^5 F_i(c) \right).$$

CLAIM 2.

$$E[f_5(R_{T^{**}}^5)] \geq \phi(F_1(c), \dots, F_5(c)) - F_1(c)F_2(c)F_3(c)P(A).$$

Claim 2 follows by noting that

$$\begin{aligned}
 P(R_{T^{**}}^5 = 5) &\leq \prod_{i=1}^5 (1 - F_i(c)) + P\left(\bigvee_{i=1}^3 X_i \leq c, X_4 < X_1 \wedge X_2 \wedge X_3 \wedge X_5\right) \\
 &\leq \prod_{i=1}^5 (1 - F_i(c)) + F_1(c)F_2(c)F_3(c)P(X_4 \leq c, X_4 < X_5) \\
 &= \prod_{i=1}^5 (1 - F_i(c)) + (1 - F_5(c)) \prod_{i=1}^4 F_i(c) \\
 &\quad + F_1(c)F_2(c)F_3(c)P(A)
 \end{aligned}$$

and that

$$\begin{aligned}
 P(R_{T^{**}}^5 = 1) &\geq \sum_{j=1}^3 \left((1 - F_j(c)) \prod_{\substack{i=1 \\ i \neq j}}^5 F_i(c) \right) \\
 &\quad + P\left(\bigvee_{i=1}^3 X_i \leq c, X_1 \vee X_2 \vee X_3 \vee X_5 < X_4\right) \\
 &\geq \sum_{j=1}^4 \left((1 - F_j(c)) \prod_{\substack{i=1 \\ i \neq j}}^5 F_i(c) \right).
 \end{aligned}$$

To complete the proof of Proposition 4.1, choose ε , $0 < \varepsilon < \min\{(1 - p)/2, (B^* - B)/3\}$, so that

$$|q_i - (1 - p)| < \varepsilon \text{ for all } i \text{ implies that } |\phi(q_1, \dots, q_5) - B^*| < \frac{B^* - B}{3},$$

and using the continuity of ψ , compactness of $[0, 1]^5$ and the uniqueness conclusion in (9), let $\delta > 0$ be such that

$$|q_i - (1 - p)| \geq \varepsilon \text{ for some } i \text{ implies that } \psi(q_1, \dots, q_5) > B + \delta.$$

CASE 1. Suppose that $|F_i(c) - (1 - p)| < \varepsilon$ for all i and that $P(A) < \varepsilon$. Then by Claim 2,

$$\begin{aligned}
 E[f_5(R_{T^{**}}^5)] &\geq B^* - \left(\frac{B^* - B}{3}\right) - \varepsilon \\
 &\geq B^* - 2\left(\frac{B^* - B}{3}\right) = B + \frac{B^* - B}{3}.
 \end{aligned}$$

CASE 2. Suppose that $|F_i(c) - (1 - p)| \geq \varepsilon$ for some i . Then by Claim 1,

$$E[f_5(R_{T^*}^5)] > B + \delta.$$

CASE 3. Suppose that $|F_i(c) - (1 - p)| < \varepsilon$ for all i and that $P(A) \geq \varepsilon$. Then by (9) and Claim 1,

$$E[f_5(R_{T^*}^5)] \geq B + \left(\frac{1-p}{2}\right)^3 \varepsilon.$$

Thus in each case there exists a stopping time $T \in \mathcal{S}_5$ with

$$E[f_5(R_T^5)] \geq B + \delta^*,$$

where $\delta^* = \min\{(B^* - B)/3, \delta, \varepsilon((1-p)/2)^3\} > 0$ does not depend on the distribution of the random variables X_1, \dots, X_5 , and this completes the proof. \square

Acknowledgment. The authors are grateful to the referees and to the Editors for their helpful suggestions, which led to an improvement of this paper.

REFERENCES

- FERGUSON, T. S. (1989). Who solved the secretary problem? *Statist. Sci.* **4** 282–296.
- FRANK, A. Q. and SAMUELS, S. M. (1980). On an optimal stopping problem of Gusein-Zade. *Stochastic Process Appl.* **10** 299–311.
- FREEMAN, P. R. (1983). The secretary problem and its extensions: A review. *Internat. Statist. Rev.* **51** 189–206.
- GILBERT, J. and MOSTELLER, F. (1966). Recognizing the maximum of a sequence. *J. Amer. Statist. Assoc.* **61** 35–73.
- KRENGEL, U. and SUCHESTON, L. (1978). On semiamarts, amarts, and processes with finite value. In *Probability on Banach Spaces* (J. Kuelbs, ed.). Dekker, New York.
- MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic, New York.
- SAKAGUCHI, M. (1984). Bilateral sequential games related to the no-information secretary problem. *Math. Japon.* **29** 961–973.

SCHOOL OF MATHEMATICS
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GEORGIA 30332

DEPARTMENT OF PURE MATHEMATICS
AND MATHEMATICAL STATISTICS
UNIVERSITY OF CAMBRIDGE
16 MILL LANE
CAMBRIDGE CB2 15B
ENGLAND