A Prophet Inequality Related to the Secretary Problem

THEODORE P. HILL AND ULRICH KRENGEL

ABSTRACT. Let $Z_1, Z_2, \ldots, Z_n$ be independent 0-1-valued random variables. A gambler gets a reward 1 if he stops at the time of the last success and otherwise gets no reward. A simple comparison with a Poisson process is used to show that a prophet can do at most $e$ times as well as the gambler using an optimal stopping time. For fixed $n$, the best constant is $(n/(n-1))^{n-1}$.

§1. Introduction

In the classical secretary problem one observes the relative ranks $R_1, R_2, \ldots, R_n$ of $n$ rankable objects presented one by one in random order. The $R_i$ are independent with $P(R_i = j) = 1/i$ ($j = 1, \ldots, i$). The aim is to find a stopping rule maximizing the probability of stopping when the last rank 1 object appears. The only relevant information about $R_i$ is whether $R_i$ takes the value 1 or not, so observing $Z_i = 1_{\{R_i = 1\}}$ is sufficient. Thus, the secretary problem is the special case $p_i = 1/i$ of the problem of finding an optimal stopping time for a sequence of independent observations $Z_1, Z_2, \ldots, Z_n$ with

$$p_i = P(Z_i = 1) = 1 - P(Z_i = 0)$$

when the reward function $X_i$ is the indicator function of the event

$$A_i = \{Z_i = 1, Z_j = 0 \text{ for all } j > i\}.$$  

Let $\mathcal{F}_i$ denote the $\sigma$-algebra generated by $Z_1, \ldots, Z_i$, and let $T$ be the family of stopping times relative to $(\mathcal{F}_i)$. The functional

$$V_n = \sup \{EX_t : t \in T\}$$


Key words and phrases. Secretary problem, best choice problem, prophet inequality, optimal stopping.

The first author was supported in part by NSF Grant # DMS-89-01267.

This paper is in final form and no version of it will be submitted for publication elsewhere.

© 1992 American Mathematical Society
0271-4132/92 $1.00 + .25 per page
is called the value of the process \( (X_t)_{t=1}^n \) for the gambler, and
\[
U_n = E \left( \max_{t \leq n} X_t \right) = 1 - \prod_{i=1}^n (1 - p_i)
\]
is often called the value for the prophet. We shall show that a simple comparison with the corresponding problem for the Poisson process leads to the following inequality.

**Theorem 1.** For all \( n \) and \((p_i)\), \( U_n \leq e V_n \). Moreover, the bound "\( e \)" is sharp.

In particular this shows, without the effort of finding the optimal stopping time, that it is possible to find the best object with probability \( \geq 1/e \). For fixed \( n \), the constant \( e \) in Theorem 1 can be replaced by \( (n/(n - 1))^{n-1} \), and this constant is optimal. After reducing the proof to the case \( p_1 = 1 \), this could also be deduced from a recent (earlier) result of Hill and Kennedy [HK]. However, we present an independent direct argument below.

We refer to Chow-Robbins-Siegmund [CRS] for general background on optimal stopping and to the papers of Freeman [Fr] and Ferguson [Fe] for the history of the secretary problem.

§2. Proof of Theorem 1 — A Comparison with a Poisson Process

Let \( \{N_t, t \geq 0\} \) be a right continuous homogeneous Poisson process with parameter \( \lambda = 1 \) and \( N_0 = 0 \). We consider the problem of stopping the process in the interval \((0, T]\). The reward is 1 if stopping is done at the time of the last jump of the Poisson process within the interval \([0, T]\) and zero otherwise. In other words, the reward function \( Y_t \) for stopping at time \( t \) is \( 1 \) when \( \lim_{s \rightarrow t} N_s < N_t = N_T \) and 0 otherwise. (If there is no jump in \([0, T]\), no reward is obtained.)

For \( \beta \leq T \) consider the following stopping time \( \tau_\beta \): stop at the time of the first jump in the interval \([\beta, T]\), or at time \( T \), whichever occurs first (so \( P(\tau_\beta < \infty) = 1 \)).

The probability that there is a jump in an interval \([t, t + dt]\) is \( dt \). The probability that it is the last one in the interval \((0, T]\) is \( r(t) = P(N_t = N_T) = e^{-(T-t)} \), and the probability that it is the first one in the interval \([\beta, T]\) is \( e^{-(T-\beta)} \). Hence
\[
EV_{\tau_\beta} = \int_\beta^T e^{-(T-t)} e^{-(T-\beta)} dt = (T-\beta) e^{-(T-\beta)}.
\]
Since the function \( f(x) = xe^{-x} \) has its maximum at \( x = 1 \) it is easy to see that \( EV_{\tau_\beta} \) is maximal for \( \beta = T-1 \) when \( T \geq 1 \), and for \( \beta = 0 \) when \( T < 1 \). \( U = P(N_T \geq 1) = 1 - e^{-T} \) can be considered to be the value for the prophet, since he always knows if a jump is the last one. Thus he can collect the reward 1 whenever there is at least one jump. Clearly
\[
V := \sup \{ EV_\tau : \tau \text{ is a stopping time} \} \geq EV_{\tau_\beta}.
\]
It is now simple to check that
\begin{equation}
U \leq eE\tau_{\beta} \leq eV.
\end{equation}

**Proof of Theorem 1.** As \(U_n\) and \(V_n\) depend continuously on the \(p_i\) we may assume \(0 < p_i < 1\). Let \(u_i = -\ln(1 - p_i)\), so \(P(N_{u_i} = 0) = 1 - p_i\).

Let \(t_0 = 0, t_i = u_1 + \cdots + u_i\) \((i = 1, \ldots, n)\), and \(T = t_n\). The random variables
\[
Z_i = 1_{\{N_{t_i} - N_{t_{i-1}} > 0\}} \quad (i = 1, \ldots, n)
\]
have the same joint distribution as the \(Z_i\). We can therefore assume \(Z_i = Z_i'\) for each \(i\).

We have \(U_n = P(Z_i \geq 1\) for at least one \(i \leq n) = P(N_T \geq 1) = U\). By (2.1), it now suffices to prove the existence of a stopping time \(\tau\) for the \((Z_i)\) process with \(E\tau \geq E\tau_{\beta}\).

The argument is very simple if there exists an \(m\) with \(0 \leq m < n\) and \(\beta = T - 1 = t_m\). Let

\[
\tau(\omega) = \inf\{i > m : Z_i(\omega) = 1\}.
\]
(If no such \(i\) exists, put \(\tau(\omega) = n + 1, X_\tau(\omega) = 0\).) For \(m < i \leq n\),
\[
P(\tau = i) = p_i \prod_{j=m+1}^{i-1} (1 - p_j) = P(\tau_\beta \in [t_{i-1}, t_i)).
\]
If \(\tau = i\), the probability that \(i\) is the last index \(j\) with \(Z_j = 1\) is \(r(t_i)\). When \(\tau_\beta = t \in [t_{i-1}, t_i)\), the probability that the Poisson process has no further jump in \([t, T]\) is \(r(t) \leq r(t_i)\). Hence \(E\tau \geq E\tau_{\beta}\). (The expected reward for the Poisson process is a lower bound since there may be several jumps in the interval where \(\tau\) stops.)

Since the case \(T < 1\) falls under the case \(\beta = t_m\) with \(m = 0\), it remains only to consider the case \(t_{m-1} < \beta = T - 1 < t_m\) for some \(m\) with \(1 \leq m \leq n\). In this case, take

\[
\tau'(\omega) = \inf\{i \geq m : Z_i(\omega) = 1\}.
\]
Then
\begin{equation}
E\tau' \geq p_m r(t_m) + (1 - p_m)(T - t_m)e^{-(T - t_m)},
\end{equation}
so since \((T - t_m)e^{-(T - t_m)} = E\tau_{t_m}\) is a lower bound for \(E\tau\) by the argument used in the case \(\beta = t_m\). It now suffices to show that the right hand side of (2.2) is \(\geq e^{-1} = E\tau_{t_m}\). Put \(b = t_m - \beta, c = \beta - t_{m-1}\). We have \(1 - p_m = e^{-(b+c)}\) and \(T - t_m = 1 - b\). So, we have to check whether
\[
p_m + (1 - p_m)(1 - b) \geq e^{-1+T-t_m} = e^{-b}
\]
holds. This is equivalent to
\begin{equation}
1 - e^{-(b+c)} + e^{-(b+c)}(1 - b) = 1 - be^{-(b+c)} \geq e^{-b}
\end{equation}
It follows from $e^b \geq 1 + b$, that $1 \geq e^{-b}(1 + b)$ and hence $1 - be^{-b} \geq e^{-b}$, which implies (2.3). \hfill \Box

**Remarks.** It may be seen by a continuity argument that the stopping time $\tau'$ can also be used in the case $\beta = t_m$.

Note that $r(t_k) = (p_{k+1})(1-p_{k+2}) \cdots (1-p_k)$. Thus, for practical purposes, one only has to compute these products, and to determine $m$ with $r(t_{m-1}) < e^{-1} \leq r(t_m)$. $\tau'$ is the "pass $m - 1$ rule" $\inf\{i \geq m : Z_i = 1\}$. Again by a continuity argument, one can see that the same rule can be used even if some of the $p_j$ are equal to 1.

In the classical secretary problem the optimal stopping time is the "pass $k$ rule" where $p_k + p_{k+1} + \cdots + p_{n-1} \geq 1 > p_{k+1} + \cdots + p_{n-1}$. It is possible that $k \neq m - 1$. For example, for $n = 5$ we find $k = 2$ and $m - 1 = 1$. Thus, the rule $\tau'$ used above need not be optimal.

For general $p_i$ the rule used in the secretary problem (like any rule which does not depend on $p_n$) may of course be bad. (For example, if $n = 3$, $p_1 = .5$ and $p_2 = p_3 = .7$, then the optimal rule used in the secretary problem in this case, namely "stop with the first relative rank one occurring at time two or later" yields an expected return of only .42, whereas the simple rule $\tau = 3$ returns .7.) The rule $\tau'$ above, though sometimes not optimal, always seems to be a fairly good one.

### §3. The Optimal Stopping Rule

The point of the above considerations has been to obtain a general estimate by a simple argument. However, it is also not hard to give a recursive description of the optimal stopping rule needed in §4 below:

Let $\mathcal{F}(i, k)$ denote the $\sigma$-algebra generated by $Z_i, Z_{i+1}, \ldots, Z_k$. Put $r(n) = n$. If $r(k + 1)$ has been defined, put $r(k) = k$ if $E(\mathcal{X}_k | \mathcal{F}(1, k)) \geq E(\mathcal{X}_{r(k+1)} | \mathcal{F}(1, k))$, and $r(k + 1)$ otherwise. Backward induction tells us that $r(1)$ is the optimal stopping rule. Let us make this construction more explicit. Put

\[
\xi_n = 1 \quad \text{and} \quad \xi_k = \prod_{j=k+1}^{n} (1 - p_j) \quad (k < n).
\]

We have $X_n = Z_n$ and $X_k = Z_k(1 - Z_{k+1}) \cdots (1 - Z_n)$ for $k < n$. The independence of $Z_1, \ldots, Z_n$ yields $E(X_k | \mathcal{F}(1, k)) = \xi_k Z_k$. Clearly $X_{r(n)} = Z_n$ is independent of $\mathcal{F}(1, n - 1)$ and $v_n = E(X_{r(n)} | \mathcal{F}(1, n - 1))$ a constant. By induction, assume that $X_{r(k+1)}$ is $\mathcal{F}(k+1, n)$ measurable, and hence $v_{k+1} = E(X_{r(k+1)} | \mathcal{F}(1, k))$ constant. Then

\[
X_r(k) = \mathbf{1}_{\{Z_k \xi_k \geq v_{k+1}\}} Z_k(1 - Z_{k+1}) \cdots (1 - Z_n) + \mathbf{1}_{\{Z_k \xi_k < v_{k+1}\}} X_{r(k+1)}
\]

is $\mathcal{F}(k, n)$ measurable, and $v_k = E(X_r(k) | \mathcal{F}(1, k - 1))$ constant.

The optimal stop rule $\tau(1)$ requires that you stop at time $k$ if you have not stopped before, and $k = n$ or $\xi_k \geq v_{k+1}$ and $Z_k = 1$. It follows from $v_n \leq v_{n-1} \leq \ldots \leq v_1$ that

\[
\text{if } v_k \geq Z_k, \text{ then } v_k \geq Z_{k+1} \geq \ldots \geq Z_n.
\]
\[ v_1 \leq v_2 \leq \cdots \leq v_n = 1 \] that there exist an integer \( h \) with \( 1 \leq h \leq n \) such that
\[
\tau(1) = \inf \{ k \geq h : Z_k = 1 \}
\]
which is the “pass \( h - 1 \) rule.” Setting \( v_{n+1} = 0 \), \( h \) is determined by \( h = \inf \{ k : \xi_k \geq v_{k+1} \} \). Clearly \( v_n = p_n \). (3.1) implies that the other \( v_k \) can be recursively calculated: For \( k \geq h \) we find
\[ v_k = p_k \xi_k + (1 - p_k) v_{k+1}. \]

For \( k < h \), \( v_k = v_h \).

§4. The Prophet Constant for Fixed \( n \)

We now write \( U_n(p) \) and \( V_n(p) \) for \( U_n \) and \( V_n \) when \( P(Z_i = 1) = p_i \) \((i = 1, \ldots, n)\), and \( p = (p_1, \ldots, p_n) \). We shall determine the minimal constant \( c_n \) such that \( U_n(p) \leq c_n V_n(p) \) holds for all \( p \).

Clearly \( c_1 = 1 \). For \( n \geq 2 \), \( c_n \geq 2 \) as may be seen by looking at \( p = (1, 1/2, 0, 0, \ldots, 0) \).

**Lemma 4.1.** If \( U_n(p) \leq c_n V_n(p) \) holds for all \( p \) with \( p_1 = 1 \), then this inequality holds for all \( p \).

**Proof.** If \( \epsilon > 0 \) is sufficiently small and \( p_1 + p_2 + \cdots + p_n < \epsilon \), then \( U_n(p) \leq 2V_n(p) \), as may be seen by looking at \( \tau = \inf \{ k : Z_k = 1 \} \). Hence we can restrict attention to \( D_\epsilon = \{ p : p_1 + \cdots + p_n \geq \epsilon \} \). On this compact set, by Theorem 1, \( V_n(p) \geq U_n(p)/\epsilon \geq c/nc \) is bounded away from 0. Hence \( c_n(p) = U_n(p)/V_n(p) \) assumes its maximum \( c_n \) by continuity. Let \( q = (q_1, \ldots, q_n) \) be a vector with \( c_n = c_n(q) \). We can assume \( q_1 > 0 \). (If \( q_1 = 0 \) replace \( q \) by \((q_2, q_3, \ldots, q_n, 0)\) without changing \( U_n(q) \) and \( V_n(q) \); if also \( q_2 = 0 \) rotate the vector once more and so forth.)

Put \( q' = (1, q_2, \ldots, q_n) \) and \( q'' = (0, q_2, \ldots, q_n) \), then \( q = q_1 q' + (1 - q_1) q'' \). It is simple to see that
\[
U_n(q) = q_1 U_n(q') + (1 - q_1) U_n(q'').
\]
Write \( \xi_1(p) \) and \( v_1(p) \) for the \( \xi_i \) and \( v_i \) in Section 3. The numbers \( v_2(p) \) and \( \xi_1(p) \) do not depend on \( p_1 \). If \( \xi_1(q) < v_2(q) \), the optimal rule for all three vectors \( q, q', q'' \) says that you should never stop at time 1. In this case \( v_1(q) = v_2(q) \), \( v_1(q') = v_2(q') = v_2(q) \) and \( v_1(q'') = v_2(q'') = v_2(q) \). If \( \xi_1(q) \geq v_2(q) \), then \( v_1(q) = q_1 \xi_1(q) + (1 - q_1) v_2(q) \), \( v_1(q') = \xi_1(q) \), and \( v_1(q'') = v_2(q) \). In all cases we obtain
\[
V_n(q) = q_1 V_n(q') + (1 - q_1) V_n(q'').
\]
If \( U_n(q') < c_n V_n(q') \), then using \( U_n(q'') \leq c_n V_n(q'') \), we would obtain \( U_n(q) < c_n V_n(q) \), contrary to the choice of \( q \). Hence, \( c_n = c_n(q) \).

We now know that \( c_n(p) = U_n(p)/V_n(p) \) is maximal for a \( p \) with \( p_1 = 1 \). For such a \( p \), \( U_n(p) = 1 \). Hence to maximize \( c_n(p) \) we have to minimize \( V_n(p) \) in
the set \( D \) of all \( p \) with \( p_1 = 1 \). We sketch a direct argument, although this particular result also follows from a more general result of Hill and Kennedy [HK] who studied the problem of finding universal bounds for an optimal return of a non-increasing function of the relative ranks of an arbitrary (i.e., not only 0–1 valued) sequence of independent random variables \( W_1, \ldots, W_n \). In particular, Hill and Kennedy determined an optimal bound for the probability of stopping at any moment when the maximum of finitely many independent random variables (rather than the last maximum) occurs. Our problem reduces to theirs by considering \( Z'_i = (1 - (i + 1)^{-1}) Z_i \). Their proof rests on general results on Schur convexity, and uses threshold rules rather than backward induction, so it does not attempt to determine the optimal rule.

**Lemma 4.2.** When \( p \) is a vector in \( D \) minimizing \( V_n(p) \), then \( 0 < p_i < 1 \) for \( 2 \leq i \leq n \).

**Proof.** Suppose not. Then there exists a shorter vector \( q = (q_1, q_2, \ldots, q_m) \) with \( m < n \), \( q_1 = 1 \), \( 0 < q_i < 1 \) for \( i = 2 \leq i \leq m \), and \( V_n(p) = V_m(q) \). Let \( q'_m, q'_{m+1} \) be any numbers strictly between 0 and 1 with \( q_m = q'_m + (1 - q'_m)q'_{m+1} \). Put \( q'_i = q_i \) for \( 1 \leq i < m \), and \( q' = (q'_1, \ldots, q'_{m+1}) \).

It is simple to check that \( v_i(q') < v_i(q) \) for \( i = 2, \ldots, m \) and \( \xi_i(q') = \xi_i(q') \) for \( i = 1, \ldots, m \).

If \( \xi_1(q') \leq v_2(q') \), then \( v_1(q') = v_2(q') \) and \( v_1(q) = v_2(q) \) and hence \( v_1(q') > v_1(q) = V_n(p) \), contradicting the minimizing property of \( p \). If \( v_2(q') < \xi_1(q') \), then \( v_1(q') = \xi_1(q') < v_2(q) \), again contradicting the choice of \( p \). Hence, \( \xi_1(q') \geq v_2(q) \).

Pass to a vector \( q'' = (q''_1, \ldots, q''_{m+1}) \) with \( q''_1 = 1 \) and \( q''_i > q'_i \) \( (i = 2, \ldots, m + 1) \). If \( q''_i \) differs sufficiently little from \( q'_i \) we have \( v_2(q''_i) < v_2(q) \). But \( \xi_1(q''_i) < \xi_1(q') = \xi_i(q) \). Now \( \xi_1(q') \geq v_2(q) \) and \( q_1 = 1 \) yields \( v_1(q) = \xi_1(q) \). We obtain \( v_1(q''_i) = \max\{\xi_1(q''_i), v_2(q''_i)\} < v_1(q) = v_1(p) \). Again we have a contradiction. \( \square \)

**Lemma 4.3.** If \( p \in D \) minimizes \( V_n(p) \) in \( D \) then \( \xi_1(p) = v_2(p) \).

**Proof.** Assume \( h := \inf\{i \geq 1 : \xi_i(p) \geq v_{i+1}(p)\} > 1 \). We write \( \xi_i \) for \( \xi_i(p) \) and \( v_i \) for \( v_i(p) \). If \( \xi_h > v_{h+1} \) we can lower \( V_n(p) \) by making \( p_h \) slightly smaller. Hence \( \xi_h = v_{h+1} \), and therefore \( v_h = v_{h+1} \). As \( h + 1 \geq 3 \) and \( V_n(p) = v_{h+1} \), \( V_n(p) \) is not changed if we replace \( p_2 \) by 0. Then there would exist a minimizing vector with \( p_2 = 0 \), contradicting Lemma 4.2. \( \square \)

**Lemma 4.4.** \( p = (1, 1/n, 1/n, \ldots, 1/n) \) minimizes \( V_n(p) \) in \( D \).

**Proof.** Let \( p \in D \) be a vector minimizing \( V_n(p) \). By Lemma 4.3, \( \xi_1 \geq v_2 = p_2 q_2 + (1 - p_2) q_3 \). It follows from \( \xi_2 > \xi_1 \geq v_2 \) that \( v_2 < v_3 \). If \( \xi_1 > v_2 \) holds, we can decrease \( \xi_1 \) a little by increasing \( p_2 \) slightly, which would decrease \( V_n(p) \). As this is impossible, \( \xi_1 = v_2 \).
The optimal rule $\xi_1 = v_2$ stops as soon as the first success at some time $\geq 2$ is observed (since $h = 1$). Hence

$$v_2 = p_2 \xi_2 + (1 - p_2)p_3 \xi_3 + (1 - p_2)(1 - p_3)p_4 \xi_4 + \cdots = \sum_{i=2}^{n} \frac{p_i \xi_i}{1 - p_i}.$$ 

Thus the condition $\xi_1 = v_2$ leads to the condition

$$\sum_{i=2}^{n} \frac{p_i}{1 - p_i} = 1.$$ (4.1)

We have to minimize $\xi_1 = v_1 = \prod_{i=2}^{n} (1 - p_i)$ subject to (4.1). Assume there exist indices $2 \leq i, j \leq n$ with $p_i \neq p_j$. We replace both these probabilities by two equal numbers such that the sum in the corresponding terms in (4.1) is not altered. In other words: Put $\beta = p_j/(1 - p_i) + p_j(1 - p_i), 2r/(1 - r) = \beta$ and $p'_i = p'_j = r$. It may be shown by simple algebra that $(1 - p'_i)(1 - p'_j) = (1 - r)^2 < (1 - p_i)(1 - p_j)$. Hence, if $p'(1, \ldots, p'_n)$ with $p'_\nu = p_\nu$ for $\nu \notin \{i, j\}$, then $V_n(p') < V_n(p)$. As this contradicts the choice of $p$, we see that $p_2 = p_3 = \cdots = p_n$. It now follows from (4.1) that $p_\nu = 1/n$ for $\nu = 2, \ldots, n$. $\square$

Theorem 2. $U_n(p) \leq (n/(n - 1))^n V_n(p)$ for all $p = (p_1, \ldots, p_n)$, and the constant $c_n = (n/(n - 1))^n$ is optimal.

Proof. For the $p$ in Lemma 4 we have $V_n(p) = \xi_1 = \prod_{i=2}^{n} (1 - \frac{1}{n}) = (\frac{n-1}{n})^{n-1}$. The conclusion follows since $U_n(p) = 1$. $\square$

Remark. Let $p = p(n) = (1, \frac{1}{n}, \ldots, \frac{1}{n})$ and let $p'(n) = (1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n})$ be the distribution in the secretary problem. It is well known (e.g., [Fr]) that $c \leq c(p'(n)) \rightarrow e$. By Lemma 4.4, $c_n(p(n)) \geq c_n(p'(n))$ and clearly $e \leq c_n(p(n)) = (n/(n - 1))^n \rightarrow e$. For moderate values of $n, c_n(p(n))$ may be noticeably closer to $e$. For example, $c_6(p(6)) = 2.49, c_6(p'(6)) = 2.34, c_{10}(p(10)) = 2.58, c_{10}(p'(10)) = 2.51$.

Acknowledgements. This joint research was made possible by a visit of the second author to the School of Mathematics at the Georgia Institute of Technology, which was supported in part by the Deutsche Forschungsgemeinschaft. The authors are also grateful to the referee for several useful suggestions and corrections.

References


School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332

E-mail: th5d@gatech.edu

Institut für Mathematische Stochastik, Lotzestrasse 13, W-3400 Göttingen, Federal Republic of Germany