4-1981

Optimal Sequential Selection of a Monotone Sequence From a Random Sample

Stephen Samuels
Purdue University

J Michael Steele
University of Pennsylvania

Recommended Citation

Samuels, S., & Steele, J. M. (1981). Optimal Sequential Selection of a Monotone Sequence From a Random Sample. *The Annals of Probability, 9* (6), 937-947. http://dx.doi.org/10.1214/aop/1176994265

At the time of publication, author J. Michael Steele was affiliated with Stanford University. Currently, he is a faculty member at the Statistics Department at the University of Pennsylvania.

Optimal Sequential Selection of a Monotone Sequence From a Random Sample

Abstract
The length of the longest monotone increasing subsequence of a random sample of size n is known to have expected value asymptotic to $2n^{1/2}$. We prove that it is possible to make sequential choices which give an increasing subsequence of expected length asymptotic to $(2n)^{1/2}$. Moreover, this rate of increase is proved to be asymptotically best possible.

Keywords
monotone subsequence, optimal stopping, subadditive process

Disciplines
Physical Sciences and Mathematics

Comments
At the time of publication, author J. Michael Steele was affiliated with Stanford University. Currently, he is a faculty member at the Statistics Department at the University of Pennsylvania.

The Annals of Probability
1981, Vol. 9, No. 6, 937–947

OPTIMAL SEQUENTIAL SELECTION OF A MONOTONE SEQUENCE FROM A RANDOM SAMPLE[1]

BY STEPHEN M. SAMUELS AND J. MICHAEL STEELE

Purdue University and Stanford University

The length of the longest monotone increasing subsequence of a random sample of size n is known to have expected value asymptotic to $2n^{1/2}$. We prove that it is possible to make sequential choices which give an increasing subsequence of expected length asymptotic to $(2n)^{1/2}$. Moreover, this rate of increase is proved to be asymptotically best possible.

1. Introduction. A central theme in the theory of optimal stopping is that many stochastic tasks can be performed almost as well by someone unable to foresee the future as by a prophet. In one classic example, the "secretary problem", the task is to stop at the largest of n sequentially observed independent identically distributed observations $X_1, X_2, \cdots, X_n$. Without clairvoyance one attains X_τ where τ is some stopping time, but a prophet is always able to attain $\max_{1\le i\le n} X_i$.

That the prophet's advantage is rather modest follows from the well-known fact that no matter how large n is there is a stopping time τ_n for which

$$P(X_{\tau_n} = \max_{1\le i\le n} X_i) > e^{-1}, \tag{1.1}$$

(see, e.g., Gilbert and Mosteller (1966)).

The stochastic task we consider here is more complex than the secretary problem and the central theme is illustrated in a different way from (1.1). To set the problem, let $\{X_i: 1 \le i < \infty\}$, denote independent random variables with continuous distribution F. The basic object of interest is

$$L_n = \max\{k: X_{i_1} > X_{i_2} > \cdots > X_{i_k} \quad \text{with} \quad 1 \le i_1 < i_2 < \cdots < i_k \le n\},$$

the length of the longest monotone decreasing subsequence of the sample $\{X_1, X_2, \ldots, X_n\}$. (We could equally well have considered increasing subsequences but the notation will be simpler this way.) The variable L_n has been studied extensively and it is now known that

$$EL_n \sim 2n^{1/2}. \tag{1.2}$$

The first result, $EL_n \sim cn^{1/2}$, was obtained by Hammersley (1972) via an ingenious use of the planar Poisson process. Baer and Brock (1968) had conjectured earlier on the basis of computer simulations that $c = 2$. By a delicate variational argument Logan and Shepp (1977) proved that c is at least 2 and by a similar method Veršik and Kerov (1977) established that c equals 2.

How well can one sequentially choose a monotone decreasing subsequence using only stopping times? Formally, we call a sequence of stopping times $\tau_1, \tau_2, \cdots$ a *policy* if (1) they are adapted to $\{X_i: 1 \le i < \infty\}$, (2) $1 \le \tau_1 < \tau_2 < \cdots$, and (3) $X_{\tau_1} > X_{\tau_2} > \cdots$. The class of all policies is denoted by $\mathcal{S}$ and our main problem is to determine

$$u_n = \sup_{\tau\in\mathcal{S}} E(\max\{k: \tau_k \le n\}).$$

The quantity u_n is the largest expected length of a monotone decreasing subsequence

Received April 23, 1979; revised April 23, 1980.
[1] The research was supported in part by Grant N0014-76-C-0475 (Office of Naval Research).
AMS 1970 *subject classifications.* Primary 62L15; secondary 60G40.
Key words and phrases. Monotone subsequence, optimal stopping, subadditive process.

which can be achieved by sequential selection. Although u_n would *a priori* depend on F, one can easily check that it is the same for all continuous F. Moreover the optimal policy for a given F can be obtained via the probability integral transform from the policy for the uniform distribution on [0, 1].

Our main result is the following

THEOREM.

$$u_n \sim (2n)^{1/2}. \tag{1.3}$$

The intuitive content of this result is another illustration of the central theme; the prophet asymptotically outperforms the intelligent (but nonclairvoyant) individual by only a factor of $2^{1/2}$. One should also note that the naive individual who too eagerly reports each successive record low achieves an expected length of only $\sum_{k=1}^{n} 1/k \sim \log n$ and thus does much worse than the prophet or intelligent individual.

Our proof that $u_n \sim (2n)^{1/2}$ begins with a simple algorithm for computing u_n based on an integral equation obtained by dynamic programming. Standing alone, the integral equation seems to be ineffective, so in Section 3 we prove by a subadditivity argument that $u_n/n^{1/2}$ has a limit. A sequence of efficient, but suboptimal, policies are then given in Section 4 which show $\lim u_n/n^{1/2} \geq 2^{1/2}$. The crux of the proof is in Section 5, where the integral equation is finally used to show $\lim u_n/n^{1/2} \leq 2^{1/2}$ after first establishing an essential regularity property of the solution by a probabilistic argument.

The proof outlined above yields several results *en route*. In particular, we obtain results on optimal selection when the sample size is random. These results as well as comments on a related problem are collected in Section 6.

In the final section we are fortunate to be able to include a result due to Burgess Davis on the sequential selection of a decreasing subsequence from a random permutation. This result teams up with the main theorem of this paper to settle a second conjecture given in the computational paper of Baer and Brock (1968). We would like to thank Professor Davis for his kind suggestion that his result be included in the present paper.

2. An algorithm for computing the optimal expected length. First of all, as we remarked in the introduction, we may assume without loss of generality that the common distribution of the observations $X_1, X_2, \cdots$ is uniform on (0, 1).

Let us define, for each $t \in (0, 1]$,

$$\mathscr{S}_t = \{\tau = (\tau_1, \tau_2, \cdots) \in \mathscr{S} : X_{\tau_i} < t,\ i = 1, 2, \cdots\}, \tag{2.1}$$

the class of policies which only select observations smaller than t. We also let

$$u_n(t) = \sup_{\tau \in \mathscr{S}_t} E\{\max\{k : \tau_k \leq n\}\}. \tag{2.2}$$

Clearly

$$u_n = u_n(1) \tag{2.3}$$

and

$$u_1(t) = P(X_1 \leq t) = t. \tag{2.4}$$

We also record the trivial fact: $u_0(t) \equiv 0$.

Now fix t and consider $n + 1$ available observations. Because of the stationarity of the X_i's, the maximal conditional expected subsequence length, given X_1, will be just $u_n(t)$ if X_1 is not selected and $1 + u_n(X_1)$ if X_1 is selected (in which case necessarily $X_1 < t$). Since the optimal policy must do whichever maximizes the conditional expectation, we have the algorithm:

$$\begin{aligned} u_{n+1}(t) &= u_n(t)P(X_1 \geq t) + E\max\{u_n(t), 1 + u_n(X_1)\}I_{\{X_1 < t\}} \\ &= (1 - t)u_n(t) + \int_0^t \max\{u_n(t), 1 + u_n(s)\}\, ds. \end{aligned} \tag{2.5}$$

It follows easily that for each n and t, the policy which achieves $u_n(t)$ is

$$
\begin{aligned}
\tau_1 &= \min\{i: X_i < t \text{ and } 1 + u_{n-i}(X_i) \geq u_{n-i}(t)\} \\
&> n \text{ (arbitrary) if no such } i \leq n \\
\tau_{k+1} &= \min\{i > \tau_k: X_i < X_{\tau_k} \text{ and } 1 + u_{n-i}(X_i) \geq u_{n-i}(X_{\tau_k})\} \\
&> n \text{ (arbitrary) if no such } i \leq n \text{ or if } \tau_k > n.
\end{aligned} \tag{2.6}
$$

Since (2.4) and (2.5) imply that each $u_n(\cdot)$ is strictly increasing on (0, 1), we can implicitly define the functions $t_n^*(t)$ by the following relations:

$$
\begin{aligned}
u_n(t_n^*(t)) &= u_n(t) - 1 \quad \text{if} \quad u_n(t) \leq 1 \\
t_n^*(t) &= 0 \qquad\qquad\quad \text{if} \quad u_n(t) \leq 1.
\end{aligned} \tag{2.7}
$$

Equation (2.6) now simplifies to the following:

$$\tau_1 = \min\{i: t_{n-i}^*(t) \leq X_i < t\} \tag{2.8a}$$

$$\tau_{k+1} = \min\{i > \tau_k: t_{n-i}^*(X_{\tau_k}) \leq X_i < X_{\tau_k}\}. \tag{2.8b}$$

One would naturally like to show $u_n \sim (2n)^{1/2}$ directly from (2.5), but this does not seem possible. The integral equation (2.5) becomes effective only after substantial quantitative information about $u_n(t)$ is obtained.

Here we should remark that

$$u_n(t) = \sum_{k=1}^{n} k^{-1} \sum_{j=k}^{n} \binom{n}{j} t^j (1-t)^{n-j} \tag{2.9}$$

for all t small enough so that the right side of (2.9) is less than or equal to one. This can be shown either directly from the integral equation or by noting that when $u_n(t) \leq 1$ the optimal policy is to select all successive minima among those X_i's, $i = 1, 2, \ldots, n$, which are smaller than t. The right side of (2.9) is the expected number of such minima.

3. Existence of the limit. To show that $u_n/n^{1/2}$ has a limit, we shall first prove that a limit exists for an analogous planar Poisson process problem and then show that the two problems are asymptotically similar. The proof is a version of Hammersley's subadditivity idea made somewhat simpler because we deal only with expectations rather than with random variables.

3a. *The planar Poisson process problem.* Let $Z_1, Z_2, \cdots$, be i.i.d., each exponentially distributed with mean one, and independent of the X_i's which are i.i.d. uniform on (0, 1). Let $\mathscr{S}_Z$ be the class of policies $\tau = (\tau_1, \tau_2, \cdots)$ with

(a) each τ_i adapted to $\{(Z_i, X_i): 1 \leq i < \infty\}$,

(b) $1 \leq \tau_1 < \tau_2 < \cdots$ and $X_{\tau_1} > X_{\tau_2} > \cdots$;

and let

$$w(\lambda) = \sup_{\tau \in \mathscr{S}_Z} E\{\max k: \textstyle\sum_{i=1}^{\tau_k} Z_i < \lambda\}.$$

In other words, we observe a Poisson process with arrival rate one, on an interval of length λ. At each arrival time we are allowed to observe a random variable uniform on (0, 1) and independent of its predecessors, and the object is to select a decreasing subsequence of maximal expected length.

What makes this problem so appealing is the well-known fact that, if we choose p and t, each in (0, 1), then the following two processes are also Poisson:

(a) those arrival times in $(0, p\lambda)$ for which the corresponding X_i's are $\geq t$;

(b) those arrival times in $(p\lambda, \lambda)$ for which the corresponding X_i's are $< t$.

Those processes have expected numbers of arrivals $p(1-t)\lambda$ and $(1-p)t\lambda$ respectively. It follows, by considering the subclass of $\mathscr{S}_Z$ consisting of those τ's with

$$
\begin{aligned}
X_{\tau_i} &\geq t \quad \text{if} \quad \textstyle\sum_{j=1}^{\tau_i} Z_j < p\lambda \\
&< t \quad \text{if} \quad \textstyle\sum_{j=1}^{\tau_i} Z_j > p\lambda,
\end{aligned}
$$

that

$$w(\lambda) \geq w(p(1-t)\lambda) + w((1-p)t\lambda). \tag{3.1}$$

If we now define

$$\rho(x) = w(x^2)$$

and choose $\lambda = (r+s)^2, p = 1 - t = r/(r+s)$, (3.1) becomes

$$\rho(r+s) \geq \rho(r) + \rho(s). \tag{3.2}$$

By the elementary lemma on subadditive sequences (Fekete (1923), or Khinchin (1957)), this implies the existence of

$$\gamma = \lim_{t\to\infty} \rho(t)/t = \limsup \rho(t)/t \geq 1. \tag{3.3}$$

The finiteness of γ follows from the inequality

$$w(\lambda) \leq \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} EL_k \leq M \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} k^{1/2} \leq M \lambda^{1/2};$$

the second inequality holds for some constant M by (1.2). The last inequality is just an application of the familiar inequality $E|X| \leq (EX^2)^{1/2}$.

Thus we have shown that there is a finite $\gamma > 0$ such that

$$\begin{aligned} \gamma &= \lim_{t\to\infty} w(t^2)/t \\ &= \lim_{t\to\infty} w(t)/t^{1/2}. \end{aligned} \tag{3.4}$$

3b. *Asymptotic similarity of the two processes.* We will now use some elementary estimates to show that $u_n \sim w(n)$. With the usual notation

$$P(k; \lambda) = \sum_{j=0}^{k} \lambda^j e^{-\lambda}/j!$$

and

$$\bar{P}(k; \lambda) = 1 - P(k-1; \lambda)$$

we will prove the inequalities:

$$u_{[(1-\epsilon)\lambda]} \bar{P}([(1-\epsilon)\lambda]; \lambda) \leq w(\lambda) \tag{3.5}$$

and

$$w(\lambda) \leq u_{[(1+\epsilon)\lambda]} P([(1+\epsilon)\lambda]; \lambda) + \lambda \bar{P}([(1+\epsilon)\lambda]; \lambda). \tag{3.6}$$

The first inequality holds since the optimal policy for $n = [(1-\epsilon)\lambda]$ observations used on the Poisson process (paying no heed to the Z_i's) would yield an expected length of at least $u_{[(1-\epsilon)\lambda]}$ whenever there are at least $[(1-\epsilon)\lambda]$ arrivals.

For the second inequality let N be the number of arrivals and let L be the subsequence length obtained when the optimal policy in $\mathscr{S}_Z$ is used. We then have

$$w(\lambda) = E(L \mid N \leq [(1+\epsilon)\lambda]) P([(1+\epsilon)\lambda]; \lambda) + E\{E(L \mid N) I_{\{N > [(1+\epsilon)\lambda]\}}\}.$$

We note $E(L \mid N \leq [(1+\epsilon)\lambda]) \leq u_{[(1+\epsilon)\lambda]}$ and trivially $E(L \mid N) \leq N$. Inequality (3.6) then follows since

$$ENI_{\{N > [(1+\epsilon)\lambda]\}} = \lambda \bar{P}([(1+\epsilon)\lambda]; \lambda).$$

From the fact (3.4) that $w(\lambda) \sim \gamma\lambda^{1/2}$ and elementary bounds on the Poisson distribution one now easily deduces from (3.5) and (3.6) that $u_n \sim \gamma n^{1/2}$.

We remark that it is not hard to extend this result to show that for each $t \in (0, 1]$ one has $u_n(t)/(nt)^{1/2} \to \gamma$.

4. A lower bound for the limit. For any constant $\alpha > 0$, and for each n, we consider the policy $\tau(n) = \tau(n; \alpha)$, where

$$\tau_1(n) = \min\{j: X_j \geq 1 - \alpha/n^{1/2}\}$$

$$\tau_{k+1}(n) = \min\{j > \tau_k(n): X_j \in [X_{\tau_k(n)} - \alpha/n^{1/2}, X_{\tau_k(n)})\}.$$

We shall show that for these policies

$$\liminf_{n\to\infty} E \max\{k: \tau_k(n) \leq n\} \geq n^{1/2} \min(\alpha, 2/\alpha). \tag{4.1}$$

The right side of (4.1) is maximized for $\alpha = 2^{1/2}$, which shows that $\liminf_{n\to\infty} u_n/n^{1/2} \geq 2^{1/2}$. (Of course when we complete the proof of the theorem, (4.1) with $\alpha = 2^{1/2}$ will imply that the policies $\tau(n; 2^{1/2})$ are asymptotically optimal.)

What makes the policies $\tau(n)$ easy to evaluate is the fact that if $X_{\tau_k(n)} \geq \alpha/n^{1/2}$, then $\tau_{k+1}(n) - \tau_k(n)$ and $X_{\tau_k(n)} - X_{\tau_{k+1}(n)}$ are conditionally independent and independent of $\{\tau_1(n), \cdots, \tau_k(n); X_{\tau_1(n)}, \cdots, X_{\tau_k(n)}\}$, with geometric ($p = \alpha/n^{1/2}$) and uniform on $(0, \alpha/n^{1/2})$ distributions respectively. Hence we now let $\{Y_k: k = 1, 2, \cdots\}$ and $\{Z_k: k = 1, 2, \cdots\}$ be independent sequences of i.i.d. random variables with these geometric and uniform distributions respectively. Also set $S_k = Y_1 + \cdots + Y_k$ and $S'_k = Z_1 + \cdots + Z_k$, and define

$$M_n = \max\{k: S_k \leq n \text{ and } S'_k \leq 1 - \alpha/n^{1/2}\}.$$

We first observe

$$EM_n < E \max\{k: \tau_k(n) \leq n\}. \tag{4.2}$$

Now, for any $\epsilon > 0$, Chebyshev's inequality gives

$$\begin{aligned} P(S_k \leq n) &= 1 + O(n^{-1/2}) \quad && \text{if} \quad k \leq (1-\epsilon)\alpha n^{1/2} \\ &= O(n^{-1/2}) && \text{if} \quad k \geq (1+\epsilon)\alpha n^{1/2} \end{aligned}$$

and

$$\begin{aligned} P(S'_k \leq 1 - \alpha/n^{1/2}) &= 1 + O(n^{-1/2}) \quad && \text{if} \quad k \leq (1-\epsilon)(2/\alpha) n^{1/2} \\ &= O(n^{-1/2}) && \text{if} \quad k \geq (1+\epsilon)(2/\alpha) n^{1/2}. \end{aligned}$$

The $O(n^{-1/2})$ terms are uniformly small in the indicated range, so, by the independence of the two sequences,

$$EM_n = \sum_{k=1}^{n} P(S_k \leq n) P(S'_k \leq 1 - \alpha/n^{1/2}) \sim n^{1/2} \min(\alpha, 2/\alpha).$$

By (4.2), this shows that (4.1) holds, completing the proof.

5. An upper bound for the limit. Now that we know that $\lim u_n/n^{1/2}$ exists and is at least $2^{1/2}$, to complete the proof of the theorem it will suffice to show that

$$\liminf u_n/n^{1/2} \leq 2^{1/2}. \tag{5.1}$$

Our proof of (5.1) hinges on showing that

$$u_n(t)/t^{1/2} \uparrow \text{ in } t \text{ for each } n. \tag{5.2}$$

The derivative of $u_n(t)/t^{1/2}$ is

$$t^{-1/2}\{u'_n(t) - (2t)^{-1} u_n(t)\},$$

so, to prove (5.2), we must show that for each $t \in (0, 1)$

$$u_n(t + \delta) - u_n(t) > (½)(\delta/t) u_n(t) + o(\delta) \text{ as } \delta \downarrow 0. \tag{5.3}$$

This inequality will be proved by selecting a suboptimal member of $\mathscr{S}_{t+\delta}$ (as defined in

(2.1)) and showing that this policy improves on the optimal policy in $\mathscr{S}_t$ by an amount equal to the right side of (5.3).

What we actually do is a bit more complicated than this and involves showing that $u_n(t)$ is also the maximal expected subsequence length in a problem where the number of available observations is random with a binomial distribution.

5a. *Optimal selection with binomially many observations.* Since the policy in $\mathscr{S}_t$ which achieves $u_n(t)$, as given by (2.8), ignores the actual values of all X_i's which are greater than t, and since the other X_i's are, conditionally, i.i.d. uniform on $(0, t)$, we could just as well replace the X_i's by a sequence of observable coin tosses with probability t of heads, letting each toss which gives heads be accompanied by the next in a sequence of i.i.d. random variables uniform on $(0, t)$.

To exploit this observation let $Y_1, Y_2, \cdots$ and $X_1, X_2, \cdots$ be independent sequences of i.i.d. random variables, the Y's Bernoulli (t), and the X's uniform on (0, 1). Let $\mathscr{U}_t$ be the class of policies adapted to the (Y_i, X_i)'s which select a monotone decreasing subsequence by selecting only X_i's for which $Y_i = 1$ and which totally ignore all X_i's for which $Y_i = 0$. Then for the $u_n(t)$ defined by (2.2) we have the representation

$$\sup_{\tau \in \mathscr{U}_t} E\{\max\{k : \tau_k \le n\}\} = u_n(t).$$

Since we have made the X_i's uniform on (0, 1) rather than on $(0, t)$—this is to avoid confusion in what follows—the policy $\tau = (\tau_1, \tau_2, \cdots)$ which achieves $u_n(t)$ becomes

$$\tau_1 = \min\{i : Y_i = 1 \quad \text{and} \quad t^*_{n-i}(t) \le tX_i < t\}$$

$$\tau_{k+1} = \min\{i > \tau_k : Y_i = 1 \quad \text{and} \quad t^*_{n-i}(tX_{\tau_k}) \le tX_i < tX_{\tau_k}\}.$$

Now suppose we introduce a second coin toss at each stage—letting $Y_1', Y_2', \cdots$ be i.i.d. Bernoulli (p') and independent of the Y_i's and the X_i's—and we allow policies adapted to the $\{Y_i'\}$ as well, but maintain the requirement that all X_i's for which $Y_i = 0$ must be ignored. Then clearly what we have introduced is external randomization; those policies which depend in some way on the $\{Y_i'\}$ are simply randomized policies, and, of course, none of these can improve on the best nonrandomized policies. In particular, any policy which ignores all X_i's for which either $Y_i = 0$ or $Y_i' = 0$ is really a policy in $\mathscr{U}_{tt'}$; hence the expected length of the subsequence of $X_1, \cdots, X_n$ which it selects is no greater than $u_n(tt')$. Just such a policy will be needed in proving (5.3).

5b. *Monotonicity of* $u_n(t)/t^{1/2}$. We now fix n and let the $\{Y_i\}$ be Bernoulli $(t + \delta)$ and the $\{Y_i'\}$ be Bernoulli $(t/(t + \delta))$. First note that $P(Y_i = Y_i' = 1) = t$. We consider two randomized policies τ and τ' in $\mathscr{U}_{t+\delta}$. The first is to be equivalent to the optimal policy (for given n) in $\mathscr{U}_t$ while the second is to be a slight modification of the first. Specifically, we let $\tau = (\tau_1, \tau_2, \cdots)$ with

$$\tau_1 = \min\{i : Y_i = Y_i' = 1 \quad \text{and} \quad t^*_{n-i}(t) \le tX_i\}$$

$$\tau_{k+1} = \min\{i > \tau_k : Y_i = Y_i' = 1 \quad \text{and} \quad t^*_{n-i}(tX_{\tau_k}) \le tX_i < tX_{\tau_k}\}.$$

We want τ' to agree with τ up to the first $i \le n$, if any, at which $Y_i = 1$, $Y_i' = 0$, and at which X_i would have been selected by τ if Y_i' had been 1. We want τ' to select this X_i, but thereafter to continue to behave like τ. We thus define

$$I = \min\{i : Y_i = 1,\ Y_i' = 0,\ t^*_{n-i}(tX_{\sigma_i}) \le tx_i < tX_{\sigma_i}\}$$

$$= \infty \text{ if no such } i \le n,$$

where

$$\sigma_i = \max\{\tau_k : \tau_k < i\}$$

$$= 0 \text{ (and } X_0 \equiv 1) \quad \text{if} \quad \tau_1 \ge i.$$

We then let $\tau' = (\tau_1', \tau_2', \cdots)$ where

$$\tau_k' = \tau_k \quad \text{if} \quad \tau_k < I$$

$$\tau_k' = I \quad \text{if} \quad \tau_k < I < \tau_{k+1}$$

$$\tau_{k+1}' = \min\{i > \tau_k' : Y_i = Y_i' = 1 \quad \text{and} \quad t_{n-i}^*(tX_{\tau_k'}) \le tX_i < tX_{\tau_k'}\} \qquad \text{if} \quad \tau_k' \ge I.$$

Now, for convenience, we let L and L' be the lengths of the subsequences of $X_1, \cdots, X_n$ selected by τ and τ' respectively. Then $L = L'$ on $\{I > n\}$,

$$EL = u_n(t),$$

and

$$EL' \le u_n(t + \delta)$$

so

$$u_n(t + \delta) - u_n(t) \ge E(L' - L \mid I \le n)P(I \le n). \tag{5.4}$$

Furthermore, from (2.2) and the definitions of τ and τ',

$$E(L' - L \mid I = i, X_{\sigma_i} = x) = 1 + E\{u_{n-i}(tX_i) \mid I = i, X_{\sigma_i} = x\} - u_{n-i}(tx). \tag{5.5}$$

(Note: (5.5) is valid even when $i = n$, since we have set $u_0(t) \equiv 0$.)

From (5.4) and (5.5) we see that to establish (5.3) it suffices to prove

$$P(I \le n) \le (\delta/t)u_n(t) + o(\delta) \quad \text{as} \quad \delta \downarrow 0 \tag{5.6}$$

and, for each $i \le n$,

$$E\{u_{n-i}(tX_i) \mid I = i, X_{\sigma_i} = x\} \ge u_{n-i}(tx) - ½. \tag{5.7}$$

To prove (5.6) we first remark that, since $t/(t + \delta) \to 1$ as $\delta \downarrow 0$, we have

$$P(I \le n) = \textstyle\sum_{i=1}^n P(Y_i = 1, Y_i' = 0, A_i) + o(\delta) \tag{5.8}$$

where A_i denotes the event $\{t_{n-i}^*(tX_{\sigma_i}) \le tX_i < tX_{\sigma_i}\}$. A_i is independent of $\{Y_i = 1, Y_i' = 0\}$ so

$$\begin{aligned} P(Y_i = 1, Y_i' = 0, A_i) &= P(Y_i = 1, Y_i' = 0)P(A_i) \\ &= (\delta/t)P(Y_i = Y_i' = 1, A_i) \\ &= (\delta/t)P(X_i \text{ is selected by } \tau). \end{aligned}$$

Putting this back into (5.8) we have

$$P(I \le n) = (\delta/t) \textstyle\sum_{i=1}^n P(X_i \text{ is selected by } \tau) + o(\delta) = (\delta/t)u_n(t) + o(\delta),$$

which is (5.6).

To prove (5.7), we first remark that the conditional distribution of tX_i given $X = i$ and $X_{\sigma_i} = x$ is uniform on $(t_{n-i}^*(tx), tx)$. Also we note that

$$u_{n-i}(t_{n-i}^*(tx)) \ge u_{n-i}(tx) - 1 \tag{5.9}$$

with equality holding unless $u_{n-i}(tx) < 1$. Hence if $u_{n-i}(\cdot)$ were *linear* on the interval $(t_{n-i}^*(tx), tx)$, (5.7) would hold and would in fact be an *equality* if $u_{n-i}(tx) \ge 1$. So the most natural way to establish (5.7) is to prove the following lemma:

LEMMA 5.2. *For each n, $u_n(\cdot)$ is concave.*

PROOF. We proceed by induction and first note the lemma is true for $n = 1$ because $u_1(t) = t$.

Using (2.7), we rewrite (2.5) as

$$u_{n+1}(t) = t + (1-t)u_n(t) + t_n^*(t)(u_n(t) - 1) + \int_{t_n^*(t)}^{t} u_n(s)\, ds.$$

Formally we have

$$u'_{n+1}(t) = 1 - u_n(t) + (1-t)u'_n(t) - t_n^{*\prime}(t)(u_n(t) - 1)$$
$$+ t_n^*(t)u'_n(t) + u_n(t) - t_n^{*\prime}(t)u_n(t_n^*(t)).$$

Now $u_n(t_n^*(t)) = (u_n(t) - 1)^+$ and $t_n^{*\prime}(t) = 0$ if $u_n(t) < 1$. Hence whether or not $u_n(t) \geq 1$ we have

$$u'_{n+1}(t) = 1 + \{1 - (t - t_n^*(t))\}u'_n(t)$$

and

$$u''_{n+1}(t) = \{1 - (t - t_n^*(t))\}u''_n(t) - (t - t_n^*(t))'u'_n(t).$$

If $u_n(\cdot)$ is concave (i.e., $u''_n < 0$), then $(t - t_n^*(t))$ is increasing. Since, $t - t_n^*(t) < 1$ and $u'_n(\cdot) > 0$, we conclude that $u''_{n+1}(t) < 0$, i.e., $u_{n+1}(\cdot)$ is concave.

The validity of the foregoing differentiations are also easily established by induction. To begin note $u_1(t) = t$ and $t_1^*(t) \equiv 0$. Next the differentiability of u_n and u'_n implies $t_n^*(\cdot)$ is differentiable on $\{t: u_n(t) > 1\}$; in fact, we have $t_n^{*\prime}(t) = u'_n(t)/u'_n(t_n^*(t))$. By means of (2.5) one even more easily sees the required differentiability of $u_n(\cdot)$.

This completes the proof of the lemma, from which we obtain (5.7).

5c. *Completion of the proof.* At last we are ready to prove (5.1). We define

$$c_n = u_n/n^{1/2} = u_n(1)/n^{1/2}, \tag{5.10}$$

and note that it only remains to show $\liminf c_n \leq 2^{1/2}$. By (5.2),

$$c_n(nt)^{1/2} \geq u_n(t) \qquad t \in (0, 1]. \tag{5.11}$$

Abbreviate

$$t_n^* \equiv t_n^*(1)$$

and define s_n^* analogously by

$$c_n(ns_n^*)^{1/2} = c_n n^{1/2} - 1$$

so

$$s_n^* = 1 - 2c_n^{-1}n^{-1/2} + c_n^{-2}n^{-1}. \tag{5.12}$$

Now (5.10) and (5.11) imply that

$$s_n^* \leq t_n^* \tag{5.13}$$

so, rewriting (2.5), with $t = 1$, as

$$u_{n+1} = u_n + \int_{t_n^*}^{1} \{u_n(t) - (u_n - 1)\}\, dt,$$

we conclude from (5.10), (5.11), and (5.13) that

$$u_{n+1} \leq c_n n^{1/2} + \int_{s_n^*}^{1} \{c_n(nt)^{1/2} - (c_n n^{1/2} - 1)\}\, dt. \tag{5.14}$$

This is perhaps the central inequality in the proof, and it is made possible by (5.2). The remainder of the proof demands only straightforward analytical manipulation of the right side of (5.14).

Evaluating the right side of (5.14) we get

$$c_{n+1} \le c_n n^{1/2} + (2/3)c_n n^{1/2}(1 - s_n^{*3/2}) - (1 - s_n^*)(c_n n^{1/2} - 1). \tag{5.15}$$

Substituting (5.12) and the Taylor series expansion

$$1 - s_n^{*3/2} = (3/2)(1 - s_n^*) - (3/8)(1 - s_n^*)^2 + O((1 - s_n^*)^3),$$

into the right side of (5.15) gives

$$\begin{aligned} u_{n+1} &\le c_n n^{1/2} + c_n^{-1} n^{-1/4} + O(n^{-1}) \\ &= (n+1)^{1/2}\{c_n n^{1/2}(n+1)^{-1/2} + c_n^{-1} n^{-1/2}(n+1)^{-1/2} + O(n^{-3/2})\}. \end{aligned} \tag{5.16}$$

Now direct computation shows that

$$n^{1/2}(n+1)^{-1/2} = 1 - (1/2 - \delta_n)/n^{1/2}(n+1)^{1/2}$$

where $\delta_n > 0$ and $\delta_n \to 0$ as $n \to \infty$. Hence (5.16) is

$$u_{n+1} \le (n+1)^{1/2}\{c_n + n^{-1/2}(n+1)^{-1/2}(c_n^{-1} - \left(\frac{1}{2}\right)c_n + o(1))\}.$$

From (5.10), with $n + 1$ instead of n, we have

$$c_{n+1} \le c_n + n^{-1/2}(n+1)^{-1/2}\{c_n^{-1} - \left(\frac{1}{2}\right)c_n + o(1)\}. \tag{5.17}$$

This is exactly what we need to show that $\liminf c_n \le 2^{1/2}$ and thereby complete the proof.

We use the fact that $c^{-1} - c/2$ is decreasing in c and zero for $c = 2^{1/2}$. Choose $\epsilon > 0$ and n_ϵ large enough so that for all $n \ge n_\epsilon$ the $o(1)$ in (5.17) satisfies

$$o(1) < (1/2)\,|\,(2^{1/2} + \epsilon)^{-1} - (1/2)(2^{1/2} + \epsilon)\,| \equiv \delta_\epsilon.$$

Then for all $n \ge n_\epsilon$,

$$c_n > 2^{1/2} + \epsilon \Rightarrow c_{n+1} < c_n - \delta_\epsilon n^{-1/2}(n+1)^{-1/2}.$$

But

$$\sum n^{-1/2}(n+1)^{-1/2} = \infty$$

so $c_{n_0} > 2^{1/2} + \epsilon$ for some $n_0 \ge n_\epsilon$ implies $c_{n_0+m} < 2^{1/2} + \epsilon$ for some m. Since $\epsilon > 0$ is arbitrary this shows that $\liminf c_n \le 2^{1/2}$ as required to complete the proof.

We should remark that exactly the same argument can be used to prove that

$$u_n(t)/(nt)^{1/2} \to 2^{1/2}$$

for every t in (0, 1].

6. Random sample size and an open problem.

As an easy consequence of $u_n \sim (2n)^{1/2}$ one can obtain several results on subsequence selection when the underlying sample size N is random. In particular we now define u_N by

$$u_N = \sup_{\tau \in \mathscr{S}} E\{\max\{k : \tau_k \le N\}\}, \tag{6.1}$$

where $\mathscr{S}$ consists of those strategies adapted to $\{X_i\}_{i=1}^\infty$ but not adapted to N. When N is Poisson or binomial (with fixed p) one can easily show that as $EN \to \infty$ we have

$$u_N \sim (2EN)^{1/2}. \tag{6.2}$$

In fact one can check that the same result holds whenever $EN \to \infty$ and $\operatorname{Var} N = O(EN)$. (To compare these results with the asymptotic relations of Sections 3a and 5a one needs to note that the class of policies applied there were quite different from those used in (6.1) since they were also adapted to the relevant Poisson or binomial processes.)

We now consider the next most complex case where N has the geometrical distribution

($P(N = k) = p(1-p)^{k-1}$, $k = 1, 2, \cdots$). The condition Var $N = O(EN)$ does not hold, so as before the natural analysis begins with dynamic programming.

Analogously to (2.2) we define

$$u_N(t) = \sup_{\tau \in \mathcal{S}_t} E\{\max\{k : \tau_k \le N\}\}.$$

One can easily check in this case that

$$u_{N(p)}(t) = u_{N(p')}(1)$$

where

$$p' = p/\{p + t(1-p)\}.$$

If we define $f(p) = u_{N(p)}(1)$ we are led to a *single* integral equation:

$$f(p) = (1-p)\int_0^1 \max\{f(p), 1 + f(1/\{p + t(1-p)\})\}\, dt.$$

As we similarly noted at the end of Section 2 we can solve the equation for sufficiently extreme values. In this case we know

$$f(p) = \log(p^{-1}) \qquad \text{if} \quad p > e^{-1}.$$

This observation just says that if $f(p) \le 1$ the optimal policy is to select all record values exactly as one does in the fixed sample size problem when $u_n(t) \le 1$.

One would like to determine the asymptotic behavior of $f(p)$ as $p \to 0$. We conjecture, but have not been able to prove, that as $p \to 0$

$$f(p) \sim cp^{-1/2}$$

for a constant

$$c < 2^{1/2}.$$

7. A result of Burgess Davis on selections from permutations. If one considers a random permutation of the set $\{1, 2, \cdots, n\}$, then the distribution of the length of the longest decreasing subsequence is the same as that in a random sample of size n from a uniform distribution. In contrast, the length of the optimal *sequentially* selected decreasing subsequence is stochastically larger in the first case.

We let l_n denote the expected length of the longest decreasing subsequence which can be chosen sequentially from a random permutation. The main result of this paper immediately implies that

$$\liminf l_n/n^{1/2} \ge 2^{1/2}. \tag{7.1}$$

Part of the interest of this observation stems from the fact that the study of the l_n's was already a primary objective in Baer and Brock (1968), where "natural" is used as a synonym for "sequential." On the basis of substantial computation Baer and Brock even conjectured that $l_n \sim (2n)^{1/2}$. The truth of this conjecture is an immediate consequence of our main result $u_n \sim (2n)^{1/2}$ and the previously unpublished theorem due to Burgess Davis which is proved below.

THEOREM. (Burgess Davis).

$$l_n \sim u_n.$$

PROOF. First suppose X_i, $1 \le i \le \infty$, are i.i.d. uniform on [0, 1]. For any $\epsilon > 0$ and $0 < \delta < 1$, let $I_k^{(n)}$ denote the interval $((k-1)\epsilon/n^{1/2}, k\epsilon/n^{1/2}]$ and let $Y_k^{(n)}$ denote the cardinality of the set $\{i : 1 \le i \le n, X_i \in I_k^{(n)}\}$. Consider the events

$$A_n = \{\min_{1 \le k \le n^{1/2}/\epsilon} Y_k^{(n)} \ge (1-\delta)\epsilon n^{1/2}\}$$

and note that elementary binomial estimates show $P(A_n) \to 1$ as $n \to \infty$.

Now assign "pretend ranks" as follows: if $X_i \in I_k^{(n)}$ and X_i is one of the first $[(1 - \delta)\epsilon n^{1/2}]$ elements of $I_k^{(n)}$ we chose its "pretend rank" at random from those integers between $(k-1)(1-\delta)\epsilon n^{1/2}$ and $k(1-\delta)\epsilon n^{1/2}$ which are not already assigned. We then ignore all X_i's in $I_k^{(n)}$ after the first $[(1-\delta)\epsilon n^{1/2}]$. We note that on A_n the sequence of "pretend ranks" is simply a permutation on the integers $\{k : 1 \le k \le (1-\delta)n\}$.

Now use the optimal random permutation policy on the set of "pretend ranks" to select a subsequence with decreasing pretend ranks. Delete from the subsequence all X_i's which are not smaller than all their predecessors in the subsequence. This gives a decreasing subsequence.

We now claim that the expected cardinality of the resulting subsequence is at least $P(A_n)(l_m - \epsilon n^{1/2})$ with $m = [(1-\delta)n]$. First note, on A_n the expected length before deletions is l_m. Let $J_i^{(n)}$ denote the interval $I_k^{(n)}$ which contains the smallest *selected* element from $\{X_1, X_2, \cdots, X_i\}$. For X_{i+1} to be an observation deleted from the selected subsequence it is necessary that $X_{i+1} \in J_i^{(n)}$. By Boole's inequality and the independence of X_{i+1} and $J_i^{(n)}$ we have that the expected number of deleted observations is at most

$$\sum_{i=0}^{n-1} P(X_{i+1} \in J_i^{(n)}) = n(\epsilon/n^{1/2}) = \epsilon n^{1/2}.$$

This proves $l_m/n^{1/2} \le P(A_n)^{-1}u_n/n^{1/2} + \epsilon$. By (7.1) and the arbitrariness of ϵ and δ the theorem is proved.

REFERENCES

BAER, R. M. and BROCK, P. (1968). Natural sorting over permutation spaces. *Math. Comp.* **22** 385–410.

FEKETE, M. (1923). Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten. *Math. Z.* **17** 228–249.

GILBERT, J. P. and MOSTELLER, F. (1966). Recognizing the maximum of a sequence. *J. Amer. Statist. Assoc.* **61** 35–73.

HAMMERSLEY, J. M. (1972). A few seedlings of research. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 345–394.

KHINCHIN, A. I. (1957). *Mathematical Foundations of Information Theory.* 48. Dover, New York.

LOGAN, B. F. and SHEPP, L. A. (1977). A variational problem for random Young tableaux. *Advances in Math.* **26** 206–222.

VERŠIK, A. M. and KEROV, S. V. (1977). Asymptotics of the Plancherel measure of the symmetric group and the limiting form of the Young tables. *Soviet Math. Dokl.* **18** 527–531.

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305