

האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

**SECRETARY PROBLEM OF MINIMIZING
EXPECTED RANK: A SIMPLE SUBOPTIMAL
APPROACH WITH GENERALIZATIONS**

by

ABBA M. KRIEGER and ESTER SAMUEL-CAHN

Discussion Paper # 502

January 2009

מרכז לחקר הרציונליות

**CENTER FOR THE STUDY
OF RATIONALITY**

Feldman Building, Givat-Ram, 91904 Jerusalem, Israel

PHONE: [972]-2-6584135 FAX: [972]-2-6513681

E-MAIL: ratio@math.huji.ac.il

URL: <http://www.ratio.huji.ac.il/>

The Secretary Problem of Minimizing Expected Rank: A Simple Suboptimal Approach with Generalizations

Abba M. Krieger ^{*} and Ester Samuel-Cahn [†]
University of Pennsylvania and Hebrew University

January 15, 2009

Abstract

The secretary problem for selecting one item so as to minimize its expected rank, based on observing the relative ranks only, is revisited. A simple suboptimal rule, which performs almost as well as the optimal rule, is given. The rule stops with the smallest i such that $R_i \leq ic/(n+1-i)$ for a given constant c , where R_i is the relative rank of the i^{th} observation, and n is the total number of items. This rule has added flexibility. i) A curtailed version thereof can be used to select an item with a given probability P , $P < 1$. ii) The rule can be used to select two or more items. The problem of selecting a fixed proportion, α , $0 < \alpha < 1$, of n , is also treated. Numerical results are included to illustrate the findings.

2000 Mathematical Subject Classification: Primary 62L99 ; Secondary 62F07; 60F15; 60F15.

Keywords: Secretary problem, group selection rules, suboptimal rule, near optimal rule, absolute rank, relative rank.

^{*}Postal address: Department of Statistics of the Wharton School of the University of Pennsylvania, Philadelphia 19104, PA, USA. Email address: krieger@wharton.upenn.edu. Supported by funds from the Marcy Bogen Chair of Statistics at the Hebrew University of Jerusalem.

[†]Postal address: Department of Statistics (and Center for Rationality), The Hebrew University of Jerusalem, Jerusalem 91905, Israel. Email address: scahn@mscc.huji.ac.il. Supported by the Israel Science Foundation Grant 467/04.

1 Introduction

Consider a group of n items that arrive sequentially and are rankable without ties. The number of items n is known. A rank of 1 denotes the best item, a rank of 2 the second best, etc. It is assumed that all $n!$ permutations of the ranks $\{1, \dots, n\}$ are equally likely.

The problem is to establish a rule that decides on whether item i is to be chosen based on its relative rank, R_i , i.e., its rank among the first i items. Just as in the classical secretary problem, once a decision is made on an item, there is no recourse.

The performance of the rule depends on the absolute rank of item i , A_i^n , i.e., its rank among all n items, if it is chosen. In the classical secretary problem the objective is to choose a single item by means of an optimal stopping rule τ_n that maximizes $P(A_{\tau_n}^n = 1)$.

Chow et. al. [2] consider the problem of finding an optimal stopping rule τ_n for choosing one item minimizing the expected value of the absolute rank of the item selected. They show that the optimal stopping rule, satisfies

$$\lim_{n \rightarrow \infty} E(A_{\tau_n}^n) = \prod_{j=1}^{\infty} \left(1 + \frac{2}{j}\right)^{1/(j+1)} = 3.869$$

This is an astonishing result since it shows that, for example, from one million items one can sequentially select an item with expected rank less than 4. Note that since one item *must* be chosen, if one has not stopped earlier, one must take one of the last few items. Since the last few items have expected absolute rank $(n+1)/2$, the probability of this happening must be at most $O(\frac{1}{n})$, for the expected rank of the chosen item not to tend to infinity. The Chow et. al. [2] paper is heavily cited and has spawned many variations.

The optimal stopping rule in Chow et. al. for keeping one item, can in principle be determined by dynamic programming and so in practice is hard to implement for large n . The asymptotic form of the rule is to pass on the first $\gamma_0 n$ items; choose the first item with $R_i = 1$ after $i > \gamma_0 n$ and up to $\gamma_1 n$; an item with $R_i \leq 2$ after $\gamma_1 n$ (if none has been chosen before) and up to $\gamma_2 n$ etc. The values of the γ_i 's can be computed asymptotically. For example, from Ferguson [4] (Page 2.6), $\gamma_0 = .2584$, $\gamma_1 = .4476$, and $\gamma_2 = .5640$.

We consider finding *simple* stopping rules that perform well in minimizing the sum of the expected value of the absolute ranks of the items selected, when one or more items are desired. In Section 2, we pose a class of rules

which are easy to implement, and specifically a rule that achieves asymptotic expected absolute rank of 3.928, nearly the same as the optimal rule. The rules that are posed have added flexibility. They allow for stopping rules that select an item only with high probability, P , of stopping. One can find rules that do substantially better than the optimal expected absolute rank of 3.89 by relaxing the "must" select an item only slightly. For example when the probability of stopping is at least 95% the asymptotic expected absolute rank (conditional on stopping) is 2.987. Yeo and Yeo [7] also consider stopping rules for secretary problems which do not necessarily stop with probability one, and include some asymptotic evaluations. Our approach is different from theirs in that we are considering simple suboptimal rules that perform well.

In Section 3, we consider the problem of selecting a fixed number, r , of items, $r > 1$. The optimal stopping rule can again be determined by dynamic programming. The simple rule we establish for $r = 1$ can be adapted to this problem as well, again producing nearly optimal results.

The above rules fix the number of items selected. Other rules select a random number of items satisfying a condition on the average number of items retained, for example the average number that is retained is a given proportion of n . The problem of sequentially selecting a group of items on the basis of their relative ranks, has recently been described in Krieger et. al. [5]. They consider a class of rules, termed, "percentage rules", for $0 < p \leq 1$, with the property that they retain on average an order of n^p items, most of which are good. Simple rules that perform asymptotically optimally in terms of the sum of the absolute ranks and retain (approximately) a fixed given fraction α are described in Section 4. For the rules in Krieger et. al. [5], as well as for the rules in Section 4, no prior knowledge of n is needed. In Section 5 we supply tables illustrating the results obtained in the earlier sections.

2 Retaining One Item

The optimal rule to retain one item that minimizes the expected value of the absolute rank can be found using dynamic programming for any n . This rule can be described by a sequence of integers $r_n(j)$; $j = 0, \dots, n$ with $1 \leq r_n(0) \leq r_n(1) \leq \dots \leq r_n(n) = n + 1$. Let i denote the index of the item. If $i < r_n(0)$ then the item is never accepted. In general, item i is chosen when its relative rank $R_i \leq j$ and $i < r_n(j)$, provided no item has been retained earlier.

In this section we consider simple rules that are easy to implement, (i.e., with values of $r_n(j)$ that are readily available). To this end, let $c \geq 1$ be a fixed constant, and let the stopping rule be

$$t_n(c) = \inf \left\{ i : R_i \leq \frac{ic}{n+1-i} \right\}. \quad (2.1)$$

Note that the stopping rule $t_n(c)$ satisfies $P(t_n(c) \leq n) = 1$.

We shall suppress the dependence on c in some of the following notation, where we consider c as fixed. Let $a_n(k)$ be the smallest integer greater than or equal to $\frac{(n+1)k}{(c+k)}$, i.e.,

$$a_n(k) = \lceil \frac{(n+1)k}{(c+k)} \rceil, \quad k = 1, \dots, n.$$

Note that for small values of k (as compared to n) the values of $a_n(k)$ will be distinct, but for large values of k some consecutive $a_n(k)$ may coincide. Clearly $t_n(c)$ of (2.1) can be written as

$$t_n(c) = \inf \{ i : R_i = k, \quad \text{for some } i \geq a_n(k), \quad k = 1, \dots, n \}. \quad (2.2)$$

One of our main interests will be in a "curtailed" version of (2.2), which stops with probability less than one. Let k_s be a fixed integer with $k_s < n$. Define

$$t_{k_s, n}(c) = \inf \{ i : R_i = k, \quad \text{for some } i \geq a_n(k), \quad k = 1, \dots, k_s \}. \quad (2.3)$$

Clearly $t_{k_s, n}(c)$ will stop with probability less than 1. For fixed c and k_s we are interested in finding the limiting probability of stopping, and the expected absolute rank upon stopping, as $n \rightarrow \infty$. This is used later to determine c and k_s to attain a desired probability P of stopping.

Let $b_n(k) = a_n(k+1) - 1$ for $k = 1, \dots, k_s - 1$, and let $b_n(k_s) = n$. Assume that n is large, so that $b_n(k) > a_n(k)$ for $k = 1, \dots, k_s$. Let $Q_n(k) = \{ R_i > j \text{ for all } i \leq b_n(j) \text{ and all } j = 1, \dots, k \}$. Thus,

$$P(Q_n(k)) = P \{ t_{k_s, n}(c) > b_n(k) \} \text{ for } k = 1, \dots, k_s - 1,$$

(where not stopping is considered $t = \infty$) and

$$P_n(k_s, c) = P \{ t_{k_s, n}(c) < \infty \} = 1 - P(Q_n(k_s)), \quad (2.4)$$

where $Q_n(k_s)$ is defined by the above expression and let $Q_n(0)$ be the entire space.

Lemma 2.1. For $k \leq k_s$, let

$$f(k) := \lim_{n \rightarrow \infty} P(Q_n(k)|Q_n(k-1)) = \lim_{n \rightarrow \infty} \prod_{j=a_n(k)}^{b_n(k)} \left(1 - \frac{k}{j}\right). \quad (2.5)$$

Then for $k = 1, \dots, k_s - 1$

$$f(k) = \left\{ \frac{k(k+c+1)}{(k+c)(k+1)} \right\}^k \quad (2.6)$$

and

$$f(k_s) = \left\{ \frac{k_s}{k_s+c} \right\}^{k_s}. \quad (2.7)$$

Proof. To see (2.6) note that

$$\begin{aligned} \lim_{n \rightarrow \infty} \left[\log \prod_{j=a_n(k)}^{b_n(k)} \left(1 - \frac{k}{j}\right) \right] &= \lim_{n \rightarrow \infty} \sum_{j=a_n(k)}^{b_n(k)} \log \left(1 - \frac{k}{j}\right) = \lim_{n \rightarrow \infty} \left(-k \sum_{j=a_n(k)}^{b_n(k)} \frac{1}{j} \right) \\ &= -k \lim_{n \rightarrow \infty} [\log b_n(k) - \log a_n(k)] = \lim_{n \rightarrow \infty} \log \left\{ \frac{a_n(k)}{b_n(k)} \right\}^k = \log \left\{ \frac{k(k+c+1)}{(k+c)(k+1)} \right\}^k. \end{aligned}$$

Equation (2.7) follows similarly. \square

Proposition 2.1.

$$\begin{aligned} P(k_s, c) &:= \lim_{n \rightarrow \infty} P_n(k_s, c) = \lim_{n \rightarrow \infty} P\{t_{k_s, n}(c) < \infty\} \\ &= 1 - \prod_{k=1}^{k_s} f(k) = 1 - \frac{k_s!}{\prod_{k=1}^{k_s} (k+c)}. \end{aligned} \quad (2.8)$$

Proof. This follows from (2.4), (2.5) and Lemma 2.1. \square

To compute the expected absolute rank upon stopping note that

$$E(A_i^n | R_i) = \frac{n+1}{i+1} R_i, \quad (2.9)$$

where A_i^n is the absolute rank, among n , of the i^{th} observations. It follows that $E(A_i^n | R_i \leq k) = \frac{(n+1)(k+1)}{2(i+1)}$. We are interested in computing $E\left(A_{t_{k_s, n}^n}^n(c) | t_{k_s, n}(c) < \infty\right) := E(n, k_s, c)$, as well as its limiting value, to be

denoted by $E(k_s, c)$. Let $V_n(k) = P(t_{k_s, n}(c) > b_n(k-1))$ for $k = 2, \dots, k_s$, and $V_n(1) = 1$. Note that for $k = 2, \dots, k_s$, from (2.6)

$$\lim_{n \rightarrow \infty} V_n(k) := V(k) = \prod_{j=1}^{k-1} f(j) = \frac{(k-1)!}{\prod_{j=1}^{k-1} (j+c)} \left(\frac{k+c}{k} \right)^{k-1}. \quad (2.10)$$

We now find the conditional (on an item being retained) expected absolute rank, $E(n, k_s, c)$ and its limit as $n \rightarrow \infty$, $E(k_s, c)$ in

Theorem 2.1.

$$E(k_s, c) = \left\{ \sum_{j=1}^{k_s} \alpha_{j-1} \beta_j - k_s \alpha_{k_s} \right\} / 2P(k_s, c) \quad (2.11)$$

and when $c > 1$

$$\lim_{k_s \rightarrow \infty} E(k_s, c) := E(c) = \frac{1}{2} \sum_{j=1}^{\infty} \alpha_{j-1} \beta_j, \quad (2.12)$$

where

$$\alpha_m = \frac{m!}{\prod_{j=1}^m (j+c)} = \left[\prod_{j=1}^m \left(1 + \frac{c}{j} \right) \right]^{-1},$$

and

$$\beta_m = \left(\frac{m+c}{m} \right)^m = \left(1 + \frac{c}{m} \right)^m$$

with $\alpha_0 = 1$.

Proof. We shall compute the expected value $E(n, k_s, c)$ by considering in which of the disjoint intervals, $[a_n(k), b_n(k)]$ for $k = 1, \dots, k_s$, and at which value i inside that interval the rules stop. It is easily seen that

$$E(n, k_s, c) = \sum_{k=1}^{k_s} V_n(k) \left\{ \sum_{i=a_n(k)}^{b_n(k)} \left[\prod_{m=a_n(k)}^{i-1} \left(1 - \frac{k}{m} \right) \right] \frac{k(n+1)(k+1)}{i \cdot 2(i+1)} \right\} / P_n(k_s, c). \quad (2.13)$$

To evaluate the limit, note that for $k \leq k_s$ the value of i will be a fraction of n . Thus denote $i/n = x$. The term $\prod_{m=a_n(k)}^{i-1} \left(1 - \frac{k}{m} \right) = \frac{(i-1-k) \dots (a_n(k)-k)}{(i-1) \dots a_n(k)}$. Dividing the numerator and denominator by n and letting $n \rightarrow \infty$ yields

$\left(\frac{k}{k+c}\right)^k$ in the numerator and x^k in the denominator. Then it follows that for $k < k_s$

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sum_{i=a_n(k)}^{b_n(k)} \left[\prod_{m=a_n(k)}^{i-1} \left(1 - \frac{k}{m}\right) \right] \frac{n+1}{i(i+1)} = \left[\frac{k}{k+c}\right]^k \int_{\frac{k}{k+c}}^{\frac{k+1}{k+1+c}} x^{-(k+2)} dx \\ & = \left[\frac{k}{k+c}\right]^k \left[\left(\frac{k+c}{k}\right)^{k+1} - \left(\frac{k+1+c}{k+1}\right)^{k+1} \right] / (k+1) \end{aligned} \quad (2.14)$$

and similarly, the corresponding limit for k_s is

$$\left[\frac{k_s}{k_s+c}\right]^{k_s} \int_{\frac{k_s}{k_s+c}}^1 x^{-(k_s+2)} dx = \left[\frac{k_s}{k_s+c}\right]^{k_s} \left[\left(\frac{k_s+c}{k_s}\right)^{k_s+1} - 1 \right] / (k_s+1). \quad (2.15)$$

Taking limits of (2.13) by substituting the limits in (2.10), (2.14) and (2.15) and simplifying somewhat, we get

$$\begin{aligned} E(k_s, c) &= \left\{ \sum_{k=1}^{k_s-1} \frac{k!k}{\prod_{j=1}^k (j+c)} \left[\left(\frac{k+c}{k}\right)^{k+1} - \left(\frac{k+1+c}{k+1}\right)^{k+1} \right] \right. \\ & \quad \left. + \frac{k_s!k_s}{\prod_{j=1}^{k_s} (j+c)} \left[\left(\frac{k_s+c}{k_s}\right)^{k_s+1} - 1 \right] \right\} / 2P(k_s, c), \end{aligned} \quad (2.16)$$

where $P(k_s, c)$ is given in (2.8). Opening the square bracket and canceling yields

$$E(k_s, c) = \left\{ \sum_{k=1}^{k_s} \frac{(k-1)!}{\prod_{j=1}^{k-1} (j+c)} \left(\frac{k+c}{k}\right)^k - \frac{k_s!k_s}{\prod_{j=1}^{k_s} (j+c)} \right\} / 2P(k_s, c).$$

The result (2.11) follows from the definitions of α_m and β_m .

Note that the negative term in (2.11), for $c = 1$ equals $k_s/(k_s+1) < 1$, and is a decreasing function of c . We are interested in $E(c) = \lim_{k_s \rightarrow \infty} E(k_s, c)$. Clearly, from (2.8), $\lim_{k_s \rightarrow \infty} P(k_s, c) = 1$. Since $\alpha_m m^c \rightarrow \text{constant}$ as $m \rightarrow \infty$ and $\beta_m \rightarrow e^c$ as $m \rightarrow \infty$, it follows that for $c > 1$, $\lim_{k_s \rightarrow \infty} k_s \alpha_{k_s} \rightarrow 0$, and we have (2.12). \square

When $0 < c \leq 1$, however, the sum in (2.12) is infinite.

An important issue is to show that the rule in (2.2) that is not curtailed behaves like the curtailed rule with $k_s \rightarrow \infty$, as given in (2.12), when $n \rightarrow \infty$. It is sufficient to consider the curtailed rule with $n = k_s$ as $n \rightarrow \infty$.

Theorem 2.2. *Let $t_n(c)$ be as in (2.1) and $c > 1$. Then*

$$\lim_{n \rightarrow \infty} EA_{t_n(c)}^n = E(c)$$

where $E(c)$ is given in (2.12).

Proof. In order to show that the convergence of $E(n, k_s, c)$ is uniform in k_s as $n \rightarrow \infty$, we need to consider the remainder term from (2.13) (for $n > k_0$ large)

$$R_n(k_0) = \sum_{k=k_0}^{k_s} V_n(k) \left\{ \sum_{i=a_n(k)}^{b_n(k)} \left[\prod_{m=a_n(k)}^{i-1} \left(1 - \frac{k}{m}\right) \right] \frac{k(n+1)(k+1)}{i \cdot 2(i+1)} \right\}. \quad (2.17)$$

We want to show that $R_n(k_0)$ can be made arbitrarily small by choosing k_0 to be sufficiently large. We consider (2.17) in three parts:

1. For $V_n(k)$ we have

$$V_n(k) = \prod_{j=1}^{k-1} \sum_{m=a_n(j)}^{b_n(j)} \left(1 - \frac{j}{m}\right) < \prod_{j=1}^{k-1} \left(1 - \frac{j}{b_n(j)}\right)^{b_n(j) - a_n(j)}.$$

But $b_n(j) - a_n(j) \geq \frac{(n+1)c}{(c+j+1)(c+j)} - 1$. Hence with k and n fixed, for A , B and D (used later) positive finite constants,

$$\begin{aligned} V_n(k) &< \exp \left[- \sum_{j=1}^{k-1} \frac{j(c+j+1)}{(n+1)(j+1)} \left(\frac{(n+1)c}{(c+j+1)(c+j)} - 1 \right) \right] \\ &< A \exp \left[- c \sum_{j=1}^{k-1} \frac{j}{(j+1)(c+j)} \right] = A \exp \left[- c \left(\sum_{j=1}^{k-1} \frac{1}{c+j} - \sum_{j=1}^{k-1} \frac{1}{(j+1)(c+j)} \right) \right] \\ &< B \exp \left[- c \log(c+k-1) \right] = B(c+k-1)^{-c}. \end{aligned} \quad (2.18)$$

2. We next consider the middle term in (2.17), which refers to the probability of obtaining a value in $[a_n(k), b_n(k)]$ which would force us to stop for a given k . Note that for k_s we can bound this probability trivially by 1. For $k < k_s$ an upper bound is $\frac{c}{k+1}$. The upper bound follows from

$$\sum_{i=a_n(k)}^{b_n(k)} \prod_{m=a_n(k)}^{i-1} \left(1 - \frac{k}{m}\right) \frac{k}{i} = 1 - \prod_{m=a_n(k)}^{b_n(k)} \left(1 - \frac{k}{m}\right) \leq 1 - \left(1 - \frac{k(c+k)}{(n+1)k}\right)^{\frac{(n+1)c}{(c+k+1)(c+k)}}$$

$$\leq \left(\frac{k(c+k)}{(n+1)k} \right) \left(\frac{(n+1)c}{(c+k)(c+k+1)} \right) = \frac{c}{c+k+1} \leq \frac{c}{k+1}. \quad (2.19)$$

3. Finally, we bound the expected absolute rank, for i in $[a_n(k), b_n(k)]$, by

$$\frac{(n+1)(k+1)}{2(i+1)} \leq \frac{(n+1)(k+1)(c+k)}{2(n+1)k} < (c+1)(k+1). \quad (2.20)$$

Substituting (2.20), (2.18) and (2.19) into (2.17) yields

$$R_n(k_0) \leq D \sum_{k=k_0}^{\infty} (c+k-1)^{-c} + (c+1)(k_s+1)B(c+k_s-1)^{-c}. \quad (2.21)$$

It is clear for $c > 1$, that the sum in (2.21) converges, and hence that sum can be made arbitrarily small by choosing k_0 sufficiently large. Likewise the last term, involving k_s , tends to 0 as $k_s \rightarrow \infty$. So for any $\epsilon > 0$, k_0 can be chosen to be suitably large to ensure that $R_n(k_0) < \epsilon$, as desired. \square

We have a class of rules that depends on the value of c . Since we are free to choose any c , the issue is to find the value of c that minimizes $E(c)$ in (2.12). The following Theorem shows that this minimum exists and is unique.

Theorem 2.3. $E(c) = \frac{1}{2} \sum_{i=1}^{\infty} \alpha_{i-1} \beta_i$ is a strictly convex function of c .

Proof. Let $g_i = \log(\alpha_{i-1} \beta_i)$. Then, $E(c) = \frac{1}{2} \sum_{i=1}^{\infty} e^{g_i}$, $E'(c) = \frac{1}{2} \sum_{i=1}^{\infty} e^{g_i} g'_i$, and $E''(c) = \frac{1}{2} \sum_{i=1}^{\infty} e^{g_i} [(g'_i)^2 + g''_i]$.

It is therefore sufficient to show that $(g'_i)^2 + g''_i \geq 0$ for all i , with at least one inequality strict. It is straightforward to show that it is true for $i = 1$ since $g_1 = \log(c+1)$. For $i > 1$

$$g_i = i \log(c+i) - i \log i + \sum_{j=1}^{i-1} \log j - \sum_{j=1}^{i-1} \log(c+j).$$

Hence

$$g'_i = \frac{i}{c+i} - \sum_{j=1}^{i-1} \frac{1}{c+j} = \left[1 - \sum_{j=1}^{i-1} (h_{i,j} - 1) \right] / (c+i),$$

where $h_{i,j} = \frac{c+i}{c+j} > 1$ for $1 \leq j \leq i-1$. Similarly,

$$g''_i = \frac{-i}{(c+i)^2} + \sum_{j=1}^{i-1} \frac{1}{(c+j)^2} = \left[-1 + \sum_{j=1}^{i-1} (h_{i,j}^2 - 1) \right] / (c+i)^2.$$

Hence

$$\begin{aligned}
(c+i)^2 [(g'_i)^2 + g''_i] &= \left[1 - \sum_{j=1}^{i-1} (h_{i,j} - 1) \right]^2 - 1 + \sum_{j=1}^{i-1} (h_{i,j} - 1)(h_{i,j} + 1) \\
&= 1 - 2 \sum_{j=1}^{i-1} (h_{i,j} - 1) + \left[\sum_{j=1}^{i-1} h_{i,j} - 1 \right]^2 - 1 + \sum_{j=1}^{i-1} (h_{i,j} - 1)(h_{i,j} + 1) > 0.
\end{aligned}$$

The inequality holds because $h_{i,j} + 1 > 2$. \square

In Section 5 we list some values of the expected rank of the optimal rule, and compare them with the corresponding values of our simple rule, both when the latter stops with probability one, and with a given probability less than one. We find, numerically, the optimal c -values, which when stopping with probability one is about 4.2 and is not very sensitive to n , as seen in Tables 1 and 2.

3 Retaining More than One Item

In this section, we consider the performance of the simple rules given by (2.3) where we allow the rules to continue until a given number of items, r , are retained. To this end, we extend the formula (2.8) for $P(k_s, c)$, which provides the probability that at least one item is retained when $n \rightarrow \infty$, to the entire distribution for the number of items that these rules retain, asymptotically in n . Equation (2.11) that provides the expected absolute rank for the item that is retained by this rule is extended to the corresponding value when $r > 1$. Finally, a discussion of the implementation of the optimal dynamic rule and its behavior (in terms of the relative rank that is required for retention as a function of the item number) for fixed n is provided.

Let Z_n denote the probability distribution for the number of items that could be retained using the simple rule. Clearly, $Z_n = \sum_{i=1}^n I_i$, where I_1, \dots, I_n are a sequence of independent 0,1 random variables with $P(I_i = 1) = p_{in} = \min\left(\lfloor \frac{ic}{n+1-i} \rfloor, k_s\right) / i$ where $\lfloor x \rfloor$ is the largest integer less than or equal to x . Then Z_n converges uniformly to a Poisson random variable (c.f., Barbour et. al. [1], page 3). Let $p_{(n)} = \max p_{in}$, $i = 1, \dots, n$. Then

$$P(Z_n = r) = \exp(-\lambda_n) \lambda_n^r / (r!) \exp\{O(\lambda_n p_{(n)}, r^2 \lambda_n^{-1} p_{(n)})\},$$

where $\lambda_n = \sum_{i=1}^n p_{in}$.

It remains to find the parameter λ^* of the limiting Poisson distribution as $n \rightarrow \infty$. Following arguments similar to those in the previous section

$$\begin{aligned}\lambda^* &= \lim_{n \rightarrow \infty} \sum_{k=1}^{k_s} \sum_{i=a_n(k)}^{b_n(k)} \frac{k}{i} \\ &= - \sum_{k=1}^{k_s} k \log(f_k) = - \sum_{k=1}^{k_s} \log\left(\frac{k}{k+c}\right).\end{aligned}$$

Specifically, it follows from Barbour et. al. [1] that

Proposition 3.1.

$$\lim_{n \rightarrow \infty} P(Z_n = r) = \exp^{\lambda^*} \frac{\lambda^{*r}}{r!}.$$

The probability of retaining one item as $n \rightarrow \infty$ is $1 - \lim_{n \rightarrow \infty} P(Z_n = 0) = 1 - \exp(-\lambda^*)$ which agrees with the result (2.8).

We now turn to the limiting expected sum of the absolute ranks, denoted by $E_2(k_s, c)$, for keeping $r = 2$ items, conditional on two items being kept, for simple rules given by (2.3), for fixed k_s , as $n \rightarrow \infty$. Let $A_2(k_s, c)$ be the limiting expected absolute rank of the second item that is kept, assuming that the first two available items (by the rule) are kept. Then the expected absolute sum of ranks from the two items kept (conditional on two items being kept), is

$$E_2(k_s, c) = [\bar{F}(0)E(k_s, c) + A_2(k_s, c)] / \bar{F}(1) \quad (3.1)$$

where $\bar{F}(x) = 1 - F(x)$ and F is the cumulative distribution function of a Poisson random variable with parameter λ^* .

The expected absolute rank from the second item that is kept for fixed n can be written in a form similar to (2.13) by summing over instances that item k is the first one that is retained. Similar limiting arguments as $n \rightarrow \infty$ yield

$$\begin{aligned}A_2(k_s, c) &= \sum_{k=1}^{k_s} V(k) \left\{ \left[\frac{k^2}{2(k+1)} \left(\frac{c+k}{k} - f(k) \frac{c+k+1}{k+1} \right) \right] + f(k) \frac{k}{2} \frac{c+k+1}{k+1} \log f(k) \right\} \\ &\quad - \sum_{k=1}^{k_s-1} \log f(k) \sum_{j=k+1}^{k_s} V(j) \frac{j}{2} \left[\frac{c+j}{j} - f(j) \frac{c+j+1}{j+1} \right].\end{aligned} \quad (3.2)$$

Substituting (2.16) and (3.2) into (3.1) yields the desired result.

The rule can be modified so that the value of c until the first item is retained, c_1 , is different from the value of c for selecting the second item, c_2 . The numerical results in Table 5 show that for the optimal choice of these values, $c_1 > c_2$. This makes sense as this implies that the first item will be chosen rather early to allow for a sufficient number of items from which to make the second choice. This extends to retaining more than two items by means of different c values.

We now turn to the dynamic programming required to find the optimal rule of selecting r items, with probability one, $r \leq n$. Let (m, j) be the state that the relative rank of item m is being observed and j additional items are required. Let $C(m, j)$ be the minimum over all stopping rules of the sum of the expected absolute ranks from the last j items that are retained given the state is (m, j) . The optimal solution to retain r is then $C(1, r)$. This implies that the optimal solution for retaining r out of n is obtained simultaneously for all $r = 1, \dots, n$ by one run of the dynamic program. Let $h(m, j)$ be the largest relative rank of an item to be retained in the optimal solution given that the state is (m, j) .

The values of $C(m, j)$ and $h(m, j)$ are determined recursively as follows:

$$h(n, 1) = n$$

and

$$C(n, 1) = \frac{n+1}{2}; \quad C(n, j) = \infty \text{ for } j > 1.$$

Furthermore, since it is preferred to retain item m with $R_m = k$ in state (m, j) if

$C(m+1, j-1) + \frac{(n+1)k}{(m+1)} \leq C(m+1, j)$ it follows that

$$h(m, j) = \left\lfloor \frac{(C(m+1, j) - C(m+1, j-1))(m+1)}{n+1} \right\rfloor.$$

This implies that

$$C(m, j) = \frac{h(m, j)}{m} \left[C(m+1, j-1) + \frac{(n+1)(h(m, j) + 1)}{2(m+1)} \right] + \left(1 - \frac{h(m, j)}{m} \right) C(m+1, j).$$

Some values of $C(1, r)$ for various values of n are given in Table 4.

4 Rules That Retain a Given Percentage of Items

In this section we consider simple rules which, for a fixed given fraction α , $0 < \alpha < 1$, retain (approximately) the fraction α of the population of "good"

items. More specifically, if for fixed α , we denote by $L_n(\alpha)$ the number of items retained by the rules considered, we show that

$$\lim_{n \rightarrow \infty} EL_n(\alpha)/n = \alpha \quad (4.1)$$

and

$$L_n(\alpha)/n \rightarrow \alpha \text{ a.s. as } n \rightarrow \infty. \quad (4.2)$$

Let $S_n(\alpha)$ denote the sum of the absolute ranks of the items retained by the rule. We furthermore show that

$$\lim_{n \rightarrow \infty} E \frac{S_n(\alpha)}{nE(L_n(\alpha))} = \alpha/2. \quad (4.3)$$

Note that this is optimal in the sense that if one could inspect all the items first, and pick the $\lceil \alpha n \rceil$ best items, then the sum of the ranks, B_n would satisfy

$$\lim_{n \rightarrow \infty} \frac{B_n}{n \lceil \alpha n \rceil} = \lim_{n \rightarrow \infty} \frac{\lceil \alpha n \rceil + 1}{2n} = \frac{\alpha}{2}.$$

Let $0 < \alpha < 1$ be fixed. We omit the index α . Consider

Rule 1. Keep item j if and only if $R_j \leq \lceil \alpha j \rceil$.

Note that knowledge of n is not required to implement the rule. Let $I_j = 1$ or 0, according to whether or not the j^{th} item is kept. Then $L_n = \sum_{j=1}^n I_j$, and

$$\alpha \leq EI_j < \alpha + \frac{1}{j}, \quad j = 1, \dots, n$$

Thus,

$$n\alpha \leq EL_n < \alpha n + \log n + 1, \quad (4.4)$$

and (4.1) follows. The second limit, (4.2), follows by noting that, by the independence of the R_j 's, the I_j 's are independent (almost identically distributed) Bernoulli variables, and applying the SLLN.

Here $S_n = \sum_{j=1}^n A_j^n I_j$, thus

$$\begin{aligned} ES_n &= \sum_{j=1}^n E(A_j^n I_j) = \sum_{j=1}^n E(EA_j^n I_j | R_j) = (n+1) \sum_{j=1}^n \frac{1}{j(j+1)} \sum_{k=1}^{\lceil j\alpha \rceil} k \\ &= \frac{n+1}{2} \sum_{j=1}^n \frac{\lceil j\alpha \rceil (\lceil j\alpha \rceil + 1)}{j(j+1)} = \frac{n+1}{2} [n\alpha^2 + o(n)]. \end{aligned} \quad (4.5)$$

Combining (4.5) and (4.4) yields (4.3).

We now show that the sum of the absolute ranks, S_n , suitably normalized, converges a.s. We first prove a result that shows that for any α the normalized sum of the relative ranks of kept items converges a.s. to half of the sum of the resulting absolute ranks.

Lemma 4.1. *Let $T_n = \sum_{i=1}^n R_i I_i$. Then*

$$\frac{T_n}{nL_n} \rightarrow \frac{\alpha}{4} \text{ almost surely as } n \rightarrow \infty. \quad (4.6)$$

Proof. Let $U_i = R_i I_i - \frac{v_i(v_i+1)}{2i}$ where $v_i = \lceil \alpha i \rceil$. Then U_i are independent random variables with mean zero. Let $b_i = i^2$. Since

$$\sum_{i=1}^{\infty} E\left(\frac{U_i^2}{b_i^2}\right) < \sum_{i=1}^{\infty} \frac{v_i(v_i+1)(2v_i+1)}{6i^5} < \infty,$$

it follows from Feller [3] (Theorem 2, page 239) that $b_n^{-1} \sum_{i=1}^n U_i \rightarrow 0$ a.s. as $n \rightarrow \infty$. This implies that $T_n/n^2 \rightarrow \alpha^2/4$ a.s. because $n^{-2} \sum_{i=1}^n [v_i(v_i+1)]/(2i) \rightarrow \alpha^2/4$. (4.6) follows from (4.2). \square

We are now prepared to show

Theorem 4.1.

$$\frac{S_n}{nL_n} \rightarrow \frac{\alpha}{2} \text{ a.s. as } n \rightarrow \infty.$$

Proof. Let G_n be the sum over $j = 1, \dots, n$ of the contribution to the absolute rank A_j^n above R_j for the items retained, i.e., $G_n = \sum_{j=1}^n (A_j^n - R_j) I_j$. Then by (2.9)

$$E(G_n | \mathcal{F}_{n-1}) = (G_{n-1} + R_{n-1} I_{n-1}) \frac{(n+1)}{n} \quad (4.7)$$

where \mathcal{F}_{n-1} is the σ -algebra generated by the relative ranks of the first $n-1$ items. Let $Z_n = \frac{G_n}{n(n+1)}$. (4.7) implies that

$$\begin{aligned} E(Z_n | \mathcal{F}_{n-1}) &= \frac{Z_{n-1}(n-1)n(n+1)}{n(n+1)n} + \frac{R_{n-1}I_{n-1}}{n^2} \\ &= \frac{n-1}{n} Z_{n-1} + \frac{R_{n-1}I_{n-1}}{n^2}. \end{aligned}$$

Note that by Lemma 4.1 $\sum_{n=2}^{\infty} \frac{I_{n-1}R_{n-1}}{n^2} \rightarrow \alpha^2/4$ a.s. Hence G_n/n^2 a.s. converges from a theorem in Robbins and Siegmund [6]. But T_n/n^2 converges a.s. from Lemma 4.1 and $S_n/n^2 = (G_n + T_n)/n^2$. Therefore, S_n/n^2 converges a.s.

To see that the a.s. limit of S_n/n^2 is $\alpha^2/2$, note that $S_n \geq L_n(L_n + 1)/2$ and thus by (4.2) $\liminf S_n/n^2 \geq \alpha^2/2$ a.s. Since S_n/n^2 is bounded, it follows that $E(\lim S_n/n^2) = \lim E(S_n/n^2) = \alpha^2/2$, where the last equality follows from (4.3) and (4.1). Thus if $P(\lim S_n/n^2 > \alpha^2/2) > 0$ we would get a contradiction. \square

The simple rule discussed above satisfies $L_n/\lceil \alpha n \rceil \rightarrow 1$ a.s. Suppose however, that it is desired that a rule keep exactly $\lceil \alpha n \rceil$ items. Can this be achieved? Below we exhibit a rule for which $P(L_n = \lceil \alpha n \rceil) \rightarrow 1$, and for which (4.3) holds. This is achieved by increasing the cutoff level for which an item is kept, somewhat.

Rule 2. Fix $0 < \alpha < 1$ and $0 < \omega < .5$. Retain item j if

$$R_j \leq \lceil (\alpha + n^{-\omega})j \rceil, \quad (4.8)$$

and if not already $\lceil \alpha n \rceil$ items have been chosen. Note that to implement this rule we need to know n .

Proposition 4.1. *For Rule 2*

$$P(L_n = \lceil \alpha n \rceil) \rightarrow 1, \text{ as } n \rightarrow \infty. \quad (4.9)$$

Proof. Let X_n be the number of items j , for which (4.8) holds, $j = 1, \dots, n$. We shall show that $P(X_n \geq \lceil \alpha n \rceil) \rightarrow 1$. Let $I_j = 1$ if (4.8) holds for j , $I_j = 0$ otherwise. Then $P(I_j = 1) = p_{nj} \geq p_n = \alpha + n^{-\omega}$. Let Y_n be a Binomial(n, p_n) random variable. Then X_n is stochastically larger than or equal to Y_n . Hence

$$\begin{aligned} \lim_{n \rightarrow \infty} P(X_n \geq \lceil \alpha n \rceil) &\geq \lim_{n \rightarrow \infty} P(Y_n \geq \lceil \alpha n \rceil) = \lim_{n \rightarrow \infty} \left(1 - \Phi \left(\frac{\lceil \alpha n \rceil - np_n}{\sqrt{np_n(1-p_n)}} \right) \right) \\ &\geq \lim_{n \rightarrow \infty} \left(1 - \Phi \left(\frac{1 - n^{1-\omega}}{\sqrt{np_n(1-p_n)}} \right) \right) = 1, \end{aligned}$$

since $0 < \omega < .5$, where Φ is the standard normal distribution. Thus (4.9) holds. \square

Proposition 4.2. *(4.3) holds for Rule 2.*

Proof.

$$E(A_j^n | I_j = 1) \leq \frac{n+1}{j+1} \left(\frac{(\alpha + n^{-\omega})j + 1}{2} \right) \leq (n+1) \left(\frac{\alpha + n^{-\omega}}{2} + \frac{1}{2(j+1)} \right).$$

Let j_1^*, \dots, j_s^* be the at most $\lceil \alpha n \rceil$ smallest indices $j \leq n$ for which (4.8) holds. Thus $S_n = \sum_{i=1}^n A_{j_i^*}^n I_{j_i^*}$. Hence,

$$ES_n \leq \lceil \alpha n \rceil (n+1)(\alpha + n^{-\omega})/2 + (n+1) \log(n+2)/2.$$

It follows that $ES_n/\lceil \alpha n \rceil n \rightarrow \alpha/2$ as $n \rightarrow \infty$, i.e., (4.3) holds. \square

A result corresponding to Theorem 4.1, for Rule 2, follows in a similar way as for Rule 1.

5 Numerical Results

This section demonstrates the results from the previous sections. We first illustrate the findings from Section 2 where one item is retained. Table 1 presents the expected value of the absolute rank of the retained item of the optimal rule (based on dynamic programming for each given n) and the corresponding value and the optimal choice of c for the simple rule, which is evaluated directly for each given c .

It is apparent from Table 1 that the relative advantage of the using the simple rule is greater for large n . It is interesting that the performance of the simple rule is less sensitive to the value of n . Hence, when the number of items becomes large, the case where it is hard to implement dynamic programming, is when the simple rule performs almost as well as the optimal rule.

The choice of c in the simple rule should be taken to be about 2.4. This value of c is nearly optimal for a wide range of n . Furthermore, the actual performance does not vary much for different c near 2.4, as seen in Table 2. Note from Table 2 that for $n = 100$, $c = 2.30$ performs better than $c = 2.40$, consistent with the findings in Table 1.

In Table 3 we illustrate the performance of the rules which select an item with probability $P < 1$, along with their optimal c and k_s . It is observed that even for $P = .99$ one does not need k_s bigger than 11, for all $n \leq 10^6$. It is noted that for $n \geq 1000$ the optimal values of c and k_s do not change with n (up to the precision given).

Table 4 lists the optimal sum of the ranks for keeping r , standardized by $r(r+1)/2$ (which is the lowest possible attainable sum of ranks). Those values were obtained by the dynamic programming described in Section 3.

In Table 5 we list the performance of the simple rule which selects two items, with one c -value, and with $c_1 \neq c_2$ respectively. It is seen that $c_1 \neq c_2$ yields a large improvement. The optimal values are listed for comparison.

In Table 6 we list the corresponding conditional values when a selection is only required with given probability P . The same c is used for both choices. A comparison with the corresponding values in Table 3 shows that the second term improves the average (when divided by $r(r + 1)/2 = 3$).

Finally, in Table 7 shows the performance for Rule 1 of Section 4, for $\alpha = .1, .2, .3, .4$ and $.5$ and $n = 100, 1000, 10000$. This is the only table that is based on simulations, with 1000 replications for each n and α . Since the probability of retaining an item is usually greater than α and at least equal to α , it is not surprising that the average number kept, when standardized by $n\alpha$, is somewhat greater than 1. This effect is larger for small values of n . Similarly, the average absolute rank standardized by $n/2$ times the number kept, is slightly above α , especially when n is small .

Acknowledgement

We would like to thank Moshe Pollak for many helpful suggestions.

Table 1: Optimal rule and simple rule for retaining one item.

| n | Optimal Exp. Rank | Simple rule Exp. Rank | Optimal c |
|-----------|----------------------|--------------------------|-------------|
| 100 | 3.6032 | 3.7348 | 2.30 |
| 1,000 | 3.8324 | 3.9062 | 2.36 |
| 10,000 | 3.8649 | 3.9258 | 2.41 |
| 100,000 | 3.8690 | 3.9279 | 2.41 |
| 1,000,000 | 3.8695 | 3.9281 | 2.42 |

Table 2: Sensitivity of one-choice simple rule to c

| c | $n = 100$ | $n = 1,000$ | $n = 10,000$ | $n = 100,000$ | $n = 1,000,000$ |
|------|-----------|-------------|--------------|---------------|-----------------|
| 2.00 | 3.7706 | 4.0019 | 4.0397 | 4.0450 | 4.0457 |
| 2.10 | 3.7550 | 3.9543 | 3.9857 | 3.9894 | 3.9899 |
| 2.20 | 3.7418 | 3.9251 | 3.9513 | 3.9543 | 3.9546 |
| 2.30 | 3.7348 | 3.9109 | 3.9325 | 3.9349 | 3.9352 |
| 2.40 | 3.7405 | 3.9066 | 3.9259 | 3.9280 | 3.9282 |
| 2.50 | 3.7540 | 3.9111 | 3.9292 | 3.9311 | 3.9313 |

Table 3: One item selected with given P , Optimal c , k_s and conditional expected absolute rank, and actual value of P obtained.

| P | Act Prob | c | k_s | Exp Abs Rank |
|-------------|----------|------|-------|--------------|
| n=100 | | | | |
| 0.50 | 0.5051 | 1.06 | 1 | 1.5049 |
| 0.60 | 0.6061 | 1.58 | 1 | 1.7549 |
| 0.70 | 0.7013 | 1.15 | 2 | 1.8982 |
| 0.80 | 0.8032 | 1.25 | 3 | 2.2058 |
| 0.90 | 0.9013 | 1.61 | 4 | 2.6028 |
| 0.95 | 0.9506 | 1.74 | 6 | 2.9378 |
| 0.99 | 0.9906 | 2.16 | 11 | 3.4140 |
| n=1000 | | | | |
| 0.50 | 0.5015 | 1.01 | 1 | 1.5025 |
| 0.60 | 0.6016 | 1.51 | 1 | 1.7536 |
| 0.70 | 0.7021 | 1.14 | 2 | 1.9176 |
| 0.80 | 0.8010 | 1.22 | 3 | 2.2254 |
| 0.90 | 0.9001 | 1.60 | 4 | 2.6346 |
| 0.95 | 0.9502 | 1.96 | 5 | 2.9817 |
| 0.99 | 0.9901 | 2.15 | 11 | 3.4964 |
| n=10,000 | | | | |
| 0.50 | 0.5025 | 1.01 | 1 | 1.5049 |
| 0.60 | 0.6016 | 1.51 | 1 | 1.7548 |
| 0.70 | 0.7023 | 1.14 | 2 | 1.9193 |
| 0.80 | 0.8011 | 1.22 | 3 | 2.2277 |
| 0.90 | 0.9004 | 1.60 | 4 | 2.6391 |
| 0.95 | 0.9502 | 1.96 | 5 | 2.9868 |
| 0.99 | 0.9900 | 2.15 | 11 | 3.5035 |
| n=100,000 | | | | |
| 0.50 | 0.5025 | 1.01 | 1 | 1.5050 |
| 0.60 | 0.6016 | 1.51 | 1 | 1.7550 |
| 0.70 | 0.7000 | 1.13 | 2 | 1.9142 |
| 0.80 | 0.8011 | 1.22 | 3 | 2.2279 |
| 0.90 | 0.9005 | 1.60 | 4 | 2.6395 |
| 0.95 | 0.9502 | 1.96 | 5 | 2.9873 |
| 0.99 | 0.9900 | 2.15 | 11 | 3.5043 |
| n=1,000,000 | | | | |
| 0.50 | 0.5025 | 1.01 | 1 | 1.5050 |
| 0.60 | 0.6016 | 1.51 | 1 | 1.7550 |
| 0.70 | 0.7000 | 1.13 | 2 | 1.9142 |
| 0.80 | 0.8011 | 1.22 | 3 | 2.2280 |
| 0.90 | 0.9005 | 1.60 | 4 | 2.6396 |
| 0.95 | 0.9502 | 1.96 | 5 | 2.9874 |
| 0.99 | 0.9900 | 2.15 | 11 | 3.5044 |

Table 4: Expected absolute rank/ $[r(r + 1)/2]$ obtained by the Optimal rule for keeping r .

| r | $n = 100$ | $n = 1,000$ | $n = 10,000$ | $n = 100,000$ | $n = 1,000,000$ |
|-----|-----------|-------------|--------------|---------------|-----------------|
| 1 | 3.6032 | 3.8324 | 3.8649 | 3.8690 | 3.8695 |
| 2 | 2.7937 | 2.97807 | 3.0039 | 3.0072 | 3.0076 |
| 3 | 2.3774 | 2.5386 | 2.5612 | 2.5640 | 2.5644 |
| 4 | 2.1216 | 2.2688 | 2.2893 | 2.2919 | 2.2922 |
| 5 | 1.9476 | 2.0851 | 2.1042 | 2.1066 | 2.1069 |
| 6 | 1.8209 | 1.9513 | 1.9694 | 1.9716 | 1.9719 |
| 7 | 1.7242 | 1.8491 | 1.8664 | 1.8686 | 1.8688 |
| 8 | 1.6475 | 1.7683 | 1.7850 | 1.7871 | 1.7873 |
| 9 | 1.5849 | 1.7026 | 1.7188 | 1.7208 | 1.7211 |
| 10 | 1.5331 | 1.6481 | 1.6639 | 1.6659 | 1.6661 |

Table 5: Two items selected with probability one for equal c and unequal c_1 and c_2 , Optimal c , c_1, c_2 and expected absolute rank/3.

| n | c | c_1 | c_2 | Exp Abs Rank Equal/3 | Exp Abs Rank Unequal/3 | Exp Abs Rank Optimal/3 |
|--------|------|-------|-------|-------------------------|---------------------------|---------------------------|
| 100 | 3.21 | 3.81 | 2.16 | 3.0857 | 2.8667 | 2.7937 |
| 1,000 | 3.19 | 3.79 | 2.20 | 3.2198 | 3.0078 | 2.9798 |
| 10,000 | 3.15 | 3.80 | 2.22 | 3.2342 | 3.0259 | 3.0039 |

Table 6: Two items selected with given P , Optimal k_s and conditional expected absolute rank/3.

| P | Act Prob | c | k_s | Exp Abs Rank/3 |
|----------|----------|------|-------|----------------|
| n=100 | | | | |
| 0.50 | 0.5365 | 2.04 | 2 | 1.9407 |
| 0.60 | 0.6030 | 2.48 | 2 | 2.0179 |
| 0.70 | 0.7037 | 2.20 | 3 | 2.0724 |
| 0.80 | 0.8024 | 2.31 | 4 | 2.2093 |
| 0.90 | 0.9007 | 2.42 | 6 | 2.4414 |
| 0.95 | 0.9502 | 2.62 | 8 | 2.6107 |
| 0.99 | 0.9903 | 3.06 | 13 | 2.8939 |
| n=1,000 | | | | |
| 0.50 | 0.5019 | 1.82 | 2 | 1.9436 |
| 0.60 | 0.6003 | 1.66 | 3 | 2.0209 |
| 0.70 | 0.7009 | 2.18 | 3 | 2.1091 |
| 0.80 | 0.8013 | 2.32 | 4 | 2.2568 |
| 0.90 | 0.9004 | 2.48 | 6 | 2.4940 |
| 0.95 | 0.9503 | 2.70 | 8 | 2.6919 |
| 0.99 | 0.9901 | 2.95 | 15 | 2.9935 |
| n=10,000 | | | | |
| 0.5000 | 0.5016 | 1.82 | 2 | 1.9468 |
| 0.6000 | 0.6002 | 1.66 | 3 | 2.0235 |
| 0.7000 | 0.7002 | 2.18 | 3 | 2.1119 |
| 0.8000 | 0.8009 | 2.32 | 4 | 2.2607 |
| 0.9000 | 0.9007 | 2.49 | 6 | 2.5011 |
| 0.9500 | 0.9504 | 2.71 | 8 | 2.7002 |
| 0.9900 | 0.9902 | 2.96 | 15 | 3.0023 |

Table 7: Rule 1 for selecting a proportion α . Average number kept normalized by $n\alpha$. Average Abs Rank normalized by (Number kept times $n/2$).

| n | α | Avearge Abs Rank/ Number kept $\cdot \frac{n}{2}$ | Std Dev Abs Rank | Avearge Number kep/ $n\alpha$ | Std Dev Number kept |
|-------|----------|---|---------------------|-------------------------------------|------------------------|
| 100 | 0.1 | 0.1760 | 0.06500 | 1.2291 | 0.3305 |
| 100 | 0.2 | 0.2579 | 0.05604 | 1.1114 | 0.2101 |
| 100 | 0.3 | 0.3457 | 0.05306 | 1.0582 | 0.1621 |
| 100 | 0.4 | 0.4457 | 0.05276 | 1.0469 | 0.1253 |
| 100 | 0.5 | 0.5572 | 0.05041 | 1.0572 | 0.0967 |
| 1000 | 0.1 | 0.1127 | 0.01351 | 1.0386 | 0.0964 |
| 1000 | 0.2 | 0.2092 | 0.01435 | 1.0171 | 0.0651 |
| 1000 | 0.3 | 0.3078 | 0.01636 | 1.0102 | 0.0514 |
| 1000 | 0.4 | 0.4080 | 0.01566 | 1.0095 | 0.0376 |
| 1000 | 0.5 | 0.5091 | 0.01692 | 1.0098 | 0.0327 |
| 10000 | 0.1 | 0.1017 | 0.00325 | 1.0055 | 0.0302 |
| 10000 | 0.2 | 0.2013 | 0.00396 | 1.0025 | 0.0195 |
| 10000 | 0.3 | 0.3012 | 0.00476 | 1.0019 | 0.0156 |
| 10000 | 0.4 | 0.4012 | 0.00495 | 1.0015 | 0.0123 |
| 10000 | 0.5 | 0.5012 | 0.00502 | 1.0014 | 0.0100 |

References

- [1] Barbour, A.D., Holst, L., and Janson, S. (1992). *Poisson Approximation*. Oxford University Press, Oxford.
- [2] Chow, Y.S. Moriguti, S. Robbins, H. and Samuels, S.M. (1964). Optimal selection based on relative rank (the "Secretary Problem"). *Israel Journal of Mathematics*, **2**, pp. 81-90.
- [3] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* Volume 2, Second Edition. John Wiley & Sons, Inc., New York.
- [4] Ferguson, T.S. (2002). *Optimal Stopping and Applications*. <http://www.math.ucla.edu/~tom/stopping?/contents.html>
- [5] Krieger, A.M., Pollak, M. and Samuel-Cahn, E. (2007). Select sets: rank and file. *Annals of Applied Probability*, **17**, 360-385.
- [6] Robbins, H. and Siegmund, D. (1971). A convergence theorem for non-negative almost supermartingales and some applications. In *Optimizing Methods in Statistics*, J.S. Rustagi Editor, pp. 233–257. Academic Press, New York.
- [7] Yeo, A.J. and Yeo, G.F. (1994). Selecting Satisfactory Secretaries. *Australian Journal of Statistics*, **36** (2), pp. 185-198.