האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

BEAT THE MEAN: BETTER THE AVERAGE

by

ABBA M. KRIEGER, MOSHE POLLAK
and ESTER SAMUEL-CAHN

Discussion Paper # 469 November 2007

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/

Beat the Mean: Better the Average

Abba M. Krieger[*], Moshe Pollak[†] and Ester Samuel-Cahn[‡]
University of Pennsylvania, Hebrew University and Hebrew University

November 21, 2007

We consider a sequential rule, where an item is chosen into the group, such as a university faculty member, only if his score is better than the average score of those already belonging to the group. We study four variables: The average score of the members of the group after k items have been selected, the time it takes (in terms of number of observed items) to assemble a group of k items, the average score of the group after n items have been observed, and the number of items kept after the first n items have been observed. We develop the relationships between these variables, and obtain their asymptotic behavior as k (respectively, n) tends to infinity. The assumption throughout is that the items are independent, identically distributed, with a continuous distribution. Though knowledge of this distribution is not needed to implement the selection rule, the asymptotic behavior does depend on the distribution. We study in some detail the Exponential, Pareto and Beta distributions. Generalizations of the "better than average" rule to the β better than average rules are also considered. These are rules where an item is admitted to the group only if its score is better than β times the present average of the group, where $\beta > 0$.

AMS 2000 subject classfication: Primary 62G99, secondary 62F07, 60F15
Keywords and phrases: Selection rules, averages, nonparametrics, sequential observations, asymptotics

[*]Supported by funds from the Marcy Bogen Chair of Statistics at the Hebrew University of Jerusalem.
[†]Supported by funds from the Marcy Bogen Chair of Statistics at the Hebrew University of Jerusalem.
[‡]Supported by the Israel Science Foundation Grant 467/04.

1 Introduction

In many practical situations, when it is desired to assemble sequentially a group of good quality, the "better than average" rule is used. This rule selects an individual into the group if and only if she is better than the average quality of the previously selected members of the group. Individuals are considered sequentially and a decision of retention or rejection must be made without recourse. Such a policy may be used when recruiting new faculty to a university department, selecting students to a university for graduate study, or selecting a sports team.

In this paper we consider the asymptotic behavior, both in number and average quality, of the selected group under the better than average rule under the assumption that the observations are i.i.d. We also extend these results to a class of related selection rules, which we term β better than average rules. Under these rules an item is selected if and only if it is better than β times the average quality of the items that already belong to the group. For $\beta = 1$, this is the usual better than average rule. When "better" is "larger than", for $\beta > 1$ (the more typical case), the β better than average rule is more stringent and for $\beta < 1$ it is less stringent than the better than average rule.

The better than average selection rule was first considered by Preater (2000). He dealt with exponentially generated values and related cases. The focus of his paper is on the average quality of the retained items.

Sequential rules that retain an observation based on how its relative rank relates to the ranks of those already retained are considered by Krieger, Pollak and Samuel-Cahn (2007). The behavior of the number of items kept and their average quality are determined in that paper.

The next section is devoted to general results. In essence we consider the asymptotic behavior of four quantities: the average quality and the number of items observed until k are retained; the average quality and the number of items retained after n items have been observed. In Section 3, the relationships in asymptotic behavior of the other three quantities to that of the average quality when k items are retained are obtained under certain conditions. The exponential distribution is used as an example to illustrate these results and relate them to the findings of Preater (2000). Many results depend heavily on the underlying distribution of the i.i.d. random variables. The Pareto family is explored in Section 4 and the Beta family in Section 5. These three families of distributions are representatives of the three domains of attraction of extreme values. Finally, related results and observations are presented in the form of remarks in the concluding section.

2 Notation and Preliminary Results

This article is about a sequence of observations that are independent and identically distributed. The decision of whether to retain an item needs to be made when it is observed, based only on its score and the average score of those items already seen. The observations are denoted by $X_1, X_2, \ldots$ and are i.i.d. random variables from a common distribution F. Knowledge of F is not necessary to implement the rule under investigation. However, the asymptotic behavior of the rule does depend on the specification of F.

The behavior of rules will be characterized by considering four quantities:

- $T_k = T_k(F) = T_k(X) =$ The number of observations inspected (including that item) until the k^{th} item is retained.
- $\overline{Y}_k = \overline{Y}_k(F) = \overline{Y}_k(X) =$ The average of the first k observations that are kept.
- $M_n = M_n(F) = M_n(X) =$ The number of items that are kept by time n.
- $A_n = A_n(F) = A_n(X) =$ the average of the items that are kept by time n.

We use either of the three forms of notation (e.g., T_k, $T_k(F)$ or $T_k(X)$) as convenient.

The β better than average rule is defined as follows: For fixed β (which is suppressed in the notation) and T_k as defined above as the number of items observed until the k^{th} item is selected, let $Y_1 = X_1$. Thus $T_1 = 1$. Now T_k and Y_k are defined inductively by

$$T_{k+1} = \inf\{i > T_k : X_i > \beta\overline{Y}_k\} \quad , \quad k = 1, 2, \ldots$$

($T_{k+1} = \infty$ if $X_i \leq \beta\overline{Y}_k$ for all $i > T_k$), and if $T_{k+1} < \infty$,

$$Y_{k+1} = X_{T_{k+1}} \quad , \quad k = 1, 2, \ldots .$$

It is clear that $\overline{Y}_k$ increases in k for $\beta = 1$. If $\beta > 1$ we assume non-negative X_i to avoid the situation that if $\overline{Y}_k$ is negative then the cutoff to retain an observation becomes less stringent. Below is an intuitive result, which is true generally. The proof is adapted from Preater (2000).

Let $x_F = \sup\{x : F(x) < 1\}$. When $x_F = \infty$ then clearly $T_k < \infty$ a.s. for all k and all β. When $\beta > 1$ we assume that $x_F = \infty$.

Lemma 2.1. *Let X_i be i.i.d. from any distribution. Unless F has an atom at x_F and $X_1 = x_F$ (in which case $\overline{Y}_1 = x_F$), the sequence $T_1, T_2, \ldots$ is well-defined for all $k \geq 1$, and the better than average rule satisfies*

$$\overline{Y}_k \to x_F \quad a.s., \text{ as } k \to \infty. \tag{2.1}$$

Furthermore, when $X_i \geq 0$ and $x_F = \infty$, the sequence $T_1, T_2, \ldots$ is well-defined, and (2.1) holds for all $\beta > 1$.

Proof. Consider first the better than average rule. Let $\overline{Y}_0 = 0$. It is easily seen by induction that

$$\overline{Y}_k = \sum_{j=1}^{k} (Y_j - \overline{Y}_{j-1})/j.$$

Since $\overline{Y}_k$ is monotone increasing, it follows that $Y = \lim_{k\to\infty} \overline{Y}_k$ exists. We want to show that for any $x < x_F$ one has $P(Y \leq x) = 0$. Let $x < x_F$ be given, and when $x_F < \infty$ let $0 < \epsilon < x_F - x$ while if $x_F = \infty$ let $\epsilon = 1$. Then on the event $A_x = \{Y \leq x\}$, we have

$$Y = \sum_{k=1}^{\infty} (Y_k - \overline{Y}_{k-1})/k \geq \sum_{k=1}^{\infty} (Y_k - x)^+/k \geq \sum_{k=1}^{\infty} \epsilon I(Y_k > x + \epsilon)/k.$$

But the sum on the right hand side diverges, implying that $P(A_x) = 0$. Thus (2.1) follows for the better than average rule. Now, the average $\overline{Y}_k[\beta]$ satisfies $P(\overline{Y}_k[\beta] \geq \overline{Y}_k) = 1$ for the β better than average rule for all k, hence (2.1) follows. □

We exclude the case $X_1 = x_F$ in our further considerations. We now turn to a general result that relates the number of items kept after n have been observed, when $X \sim F$ to the number of items kept when $Z \sim G$, where here the sequence of random variables considered are i.i.d. $Z_1, Z_2, \ldots$. Assume that the random variables are continuous. We can couple the items generated according to F and G in the following way: Let $Z_i = g(X_i)$ where $g = G^{-1}F$, and thus is an increasing function. Note that if and only if g is convex g^{-1} (which relates Z to X) is concave.

Denote by $M_n(X)$ the number of items kept by a β better than average rule for the X- sequence $X_1, X_2, \ldots$ after n items are observed. Similarly $M_n(Z)$ denotes the number of items kept using the Z- sequence. We then have

Theorem 2.1. *Let $X \sim F$ and $g(x)$ be an increasing concave function. Let G be the c.d.f. of a random variable distributed as $g(X)$. When $\beta = 1$, $M_n(G) \geq_{st} M_n(F)$ for all n. Thus $T_k(F) \geq_{st} T_k(G)$.*

Proof. Clearly $g(x) = G^{-1}(F(x))$. Let $Z \sim G$ and let $A_n(X)$ be the average of the X values of the kept items after n items are observed, using for a selection rule the better than average rule based on the X- sequence, and let $B_n(X)$ be the average of the X values kept using the better than average rule based on the Z- sequence. Without loss of generality couple the X_i values and Z_i values by setting $Z_i = g(X_i)$. Let $A_n(Z)$ be the average of the Z- values kept by its rule after n items are observed.

We shall prove by induction that with this coupling
i) $M_n(X) \leq M_n(Z)$
and
ii) $A_n(X) \geq B_n(X)$.
Clearly the stochastic ordering in the statement follows from this. For $n = 1$, clearly $M_1(X) = M_1(Z) = 1$ and $A_1(X) = B_1(X) = X_1$. Now suppose i) and ii) hold for $n - 1$. We consider two cases:

a) $X_n > A_{n-1}(X)$. Item n is kept using the X- sequence. Then

$$Z_n = g(X_n) > g(A_{n-1}(X)) \geq g(B_{n-1}(X)) > A_{n-1}(g(X)) = A_{n-1}(Z).$$

The last inequality follows because g is concave. This implies that the n^{th} item is also kept by its rule using the Z- sequence. Hence i) holds. Furthermore,

$$A_n(X) = \frac{M_{n-1}(X)A_{n-1}(X)}{M_{n-1}(X)+1} + \frac{X_n}{M_{n-1}(X)+1}$$

and

$$\begin{aligned}
B_n(X) &= \frac{M_{n-1}(Z)B_{n-1}(X)}{M_{n-1}(Z)+1} + \frac{X_n}{M_{n-1}(Z)+1} \\
&\leq \frac{M_{n-1}(X)B_{n-1}(X)}{M_{n-1}(X)+1} + \frac{X_n}{M_{n-1}(X)+1} \\
&\leq \frac{M_{n-1}(X)A_{n-1}(X)}{M_{n-1}(X)+1} + \frac{X_n}{M_{n-1}(X)+1} = A_n(X).
\end{aligned}$$

The first inequality holds because $M_{n-1}(X) \leq M_{n-1}(Z)$ and the last inequality holds because $A_{n-1}(X) \geq B_{n-1}(X)$. Hence ii) follows.

b) $X_n \leq A_{n-1}(X)$. Hence the n^{th} item is not kept by the rule that uses the X- sequence. i) is immediate. If the n^{th} element is not kept by the Z-

sequence ii) is also immediate. Now assume that the n^{th} item is retained using the Z sequence. Then

$$B_n(X) = \frac{M_{n-1}(Z)B_{n-1}(X)}{M_{n-1}(Z)+1} + \frac{X_n}{M_{n-1}(Z)+1} \leq A_{n-1}(X) = A_n(X)$$

as both $B_{n-1}(X)$ and X_n do not exceed $A_{n-1}(X)$. □

Remark 2.1 It is interesting to note that Theorem 2.1 requires $\beta = 1$, i.e., the better than average rule. Here is an example where the conclusion fails for $\beta > 1$. For $\beta > 1$ consider $Z = X^{1/2}$. The second item is kept for the $X-$ sequence if $X_2 > \beta X_1$. The second item is not kept for the $Z-$ sequence if $X_2^{1/2} < \beta X_1^{1/2}$. Hence when X_1 is such that $\beta X_1 < X_2 < \beta^2 X_1$ the conclusion fails. This is easily fulfilled, since $\beta > 1$.

The next result considers a sequence of distribution functions $\{F_j\}$ and sequences of i.i.d. random variables from each.

Theorem 2.2. *Let $\{F_j\}$ be a sequence of continuous cumulative distribution functions that converges weakly to a continuous distribution F. Then:*
i) For fixed n, $M_n(F_j) \to M_n(F)$ as $j \to \infty$, in distribution.
ii) For fixed n, $A_n(F_j) \to A_n(F)$ as $j \to \infty$, in distribution.
iii) For fixed k, $T_k(F_j) \to T_k(F)$ as $j \to \infty$, in distribution.
iv) For fixed k, $\overline{Y}_k(F_j) \to \overline{Y}_k(F)$ as $j \to \infty$, in distribution.

Proof. For fixed n, let $U_1, U_2, ..., U_n$ be independent Uniform[0,1]-distributed random variables and define $X_i^{(j)} = F_j^{-1}(U_i)$, $X_i = F^{-1}(U_i)$ for $i = 1, ..., n$. Since$(X_1^{(j)}, ..., X_n^{(j)}) \to (X_1, ..., X_n)$ as $j \to \infty$ almost surely, it follows that the convergence in i) and ii) (with $M_n(F_j)$ and $A_n(F_j)$ defined on the sequence $(X_1^{(j)}, ..., X_n^{(j)})$) holds a.s. on this sample space, so that in general i) and ii) hold.

To prove iii) and iv), fix k. Let $\epsilon > 0$. There exists $n_k(\epsilon)$ such that $P(T_k(F) \leq n_k(\epsilon)) > 1 - \epsilon$. Let $n = n_k(\epsilon)$ and let $\{X_i^{(j)}\}$ and $\{X_i\}$ be as above. On $\{T_k(F) \leq n_k(\epsilon)\}$, $(X_1^{(j)}, ..., X_n^{(j)}) \to (X_1, ..., X_n)$ as $j \to \infty$ almost surely, so that there exists $m_k(\epsilon)$ such that $P(T_k(F_j) = T_k(F)$ for all $j \geq m_k(\epsilon)) > 1 - \epsilon$ (on the constructed sample space as above). Hence $P(T_k(F_j) = T_k(F)$ for all $j \geq m_k(\epsilon)) > 1 - 2\epsilon$. Letting $\epsilon \to 0$ proves iii). An analogous argument accounts for iv). □

Remark 2.2 This Theorem ensures that the four quantities of interest behave similarly for similar distributions for finite n and k. For example (as appears in Remark 5.3) the exponential distribution and Beta$(1, j)$ distribution suitably scaled for large j are similar to each other. Hence as is practical for finite n, the number kept by these two distributions behave similarly. This does not imply that the rates of convergence (as $n \to \infty$, $k \to \infty$, respectively) of the four quantities of interest are approximately the same for the two distributions. We know this is true for M_n and T_k for the examples considered in this paper. We conjecture that it is true in general, but we have not been able to prove it.

3 Almost Sure Convergence

We assume observations $X_1, X_2, \ldots$ are non-negative, i.i.d. random variables from c.d.f. F. We previously defined four quantities. When interest is in the status at the instant at which the k^{th} item is kept then T_k denotes the number of items that have been considered and $\overline{Y}_k$ the average of the k items that are kept. Correspondingly, if focus is on the status after n items are observed then M_n and A_n refer to the number of items that are kept and the average of those items, respectively.

In this section, we find conditions on the behavior of $\overline{Y}_k$ from which it follows that T_k, M_n and A_n converge a.s. Sections 4 and 5 are devoted to finding particulars and establishing that these conditions on $\overline{Y}_k$ hold for various families of distributions.

Of the four quantities considered, $\overline{Y}_k$ is the easiest to handle directly, often through use of the Martingale Convergence Theorem. It is not unusual for $\overline{Y}_k$ to satisfy the conditions in Theorem 3.2. The asymptotic behavior of T_k can then also be determined, as in Theorem 3.2, by use of a theorem in Feller reproduced below as Theorem 3.1 (in the generality needed here).

The quantities $\{A_n\}$ and $\{M_n\}$ are more difficult to handle directly, as they develop as intertwined sequences. Let $\mathcal{F}_n$ be the sigma-field generated by $\{X_1, \ldots, X_n\}$. Then clearly

$$\begin{aligned} E(A_n|\mathcal{F}_{n-1}) &= A_{n-1}P(X_n \leq \beta A_{n-1}|A_{n-1}) \\ &+ \frac{M_{n-1}A_{n-1} + X_n}{M_{n-1} + 1} P(X_n > \beta A_{n-1}|A_{n-1}) \end{aligned}$$

and

$$E(M_n|\mathcal{F}_{n-1}) = M_{n-1} + P(X_n > \beta A_{n-1}|A_{n-1}).$$

It is therefore difficult to separate these two quantities. Their asymptotic behavior can nevertheless, in many instances, be derived, through an "inversion" of the asymptotic behavior of $\overline{Y}_k$ and T_k respectively, as given in Theorems 3.3 and 3.4.

Theorem 3.2, which considers the a.s. convergence of T_k relies on a Theorem in Feller (1971, Vol 2. page 239):

Theorem 3.1. *Let $Q_1, Q_2, \dots$ be independent r.v.s with $E(Q_n) = 0$, and let $S_n = \sum_{i=1}^{n} Q_i$. If*
1) $b_1 < b_2 < \cdots \to \infty$ are constants
and
2) $\sum_{n=1}^{\infty} E(Q_n^2/b_n^2) < \infty$
then

$$b_n^{-1} S_n \to 0 \quad a.s. \text{ as } n \to \infty \ .$$

We use this theorem in the following way:

Theorem 3.2. *Let $P_j = 1 - F(\beta \overline{Y}_{j-1})$ with $P_1 \equiv 1$. Suppose that*

$$j^{\omega} P_j \to W \quad a.s. \text{ as } j \to \infty, \tag{3.1}$$

for some $0 < \omega < \infty$, where $P(0 < W < \infty) = 1$. Then

$$\frac{T_k}{k^{\omega+1}} \to \frac{1}{(\omega+1)W} \quad a.s. \text{ as } k \to \infty.$$

Proof. We shall use Feller's Theorem conditionally on the sequence $\{\overline{Y}_k\}$. Let $b_j = \sum_{i=1}^{j} P_i^{-1}$ and $Q_i = T_i - T_{i-1} - P_i^{-1}$ with $T_0 \equiv 0$. Obviously, the sequence $\{b_j\}_{j=1}^{\infty}$ satisfies the first condition of Feller's Theorem.

Note that conditional on $\{P_j\}$ the distribution of $T_i - T_{i-1}$ is Geometric (P_i) and these differences are conditionally independent of each other. Hence $\{Q_n\}_{n=1}^{\infty}$ is a sequence of conditionally independent random variables with zero expectation and variance $(1 - P_n)/P_n^2$. We shall show that the second condition of Feller's Theorem holds.

Let $\epsilon > 0$. Define J_ϵ to be such that

$$W(1-\epsilon) \le j^{\omega} P_j \le W(1+\epsilon) \quad \text{for all } j \ge J_\epsilon,$$

where W is given in (3.1). Therefore, conditonal on $\{P_j\}$

$$\begin{aligned}
\sum_{n=1}^{\infty} E(Q_n^2/b_n^2) &= \sum_{n=1}^{\infty} \frac{1-P_n}{P_n^2} / (\sum_{j=1}^{n} P_j^{-1})^2 \\
&< \sum_{n=1}^{J_\epsilon - 1} \frac{1}{P_n^2} / (\sum_{j=1}^{n} P_j^{-1})^2 + \sum_{n=J_\epsilon}^{\infty} \frac{n^{2\omega}}{W^2(1-\epsilon)^2} / (\sum_{j=1}^{n} \frac{j^{\omega}}{W(1+\epsilon)})^2 \\
&\le \sum_{n=1}^{J_\epsilon - 1} \frac{1}{P_n^2} / (\sum_{j=1}^{n} P_j^{-1})^2 + \frac{(1+\epsilon)^2}{(1-\epsilon)^2} \sum_{n=J_\epsilon}^{\infty} n^{2\omega} / (\frac{n^{\omega+1}}{\omega+1})^2 < \infty.
\end{aligned}$$

Since both conditions of Feller's Theorem are satisfied,

$$b_n^{-1} S_n \to 0 \quad \text{a.s., as } n \to \infty.$$

Now

$$\begin{aligned} b_n^{-1} S_n &= (\sum_{j=1}^n P_j^{-1})^{-1}(T_n - \sum_{j=1}^n P_j^{-1}) \\ &= (\sum_{j=1}^n P_j^{-1})^{-1} T_n - 1 \to 0 \quad \text{a.s. as } n \to \infty. \end{aligned} \tag{3.2}$$

Also for $n \geq J_\epsilon$

$$\sum_{j=1}^{J_\epsilon - 1} P_j^{-1} + \sum_{j=J_\epsilon}^{n} \frac{j^\omega}{W(1+\epsilon)} \leq \sum_{j=1}^{n} P_j^{-1} \leq \sum_{j=1}^{J_\epsilon - 1} P_j^{-1} + \sum_{j=J_\epsilon}^{n} \frac{j^\omega}{W(1-\epsilon)}. \tag{3.3}$$

Therefore for n large enough

$$\frac{1}{1+\epsilon} \leq \frac{\sum_{j=1}^n P_j^{-1}}{n^{\omega+1}/(\omega+1)W} \leq \epsilon + \frac{1}{1-\epsilon}.$$

Letting $\epsilon \to 0$, it follows from (3) and (3.3) that

$$\frac{T_k}{k^{\omega+1}} \to \frac{1}{(\omega+1)W} \quad \text{a.s. as } k \to \infty. \qquad \square$$

Now that we established almost sure convergence for T_k suitably normalized, we relate this result to the almost sure convergence for M_n. For example, if T_k/k^2 converges almost surely to a well-defined random variable (as we will show in Theorem 3.5 when F is exponential and $\beta = 1$) this says that it takes on the order of k^2 observations until k items are retained. It is intuitive that if n items are observed then the order $\sqrt{n}$ of these items should be retained. The theorem below implies that for the exponential distribution with $\beta = 1$, $\frac{M_n}{\sqrt{n}}$ converges a.s.

More precisely, let c_k be an increasing sequence of positive constants (that later will denote the rate at which T_k goes to infinity) with $\lim c_k = \infty$. We define a sequence of increasing positive constants (for the rate of M_n) by

$$d_n = \sup\{j | c_j \leq n\} \quad \text{for } n = 1, 2, \ldots.$$

Since $c_k \to \infty$, d_n is defined for all n. Also assume that

$$\lim_{n\to\infty} \frac{d_{\lfloor n/x \rfloor}}{d_n} = h(x) \tag{3.4}$$

exists for every $0 < x < \infty$, and h is continuous.

The result for the almost sure convergence of M_n is explored in

Theorem 3.3. *Let*

$$\frac{T_k}{c_k} \to Q \quad a.s.\ for\ k \to \infty,\ where\ 0 < Q < \infty.$$

Assume c_k is such that (3.4) holds, with continuous h. Let $M_n = \sup\{j|T_j \leq n\}$, $n = 1, 2, \ldots$ Then

$$\frac{M_n}{d_n} \to h(Q) \quad a.s.\ as\ n \to \infty\ .$$

Proof. It clearly suffices to show that for any fixed sequence t_k such that $\frac{t_k}{c_k} \to x$ as $k \to \infty$, the sequence $m_n = \sup\{j|t_j \leq n\}$ satisfies $\frac{m_n}{d_n} \to h(x)$ as $n \to \infty$.

For a given $0 < \epsilon < x$, let j_0 be so large that $x - \epsilon < \frac{t_j}{c_j} < x + \epsilon$ for all $j \geq j_0$. Let n_0 be such that $m_{n_0} \geq j_0$. Then $m_n \geq j_0$ for all $n \geq n_0$, and

$$\begin{aligned} m_n &= \sup\{j|t_j \leq n\} = \sup\{j|\frac{t_j}{c_j} \leq \frac{n}{c_j}\} \\ &\leq \sup\{j|(x-\epsilon) \leq \frac{n}{c_j}\} = \sup\{j|c_j \leq \frac{n}{x-\epsilon}\} = d_{\lfloor \frac{n}{x-\epsilon} \rfloor}. \end{aligned}$$

Similarly, $m_n \geq d_{\lfloor \frac{n}{x+\epsilon} \rfloor}$. Thus for all $n \geq n_0$

$$\frac{d_{\lfloor \frac{n}{x+\epsilon} \rfloor}}{d_n} \leq \frac{m_n}{d_n} \leq \frac{d_{\lfloor \frac{n}{x-\epsilon} \rfloor}}{d_n}.$$

Now letting $n \to \infty$ and $\epsilon \to 0$ yields

$$\lim \frac{m_n}{d_n} = h(x). \qquad \square$$

Suppose that $1 - F(\beta \overline{Y}_k)$ properly normalized converges almost surely to a possibly non-degenerate random variable. This implies from Theorem 3.2 that T_k, and ultimately from Theorem 3.3 that M_n, when properly normalized, also converge almost surely to a possibly non-degenerate random variable. We now establish conditions such that A_n, suitably normalized, converges almost surely to a non-degenerate random variable.

Theorem 3.4. *Let $\frac{\overline{Y}_k}{k^\gamma} \to Y$ and $\frac{M_n}{n^\psi} \to M$ a.s. as $k \to \infty$ and $n \to \infty$ where $P(0 < Y < \infty) = P(0 < M < \infty) = 1$. Then*

$$\frac{A_n}{n^{\gamma\psi}} \to Y M^\gamma \quad a.s.\ as\ n \to \infty.$$

Proof. For any $\epsilon > 0$, there exists random k_ϵ, n_ϵ such that for all $k \geq k_\epsilon$ and $n \geq n_\epsilon$

$$Yk^\gamma(1-\epsilon) \leq \overline{Y}_k \leq Yk^\gamma(1+\epsilon)$$

and

$$Mn^\psi(1-\epsilon) \leq M_n \leq Mn^\psi(1+\epsilon).$$

Note that $A_n = \overline{Y}_{M_n}$. Also, eventually $M_n \geq k_\epsilon$ a.s. Then

$$Y{M_n}^\gamma(1-\epsilon) \leq A_n \leq Y{M_n}^\gamma(1+\epsilon)$$

and for $n \geq n_\epsilon$

$$YM^\gamma n^{\gamma\psi}(1-\epsilon)^{\gamma+1} \leq A_n \leq YM^\gamma n^{\gamma\psi}(1+\epsilon)^{\gamma+1}.$$

This implies

$$YM^\gamma(1-\epsilon)^{\gamma+1} \leq \frac{A_n}{n^{\gamma\psi}} \leq YM^\gamma(1+\epsilon)^{\gamma+1},$$

so that

$$\frac{A_n}{n^{\gamma\psi}} \to YM^\gamma \quad \text{a.s. for } n \to \infty\ . \qquad \square$$

We now apply the previous results to the exponential distribution.

Theorem 3.5. *Assume that the observations are i.i.d. from an exponential distribution mean 1. Let $\beta = 1$ and let G denote a random variable that has the Gumbel distribution,* $\exp\{-e^{-x}\}$.
Then

i) $\overline{Y_k} - \log k$ *converges a.s. to G as $k \to \infty$,*

ii) $\frac{T_k}{k^2}$ *converges a.s. to $e^G/2$ as $k \to \infty$,*

iii) $\frac{M_n}{\sqrt{n}}$ *converges a.s. to $\sqrt{2}e^{-G/2}$ as $n \to \infty$ and*

iv) $A_n - (\log n)/2$ *converges a.s. to $(G + \log 2)/2$ as $n \to \infty$.*

Proof. i) This result is given in Preater (2000).
ii) The value of $P_j = 1 - F(\overline{Y}_{j-1}) = e^{-\overline{Y}_{j-1}}$. But from i), $\overline{Y}_j - \log j$ converges a.s. to G . Then $jP_j = je^{-\overline{Y}_{j-1}} \approx je^{-(G+\log j)} = e^{-G} = W$. Hence jP_j converges a.s., so $\omega = 1$ in Theorem 3.2, i.e., T_k/k^2 converges a.s. to $\frac{1}{(\omega+1)W} = \frac{e^G}{2}$.
iii) Since (by ii)) $\frac{T_k}{k^2}$ converges a.s. to $Q = e^G/2$, it follows that c_k is k^2 in Theorem 3.3. Hence $d_n = \lfloor\sqrt{n}\rfloor$. This implies that $h(x) = \lim_{n\to\infty} \frac{\lfloor\sqrt{n/x}\rfloor}{\lfloor\sqrt{n}\rfloor} =$

$\frac{1}{\sqrt{x}}$. Therefore, from Theorem 3.3, $\frac{M_n}{d_n}$ converges a.s. to $\frac{1}{\sqrt{Q}} = \sqrt{2}e^{-G/2}$. Finally since $\lim_{n\to\infty} \frac{d_n}{\sqrt{n}} = 1$, the result follows.
iv) $A_n = \overline{Y}_{M_n}$. Thus from i) $A_n - \log M_n \to G$ a.s. as $n \to \infty$. From iii) $\log(\frac{M_n}{\sqrt{n}})$ converges a.s to $\sqrt{2}e^{-G/2}$. But

$$A_n - \log M_n = A_n - \log(\sqrt{n}\frac{M_n}{\sqrt{n}}) = A_n - (\log n)/2 - \log(\frac{M_n}{\sqrt{n}}).$$

Hence, $A_n - (\log n)/2$ converges a.s. to $G + \log(\sqrt{2}e^{-G/2})$ from which iv) follows. □

4 Pareto Distribution

In this section and the ensuing section different families of distributions are considered. The Pareto distribution in Section 4, and the Beta distribution with parameters $(\alpha, 1)$ in Section 5, along with the exponential distribution that was previously discussed are representative of families that belong to the three domains of attraction of extreme values (see Leadbetter, Lindgren and Rootzén, 1983). The paradigm of extreme values is suitable in this context because ultimately the average of kept observations will be governed by the behavior in the tail of the distribution generating the observations.

From the theorems in the previous section, once a condition on $P_j = 1 - F(\beta\overline{Y}_j)$ is established, it follows that T_k, M_n and A_n suitably normalized converge a.s. The result in this section and the next section that follows is to show that $\overline{Y}_k$ suitably normalized converges and $j^\omega P_j \to W$ where $P(0 < W < \infty) = 1$.

Specifically, let $X_1, X_2, \ldots$ be i.i.d. from a Pareto distribution. The Pareto distribution is given by

$$F_\alpha(x) = (1 - x^{-\alpha})I(x \geq 1) \tag{4.1}$$

and hence

$$f_\alpha(x) = \alpha x^{-(\alpha+1)}I(x \geq 1). \tag{4.2}$$

Note that for $X \sim F_\alpha$ we have for $x \geq c \geq 1$

$$P(X > x|X > c) = (x/c)^{-\alpha} = P(cX > x). \tag{4.3}$$

Let $U_1, U_2, \ldots$ be i.i.d. with distribution F_α. From (4.3) it follows (taking $c = \beta\overline{Y}_{k-1}$) that Y_k can be represented as

$$Y_k = U_k \beta \overline{Y}_{k-1} \quad \text{for } k = 1, 2, \dots \tag{4.4}$$

with $\overline{Y}_0 \equiv \beta^{-1}$. This representation is justified since conditional on $Y_k \geq \beta \overline{Y}_{k-1}$ the distribution of $\overline{Y}_k$ depends on $\overline{Y}_{k-1}$ in the above multiplicative way.

Since $P_j = (\beta \overline{Y}_{j-1})^{-\alpha}$, it is sufficient to show that $\overline{Y_j}$ suitably normalized converges almost surely to a random variable W where $P(0 < W < \infty) = 1$. From (4.4)

$$\overline{Y}_k = \frac{(k-1)\overline{Y}_{k-1} + U_k \beta \overline{Y}_{k-1}}{k} = \frac{\beta U_k + k - 1}{k} \overline{Y}_{k-1}. \tag{4.5}$$

For $X \sim F_\alpha(x)$ to have finite expectation, one must have $\alpha > 1$, and we therefore at first assume this to hold. Since $E(U_k) = \alpha/(\alpha-1)$ it follows that

$$a_k \equiv E(\frac{\beta U_k + k - 1}{k}) = 1 + \frac{\beta\alpha/(\alpha-1) - 1}{k} \quad \text{for } k \geq 1.$$

Thus, $E(\overline{Y}_k | \overline{Y}_{k-1}) = a_k \overline{Y}_{k-1}$. Now let

$$b_k = (\prod_{j=1}^{k} a_j)^{-1} \quad \textit{and} \quad V_k = b_k \overline{Y}_k. \tag{4.6}$$

It follows that $\{V_k\}$ is a nonnegative martingale sequence with expectation β^{-1}, and thus, by the Martingale Convergence Theorem, converges a.s. to a finite limit, i.e.

$$b_k \overline{Y}_k \to Y^* \quad \text{a.s. as } k \to \infty. \tag{4.7}$$

We shall write (4.5) in a form that shows the rate of convergence more clearly.

$$\begin{aligned} b_k^{-1} &= \prod_{j=1}^{k} a_j = e^{\sum_{j=1}^{k} \log a_j} = e^{\sum_{j=1}^{k} \log(1 + \frac{(\beta-1)\alpha+1}{(\alpha-1)j})} \\ &= \exp(\sum_{j=1}^{k} (\frac{(\beta-1)\alpha+1}{(\alpha-1)j} + O(j^{-2}))) = \exp(\frac{(\beta-1)\alpha+1}{\alpha-1} \log k + \delta_k) \\ &= k^{\frac{(\beta-1)\alpha+1}{\alpha-1}} D_k \end{aligned}$$

where D_k converges to a positive constant. It follows that (4.7) can be written as

$$k^{-\frac{(\beta-1)\alpha+1}{\alpha-1}} \overline{Y}_k \to Y \quad \text{a.s. as } k \to \infty \tag{4.8}$$

where Y has finite expectation.

We have shown that $P(Y < \infty) = 1$. We need to show that $P(Y > 0) = 1$. It suffices to show that $E[\log(k^{-c}\overline{Y}_k)] > A$ for some constant $A > -\infty$ for all k, where $c = \frac{(\beta-1)\alpha+1}{\alpha-1}$. We use $\overline{Y}_k = \overline{Y}_{k-1}(\frac{k-1+\beta U_k}{k})$ where U_k are i.i.d. Pareto(α) of (4.1).

Let $\Delta_k = E[\log(k^{-c}\overline{Y}_k)] - E[\log(k-1)^{-c}\overline{Y}_{k-1})]$. It is sufficient to show that $\sum_{k=1}^{\infty} |\Delta_k| < \infty$.

Note that $\Delta_k = c\log\frac{k-1}{k} + E(Z_k)$ where $Z_k = \log(\frac{k-1+\beta U_k}{k})$ and:

Lemma 4.1.

$$E(Z_k) = \begin{cases} \frac{c}{k} + O(\frac{1}{k^\alpha}) & \textit{if } 1 < \alpha < 2 \\ \frac{c}{k} + O(\frac{\log k}{k^2}) & \textit{if } \alpha = 2 \\ \frac{c}{k} + O(\frac{1}{k^2}) & \textit{if } \alpha > 2. \end{cases}$$

If the lemma is true then $\sum_{k=1}^{\infty} |\Delta_k| < \infty$ because $\Delta_k = -\frac{c}{k} + \frac{c}{k} + O(\frac{1}{k^{\min(\alpha,2)}})$ and all $O()$ terms are summable.

Proof.

$$E(Z_k) = E[\log(\frac{k-1+\beta U_k}{k})] = \log(\frac{k-1}{k}) + E[\log(1 + \frac{\beta U_k}{k-1})].$$

Consider

$$\begin{aligned} E[\log(1+\frac{\beta U_k}{k-1})] &= \int_{x=1}^{\infty} \log(1+\frac{\beta x}{k-1})\alpha x^{-(\alpha+1)}dx \\ &= (\frac{\beta}{k-1})^\alpha \int_{y=\beta/(k-1)}^{\infty} \log(1+y)\alpha y^{-(\alpha+1)}dy \\ &= (\frac{\beta}{k-1})^\alpha \int_{y=\beta/(k-1)}^{1} \log(1+y)\alpha y^{-(\alpha+1)}dy \\ &+ (\frac{\beta}{k-1})^\alpha \int_{y=1}^{\infty} \log(1+y)\alpha y^{-(\alpha+1)}dy \end{aligned}$$

provided that $k \geq \beta + 1$. The last term in the above expression is of order $O(\frac{1}{k^\alpha})$. Note that

$$y - y^2/2 \leq \log(1+y) \leq y. \tag{4.9}$$

But

$$(\frac{\beta}{k-1})^\alpha \int_{y=\beta/(k-1)}^{1} y^t \alpha y^{-(\alpha+1)}dy = \begin{cases} \frac{\alpha}{t-\alpha}[(\frac{\beta}{k-1})^\alpha - (\frac{\beta}{k-1})^t] & \text{if } \alpha \neq t \\ \alpha(\frac{\beta}{k-1})^\alpha(\log(k-1) - \log\beta) & \text{if } \alpha = t. \end{cases} \tag{4.10}$$

The leading term is when $t = 1$ ($\alpha \neq t$, always in this case) and that gives the leading term in the approximation of the first integral, of $\frac{\alpha\beta}{\alpha-1} \cdot \frac{1}{k-1} + O(\frac{1}{k^\alpha}) = \frac{\alpha\beta}{(\alpha-1)k} + O(\frac{1}{k^2}) + O(\frac{1}{k^\alpha})$.

- If $t = 2$ and $\alpha \neq 2$ (4.10) becomes $O(\frac{1}{k^2}) + O(\frac{1}{k^\alpha})$.
- If $t = 2$ and $\alpha = 2$ (4.10) becomes $O(\frac{\log k}{k^\alpha})$.

Taking $\frac{\alpha\beta}{(\alpha-1)k} + \log(\frac{\{k-1}{k})$ yields $\frac{c}{k} + O(\frac{1}{k^2})$. This proves the lemma as the remainder terms are either $O(\frac{1}{k^\alpha})$ if $1 < \alpha < 2$, $O(\frac{\log k}{k^2})$ if $\alpha = 2$ and $O(\frac{1}{k^2})$ if $\alpha > 2$. $\square$

Theorem 4.1. *If F is Pareto(α) with $\alpha > 1$ and $\beta > \frac{\alpha-1}{\alpha}$ then*

$$i)\frac{\overline{Y}_k}{k^{\frac{(\beta-1)\alpha+1}{\alpha-1}}}$$

converges a.s. as $k \to \infty$ to a positive finite r.v.

$$ii)\frac{T_k}{k^{\frac{(\beta-1)\alpha^2+2\alpha-1}{\alpha-1}}}$$

converges a.s as $k \to \infty$ to a positive finite r.v.

$$iii)\frac{M_n}{n^{\frac{\alpha-1}{(\beta-1)\alpha^2+2\alpha-1}}}$$

converges a.s. as $n \to \infty$ to a positive finite r.v.

$$iv)\frac{A_n}{n^{\frac{(\beta-1)\alpha+1}{(\beta-1)\alpha^2+2\alpha-1}}}$$

converges a.s. as $n \to \infty$ to a positive finite r.v.

Proof. First assume that $\beta \geq 1$.
i) This follows from (4.8).
ii) Since $P_j = (\beta\overline{Y}_{j-1})^{-\alpha}$, ω in (3.1) is $\frac{(\beta-1)\alpha+1}{\alpha-1}\alpha$ and the result follows from Theorem 3.2.
iii) This follows from Theorem 3.3 with $d_n = n^{\frac{\alpha-1}{(\beta-1)\alpha^2+2\alpha-1}}$.
iv) Since $\gamma = \frac{(\beta-1)\alpha+1}{\alpha-1}$ and $\psi = \frac{\alpha-1}{(\beta-1)\alpha^2+2\alpha-1}$ (hence $\gamma\psi = \frac{(\beta-1)\alpha+1}{(\beta-1)\alpha^2+2\alpha-1}$) the result follows from Theorem 3.4.
If $\frac{\alpha-1}{\alpha} < \beta < 1$ then eventually $\overline{Y}_k$ will exceed $1/\beta$ almost surely, and henceforth the representation (4.4) is valid and the proof follows as above. $\square$

In particular, when $\beta = 1$, $\frac{\overline{Y}_k}{k^{\frac{1}{\alpha-1}}}$, $\frac{T_k}{k^{\frac{2\alpha-1}{\alpha-1}}}$, $\frac{M_n}{n^{\frac{\alpha-1}{2\alpha-1}}}$ and $\frac{A_n}{n^{\frac{1}{2\alpha-1}}}$ converge a.s.

Remark 4.1 Consider Theorem 2.1 which relates M_n for two distributions that are connected to each other in a particular way. For the Pareto distribution $F(x) = (1 - x^{-\alpha})I(x \geq 1)$ with $\alpha > 0$. This implies that if $X \sim$ Pareto(α) then $Z = X^{\alpha/\alpha^*} \sim$ Pareto(α^*). It follows that the number of items retained after n are observed will be stochastically at least as large for α^* as it is for α when $\alpha^* > \alpha$. This is consistent with the previous theorem which says that M_n is of order $n^{\frac{\alpha-1}{2\alpha-1}}$.

Remark 4.2 Let X_{ij}, $i = 1, 2, \ldots$, be i.i.d. from a Pareto distribution with parameter j for $j = 1, 2, \ldots$ and consider $V_{ij} = j(X_{ij} - 1)$. It is easy to verify that the c.d.f of V_{ij} is $F_j(v) = 1 - (1 + \frac{v}{j})^{-j}$. In accordance with Theorem 2.2, $\{F_j\}$ is a sequence of cumulative distribution functions that converges to $F(v) = 1 - e^{-v}$ which is the c.d.f of an exponential random variable. Hence when observations are generated according to a Pareto distribution, as the parameter becomes large, the number of items retained after n items are observed or the number of items required until k are kept behave in the limit (as $j \to \infty$) as if the observations were generated from an exponential distribution.

Even though the Pareto distribution with $\alpha \leq 1$ does not have finite mean, nevertheless, $\overline{Y}_k$ suitably normalized still converges almost surely. It is apparent from equation (4.5) that

$$\log \overline{Y}_k = \log \overline{Y}_{k-1} + Z_k$$

where $Z_1 = U_1$ and $Z_k = \log \frac{k-1}{k} + \log(1 + \frac{\beta U_k}{k-1})$ when $k > 1$, where U_i are i.i.d. Pareto(α). Hence,

$$\log \overline{Y}_k = \sum_{j=1}^{k} Z_j.$$

The following lemma provides a handle on the rate at which $E(Z_k)$ and $E(Z_k^2)$ grow, which is needed along with Theorem 3.1 to obtain the desired results.

Lemma 4.2. *i) If* $0 < \alpha < 1$ *then for all* $\beta > 0$

$$k^{\alpha} E(Z_k) \to c_{\alpha,\beta} \quad \textit{as } k \to \infty$$

and

$$k^{\alpha} E({Z_k}^2) \to d_{\alpha,\beta} \quad \textit{as } k \to \infty,$$

where $c_{\alpha,\beta}$ and $d_{\alpha,\beta}$ are positive constants.
ii) If $\alpha = 1$ then for all $\beta > 0$

$$\frac{E(Z_k)}{(\log k)/k} \to c_{1,\beta} \quad \text{as } k \to \infty$$

and

$$kE({Z_k}^2) \to d_{1,\beta} \quad \text{as } k \to \infty.$$

Proof. The results follow by substituting in the bounds and the bounds squared on $\log(1+y)$ in (4.9) into the results in (4.10) and realizing that $\log \frac{k-1}{k}$ is of lower order. $\square$

Theorem 4.2. *i) If $0 < \alpha < 1$ then for all $\beta > 0$*

$$\frac{\log \overline{Y}_k}{k^{1-\alpha}} \quad \text{converges a.s. as } k \to \infty.$$

ii) If $\alpha = 1$ then for all $\beta > 0$

$$\frac{\log \overline{Y}_k}{(\log k)^2} \quad \text{converges a.s. as } k \to \infty.$$

Proof. i) The proof applies Theorem 3.1. Specifically, let $Q_i = Z_i - E(Z_i)$ and $b_n = \sum_{i=1}^n E(Z_i)$. Then b_n is increasing for large enough n, since by Lemma 4.2 $E(Z_i) > 0$ for large enough n. Also, from Lemma 4.2, the sum is of order $n^{1-\alpha}$ which goes to infinity. But $E(Q_n^2)$ is of order $1/n^\alpha$ and b_n^2 is of order $n^{2-2\alpha}$. Hence $E(Q_n^2)/b_n^2$ is $O(\frac{1}{n^{2-\alpha}})$ which is summable. Theorem 3.1 then implies that $\frac{\sum_{i=1}^n Z_i}{b_n}$ converges almost surely to 1. But b_n is of order $n^{1-\alpha}$ and $\sum_{i=1}^n Z_i = \log \overline{Y}_n$.
ii) The proof is the same in form as i). The only difference is that since $E(Z_n)$ is $O(\frac{\log n}{n})$ that implies that b_n is $O(\log^2 n)$. It follows that $E(Q_n^2)/b_n^2$ is $O(\frac{1}{n \log^2 n})$ which is summable. $\square$

Since $\frac{\log \overline{Y}_k}{k^{1-\alpha}}$ converges almost surely for $\alpha < 1$ it follows that $\overline{Y_k}^{\frac{1}{k^{1-\alpha}}}$ converges almost surely. Similarly, when $\alpha = 1$, $\overline{Y_k}^{\frac{1}{\log^2 k}}$ converges almost surely. The behavior of the other quantities of interest, T_k, M_n, and A_n, for the Pareto distribution with $\alpha \leq 1$, is more complicated and hence omitted from this discussion.

5 Beta Distribution

We assume observations X_i are i.i.d. from a Beta distribtion with parameters $(\alpha, 1)$, i.e.,

$$F_\alpha(x) = x^\alpha I(0 \leq x \leq 1) + I(x \geq 1), \quad \alpha > 0. \tag{5.1}$$

Here we consider the case where "better" means "smaller". In this example, we retain X_n if it is smaller than βA_{n-1} where A_{n-1} is the average of the items retained after $n-1$ items are observed.

The reason we frame "better" to be "smaller" for Beta distributions is that we are confronted with anomalous situations for any random variable with support $[0, L)$ with $L < \infty$, if better is larger, if $\beta > 1$. For example, if $\beta = 2$, then here once $A_{n-1} > 1/2$, since no observation is greater than 1, all ensuing observations will not be kept. Note that if better than average is chosen as the criterion and $\beta = 1$ then the problems of "smaller than average" and "larger than average" are related by letting $X_i^* = 1 - X_i$ and considering the distribution of $Beta(1, \alpha)$ for X_i^*.

Clearly here smaller β values result in more stringent rules. For this reason we assume that β is small throughout the present section. Let W_k be the value of the k^{th} item that is retained and $\overline{W}_k$ the average of the first k items that are retained. Since for $F_\alpha(x)$ in (5.1) for $0 \leq x \leq c \leq 1$

$$P(X \leq x | X \leq c) = (\frac{x}{c})^\alpha = P(cX \leq x)$$

we can write

$$W_k = \beta \overline{W}_{k-1} U_k \tag{5.2}$$

where U_k are i.i.d. with c.d.f. $F_\alpha(x)$ as in (5.1).

It follows that

$$\overline{W}_k = \frac{(k-1)\overline{W}_{k-1} + U_k \overline{W}_{k-1}\beta}{k} = (1 + \frac{U_k\beta - 1}{k})\overline{W}_{k-1}$$

with $W_0 = 1/\beta$.

Hence, $E(\overline{W}_k | \mathcal{F}_{k-1}) = a_k \overline{W}_{k-1}$, where $a_k = 1 - \frac{1-\beta\alpha/(\alpha+1)}{k}$, since $E(U_k) = \alpha/(\alpha+1)$.

Let

$$b_k = [\prod_{j=1}^{k} a_j]^{-1}, \tag{5.3}$$

and consider $R_k = \overline{W}_k b_k$ for $k \geq 1$. It follows that R_k is a non-negative martingale, with expectation $1/\beta$, and hence

$$R_k \to R \quad \text{a.s., as } k \to \infty.$$

We shall write (5.3) in a slightly different form which will show the rate of convergence more explicitly. For each α and β let

$$\gamma = \frac{\alpha\beta - (\alpha+1)}{\alpha+1} < 0.$$

Then

$$b_k^{-1} = \prod_{j=1}^{k} a_j = \prod_{j=1}^{k}(1+\frac{\gamma}{j}) = D_k k^{\gamma}$$

where D_k tends to a finite positive limit as $k \to \infty$. Thus for $\beta \le \frac{\alpha+1}{\alpha}$ it follows that

$$k^{1-\frac{\alpha\beta}{\alpha+1}}\overline{W}_k \to W \quad \text{a.s., for } k \to \infty,$$

where W has finite expectation.

We know $P(W < \infty) = 1$. We need to show that $P(W > 0) = 1$ to apply the results of Section 3 obtained as "smaller than".

It suffices to show that $E[\log(k^{1-\frac{\alpha\beta}{\alpha+1}}\overline{W}_k)] \ge -A$ for some positive constant A for all $k \ge 2$. Since $\overline{W}_k = W_1 \prod_{j=2}^{k}(1-\frac{1-U_j\beta}{j})$ where U_i are i.i.d. $F_\alpha(x)$, then

$$\begin{aligned} E[\log(k^{1-\frac{\alpha\beta}{\alpha+1}}\overline{W}_k)] &= (1-\frac{\alpha\beta}{\alpha+1})\log k + E[\log \overline{W}_k] \\ &= (1-\frac{\alpha\beta}{\alpha+1})\log k + E[\log U_1] + \sum_{j=2}^{k} E[\log(1-\frac{1-U_j\beta}{j})], \end{aligned}$$

$$i)E[\log U_1] = \int_0^1 (\log x)\alpha x^{\alpha-1} = (\log x)x^{\alpha}|_0^1 - \int_0^1 x^{\alpha-1}dx = -\frac{1}{\alpha}$$

and

$$ii)E[\log(1-\frac{1-U_j\beta}{j})] \ge -E[\frac{1-U_j\beta}{j}] - E[(\frac{1-U_j\beta}{j})^2] \ge -\frac{1-\alpha\beta/(\alpha+1)}{j} - \frac{1}{j^2}$$

for all $j \ge 2$. The first inequality follows because $\log(1-u) \ge -u-u^2$ for $0 \le u \le 1/2$. Hence,

$$\sum_{j=2}^{k} E[\log(1-\frac{U_j\beta}{j})] \ge -(1-\frac{\alpha\beta}{\alpha+1})\sum_{j=2}^{k}\frac{1}{j} - \frac{\pi^2}{6}. \qquad (5.4)$$

Substituting (5.4) and the result in i) into the main expression yields

$$E[\log(k^{1-\frac{\alpha\beta}{\alpha+1}}\overline{W}_k] \geq (1-\frac{\alpha\beta}{\alpha+1})(\log k - \sum_{j=2}^{k}\frac{1}{j}) - \frac{1}{\alpha} - \frac{\pi^2}{6} \geq -(\frac{1}{\alpha}+\frac{\pi^2}{6}).$$

The last inequality follows from $\sum_{j=2}^{k}\frac{1}{j} \leq \log k$.

Theorem 5.1. *If F is Beta(α,1) then for $\beta < \frac{\alpha+1}{\alpha}$*

i) $k^{1-\frac{\alpha\beta}{\alpha+1}}\overline{W}_k$ converges a.s. as $k \to \infty$.

ii) $\frac{T_k}{k^{\frac{(\alpha+1)^2-\alpha^2\beta}{\alpha+1}}}$ converges a.s. as $k \to \infty$.

iii) $\frac{M_n}{n^{\frac{\alpha+1}{(\alpha+1)^2-\alpha^2\beta}}}$ converges a.s. as $n \to \infty$.

iv) $A_n n^{\frac{\alpha+1-\alpha\beta}{(\alpha+1)^2-\alpha^2\beta}}$ converges a.s. as $n \to \infty$

In particular, when $\beta = 1$, $k^{\frac{1}{\alpha+1}}\overline{W}_k$, $\frac{T_k}{k^{\frac{2\alpha+1}{\alpha+1}}}$, $\frac{M_n}{n^{\frac{\alpha+1}{2\alpha+1}}}$, and $A_n n^{\frac{1}{2\alpha+1}}$ converge a.s.

Proof. First assume that $\beta \leq 1$. Since here $P_j = F(\beta\overline{W}_{j-1}) = (\beta\overline{W}_{j-1})^\alpha$, this implies that ω in (3.1) is $\frac{\alpha+1-\alpha\beta}{\alpha+1}\alpha$. Since $\frac{T_k}{k^{\frac{\alpha+1-\alpha\beta}{\alpha+1}\alpha+1}} = \frac{T_k}{k^{\frac{(\alpha+1)^2-\alpha^2\beta}{\alpha+1}}}$, thus (ii) follows from Theorem 3.2.

iii) This follows from Theorem 3.3.

iv) This follows from Theorem 3.4 with γ and ψ as given.

If $1 < \beta < \frac{\alpha+1}{\alpha}$ then eventually $\overline{W}_k$ will exceed $1/\beta$ almost surely, and henceforth representation (5.2) is valid and the proof follows as above. □

Remark 5.1 For $\beta = 1$ the smaller than average rule corresponds to the larger than average rule for Beta(1,α).
Let $X \sim [1-(1-x)^\alpha]I(0 \leq x \leq 1)$ and $Z \sim [1-(1-z)^{\alpha^*}]I(0 \leq z \leq 1)$ then Z is distributed like $g(X)$ where $g(x) = 1-(1-x)^{\alpha/\alpha^*}$. Clearly g is increasing and concave if $\alpha^* < \alpha$. It follows from Theorem 2.1 that $M_n(\alpha)$ is stochastically decreasing in α for every n and hence also $E(M_n(\alpha))$ is decreasing in α.

Remark 5.2 Let X have the Pareto distribution with parameter ν and consider $g(x) = 1-x^{-\nu/\alpha}$. Clearly g is increasing. It is concave for all $\nu > 0$ and $\alpha > 0$. Let $Z = g(X)$. It is easily seen that Z is Beta$(1,\alpha)$. For $\beta = 1$, Theorem 2.1 implies that $E(M_n^*(\nu)) \leq E(M_n(\alpha))$ for all α, ν and n, where $M_n^*(\nu)$ is the number retained by time n by a Pareto (ν) random variable.

Remark 5.3 Let X_{ij}, $i = 1, 2, \ldots$, be i.i.d. from a Beta distribution with parameters $(1, j)$ for $j = 1, 2, \ldots$ and consider $V_{ij} = jX_{ij}$. It is easy to verify that the c.d.f of V_{ij} is $F_j(v) = 1 - (1 - \frac{v}{j})^j, 0 < v < j$. In accordance with Theorem 2.2, $\{F_j\}$ is a sequence of cumulative distribution functions that converges to F, with $F(v) = 1 - e^{-v}$, which is the c.d.f of an exponential random variable. Hence when observations are generated from a Beta$(1, \alpha)$ distribution, as the α becomes large, the number of items retained after n items are observed or the number of items required until k items are kept behave in the limit (as $j \to \infty$) as if the observations were generated from an exponential distribution.

6 Remarks

Items are observed sequentially. At the time an item is observed it is either retained into the selected set or discarded. The class of rules that is considered in this paper retains an item if it is β times better than the average of those items that have already been selected.

Remark 6.1 This is in contrast to the rule considered in Krieger, Pollak, Samuel-Cahn (2007) in which an item is retained if its rank is sufficiently high (low) among the items already retained. In this paper and the previous one, it is assumed that items are generated in an i.i.d. fashion from a common distribution. The rules in these two papers can be implemented without knowledge of this distribution. The behavior of the rules, in terms of the number items retained and quality of the retained items, depends on the distribution when average is used as the baseline for admittance into the set in contrast to rules based on ranks where the distribution does not play a role when performance is judged in terms of ranks. For example, the number of retained items after n items are observed grows at a rate of $\sqrt{n}$, regardless of the distribution, when an item needs to be better than the median of the items already retained, in order to be selected. The rule that retains an item when it is larger than the average of those already retained has a size that grows at the rate of $\sqrt{n}$ for the exponential distribution, but, for example, at the rate of $n^{2/3}$ if items are generated from a uniform distribution.

Remark 6.2 Preater (2000) considered the behavior of the average of the first k items that are kept, $\overline{Y}_k$, when the distribution generating the observations is exponential and $\beta = 1$ in the β better than average rule. The behavior of this quantity for $\beta > 1$ is markedly different. The results in the present paper (see Theorem 3.5) provide the behavior of the number of items required until k items are kept for the better than average rule as well as behavior up to time n. If the rule is extended to β better than average, the asymptotic behavior of the number of items required depends on the choice of β. These results, which are more complicated, will appear in an ensuing paper.

Remark 6.3 The performance of the rule when the underlying distribution is normal is also more complicated. Analogous to the result for the exponential distribution that $\overline{Y}_k - \log k$ converges almost surely to a Gumbel distribution, it can be shown for the normal that $\overline{Y}_k$ less a suitably chosen function of k converges almost surely, but not to a Gumbel distribution. The results for the normal case are also left to another article.

Remark 6.4 It is not surprising that there should be some relationship between the domain of attraction to which the extremal distribution of F belongs and the limiting distribution of $\overline{Y}_k$, since the Y_k process will, on the average, select larger and larger items. Preater(2000) shows that $\overline{Y}_k$ with change of location and $\max\{X_1, \ldots, X_k\}$ with change of location and scale have the exact same limiting Gumbel distribution when the observations are i.i.d. from an exponential distribution (though $\overline{Y}_k$ converges a.s. and in L_2 while the maximum only converges in distribution). Will the limiting distribution of $\overline{Y}_k$, and $\max\{X_1, \ldots, X_k\}$ always agree, or at least have the same rate of convergence? As we have seen both for the Pareto and Beta distributions this is not necessarily the case for other distributions.

Remark 6.5 In this paper we essentially restricted β to be sufficiently large (small) when better is "larger" ("smaller"), often $\beta \geq 1$ ($\beta \leq 1$). What happens to these rules if β is small (large)? It can be shown that in certain cases when β is sufficiently small (large), particularly when the domain of X is restricted either by $0 < a < X$ (e.g., the Pareto distribution) or $X < b < \infty$ (e.g., for the Beta distribution), $\overline{Y}_k$ converges a.s. to the expected value of the underlying distribution.

Remark 6.6 Since the tail of the the distribution is all that matters in terms of the asymptotic behavior families can be extended without altering the behavior of the rules. For example, for the Pareto distributions, we can extend to the class when $1 - F(x) = x^{-\alpha}L(x)$ where $L(x)$ is slowly varying, i.e., $\lim_{x\to\infty} L(tx)/L(x) = 1$ for all $t > 0$.

Remark 6.7 In certain cases the random variables of interest converge in L_2 as well, thereby implying that the limiting distribution has finite variance. This can be shown for $\overline{Y}_k$ suitably normalized for the Pareto distribution when better is "larger" if $\alpha > 2$ and $\beta \geq 1$. It can also be shown for $\overline{W}_k$ for the Beta distribution where better is "smaller" for all α and $\beta \leq 1$.

References

[1] Feller, W. (1971). *Probability* Volume 2, Second Edition. John Wiley & Sons, Inc., New York.

[2] Krieger, A.M., Pollak, M. and Samuel-Cahn, E. (2007). Select sets: Rank and file. *Annals of Applied Probability*, 17, 360-385.

[3] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Sequences and Processes*. Springer, New York.

[4] Preater, J. (2000). Sequential selection with a better than average rule. *Statistics and Probability Letters*, 50, 187-191.