

BEAT THE MEAN: SEQUENTIAL SELECTION BY BETTER THAN AVERAGE RULES

ABBA M. KRIEGER,* **** *University of Pennsylvania*

MOSHE POLLAK ** **** AND

ESTER SAMUEL-CAHN,** ***** *The Hebrew University of Jerusalem*

Abstract

We consider a sequential rule, where an item is chosen into the group, such as a university faculty member, only if his/her 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 the 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 ∞ . The assumption throughout is that the items are independent and 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 $\beta > 0$.

Keywords: Selection rule; average; nonparametric; sequential observation; asymptotics

2000 Mathematics Subject Classification: Primary 62G99

Secondary 62F07; 60F15

1. Introduction

In many practical situations, when it is desired to sequentially assemble a group of good quality, the ‘better than average’ rule is used. This rule selects an individual into the group if and only if he/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 using the assumption that the observations are independent and identically distributed (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

Received 21 August 2007; revision received 27 November 2007.

* Postal address: Statistics Department, The Wharton School, University of Pennsylvania, 3730 Walnut Street, Philadelphia, PA 19104-6340, USA. Email address: kriegier@wharton.upenn.edu

** Postal address: Department of Statistics, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem, 91905, Israel.

*** 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.

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 in Preater (2000). He dealt with exponentially generated values and related cases. The focus of his paper was 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 in Krieger *et al.* (2007b). 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 items are retained, and 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 heavily depend on the underlying distribution of the i.i.d. random variables. The Pareto family is explored in Section 4, and the beta family is explored in Section 5. These three families of distributions are representatives of the three domains of attraction of extreme values.

2. Notation and preliminary results

In this paper we consider a sequence of observations that are i.i.d. 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, \dots 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 the 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;
- $\bar{Y}_k = \bar{Y}_k(F) = \bar{Y}_k(X)$ = 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 any of the three forms of notation as convenient.

The β better than average rule is defined as follows: for fixed β (which is suppressed in the notation) and T_k , 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

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

and

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

It is clear that \bar{Y}_k increases in k for $\beta = 1$. If $\beta > 1$, we assume nonnegative X_i to avoid the situation that if \bar{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 $x_F = \infty$, clearly $T_k < \infty$ almost surely (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 $\bar{Y}_1 = x_F$), the sequence T_1, T_2, \dots is well defined for all $k \geq 1$ and the better than average rule satisfies*

$$\bar{Y}_k \rightarrow x_F \quad \text{a.s. as } k \rightarrow \infty. \tag{2.1}$$

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

Proof. First consider the better than average rule. Let $\bar{Y}_0 = 0$. It is easily seen by induction that

$$\bar{Y}_k = \sum_{j=1}^k \frac{Y_j - \bar{Y}_{j-1}}{j}.$$

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

$$Y = \sum_{k=1}^{\infty} \frac{Y_k - \bar{Y}_{k-1}}{k} \geq \sum_{k=1}^{\infty} \frac{(Y_k - x)^+}{k} \geq \sum_{k=1}^{\infty} \frac{\varepsilon \mathbf{1}(Y_k > x + \varepsilon)}{k},$$

where $a^+ = a \mathbf{1}(a > 0)$. 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 $\bar{Y}_k[\beta]$ satisfies $P(\bar{Y}_k[\beta] \geq \bar{Y}_k) = 1$ for the β better than average rule for all k ; hence, (2.1) follows.

We exclude the case in which $X_1 = x_F$ in our further considerations. We now turn to a general result that relates the number of items kept after n have been observed when $X \sim F$. Let Z_1, Z_2, \dots be a sequence of i.i.d. random variables with distribution G . 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. Note that we are assuming that the cumulative distribution functions F and G are strictly increasing, but this can easily be extended to nonstrict monotonicity.

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, \dots after n items are observed. Similarly, $M_n(Z)$ denotes the number of items kept using the Z -sequence. We have the following result.

Theorem 2.1. *Let $X \sim F$, and let $g(x)$ be an increasing concave function. Let G be the cumulative distribution function of a random variable distributed as $g(X)$. When $\beta = 1$, $M_n(G) \geq_{st} M_n(F)$ for all n (where ‘ \geq_{st} ’ means ‘stochastically not smaller than’). 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 the 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 that (i) and (ii) hold for $n - 1$. We consider the following two cases.

(a) The case in which $X_n > A_{n-1}(X)$. 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) The case in which $X_n \leq A_{n-1}(X)$. Hence, the n th item is not kept by the rule that uses the X -sequence. Part (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 that $\beta = 1$, 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 and X_2 are 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 CDFs that converges weakly to a continuous distribution F . Then*

- (i) *for fixed n , $M_n(F_j) \rightarrow M_n(F)$ as $j \rightarrow \infty$, in distribution;*
- (ii) *for fixed n , $A_n(F_j) \rightarrow A_n(F)$ as $j \rightarrow \infty$, in distribution;*
- (iii) *for fixed k , $T_k(F_j) \rightarrow T_k(F)$ as $j \rightarrow \infty$, in distribution;*
- (iv) *for fixed k , $\bar{Y}_k(F_j) \rightarrow \bar{Y}_k(F)$ as $j \rightarrow \infty$, in distribution.*

We omit the proof. For a proof, see Krieger *et al.* (2007a).

3. Almost sure convergence

In this section we find conditions on the behavior of \bar{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 \bar{Y}_k hold for various families of distributions.

Of the four quantities considered, \bar{Y}_k is the easiest to handle directly, often through the use of the martingale convergence theorem. It is not unusual for \bar{Y}_k to satisfy condition (3.1) in Theorem 3.2, below. The asymptotic behavior of T_k can then also be determined, as in Theorem 3.2, below, using Theorem 3 of Feller (1971, p. 239), which we reproduce 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 σ -field generated by $\{X_1, \dots, X_n\}$. Then, clearly,

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

and

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

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 \bar{Y}_k and T_k , respectively, as given in Theorems 3.3 and 3.4, below.

Theorem 3.2, below, which considers the a.s. convergence of T_k relies on Theorem 3 of Feller (1971).

Theorem 3.1. (Feller’s theorem (1971, p. 239).) *Let Q_1, Q_2, \dots be independent random variables with $E(Q_n) = 0$, and let $S_n = \sum_{i=1}^n Q_i$. If $b_1 < b_2 < \dots \rightarrow \infty$ are constants and*

$$\sum_{n=1}^{\infty} E\left(\frac{Q_n^2}{b_n^2}\right) < \infty$$

then

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

We use Theorem 3.1 in the following way.

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

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

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

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

Proof. We shall use Feller’s theorem conditionally on the sequence $\{\bar{Y}_k\}$. Let

$$b_j = \sum_{i=1}^j P_i^{-1},$$

and let $Q_i = T_i - T_{i-1} - P_i^{-1}$ 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 $\varepsilon > 0$. Define J_ε to be such that

$$W(1 - \varepsilon) \leq j^\omega P_j \leq W(1 + \varepsilon) \quad \text{for all } j \geq J_\varepsilon,$$

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

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

Since both conditions of Feller’s theorem are satisfied,

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

Now

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

Also, for $n \geq J_\varepsilon$,

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

Therefore, for large enough n ,

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

Letting ε tend to 0, it follows, from (3.2) and (3.3), that

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

Now that we have 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 a.s. to a well-defined random variable (as we will show in Theorem 3.5, below, when F is exponential and $\beta = 1$), this states 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. Theorem 3.3, below, implies that, for the exponential distribution with $\beta = 1$, 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 ∞) with $\lim c_k = \infty$. We define a sequence of increasing positive constants (for the rate of M_n) by

$$d_n = \sup\{j \mid c_j \leq n\} \quad \text{for } n = 1, 2, \dots$$

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

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

exists for every $0 < x < \infty$ and that h is continuous, where $\lfloor x \rfloor$ is the largest integer less than or equal to x .

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

Theorem 3.3. *Let*

$$\frac{T_k}{c_k} \rightarrow Q \quad \text{a.s. for } k \rightarrow \infty,$$

where $0 < Q < \infty$. Assume that c_k is such that (3.4) holds, with continuous h . Let $M_n = \sup\{j \mid T_j \leq n\}$, $n = 1, 2, \dots$. Then

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

The proof is omitted. It can be found in Krieger *et al.* (2007a).

Suppose that $1 - F(\beta\bar{Y}_k)$, properly normalized, converges a.s. to a possibly nondegenerate 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 a.s. to possibly nondegenerate random variables. We now establish conditions such that A_n , suitably normalized, converges a.s. to a nondegenerate random variable.

Theorem 3.4. *Let $\bar{Y}_k/k^\gamma \rightarrow Y$ and $M_n/n^\psi \rightarrow M$ a.s. as $k \rightarrow \infty$ and $n \rightarrow \infty$, where $\gamma > 0$ and $\psi > 0$ are constants and $P(0 < Y < \infty) = P(0 < M < \infty) = 1$. Then*

$$\frac{A_n}{n^{\gamma\psi}} \rightarrow YM^\gamma \quad \text{a.s. as } n \rightarrow \infty.$$

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

$$Yk^\gamma(1 - \varepsilon) \leq \bar{Y}_k \leq Yk^\gamma(1 + \varepsilon)$$

and

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

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

$$YM_n^\gamma(1 - \varepsilon) \leq A_n \leq YM_n^\gamma(1 + \varepsilon)$$

and, for $n \geq n_\varepsilon$,

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

This implies that

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

so that

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

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 with mean 1. Let $\beta = 1$, and let G denote a random variable that has the Gumbel distribution, $\exp\{-e^{-x}\}$. Then*

- (i) $\bar{Y}_k - \log k$ converges a.s. to G as k tends to ∞ ;
- (ii) T_k/k^2 converges a.s. to $e^G/2$ as k tends to ∞ ;
- (iii) M_n/\sqrt{n} converges a.s. to $\sqrt{2}e^{-G/2}$ as n tends to ∞ ; and
- (iv) $A_n - (\log n)/2$ converges a.s. to $(G + \log 2)/2$ as n tends to ∞ .

Proof. (i) This result is given in Preater (2000).

(ii) The value of $P_j = 1 - F(\bar{Y}_{j-1}) = \exp(-\bar{Y}_{j-1})$. But, from (i), $\bar{Y}_j - \log j$ converges a.s. to G . Then $jP_j = j \exp(-\bar{Y}_{j-1}) \approx j e^{-(G+\log j)} = e^{-G} = W$. Hence, jP_j converges a.s., so $\omega = 1$ in Theorem 3.2, i.e. T_k/k^2 converges a.s. to $1/(\omega + 1)W = e^G/2$.

(iii) Since (by (ii)) T_k/k^2 converges a.s. to $Q = e^G/2$, it follows that c_k is equal to k^2 in Theorem 3.3. Hence, $d_n = \lfloor \sqrt{n} \rfloor$. This implies that

$$h(x) = \lim_{n \rightarrow \infty} \frac{\lfloor \sqrt{n/x} \rfloor}{\lfloor \sqrt{n} \rfloor} = \frac{1}{\sqrt{x}}.$$

Therefore, from Theorem 3.3, M_n/d_n converges a.s. to $1/\sqrt{Q} = \sqrt{2}e^{-G/2}$. Finally, since $\lim_{n \rightarrow \infty} d_n/\sqrt{n} = 1$, the result follows.

(iv) We have $A_n = \bar{Y}_{M_n}$. Thus, from (i), $A_n - \log M_n \rightarrow G$ a.s. as $n \rightarrow \infty$. From (iii), $\log(M_n/\sqrt{n})$ converges a.s. to $\sqrt{2}e^{-G/2}$. But,

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

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 Section 5 different families of distributions are considered. The Pareto distribution considered in this section and the beta distribution with parameters $(\alpha, 1)$ considered in Section 5, along with the exponential distribution discussed previously, are representative of families that belong to the three domains of attraction of extreme values (see Leadbetter *et al.* (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 $P_j = 1 - F(\beta\bar{Y}_j)$ is established, it follows that $T_k, M_n,$ and $A_n,$ suitably normalized, converge a.s.

Let X_1, X_2, \dots be i.i.d. from a Pareto distribution, given by

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

and, hence,

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

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

$$P(X > x \mid X > c) = \left(\frac{x}{c}\right)^{-\alpha} = P(cX > x). \tag{4.2}$$

Let U_1, U_2, \dots be i.i.d. with distribution $F_\alpha.$ From (4.2), it follows (taking $c = \beta\bar{Y}_{k-1}$) that Y_k can be represented as

$$Y_k = U_k \beta \bar{Y}_{k-1} \quad \text{for } k = 1, 2, \dots \tag{4.3}$$

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

Since $P_j = (\beta\bar{Y}_{j-1})^{-\alpha},$ it is sufficient to show that $\bar{Y}_j,$ suitably normalized, converges a.s. to a random variable $W,$ where $P(0 < W < \infty) = 1.$ From (4.3),

$$\bar{Y}_k = \frac{(k-1)\bar{Y}_{k-1} + U_k \beta \bar{Y}_{k-1}}{k} = \frac{\beta U_k + k - 1}{k} \bar{Y}_{k-1}. \tag{4.4}$$

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

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

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

$$b_k = \left(\prod_{j=1}^k a_j\right)^{-1} \quad \text{and} \quad V_k = b_k \bar{Y}_k. \tag{4.5}$$

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.

$$b_k \bar{Y}_k \rightarrow Y^* \quad \text{a.s. as } k \rightarrow \infty. \tag{4.6}$$

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

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

where D_k converges to a positive constant. It follows that (4.6) can be written as

$$k^{-((\beta - 1)\alpha + 1)/(\alpha - 1)} \bar{Y}_k \rightarrow Y \quad \text{a.s. as } k \rightarrow \infty, \tag{4.7}$$

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 $E(\log(k^{-c} \bar{Y}_k)) > A$ for some constant $A > -\infty$ for all k , where $c = ((\beta - 1)\alpha + 1)/(\alpha - 1)$. We use

$$\bar{Y}_k = \bar{Y}_{k-1} \left(\frac{k - 1 + \beta U_k}{k}\right),$$

where the U_k are the i.i.d. $\text{Pareto}(\alpha)$ of (4.1).

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

Note that $\Delta_k = c \log((k - 1)/k) + E(Z_k)$, where $Z_k = \log((k - 1 + \beta U_k)/k)$ and note the following result.

Lemma 4.1. *We have*

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

If Lemma 4.1 is true then $\sum_{k=1}^{\infty} |\Delta_k| < \infty$ because $\Delta_k = -c/k + c/k + O(1/k^{\min(\alpha, 2)})$ and all the $O(\cdot)$ terms are summable.

Proof of Lemma 4.1.

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

Consider

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

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

$$y - \frac{y^2}{2} \leq \log(1+y) \leq y. \tag{4.8}$$

But,

$$\left(\frac{\beta}{k-1}\right)^\alpha \int_{y=\beta/(k-1)}^1 y^t \alpha y^{-(\alpha+1)} dy = \begin{cases} \frac{\alpha}{t-\alpha} \left(\left(\frac{\beta}{k-1}\right)^\alpha - \left(\frac{\beta}{k-1}\right)^t \right) & \text{if } \alpha \neq t, \\ \alpha \left(\frac{\beta}{k-1}\right)^\alpha (\log(k-1) - \log \beta) & \text{if } \alpha = t. \end{cases} \tag{4.9}$$

We are interested in $t = 1, 2$. The leading term is when $t = 1$ ($\alpha \neq t$ in this case always), which gives a leading term in the approximation of the first integral of

$$\frac{\alpha\beta}{\alpha-1} \frac{1}{k-1} + O\left(\frac{1}{k^\alpha}\right) = \frac{\alpha\beta}{(\alpha-1)k} + O\left(\frac{1}{k^2}\right) + O\left(\frac{1}{k^\alpha}\right).$$

- If $t = 2$ and $\alpha \neq 2$, (4.9) becomes $O(1/k^2) + O(1/k^\alpha)$.
- If $t = 2$ and $\alpha = 2$, (4.9) becomes $O(\log k/k^\alpha)$.

Writing $\alpha\beta/(\alpha-1)k + \log((k-1)/k)$ as $c/k + O(1/k^2)$ completes the proof of the lemma as the remainder terms are either $O(1/k^\alpha)$ if $1 < \alpha < 2$, $O(\log k/k^2)$ if $\alpha = 2$, or $O(1/k^2)$ if $\alpha > 2$.

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

- (i) $\bar{Y}_k/k^{((\beta-1)\alpha+1)/(\alpha-1)}$ converges a.s. as k tends to ∞ ;
- (ii) $T_k/k^{((\beta-1)\alpha^2+2\alpha-1)/(\alpha-1)}$ converges a.s. as k tends to ∞ ;
- (iii) $M_n/n^{(\alpha-1)/((\beta-1)\alpha^2+2\alpha-1)}$ converges a.s. as n tends to ∞ to finite positive random variables; and
- (iv) $A_n/n^{((\beta-1)\alpha+1)/((\beta-1)\alpha^2+2\alpha-1)}$ converges a.s. as n tends to ∞ to finite positive random variables.

Proof. First assume that $\beta \geq 1$.

(i) This follows from (4.7).

(ii) Since $P_j = (\beta \bar{Y}_{j-1})^{-\alpha}$, ω in (3.1) is $((\beta - 1)\alpha + 1)/(\alpha - 1)\alpha$, and the result follows from Theorem 3.2.

(iii) This follows from Theorem 3.3 with $d_n = n^{(\alpha-1)/((\beta-1)\alpha^2+2\alpha-1)}$.

(iv) Since $\gamma = ((\beta - 1)\alpha + 1)/(\alpha - 1)$ and $\psi = (\alpha - 1)/((\beta - 1)\alpha^2 + 2\alpha - 1)$ (hence, $\gamma\psi = ((\beta - 1)\alpha + 1)/((\beta - 1)\alpha^2 + 2\alpha - 1)$), the result follows from Theorem 3.4.

If $(\alpha - 1)/\alpha < \beta < 1$ then eventually \bar{Y}_k will exceed $1/\beta$ a.s. and, henceforth, representation (4.3) is valid and the proof follows as above.

In particular, when $\beta = 1$,

$$\frac{\bar{Y}_k}{k^{1/(\alpha-1)}}, \quad \frac{T_k}{k^{(2\alpha-1)/(\alpha-1)}}, \quad \frac{M_n}{n^{(\alpha-1)/(2\alpha-1)}}, \quad \text{and} \quad \frac{A_n}{n^{1/(2\alpha-1)}}$$

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, $F(x) = (1 - x^{-\alpha}) \mathbf{1}(x \geq 1)$ with $\alpha > 0$. This implies that if $X \sim \text{Pareto}(\alpha)$ then $Z = X^{\alpha/\alpha^*} \sim \text{Pareto}(\alpha^*)$, where α^* is any positive constant. 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 Theorem 4.1, which states that M_n is of order $n^{(\alpha-1)/(2\alpha-1)}$.

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

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

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

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

Lemma 4.2, below, 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 $\beta > 0$,

$$k^\alpha E(Z_k) \rightarrow c_{\alpha,\beta} \quad \text{and} \quad k^\alpha E(Z_k^2) \rightarrow d_{\alpha,\beta} \quad \text{as } k \rightarrow \infty,$$

where $c_{\alpha,\beta}$ and $d_{\alpha,\beta}$ are positive constants.

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

$$\frac{E(Z_k)}{(\log k)/k} \rightarrow c_{1,\beta} \quad \text{and} \quad k E(Z_k^2) \rightarrow d_{1,\beta} \quad \text{as } k \rightarrow \infty.$$

Proof. The results follow by substituting the bounds and the bounds squared on $\log(1 + y)$ in (4.8) into the results in (4.9) and realizing that $\log((k - 1)/k)$ is of lower order.

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

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

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

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

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 ∞ . 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 $O(1/n^{2-\alpha})$, which is summable. Theorem 3.1 then implies that $\sum_{i=1}^n Z_i/b_n$ converges a.s. to 1. But, b_n is of order $n^{1-\alpha}$ and $\sum_{i=1}^n Z_i = \log \bar{Y}_n$.

(ii) The proof is the same in form as part (i). The only difference is that since $E(Z_n)$ is $O(\log n/n)$, this implies that b_n is $O(\log^2 n)$. It follows that $E(Q_n^2)/b_n^2$ is $O(1/n \log^2 n)$, which is summable.

Since $\log \bar{Y}_k/k^{1-\alpha}$ converges a.s. for $\alpha < 1$, it follows that $\bar{Y}_k^{1/k^{1-\alpha}}$ converges a.s. Similarly, when $\alpha = 1$, $\bar{Y}_k^{1/\log^2 k}$ converges a.s. 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 that observations X_i are i.i.d. from a beta distribution with parameters $(\alpha, 1)$, i.e.

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

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 $A_{n-1} > \frac{1}{2}$, all ensuing observations will not be kept. Note that if $\beta = 1$ 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 $\text{Beta}(1, \alpha)$ 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 let \bar{W}_k be the average of the first k items that are retained. Since, for $F_\alpha(x)$ in (5.1) for $0 \leq x \leq c \leq 1$,

$$P(X \leq x \mid X \leq c) = \left(\frac{x}{c}\right)^\alpha = P(cX \leq x),$$

we can write

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

where the U_k are i.i.d. with cumulative distribution function $F_\alpha(x)$, as in (5.1).

It follows that

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

with $W_0 = 1/\beta$. Hence, $E(\bar{W}_k | \mathcal{F}_{k-1}) = a_k \bar{W}_{k-1}$, where

$$a_k = 1 - \frac{1 - \beta\alpha/(\alpha + 1)}{k},$$

since $E(U_k) = \alpha/(\alpha + 1)$.

Let

$$b_k = \left(\prod_{j=1}^k a_j\right)^{-1}, \tag{5.3}$$

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

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

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

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

where D_k tends to a finite positive limit as k tends to ∞ . Thus, for $\beta \leq (\alpha + 1)/\alpha$, it follows that

$$k^{1-\alpha\beta/(\alpha+1)} \bar{W}_k \rightarrow W \quad \text{a.s. for } k \rightarrow \infty,$$

where W has finite expectation.

We know that $P(W < \infty) = 1$. We need to show that $P(W > 0) = 1$ in order to apply the results of Section 3.

It suffices to show that $E(\log(k^{1-\alpha\beta/(\alpha+1)} \bar{W}_k)) \geq -A$ for some positive constant A for all $k \geq 2$. Since

$$\bar{W}_k = W_1 \prod_{j=2}^k \left(1 - \frac{1 - U_j \beta}{j}\right),$$

where the U_i are i.i.d. $F_\alpha(x)$, then

$$\begin{aligned} E(\log(k^{1-\alpha\beta/(\alpha+1)} \bar{W}_k)) &= \left(1 - \frac{\alpha\beta}{\alpha + 1}\right) \log k + E(\log \bar{W}_k) \\ &= \left(1 - \frac{\alpha\beta}{\alpha + 1}\right) \log k + E(\log U_1) + \sum_{j=2}^k E\left(\log\left(1 - \frac{1 - U_j \beta}{j}\right)\right), \\ E(\log U_1) &= \int_0^1 (\log x) \alpha x^{\alpha-1} dx = (\log x) x^\alpha \Big|_0^1 - \int_0^1 x^{\alpha-1} dx = -\frac{1}{\alpha}, \end{aligned}$$

and

$$E\left(\log\left(1 - \frac{1 - U_j\beta}{j}\right)\right) \geq -E\left(\frac{1 - U_j\beta}{j}\right) - E\left(\left(\frac{1 - U_j\beta}{j}\right)^2\right) \geq -\frac{1 - \alpha\beta/(\alpha + 1)}{j} - \frac{1}{j^2}$$

for all $j \geq 2$. The first inequality follows because $\log(1 - u) \geq -u - u^2$ for $0 \leq u \leq \frac{1}{2}$. Hence,

$$\sum_{j=2}^k E\left(\log\left(1 - \frac{U_j\beta}{j}\right)\right) \geq -\left(1 - \frac{\alpha\beta}{\alpha + 1}\right) \sum_{j=2}^k \frac{1}{j} - \frac{\pi^2}{6}.$$

Therefore,

$$E(\log(k^{1-\alpha\beta/(\alpha+1)}\bar{W}_k)) \geq \left(1 - \frac{\alpha\beta}{\alpha + 1}\right) \left(\log k - \sum_{j=2}^k \frac{1}{j}\right) - \frac{1}{\alpha} - \frac{\pi^2}{6} \geq -\left(\frac{1}{\alpha} + \frac{\pi^2}{6}\right).$$

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

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

- (i) $k^{1-\alpha\beta/(\alpha+1)}\bar{W}_k$ converges a.s. as k tends to ∞ ;
- (ii) $T_k/k^{((\alpha+1)^2-\alpha^2\beta)/(\alpha+1)}$ converges a.s. as k tends to ∞ ;
- (iii) $M_n/n^{(\alpha+1)/((\alpha+1)^2-\alpha^2\beta)}$ converges a.s. as n tends to ∞ to finite positive random variables; and
- (iv) $A_n n^{(\alpha+1-\alpha\beta)/((\alpha+1)^2-\alpha^2\beta)}$ converges a.s. as n tends to ∞ to finite positive random variables.

In particular, when $\beta = 1$, $k^{1/(\alpha+1)}\bar{W}_k$, $T_k/k^{(2\alpha+1)/(\alpha+1)}$, $M_n/n^{(\alpha+1)/(2\alpha+1)}$, and $A_n n^{1/(2\alpha+1)}$ converge a.s. to finite positive random variables.

Proof. First assume that $\beta \leq 1$. Since here $P_j = F(\beta\bar{W}_{j-1}) = (\beta\bar{W}_{j-1})^\alpha$, this implies that ω in (3.1) is equal to $(\alpha + 1 - \alpha\beta)/(\alpha + 1)\alpha$. Since

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

part (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 $1 < \beta < (\alpha + 1)/\alpha$ then eventually \bar{W}_k will exceed $1/\beta$ a.s. and, henceforth, representation (5.2) is valid and the proof follows as above.

Remark 5.1. 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 of items retained by time n by a Pareto(ν) random variable.

Further remarks can be found in Krieger *et al.* (2007a). Among the topics discussed there are the following.

- (a) A comparison between the performance of the rules considered in Krieger *et al.* (2007b) and those of the present paper.
- (b) The behavior of the rules for β values outside the domains discussed in Theorems 3.5, 4.1, and 5.1.
- (c) The behavior of the rules under other distributions belonging to the same domains of attraction.
- (d) A comparison of the limiting behavior of $\max(X_1, \dots, X_k)$ and \bar{Y}_k belonging to the same domain.

References

- FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd edn. John Wiley, New York.
- KRIEGER, A. M., POLLAK, M. AND SAMUEL-CAHN, E. (2007a). Beat the mean: better the average. Discussion Paper 469, Center for the Study of Rationality, The Hebrew University of Jerusalem. Available at <http://www.ratio.huji.ac.il/dp.php>.
- KRIEGER, A. M., POLLAK, M. AND SAMUEL-CAHN, E. (2007b). Select sets: rank and file. *Ann. Appl. Prob.* **17**, 360–385.
- LEADBETTER, M. R., LINDGREN, G. AND ROOTZÉN, H. (1983). *Extremes and Related Sequences and Processes*. Springer, New York.
- PREATER, J. (2000). Sequential selection with a better than average rule. *Statist. Prob. Lett.* **50**, 187–191.