The Annals of Statistics
1975, Vol. 3, No. 1, 119-131

STATISTICAL INFERENCE USING EXTREME ORDER STATISTICS[1]

BY JAMES PICKANDS III

University of Pennsylvania

A method is presented for making statistical inferences about the upper tail of a distribution function. It is useful for estimating the probabilities of future extremely large observations. The method is applicable if the underlying distribution function satisfies a condition which holds for all common continuous distribution functions.

1. Introduction. Let $X_1, X_2, \cdots, X_n, \cdots$ be a sequence of mutually independent random variables with common continuous distribution function $F(x)$. In many applications we want to estimate the probability that a future observation will exceed a given high level during some specified epoch. A machine may break down if a certain noise level is reached for example or a glass store window might break under the impact of an unusually intense wind. In the book by Gumbel [5] many applications of this sort are discussed.

There are two general methods of treating problems of this kind. The first is parametric and it is given in Gumbel's book. We give an outline of it here.

In n random variables are mutually independent with common distribution function $F(x)$ and Y_n is the largest among these random variables, then the distribution function for Y_n is $F^n(x)$. Suppose there exists a pair of sequences a_n and b_n with $a_n > 0$ for all n and a distribution function $\Lambda(x)$ such that

$$\lim_{n\to\infty} P\{(Y_n - b_n)/a_n \leqq x\} = \lim_{n\to\infty} F^n(a_n x + b_n) = \Lambda(x) \tag{1.1}$$

for all x at which $\Lambda(x)$ is continuous. We say that $\Lambda(x)$ is an "extremal distribution function" and that $F(x)$ lies in its "domain of attraction." The distribution function $\Lambda(x)$ must belong to one of three parametric families. This is the extreme value trinity theorem. See Gnedenko [3], or Gumbel [5] Chapter 4. The domains of attraction have been characterized by a number of authors. The most recent and most extensive work is the book by de Haan [2]. Most "textbook" continuous families of distribution functions are discussed in [5]. They all lie in the domain of attraction of some extremal distribution function.

Suppose we have data for a period of, say, 50 years. The data might be the amount of water discharged through a river for example, or wind speeds measured every hour, or temperature readings or daily rainfall. Suppose the data are daily. Gumbel's method is to take the largest value for each year, so that we have a sample of annual maxima. The observations are assumed to be mutually

Received May 1972; revised February 1974.

[1] This research was supported by National Science Foundation grant GP 20702.

Key words and phrases. Extreme values, extreme order statistics, upper tail, generalized Pareto function.

independent with common distribution function $\Lambda(x)$ where $\Lambda(x)$ belongs to one of the three families of extreme value distribution functions. The choice is normally made intuitively. The parameters are then estimated to yield what is employed as the distribution of future maxima. These random variables are not actually mutually independent in most cases but they are nearly so and the method has been shown to be very robust against dependence.

The second method is the method of exceedances. See Gumbel and von Schelling [4] or [5] Section 2.2. This involves the probability distribution for the number of times the mth largest observation will be exceeded during a future epoch. The method is nonparametric and combinatoric. No assumption is made about the distribution function for the X_k except that it is continuous.

A third method is proposed here. Since the grouping of data into epochs is somewhat arbitrary we consider the m largest observations from the original data. For example, if we have daily data, some years may contain several among the m largest whereas others may contain none.

The proposed method is explained in Section 2. The theoretical details are contained in Sections 3 and 4, which can be omitted by the reader uninterested in the theory. A brief discussion follows in Section 5.

2. The method. In this section the proposed procedure is explained. The theory which sustains it is contained in the following two sections. Suppose that we have a sample of n mutually independent and identically distributed random variables with common but unknown continuous distribution function $F(x)$. It is assumed that for some c, $-\infty < c < \infty$,

$$\begin{aligned} \lim_{u\to x_\infty} \inf_{0<a<\infty} \sup_{0\leqq x<\infty} |([1 - F(u+x)]/[1 - F(u)]) \\ - \exp(-\textstyle\int_0^{x/a} [(1+ct)_+]^{-1}\,dt)| = 0\,, \end{aligned} \tag{2.1}$$

where $x_\infty = \text{g.l.b.}\,\{x\colon F(x) = 1\} = \text{l.u.b.}\,\{x\colon F(x) < 1\}$ and for any y, $y_+ \equiv \max(0, y)$. For any u, x, $[1 - F(u+x)]/[1 - F(u)]$ is the conditional probability that an observation is greater than $x + u$, given that it is greater than u. Our condition means that if u is large, the conditional distribution of X given that $X \geqq u$ is very nearly of the form

$$1 - G(x) = \exp - \textstyle\int_0^{x/a} [(1+ct)_+]^{-1}\,dt \tag{2.2}$$

for some a, c, $0 < a < \infty$, $-\infty < c < \infty$.

We distinguish three cases: $c > 0$, $c = 0$, and $c < 0$. If $c > 0$, then

$$1 - G(x) = (1 + cx/a)^{-1/c}$$

for all x, $0 < x < \infty$. This class of distribution functions is the well known Pareto family. If $c = 0$, then

$$1 - G(x) = \exp(-x/a)$$

for all x, $0 < x < \infty$. This is the exponential family of distributions. If $c < 0$, then

$$1 - G(x) = (1 - |c|x/a)^{1/|c|}$$

for all x, $0 < x \leqq a/|c|$ and

$$1 - G(x) = 0$$

for all x, $a/|c| \leqq x < \infty$.

If $F(x)$ is continuous, as we assume, our condition (2.1) is equivalent to the assumption that $F(x)$ lies in the domain of attraction of an extreme value distribution function. The equivalence is proved in Section 3. See Theorem 7. It is shown throughout [5] that most "textbook" continuous distribution functions lie in the domain of attraction of some extreme value distribution and so condition (2.1) is satisfied by them.

Let n be the sample size and let M be an integer much smaller than n. Intuitively the $4M$ largest observations contain information about the upper tail of the distribution function. This is useful for predictive purposes. We defer the question of choosing M.

Let $\{Z_m, m = 1, 2, \cdots, n\}$ be the descending order statistics. That is, for each m, Z_m is the mth largest observation in the sample. We treat the values $\{Z_m - Z_{4M}, m = 1, 2, \cdots, 4M - 1\}$ as though they were the descending order statistics from a sample of size $4M - 1$ from a population with a distribution function of the form (2.2) for some a, c, $0 < a < \infty$, $-\infty < c < \infty$. We estimate the parameters a and c by a simple percentile method as follows.

It is shown in Section 3 that for any y, $0 \leqq y \leqq 1$, $G^{-1}(1 - y) = a \int_0^{-\log y} e^{cu}\, du$. Therefore $G^{-1}(\frac{1}{2}) = a \int_0^{\log 2} e^{cu}\, du$ and $G^{-1}(\frac{3}{4}) = a \int_0^{\log 4} e^{cu}\, du$. Clearly, $[G^{-1}(\frac{3}{4}) - G^{-1}(\frac{1}{2})]/G^{-1}(\frac{1}{2}) = \int_{\log 2}^{\log 4} e^{cu}\, du / \int_0^{\log 2} e^{cu}\, du = 2^c$ and so $c = (\log 2)^{-1} \log ([(G^{-1}(\frac{3}{4}) - G^{-1}(\frac{1}{2})]/G^{-1}(\frac{1}{2}))$ and $a = G^{-1}(\frac{1}{2}) / \int_0^{\log 2} e^{cu}\, du$. For every pair of numbers $G^{-1}(\frac{1}{2})$, $G^{-1}(\frac{3}{4})$, $0 < G^{-1}(\frac{1}{2}) < G^{-1}(\frac{3}{4}) < \infty$, there corresponds one and only one pair of parameters a, c, $0 < a < \infty$, $-\infty < c < \infty$. To estimate c and a we replace the population quantiles $G^{-1}(\frac{1}{2})$ and $G^{-1}(\frac{3}{4})$ by the sample quantiles $\hat{G}^{-1}(\frac{1}{2}) = Z_{2M} - Z_{4M}$ and $\hat{G}^{-1}(\frac{3}{4}) = Z_M - Z_{4M}$. Then

$$\hat{c} = (\log 2)^{-1} \log ((Z_M - Z_{2M})/(Z_{2M} - Z_{4M})) \tag{2.3}$$

and

$$\hat{a} = (Z_{2M} - Z_{4M}) / \int_0^{\log 2} e^{\hat{c}u}\, du\,. \tag{2.4}$$

Now, we consider the problem of choosing M. It is intuitively clear that as n becomes large, so should M, or we fail to reap the benefit of an increasing sample size. That is we expect that

$$\lim_{n\to\infty} M = \infty$$

in probability. On the other hand, unless a portion of the upper tail is exactly of the form (2.2) we expect that

$$\lim_{n\to\infty} M/n = 0\,,$$

or the approximation of the population upper tail by means of a distribution of the form (2.2) will not improve indefinitely.

Specifically, we compute M in the following way. For each l, $l = 1, 2, \cdots, [n/4]$, let $d_l = \sup_{0 \leq x < \infty} |\hat{F}_l(x) - \hat{G}_l(x)|$ where $\hat{F}_l(x)$ is the "empirical upper tail" distribution function and $\hat{G}_l(x)$ is of the form (2.2) with c and a replaced by $\hat{c}$ and $\hat{a}$ given by (2.3) and (2.4) with $M = l$. The function $\hat{F}_l(x)$ is the usual empirical distribution function using the order statistics $\{Z_m - Z_{4l}, m = 1, 2, \cdots, 4l - 1\}$. That is, for each x, $4l(1 - \hat{F}_l(x))$ is the number of terms $Z_m - Z_{4l}$ which are greater than or equal to x. We choose M to be the smallest integer solution of

$$d_M = \min_{1 \leq l \leq [n/4]} d_l \,.$$

It is shown in Section 4 that the estimator $\hat{G}_M(x)$ is consistent in the sense that for all $\varepsilon > 0$,

$$\lim_{n \to \infty} P\{\sup_{0 \leq x < \infty} |([1 - F(Z_{4M} + x)]/[1 - F(Z_{4M})]) - [1 - \hat{G}_M(x)]| > \varepsilon\} = 0 \,.$$

3. The upper tail. The generalized Pareto upper tail is defined as an attribute of a univariate distribution function $F(x)$. It is shown that a continuous distribution function $F(x)$ has a generalized Pareto upper tail if and only if it lies in the domain of attraction of an extremal distribution function. Some definitions and properties are given.

DEFINITION 1. A function $P(x)$ is a "tail function" if $1 - P(x)$ is a distribution function with

$$P(0) = 1 \,.$$

DEFINITION 2. A tail function $Q(x)$ is a GPF (generalized Pareto function) if and only if

$$Q(x) = \exp - \int_0^{x/a} [(1 + ct)_+]^{-1} \, dt$$

for some a, c, $0 < a < \infty$, $-\infty < c < \infty$ and all x, $0 < x < \infty$.

The functions $Q(x)$ have well-defined inverse functions. For any y, $0 < y < 1$, let $Q^{-1}(y)$ be the value of x which is such that $Q(x) = y$. For all a, c, $0 < a < \infty$, $-\infty < c < \infty$.

$$Q^{-1}(y) = a \int_0^{-\log y} e^{cs} \, ds \,. \tag{3.1}$$

Let x_1 and x_2 be respectively the 50 and 75 percentiles.
We solve for a and c in terms of x_1 and x_2.
Evidently $(x_2 - x_1)/x_1 = 2^c$ and so

$$c = (\log 2)^{-1} \log ((x_2 - x_1)/x_1) \tag{3.2}$$

and

$$a = x_1 / \int_0^{\log 2} e^{cs} \, ds \,. \tag{3.3}$$

Let $Q(x, x_1, x_2) \equiv Q(x)$ where c and a are computed from x_1 and x_2 using the preceding equations. This is a reparametrization. There is one and only one function $Q(x)$ for each x_1, x_2, $0 < x_1 < x_2 < \infty$. Let $P_1(x)$ and $P_2(x)$ be any pair of tail functions. We define the metric

$$d(P_1, P_2) \equiv \sup_{0 \leqq x < \infty} |P_1(x) - P_2(x)| .$$

THEOREM 1. *Let $Q(x)$ and $Q^*(x)$ be two GPFs with parameters a and c and a^* and c^* respectively. We can write*

$$d(Q, Q^*) = \varphi(c, c^*, a/a^*) .$$

Furthermore for any c^*, $-\infty < c^* < \infty$.

$$\lim_{c \to c^*, \alpha \to 1} \varphi(c, c^*, \alpha) = 0 .$$

PROOF. Suppose that $Q(x)$ and $Q^*(x)$ are replaced by $Q(x/b)$ and $Q^*(x/b)$ respectively, where b is a real positive constant. Then $d(Q, Q^*)$ is unchanged. Let $b = a^*$. Then $d(Q, Q^*)$ has the same value if c and c^* are unchanged but a and a^* are replaced by a/a^* and 1 respectively. So $d(Q, Q^*)$ is a function of c, c^* and a/a^*.

Now the second part of the theorem is proved. Let x be an arbitrary real positive number. If $c \geqq 0$, it is clear that $Q(x)$ is a continuous function of a and c. If $c < 0$, we can write $Q(x) = \max(0, 1 - |c|x/a)^{1/|c|}$. When a and c are such that $1 - |c|x/a = 0$, $Q(x)$ is not differentiable but it is still continuous. It is clearly continuous for all other values of a and c. Since x was arbitrarily chosen, continuity with respect to the metric d follows by the usual compactness argument. □

THEOREM 2. *Let x_1 and x_2 and x_1^* and x_2^* be the 50 and 75 percentiles for $Q(x)$ and $Q^*(x)$ respectively. We can write*

$$d(Q, Q^*) = \psi(x_1/x_1^*, x_2/x_1^*, x_2^*/x_1^*) .$$

Furthermore, for any α, β, γ, $0 < \alpha, \beta, \gamma < \infty$,

$$\lim_{\alpha \to 1, \beta \to \gamma} \psi(\alpha, \beta, \gamma) = 0 .$$

PROOF. This is a consequence of Theorem 1 and the fact that x_1 and x_2 are continuous functions of a and c. □

THEOREM 3. *Let $P(x)$ be a tail function. If for some x_1, x_2, δ, $0 < x_1 < x_2 < \infty$, $0 < \delta \leqq \frac{1}{4}$,*

$$d(P, Q^*) \leqq \delta ,$$

where

$$Q^*(x) = Q(x, x_1^*, x_2^*) ,$$

and x_1 and x_2 are such that

$$P(x_i) = 1/2^i , \qquad i = 1, 2 ,$$

then

$$|x_i - x_i^*| \leqq g_i(x_1^*, x_2^*, \delta) , \qquad i = 1, 2 ,$$

and the functions g_1 and g_2 are such that for all x_1^*, x_2^*, $0 < x_1^* < x_2^* < \infty$,

$$\lim_{\delta \to 0} g_i(x_1^*, x_2^*, \delta) = 0 , \qquad i = 1, 2 .$$

PROOF. By assumption

$$\frac{1}{2^i} - \delta \leqq Q^*(x_i) \leqq \frac{1}{2^i} + \delta\,, \qquad i = 1, 2\,.$$

Recall the inverse function (3.1) for $Q^*(x)$. The theorem is true by continuity. □

DEFINITION 3. A GPF $Q(x)$ is said to "be associated with" a tail function $P(x)$ if for some x_1, x_2, $0 < x_1 < x_2 < \infty$,

$$Q(x) = Q(x, x_1, x_2)\,,$$

and

$$P(x_i) = 1/2^i\,, \qquad i = 1, 2\,.$$

THEOREM 4. *Let $Q(x)$ and $Q^*(x)$ be* GPF*s and let $P(x)$ be an arbitrary tail function. Suppose that $Q(x)$ is associated with $P(x)$ in the sense of Definition* 3. *Assume further that for some δ, $0 < \delta < \frac{1}{4}$,*

$$d(P, Q^*) \leqq \delta\,.$$

Then

$$d(Q, Q^*) \leqq \psi(x_1/x_1^*, x_2/x_1^*, x_2^*/x_1^*)\,,$$

where

$$|x_i - x_i^*| \leqq g_i(x_1^*, x_2^*, \delta)\,, \qquad i = 1, 2\,.$$

It follows that

$$\lim_{\delta \to 0} d(Q, Q^*) = 0\,.$$

PROOF. Let $Q(x) \equiv Q(x, x_1, x_2)$, and $Q^*(x) = Q(x, x_1^*, x_2^*)$. By Definition 3, $P(x_i) = 1/2^i$, $i = 1, 2$. By Theorem 2, $d(Q, Q^*) = \psi(x_1/x_1^*, x_2/x_1^*, x_2^*/x_1^*)$. But by Theorem 3, $|x_i - x_i^*| \leqq g_i(x_1^*, x_2^*, \delta)$, $i = 1, 2$. The limit result is a consequence of the conclusions of Theorems 2 and 3. □

THEOREM 5. *Under the conditions of Theorem* 4,

$$d(P, Q) \leqq \delta + \psi(x_1/x_1^*, x_2/x_1^*, x_2^*/x_1^*)$$

and so

$$d(P, Q) \to 0$$

as $\delta \to 0$.

PROOF. By the triangle inequality $d(P, Q) \leqq d(P, Q^*) + d(Q^*, Q) \leqq \delta + d(Q^*, Q)$. The result follows by Theorem 4. □

Let u be any real number such that $F(u) < 1$, or equivalently such that $-\infty < u < x_\infty$, and let

$$P_u(x) = P\{X > u + x\}/P\{X > u\} = [1 - F(u + x)]/[1 - F(u)] \tag{3.4}$$

and let $Q_u(x)$ be a GPF associated with $P_u(x)$ in the sense of Definition 3. In general $Q_u(x)$ may not exist for all u. A sufficient condition for it to exist is that $F(x)$ be continuous, because then the equations $P_u(x_i) = 1/2^i$, $i = 1, 2$, necessarily have a solution, although it need not be unique. Let

$$D(\lambda) \equiv \inf_{u:1-F(u)=\lambda} d(P_u, Q_u)\,. \tag{3.5}$$

For each λ, $0 < \lambda < 1$, let C_λ be the set of all possible values of c for $Q_u(x)$ where u is such that $1 - F(u) = \lambda$. Let

$$c^-(\lambda) \equiv \text{g.l.b.}\,\{c : c \in C_\lambda\} \tag{3.6}$$

and

$$c^+(\lambda) \equiv \text{l.u.b.}\,\{c : c \in C_\lambda\}\,. \tag{3.7}$$

If the function $D(\lambda)$ is well defined, then for each λ, $0 < \lambda \leqq 1$, the set C_λ is nonempty and so the functions $c^-(\lambda)$ and $c^+(\lambda)$ are well defined. The function $D(\lambda)$ is necessarily well defined if $F(x)$ is continuous.

THEOREM 6. *If $D(\lambda) = 0$ for some λ, $0 < \lambda \leqq 1$, then $D(\alpha) = 0$ for all α, $0 < \alpha \leqq \lambda$. Furthermore,*

$$c^-(\lambda) = c^+(\lambda) = c^-(\alpha) = c^+(\alpha)\,.$$

PROOF. There exists a value of u which is such that $1 - F(u) = \lambda$ and $d(P_u, Q_u) = 0$. So we can take $P_u(x)$ to be equal to $Q_u(x)$ and we can write $P_u(x) = \exp - \int_0^{x/a} dt/(1 + ct)$ for some a, c, $0 < a < \infty$, $-\infty < c < \infty$, for all x at which $P_u(x) > 0$. Let v be chosen so that $0 < 1 - F(u + v) < \lambda$. Then

$$\begin{aligned} P_{u+v}(x) &= [1 - F(u + v + x)]/[1 - F(u + v)] \\ &= P_u(v + x)/P_u(v) = \exp - \int_{v/a}^{(v+x)/a} dt/(1 + ct)\,. \end{aligned}$$

But

$$\begin{aligned} &\int_{v/a}^{(v+x)/a} dt/(1 + ct) \\ &\quad = \int_0^{x/a} dt\Big/\left[1 + c\left(t + \frac{v}{a}\right)\right] = \int_0^{x/a} dt\Big/\left[1 + c\frac{v}{a} + ct\right] \\ &\quad = \int_0^{x/a} dt\Big/\left(1 + c\frac{v}{a}\right)\left[1 + ct\left(1 + c\frac{v}{a}\right)^{-1}\right] = \int_0^{x/(a+cv)} dt/(1 + ct)\,. \end{aligned}$$

So $P_{u+v}(x)$ is also a GPF. The value of c is unchanged but the value of a is not. □

DEFINITION 4. A distribution function $F(x)$ is said to have a generalized Pareto upper tail if and only if $\lim_{\lambda\to 0} D(\lambda) = 0$, and there exists a constant c, $-\infty < c < \infty$ which is such that

$$\lim_{\lambda\to 0} c^-(\lambda) = \lim_{\lambda\to 0} c^+(\lambda) = c\,.$$

In [7], von Mises showed that the extremal distribution functions are of the form $\Lambda(x) \equiv \exp - \int_0^y ((1 + ct)_+)^{-1}\,dt]$ where $y = (x - b)/a$ and a, b and c are respectively the scale, location and shape parameter, with $0 < a < \infty$, and $-\infty < b, c < \infty$.

THEOREM 7. *A continuous distribution function $F(x)$ has a generalized Pareto upper tail in the sense of Definition 4 if and only if it lies in the domain of attraction of some extremal distribution function.*

REMARK. For a thorough study of the upper tail of a distribution and its relationship to the limiting extremal distribution see Resnick [6].

PROOF. By (1.1) a distribution function $F(x)$ lies in the domain of attraction of $\Lambda(x)$ if and only if for each x,

$$\lim_{n\to\infty} n \log F(a_n x + b_n) = \log \Lambda(x) . \tag{3.8}$$

But (1.1) cannot hold unless $\lim_{n\to\infty} F(a_n x + b_n) = 1$ for each x which is such that $\Lambda(x) > 0$. For any distribution function $F(x)$, $-\log F(x) \sim (1 - F(x))$ as $F(x) \to 1$. Therefore, recalling (3.8), a distribution function $F(x)$ lies in the domain of attraction of $\Lambda(x)$ if and only if

$$\lim_{n\to\infty} n(1 - F(a_n x + b_n)) = -\log \Lambda(x) \tag{3.9}$$

for all x which are such that $\Lambda(x) > 0$. It follows from (3.9) that

$$\begin{aligned} \lim_{n\to\infty} (n - 1)(1 - F(a_n x + b_n)) &= \lim_{n\to\infty} n(1 - F(a_{n+1} x + b_{n+1})) \\ &= -\log \Lambda(x) . \end{aligned}$$

The converse is also true. Let u be any real number, $u < x_\infty$, and let n be so chosen that $b_n \leqq u \leqq b_{n+1}$. This is possible since the sequence b_n can be chosen to be non-decreasing. Let $g(u) \equiv a_n$. Then, when, and only when $\lim_{n\to\infty} [1 - F(a_n x + b_n)]/[1 - F(b_n)] = -\log \Lambda(x)/[-\log \Lambda(0)]$, it follows that $\lim_{u\to x_\infty} P_u \times (xg(u)) = \lim_{u\to x_\infty} [1 - F(u + xg(u))]/[1 - F(u)] = -\log \Lambda(x)/(-\log \Lambda(0))$. The last term is a GPF in the sense of Definition 2. In other words, for each x, $0 \leqq x < \infty$,

$$\lim_{u\to x_\infty} |([1 - F(u + xg(u))]/[1 - F(u)]) - \exp - \textstyle\int_0^{x/a} [(1 + ct)_+]^{-1} dt| = 0 .$$

By the usual compactness argument, this is true if and only if

$$\begin{aligned} &\lim_{u\to x_\infty} \sup_{0\leqq x<\infty} |([1 - F(u + xg(u)]/[1 - F(u)]) - \exp - \textstyle\int_0^{x/a} [(1 + ct)_+]^{-1} dt| \\ &\quad = \lim_{u\to x_\infty} \sup_{0\leqq x<\infty} |([1 - F(u + x)]/[1 - F(u)]) \\ &\quad\quad - \exp - \textstyle\int_0^{x/ag(u)} [(1 + ct)_+]^{-1} dt| = 0 , \end{aligned}$$

which holds when and only when (2.1) does. □

4. Estimation. We consider the problem of estimating the parameter c and empirically fitting the upper tail of a distribution function. Let $X_1, X_2, \cdots, X_n, \cdots$ be mutually independent random variables with common continuous distribution function $F(x)$. Let $Z_1, Z_2, \cdots, Z_n$ be the descending order statistics. That is, for each integer m, $m = 1, 2, \cdots, n$, Z_m is the mth largest among $X_1, X_2, \cdots X_n$. The subscript n denoting the sample size is suppressed.

Let M be a positive integer-valued random variable with $1 \leqq 4M \leqq n$. Let

$$P_M(x) \equiv P_u(x) \tag{4.1}$$

where $P_u(x)$ is given by (3.4) and $u = Z_{4M}$ and let $Q_M(x)$ be a GPF associated with it in the sense of Definition 3.

DEFINITION 5. We call $\hat{P}_M(x)$ the "empirical tail function" if for each x, $0 < x < \infty$, $4M\hat{P}_M(x)$ is the number among the random variables $Z_m - Z_{4M}$, $m = 1, 2, \cdots, 4M$ which are greater than or equal to x. That is, $\hat{P}_M(x) \equiv (4M)^{-1} \sum_{m=1}^{4M} \chi_n(x)$ where $\chi_n(x) = 1$, if $Z_m - Z_{4M} \geqq x$, $= 0$, if $Z_m - Z_{4M} < x$.

DEFINITION 6. We call $\hat{Q}_M(x)$ the "empirical GPF" if $\hat{Q}_M(x)$ is associated with $\hat{P}_M(x)$ in the sense of Definition 3. In particular, we let the 50 and 75 percentiles be given by $x_1 \equiv Z_{2M} - Z_{4M}$ and $x_2 \equiv Z_M - Z_{4M}$.

It follows by definition that $\hat{P}_M(x_i) = 1/2^i$, $i = 1, 2$. Therefore $\hat{Q}_M(x) = \exp - \int_0^{x/\hat{a}} [(1 + \hat{c}t)_+]^{-1} \, dt$ where

$$\hat{c} \equiv (\log 2)^{-1} \log ([Z_M - Z_{2M}]/[Z_{2M} - Z_{4M}]) \tag{4.2}$$

and

$$\hat{a} \equiv (Z_{2M} - Z_{4M}) / \int_0^{\log 2} e^{-\hat{c}s} \, ds \, . \tag{4.3}$$

Now we introduce a Poisson process to support the proofs of some of the lemmas and theorems which follow. Let $T_1, T_2, \cdots, T_n, \cdots$ be the successive points of increase of a homogeneous one-dimensional process with unit intensity. Equivalently, the random variables $T_1, T_2 - T_1, \cdots, T_n - T_{n-1}, \cdots$ are mutually independent and have the standard exponential distribution.

LEMMA 8. *If $F(x)$ is continuous, the joint distribution of the random variables $(1 - F(Z_1)), (1 - F(Z_2)), \cdots, (1 - F(Z_n))$ is the same as the joint distribution of the random variables $T_1/T_{n+1}, T_2/T_{n+1}, \cdots, T_n/T_{n+1}$.*

PROOF. The joint distribution of the random variables $(1 - F(Z_1)), (1 - F(Z_2)), \cdots, (1 - F(Z_n))$ is the same as the joint distribution of the ascending order statistics from a population uniformly distributed on the unit interval. This is so since $F(x)$ is continuous. The joint density of $T_1/T_{n+1}, T_2/T_{n+1}, \cdots, T_n/T_{n+1}$ and T_{n+1} is $g(u_1, u_2, \cdots, u_n, u_{n+1}) = u_{n+1}^n e^{-u_{n+1}}$. But the random variable T_{n+1} has the Gamma distribution and its density

$$g(u_{n+1}) = \frac{1}{n!} u_{n+1}^n e^{-u_{n+1}} \, , \qquad 0 < u_{n+1} < \infty \, .$$

Therefore the conditional density of $T_1/T_{n+1}, T_2/T_{n+1}, \cdots, T_n/T_{n+1}$ given T_{n+1} is $g(u_1 u_2 \cdots u_n : T_{n+1}) = n!$, $0 < u_1, < u_2 < \cdots < u_n < 1$, $= 0$ otherwise. Since this does not depend upon T_{n+1} it is also the unconditional density function. But this is just the joint density of the uniform order statistics. □

LEMMA 9. *If $F(x)$ is continuous, and m is any integer $m = 1, 2, \cdots, n$, the joint distribution of the random variables $[1 - F(Z_1)]/[1 - F(Z_m)], [1 - F(Z_2)]/[1 - F(Z_m)], \cdots, [1 - F(Z_{m-1})]/[1 - F(Z_m)]$ is the same as that of the random variables $T_1/T_m, T_2/T_m, \cdots, T_{m-1}/T_m$.*

PROOF. This is an immediate consequence of Lemma 8. □

For any M, $1 \leqq 4M \leqq n$,

$$d(P_M, \hat{P}_M) = \max\Big(P_M(Z_1 - Z_{4M}), \Big|P_M(Z_1 - Z_{4M}) - \frac{1}{4M}\Big|,$$
$$\Big|P_M(Z_2 - Z_{4M}) - \frac{1}{4M}\Big|, \Big|P_M(Z_2 - Z_{4M}) - \frac{2}{4M}\Big|, \cdots,$$
$$\Big|P_M(Z_{4M-1} - Z_{4M}) - \frac{4M-2}{4M}\Big|, \Big|P_M(Z_{4M-1} - Z_{4M}) - \frac{4M-1}{4M}\Big|\Big)$$
$$= \max\nolimits_{1\leqq m\leqq 4M-1}\Big(\Big|P(Z_m - Z_{4M}) - \frac{m-1}{4M}\Big|, \Big|P(Z_m - Z_{4M}) - \frac{m}{4M}\Big|\Big),$$

since $F(x)$ is assumed to be continuous. But for each m, $1 \leqq m \leqq 4M - 1$, by Definition (3.4), $P(Z_m - Z_{4M}) = (1 - F(Z_m))/(1 - F(Z_{4M}))$ and so

$$d(P_M, \hat{P}_M) = \max\nolimits_{1\leqq m\leqq 4M-1}\Big(\Big|\frac{1 - F(Z_m)}{1 - F(Z_{4M})} - \frac{m-1}{4M}\Big|, \Big|\frac{1 - F(Z_m)}{1 - F(Z_{4M})} - \frac{m}{4M}\Big|\Big).$$

LEMMA 10. *If $F(x)$ is continuous, then for all $\varepsilon > 0$,*

$$\lim\nolimits_{l\to\infty} \sup\nolimits_{n\geqq 4l} P\{\max\nolimits_{4l\leqq 4m\leqq n} d(\hat{P}_m, P_m)(2m/\log\log 4m)^{\frac{1}{2}} > 2 + \varepsilon\} = 0\,.$$

PROOF. Rewriting the conclusion it is sufficient to prove that for all $\varepsilon > 0$,

$$\lim\nolimits_{l\to\infty} \sup\nolimits_{n\geqq 4l} P\Big\{\max\nolimits_{4l\leqq 4m\leqq n} \max\nolimits_{1\leqq k\leqq 4m-1}\Big(\Big|\frac{1 - F(Z_k)}{1 - F(Z_{4m})} - \frac{k-1}{4m}\Big|,$$
$$\Big|\frac{1 - F(Z_k)}{1 - F(Z_{4m})} - \frac{k}{4m}\Big|\Big)(2m/\log\log 4m)^{\frac{1}{2}} > 2 + \varepsilon\Big\} = 0\,.$$

By Lemma 9, it is sufficient to prove that for all $\varepsilon > 0$,

$$\lim\nolimits_{l\to\infty} \sup\nolimits_{n\geqq 4l} P\Big\{\max\nolimits_{4l\leqq 4m\leqq n} \max\nolimits_{1\leqq k\leqq 4m-1}\Big(\Big|\frac{T_k}{T_{4m}} - \frac{k-1}{4m}\Big|,$$
$$\Big|\frac{T_k}{T_{4m}} - \frac{k}{4m}\Big|\Big)(2m/\log\log 4m)^{\frac{1}{2}} > 2 + \varepsilon\Big\} = 0\,.$$

But this is true if for all $\varepsilon > 0$

$$\lim\nolimits_{l\to\infty} P\Big\{\sup\nolimits_{4m\geqq 4l} \max\nolimits_{1\leqq k\leqq 4m-1}\Big(\Big|\frac{T_k}{T_{4m}} - \frac{k-1}{4m}\Big|,$$
$$\Big|\frac{T_k}{T_{4m}} - \frac{k}{4m}\Big|\Big)(2m/\log\log 4m)^{\frac{1}{2}} > 2 + \varepsilon\Big\} = 0\,.$$

Replacing $4m$ by m,

$$\text{(4.4)} \qquad \lim\nolimits_{l\to\infty} P\Big\{\sup\nolimits_{m\geqq l} \max\nolimits_{1\leqq k\leqq m-1}\Big(\Big|\frac{T_k}{T_m} - \frac{k-1}{m}\Big|,$$
$$\Big|\frac{T_k}{T_m} - \frac{k}{m}\Big|\Big)(m/2\log\log m)^{\frac{1}{2}} > 2 + \varepsilon\Big\} = 0\,.$$

By convergence with probability 1 implies convergence in probability. So (4.4) above is true if

$$P\left\{\limsup_{m\to\infty} \max_{1\leq k\leq m-1}\left(\left|\frac{T_k}{T_m}-\frac{k-1}{m}\right|,\right.\right.$$
$$\left.\left.\left|\frac{T_k}{T_m}-\frac{k}{m}\right|\right)(m/2\log\log m)^{\frac{1}{2}}\leqq 2\right\}=1\,.$$

It is sufficient to prove that

$$(4.5)\qquad P\left\{\limsup_{m\to\infty} \max_{1\leq k\leq m-1}\left|\frac{T_k}{T_m}-\frac{(k-1)}{m}\right|(m/2\log\log m)^{\frac{1}{2}}\leqq 2\right\}=1$$

and

$$(4.6)\qquad P\left\{\limsup_{m\to\infty} \max_{1\leq k\leq m-1}\left|\frac{T_k}{T_m}-\frac{k}{m}\right|(m/2\log\log m)^{\frac{1}{2}}\leqq 2\right\}=1\,.$$

First we prove (4.6). Clearly

$$\left|\frac{T_k}{T_m}-\frac{k}{m}\right|\leqq T_k\left|\frac{1}{T_m}-\frac{1}{m}\right|+\frac{|T_k-k|}{m}$$
$$=\frac{T_k}{T_m}\frac{|T_m-m|}{m}+\frac{|T_k-k|}{m}\leqq\frac{|T_m-m|}{m}+\frac{|T_k-k|}{m}\,.$$

So

$$\max_{1\leq k\leq m-1}\left|\frac{T_k}{T_m}-\frac{k}{m}\right|(m/2\log\log m)^{\frac{1}{2}}$$
$$\leqq (|T_m-m|/(2m\log\log m)^{\frac{1}{2}})+\max_{1\leq k\leq m-1}|T_k-k|/(2m\log\log m)^{\frac{1}{2}}\,.$$

But $P\{\limsup_{m\to\infty}|T_m-m|(2m\log\log m)^{\frac{1}{2}}=1\}=1$ by the Law of the Iterated Logarithm. See Breiman [1], page 64. To prove (4.6) then it remains to prove that

$$(4.7)\qquad P\{\limsup_{m\to\infty}\max_{1\leq k\leq m-1}|T_k-k|/(2m\log\log m)^{\frac{1}{2}}\leqq 1\}=1\,.$$

But

$$\max_{1\leq k\leq m-1}|T_k-k|/(2m\log\log m)^{\frac{1}{2}}$$
$$\leqq \max_{1\leq k\leq m_0-1}|T_k-k|/(2m\log\log m)^{\frac{1}{2}}$$
$$+\max_{m_0\leq k\leq m-1}|T_k-k|/(2k\log\log k)^{\frac{1}{2}}$$

where m_0 is an arbitrary positive integer. So

$$P\{\limsup_{m\to\infty}\max_{1\leq k\leq m-1}|T_k-k|/(2m\log\log m)^{\frac{1}{2}}$$
$$\leqq \sup_{k\geq m_0}|T_k-k|/(2k\log\log k)^{\frac{1}{2}}\}=1\,.$$

But m_0 was arbitrarily chosen and

$$P\{\lim_{m_0\to\infty}\sup_{k>m_0}|T_k-k|/(2k\log\log k)^{\frac{1}{2}}=1\}=1\,,$$

again by the Law of the Iterated Logarithm. So (4.7) is true and therefore (4.6) is true. All that remains is to prove (4.5). But

$$\left|\frac{T_k}{T_m}-\frac{(k-1)}{m}\right|\leqq\left|\frac{T_k}{T_m}-\frac{k}{m}\right|+\frac{1}{m}\,.$$

So (4.5) follows from (4.6) and the fact that $\limsup_{m\to\infty}(1/2m\log\log m)^{\frac{1}{2}}=0$. □

THEOREM 11. *If $F(x)$ is continuous and $M \to \infty$ in probability as $n \to \infty$, then* $(d(P_M, \hat{P}_M)(2M/\log\log 4M)^{\frac{1}{2}} - 2)_+ \to 0$ *in probability.*

PROOF. Let ε_1 and ε_2 be arbitrarily chosen. By Lemma 10, we can chosen m_0 so that for all $n \geqq 4m_0$, $P\{\max_{4m_0 \leqq 4m \leqq n} d(P_m, \hat{P}_m)(2m/\log\log 4m)^{\frac{1}{2}} > 2 + \varepsilon_1\} \leqq \varepsilon_2$. But

$$\begin{aligned}
&P\{d(P_M, \hat{P}_M)(2M/\log\log 4M)^{\frac{1}{2}} > 2 + \varepsilon_1\} \\
&\quad \leqq P\{M \leqq m_0\} + P\{d(P_M, \hat{P}_M)(2M/\log\log 4M)^{\frac{1}{2}} > 2 + \varepsilon_1, M > m_0\} \\
&\quad \leqq P\{M \leqq m_0\} + P\{\max_{4m_0 \leqq 4m \leqq n} d(P_m, \hat{P}_m)(2m/\log\log 4m)^{\frac{1}{2}} > 2 + \varepsilon_1\} \\
&\quad \leqq P\{M \leqq m_0\} + \varepsilon_2 .
\end{aligned}$$

But by assumption $\lim_{n\to\infty} P\{M \leqq m_0\} = 0$ and ε_1 and ε_2 were arbitrarily chosen. □

THEOREM 12. *For each n, let M be a solution of*

$$d(\hat{P}_M, \hat{Q}_M) = \min_{1 \leqq m \leqq [n/4]} d(\hat{P}_m, \hat{Q}_m) . \tag{4.8}$$

If $F(x)$ is continuous and has a generalized Pareto upper tail in the sense of Definition 4, then

$$P\{\lim_{n\to\infty} M = \infty, \lim_{n\to\infty} d(\hat{P}_M, \hat{Q}_M) = 0\} = 1 .$$

REMARK. The solution of (4.8) need not be unique. A good choice for M would be the largest one.

PROOF. Let m be an integer-valued function which is such that $\lim_{n\to\infty} m/n = \lambda$ where $0 < \lambda \leqq 1$. Then $P\{\lim_{n\to\infty}(1 - F(Z_m)) = \lambda\} = 1$, and by definition $P\{\limsup_{n\to\infty} d(P_m, Q_m) \leqq D(\lambda+)\} = 1$ where $D(\lambda)$ is given by (3.5). But $d(\hat{P}_M, \hat{Q}_M) \leqq d(\hat{P}_m, \hat{Q}_m) \leqq d(\hat{P}_m, P_m) + d(P_m, Q_m) + d(Q_m, \hat{Q}_m)$. Furthermore, $P\{\lim_{n\to\infty} d(\hat{P}_m, P_m) = 0\} = 1$ by the Glivenko–Cantelli Theorem and $P\{\lim_{n\to\infty} d(\hat{Q}_m, Q_m) = 0\} = 1$, since c and a are continuous functions of the quantiles. Recall (3.2) and (3.3). Therefore $P\{\limsup_{n\to\infty} d(\hat{P}_M, \hat{Q}_M) \leqq D(\lambda)\} = 1$. By λ was arbitrarily chosen and by assumption

$$\lim_{\lambda\to 0} D(\lambda) = 0 ,$$

and so

$$P\{\lim_{n\to\infty} d(\hat{P}_M, \hat{Q}_M) = 0\} = 1 . \tag{4.9}$$

Let m_0 be any positive integer. Suppose that the probability is $q > 0$ that M is no larger than m_0 for infinitely many n. In other words, suppose that $P\{\liminf_{n\to\infty} M \leqq m_0\} = q > 0$. It follows that $P\{\limsup_{n\to\infty} d(\hat{P}_M, \hat{Q}_M) \geqq 1/8m_0\} \geqq q > 0$, since by definition $\hat{P}_M(x)$ is necessarily discrete with jumps of magnitude least $1/4m_0$ and $\hat{Q}_M(x)$ is continuous. This contradicts (4.9) and so

$$P\{\lim_{n\to\infty} M = \infty\} = 1 . \qquad \square$$

In conclusion, under the conditions of Theorems 11 and 12,

$$d(P_M, \hat{Q}_M) \leqq d(P_M, \hat{P}_M) + d(\hat{P}_M, \hat{Q}_M) \to 0$$

in probability as $n \to \infty$ where

$$d(P_M, \hat{Q}_M) \equiv \sup_{0 \leq x < \infty} |([1 - F(Z_{4M} + x)]/[1 - F(Z_{4M})]) - \exp - \int_0^{x/\hat{a}} [(1 + \hat{c}t)_+]^{-1} dt| ,$$

where

$$\hat{c} \equiv (\log 2)^{-1} \log ((Z_M - Z_{2M})/(Z_{2M} - Z_{4M}))$$

and

$$\hat{a} = (Z_{2M} - Z_{4M})/\int_0^{\log 2} e^{\hat{c}s} ds .$$

It is conjectured that under fairly general conditions on $F(x)$ for any real t, $0 < t < \infty$,

$$\limsup_{n\to\infty} n(1 - F(U_n(t)) \leqq t$$

where

$$U_n(t) \equiv Z_{4M} + \hat{a} \int_0^{\log(4M/t)} e^{\hat{c}s} ds$$

where $\hat{c}$ and $\hat{a}$ are given by (4.2) and (4.3). The function $U_n(t) - Z_{4M}$ is the $t/4M$ percentile for the random tail function $\hat{Q}_M(x)$.

5. Conclusions. In Gumbel's Classical Method, the assumption is made that the distribution function for the annual maximum is the extreme value distribution. This assumption is not precisely accurate although experience has shown that it works very well. The error, however slight, does not disappear as the sample size increases. On the other hand, the proposed method is consistent as the sample size approaches ∞. In the classical method an intuitive decision is made as to which of the three families of extreme value distribution is assumed to be the correct one. In the proposed method no such subjective decision is necessary.

The author thanks the referee for his very careful review, which begot an improvement of the exposition.

REFERENCES

[1] BREIMAN, L. (1968). *Probability*. Addison-Wesley. Reading Mass.

[2] DE HAAN, L. (1971). *On Regular Variation and its Application to the Weak Convergence of Sample Extremes*. Mathematical Center Tracts, Amsterdam.

[3] GNEDENKO, B. V. (1973). Sur la distribution limité du terme maximum d'une sèrie aleatoire. *Ann. of Math.* **44** 423–453.

[4] GUMBEL, E. J. and VON SCHELLING, H. (1950). The distribution of the number of exceedances. *Ann. Math. Statist.* **21** 247.

[5] GUMBEL, E. J. (1960). *Statistics of Extremes*. Columbia Univ. Press.

[6] RESNICK, S. I. (1971). Tail equivalence and its applications. *J. Appl. Probability* **8** 135–156.

[7] VON MISES, R. (1936). La distribution de la plus grande de n valeurs. *Selected Papers of Richard von Mises*, **2** 271–294. American Mathematical Society, Providence.

DEPARTMENT OF STATISTICS
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19174