Statistical Theory of Extremes
Homepage: http://www.gathacognition.com/book/gcb14
http://dx.doi.org/10.21523/gcb1

Part 2
Statistics for Univariate Extremes

Chapter 9

Statistical Analysis With Partial Information

J. Tiago de Oliveira
Academia das Ciências de Lisboa (Lisbon Academy of Sciences), Lisbon, Portugal.

Abstract

The chapter deals with some sparse results in which the samples are not complete. The analysis concerned with drawing inferences about quantities associated with right tail of the real $F(.)$ when n is large, using asymptotic extreme value theory. The corresponding quantiles for the lower tail can be dealt with by the usual transformation from maxima to minima. Largest observations and subsamples maxima are compared. Subsample maxima method shows definite superiority. Estimation using two or three quantiles, estimations of the parameters using block partitions of the samples and estimations using exceedances over thresholds are also discussed.

Published Online

23 June 2017

Keywords

Extreme observations;
Lower tail;
Subsample method;
Largest observations;
Subsample maxima;
Bock partitions.

Editor(s)

J.C. Tiago de Oliveira

9.1 Introduction

So far we have dealt with complete samples which were supposed to have Weibull, Gumbel or Fréchet distributions, or, otherwise, that one of them could be accepted as a good approximation to the distribution of the underlying population.

Now we will deal with some sparse results in which the samples are not complete, such as following:

1. We know only the upper (lower) part of a sample, i.e., the upper (lower) m observations, and we want to estimate the upper (lower) tail of the distribution; see Weismann(1978) and (1984) and references therein;

Originally published in 'Statistical Analysis of Extremes', 1997, 2016
http://dx.doi.org/10.21523/gcb1.1714

2. We know that the sample of maxima can be split into m subsamples and we use only the k maxima of each subsample; see Hüsler and Tiago de Oliveira (1988) for $k = 1$;
3. We use two or three quantiles to estimate the parameters; see Tiago de Oliveira (1972) and Kubat and Epstein (1980);
4. We split the sample into disjoint parts, compute averages, and estimate the parameters; see Kubat (1982), Kubat and Epstein (1980), Hüsler and Schupbach (1986), and Neves (1988);
5. We use excesses over threshold in the sequence of observations; see Smith (1984), Rösbjerg and Knudsen (1984).

9.2 The use of the extreme m observations

Let $x_1'' \geq x_2'' \geq \cdots \geq x_n''$ be the decreasing order statistics in a sample of size n from a distribution function F and suppose that $F \in \mathcal{D}(\tilde{L})$, i.e., that

$$F^n(\lambda_n + \delta_n x) \to \tilde{L}(x) \ (n \to \infty)$$

for some constants λ_n and $\delta_n > 0$. We are concerned here with drawing inferences about quantities associated with the right tail of the real $F(.)$ when n is large, using asymptotic extreme value theory.

We are thus interested in

$\overline{w} = \sup\{x|F(x) < 1\}$, the right-end point;

$F(x)$ for large $x(< \overline{w})$;

$\chi_p = Q(p) = F^{-1}(p)$ with $p = 1 - c/n$ (c constant)

and, hence, in λ_n and $\delta_n > 0$.

Similarly we can be interested in the corresponding quantiles for the lower tail which can be dealt with by the usual transformation from maxima to minima.

The estimation method used up to now, and developed at length in Gumbel (1958), is to divide the original (unordered) data into subsamples. The maxima (or minima) of the subsamples are then used to estimate the parameters of $\tilde{L}$, one of the three extreme value limit distributions. This method is natural when some period exists in the data, as in environmental series, and the natural subsample is the data period. This method is sometimes called the *yearly data* or *(natural) subsamples method*.

But another approach is possible. Instead of splitting all observations into $n|m$ samples of m observation each and taking the maximum of each sample, we can take the m largest observations of the sample of n and decide with them. Note that the underlying hypotheses are different: in the (natural) subsamples method we *only*

suppose that maxima have the same extreme value distribution, while in the largest maxima method we suppose that *all* observations are i.i.d. with the distribution $F(.)$, attracted by one of the extreme value distributions; this will be discussed in the Annex 4.

It will be seen in the next section that if the sample is i.i.d. with a Gumbel distribution then subsample method is, in general, more efficient than the method of the largest m maxima using the same number of observations (m maxima of n/m subsamples or m largest maxima, both of a sample of size n).

Let us then describe the Weissman method.

Define $x''_{ni} = (x''_i - \lambda_n)/\delta_n$. Then, as $F^n(\lambda_n + \delta_n x) \to \tilde{L}(x)$ as $n \to \infty$, the vector $(x''_{ni} : i = 1, \dots, n)^T$ converges in distribution to $(M_i = \tilde{\lambda}^{-1}(Z_1 + \cdots + Z_i) : i = 1, \dots, n)^T$, where $\tilde{\lambda}(x) = -\log \tilde{L}(x)$ and the $\{Z_i\}$ is a sequence of i.i.d. exponential random variables all with mean value 1 (David, 1981). Hence for large n, the m largest order statistics, up to a linear transformation, are distributed approximately as $(M_1, \dots, M_m)$. There are only three possible $\tilde{\lambda}(x)$ to consider:

Gumbel : $\tilde{\lambda}(x) = e^{-x}$,

Fréchet : $\tilde{\lambda}(x) = x^{-\alpha}\ (x > 0), (\alpha > 0)$,

Weibull : $\tilde{\lambda}(x) = (-x)^{\alpha}\ (x < 0), (\alpha > 0)$.

Then for large values of y, if $F^n(\lambda_n + \delta_n y) \to \tilde{L}(y)$ as $n \to \infty$, we have

$F(y) \approx \tilde{L}^{1/n}((y - \lambda_n)/\delta_n) = e^{-\tilde{\lambda}((y-\lambda_n)/\delta_n)/n}$;

estimation of $F(y)$ is thus obtained by substitution of $(\hat{\lambda}_n, \hat{\delta}_n)$ for (λ_n, δ_n).

The problem is now to estimate (λ_n, δ_n) from the m largest observations.

In the first case (Gumbel distribution), $\tilde{\lambda}^{-1}(y) = -\log y$ and thus we have that

$D_i = M_i - M_{i+1}, -\log \frac{Z_1 + \dots + Z_i}{Z_1 + \dots + Z_i + Z_{i+1}}$, and $\frac{Z_i}{i}$

have the same distribution.

The spacings D_i are thus independent exponential random variables with mean values $1/i$, independent of $M_m (i = 1, \dots, m-1)$. In the other cases the ratios M_i/M_{i+1} are independent. Independent and exponential spacings among order statistics are a well-known characterization of the exponential distribution. The latter is an important member of the domain of attraction of the Gumbel distribution.

One can plot $x_i^{"}$ vs. $M\{M_i\}$ $(i = 1, \dots, m)$ for each of the three models and decide which (if any!) fits the data, i.e., that they are approximately in a straight line. If m is at our disposal it can be determined as the largest value for which $x_1^{"} > \cdots > x_m^{"}$ fit the model. Once the model is chosen the parameters can be estimated. Maximum likelihood estimators and minimum variance estimators are discussed in Weissman (1978). For the Gumbel model the maximum likelihood estimators are

$$\widehat{\delta}_n = \frac{1}{m}\sum_{i=1}^{m-1}(x_i^{"} - x_m^{"}) = \frac{1}{m}\sum_{i=1}^{m} x_i^{"} - x_m^{"} = \bar{x}_m^{"} - x_m^{"} = \frac{1}{m}\sum_{i=1}^{m-1} i(x_i^{"} - x_{i+1}^{"}))$$

$$\hat{\lambda}_n = x_m^{"} + \widehat{\delta}_n \log m\,,$$

and the minimum variance estimators are

$$\delta_n^* = \frac{1}{m-1}\sum_{i=1}^{m-1}(x_i^{"} - x_m^{"}) = \frac{m}{m-1}\,\widehat{\delta}_n$$

$$\lambda_n^* = x_m^{"} + \delta_n^*(S_n - \gamma)$$

where $S_1 = 0$, $S_m = 1 + \frac{1}{2} + \cdots + \frac{1}{m-1}\,(m > 1)$ and $\gamma = .57722$ is Euler's constant. Then $2(m-1)\,\delta_n^*/\delta_n$ is (asymptotically) a $\chi^2(2m-2)$ variate and thus confidence intervals for δ_n can be obtained. The distribution of $U_m = (x_m^{"} - \lambda_n)/\delta_n^*$ is parameter-free and thus confidence intervals for λ_n can be obtained provided the percentage points are available. The latter are tabulated in Weissman (1978).

For the Gumbel distribution we have, as seen in the attraction conditions Chapter,

$$(Q(1 - c/n) - \lambda_n)/\delta_n \to -\log c \quad \text{as} \quad n \to \infty.$$

and so

$$\widehat{Q}(1 - c/n) = \hat{\lambda}_n + \widehat{\delta}_n(-\log c)$$

is the approximate maximum likelihood estimator for $Q(1 - c/n)$.

Similar estimates are obtained in Weissman (1981) for threshold (type 1) censoring. Particularly, only values larger than some a (fixed) are observable and then m is random. It turns out that a plays the role of $x_m^{"}$ in the formulae for $\hat{\lambda}_n$, $\widehat{\delta}_n$ and $\widehat{Q}$.

In principle, these asymptotic results hold for every distribution function $F(.) \in \mathcal{D}(\Lambda(.))$ (exponential, gamma, normal, lognormal, logistic, Weibull, to name some). But for finite n, better approximations can be obtained for distributions with exponential tail. Recently Boos (1983) used $\widehat{Q}(1 - c/n)$ to estimate large quantiles

for various distributions. He found that for some of them, as regards the mean-square error, this method is better for some distributions but worse for others; for more details see Weissman (1984).

For the Fréchet distribution the transformation $Y_i'' = \log(X_i'' - \lambda)$ transforms the model to the Gumbel one. For the Weibull distribution the transformation $Y_i'' = -\log(\lambda - X_i'')$ has the same effect. But in both cases the location parameter, unknown in general, appears and has to be estimated. An example follows.

The three-parameter case for the Weibull distribution for minima is found in life testing, strength distributions, fatigue failures, etc., the available data being

$$x_1'' \le x_2'' \le \cdots \le x_m''$$

with $m << n$.

Thus we assume that $F(x) = 1 - \exp\{-((x-\lambda)/\delta)^\alpha$ for $x > \lambda, F(x) = 0$ if $x \le \lambda$; for this distribution we have $\lambda_n = \lambda$ and $\delta_n = \delta\, n^{-1/\alpha}$. The maximum likelihood estimators for the three-parameters $(\lambda, \delta, \alpha)$ are

$$\hat{\delta} = \left(\frac{n}{m}\right)^{1/\hat{\alpha}} (x'_m - \hat{\lambda}),$$

$$\hat{\alpha} = m / \sum_{i=1}^{m} \log \frac{x'_m - \hat{\lambda}}{x'_i - \hat{\lambda}} (> 0), \text{ and}$$

$$\frac{1}{m} \sum_{i=1}^{m} \log \frac{x'_m - \hat{\lambda}}{x'_i - \hat{\lambda}} + \frac{m}{\sum_{i=1}^{m} (x'_m - \hat{\lambda})/(x'_i - \hat{\lambda})} = 1.$$

The last equation shows that $1/\hat{\alpha} < 1$ or $\hat{\alpha} > 1$ and so the method does not work for $\alpha \le 1$. We cannot use x'_1 as an estimator of λ because $x'_1 \xrightarrow{a.s.} \lambda$ and so the two other equations would give $\hat{\alpha} = 0$ and $\hat{\delta} = +\infty$.

Maximum likelihood estimation is discussed by Hall (1982) and Smith and Weissman (1985).

Confidence intervals for λ are suggested by Weissman (1981). Using the convergence in distribution of $(M_1, M_2, \ldots, M_m)^T$ conveniently adapted for minima we see that

$$\frac{x'_1 - \lambda}{x'_m - x'_1}$$

converge in distribution to a random variable whose distribution function is

$$1 - \left(1 - \left(\frac{x}{1+x}\right)^\alpha\right)^{m-1}$$

for $x \geq 0$ and 0 for $x \leq 0$.

If α is known then the quantile $\chi_{m,\alpha}(p)$ of this distribution can be obtained in a closed form. For large values of $n, (x'_1 - \lambda)/(x'_m - \lambda)$ can be used as a pivotal function to form confidence intervals for λ. If α is unknown we can use the pivotal functions

$$\log \frac{x'_m - \lambda}{x'_1 - \lambda} \Big/ \sum_{i=1}^{m-1} \log \frac{x'_m - \lambda}{x'_1 - \lambda} \quad (m \geq 3)$$

or

$$\log \frac{x'_i - \lambda}{x'_1 - \lambda} \Big/ \log \frac{x'_m - \lambda}{x'_1 - \lambda} \quad (1 < i < m \leq n)$$

to obtain confidence intervals for λ. Under Weibull distributions for minima both have limiting distributions which do not depend on any parameter. The quantiles of these distributions are tabulated in Weissman (1981). Simulation has shown good performance when $\alpha \leq 1$, which is useful since in this case ($\alpha < 1$) the maximum likelihood method fails.

Gomes (1981) investigated the maximum likelihood estimators of (λ, δ) using the m largest observations in each of n samples from $\Lambda(x)$ (Weissman (1978) treated the case $n = 1$). That is, for $m = 2$, she considered the pairs $x''_1 \geq x''_2$. The results agree with those of Tiago de Oliveira (1972) and others when $m = 1$. She then investigated the properties of these estimators for $m = 1,2,3$ and $m = 5(5)30$, by a Monte Carlo simulation based on between 4500 and 10000 replicates for each case, and found that $\hat{\lambda}$ is positively biased for $m = 1$ and negatively biased for $m > 1$ and $\hat{\delta}$ is always negatively biased. In addition, she gave a similar treatment for estimating (λ, δ) using moment estimators, simple best linear unbiased estimators and simple best linear invariant estimators. For $m = 3$, a simple linear estimator is one of the form

$$\sum_{i=1}^{m} \sum_{k=1}^{3} w_{kj} \; x''_{kj}$$

where x''_{kj} is the k-th largest order statistics in the j-th sample. The moment estimator λ^* is negatively biased for $m = 1$ and positively biased for $m > 1$ and δ^* is always positively biased. The bias properties of maximum likelihood and of moment estimators are the reverse of one another. The method of moments gave a slightly better estimator λ^* for small values of n, but yields very poor estimators δ^*.

Other references are Smith (1986) and Gomes (1984).

Finally some reference must be made to Hill's (1975) estimator for Fréchet distributions, $\Phi_\alpha(x/\delta)$ (i. e. , $\lambda = 0$) and correspondingly to an analogous estimator for the Weibull distribution $W_\alpha(x/\delta)$ (also $\lambda = 0$).

For the Fréchet distribution $\Phi_\alpha(x/\delta)$ the transformed random variable $Y = \log X$ follows the Gumbel distribution $\Lambda(\frac{y - \log \delta}{1/\alpha})$. The m largest values $x_i^{''}(i = 1, 2, \dots, m)$ correspond to the m largest values $y_i^{''} = \log x_i^{''}$. Using the ML estimators based on the largest m observations, of Weissman (1978), we have

$$1/\hat{\alpha} = \frac{1}{m}\sum_{1}^{m} y_i^{''} - y_m^{''}$$

$$\log \hat{\delta} = y_m^{''} + (1/\hat{\alpha}) \log m$$

or equivalently $\quad \hat{\alpha} = (\log \frac{(\pi_1^m x_i^{''})^{1/m}}{x_m^{''}})^{-1}$.

$$\hat{\delta} = m^{1/\hat{\alpha}}. x_m^{''}.$$

$\hat{\alpha}$ is the well-known Hill estimator of the index of Fréchet distribution. See also Diaz (1985).

For the Weibull distribution $W_\alpha(x/\delta)$ the transformed random variable $Y = -\log X$ (a decreasing transformation) has the Gumbel distribution $\Lambda(\frac{y + \log \delta}{1/\alpha})$. Then the smallest m observations $x_1^{'} \le x_2^{'} \le \cdots \le x_m^{'}$ give rise to the m largest values $y_i^{''} = -\log x_i^{'}$. In the same way we have

$$(1/\hat{\alpha}) = \frac{1}{m}\sum_{1}^{m} y_i^{''} - y_m^{''}$$

$$(-\log \hat{\delta}) = y_m^{''} + (1/\hat{\alpha})\log m$$

or equivalently $\quad \hat{\alpha} = [\log(x_m^{'}/(\pi_1^m x_i^{''})^{1/m})]^{-1}$

$$\hat{\delta} = m^{-1/\hat{\alpha}}\, x_m^{''}.$$

This result can be useful for right-censored life tests (where naturally $\lambda = 0$).

Evidently quantile estimators and predictors can be made as usual.

For more detail connected to tail estimation see the Annex 4.

9.3 Largest observations vs. m subsamples maxima methods; a comparison

In this section, corresponding to item 2 of the Introduction, we will closely follow Hüsler and Tiago de Oliveira (1988). Suppose that we have a sample of $n = km$ observations and we can use only the largest m of all observations or the m maxima of natural blocks of data — in fact we are comparing what engineers call, respectively, the *largest observations method* with the *block (yearly) maxima method*. Let us denote the $n = km$ random variables by $x_{ij}, i = 1, \dots, k$ and $j = 1, \dots, m$, where m denotes the number of possible blocks (e.g. years) and k the number of observed random variables per block (year). Let $y_j = \max\{x_{ij}, 1 \le i \le k, j = 1, \dots, m\}$ and assume that all x_{ij} have the Gumbel distribution $\Lambda(\frac{x-\lambda}{\delta})$ and so the block (yearly) distribution of the y_j is $\Lambda(\frac{x-\lambda}{\delta} - \log k) = \Lambda(\frac{x-(\lambda+\delta \log k)}{\delta})$; $\lambda + \delta \log k$ is thus the block (year) location parameter (*).

We may use the m values y_j for estimating the values λ and δ, which is the classical method, described e.g. in Gumbel (1958). As said previously, Weissman (1978) proposed an estimation based on the m-largest observations of all $n = km$ values x_{ij}, which are

$$x_1'' > x_2'' > \cdots > x_m''$$

with probability one.

At this point let us analyze the efficiency of Weissman's method with respect to the classical one (**).

We are now assuming, for the sake of comparison, that both methods can be used in a given application, i.e. that the $x_i'', i \le m$, as well as $y_j, j = 1, \dots, m$, can be observed and n is sufficiently large to the use of the Gumbel approximation. The comparison uses asymptotic results for both methods.

(*) From the practical point of view, as in the previous section, we are assuming that the x_{ij} are i. i. d. and attracted to a Gumbel distribution and so their distribution can be approximated also by a Gumbel distribution. What is essential in the subsamples (or block) maxima method is that the maxima have a Gumbel distribution, even if the observations of the subsample are dependent.

(**) Notice that there are situations where only the maxima y_j per block (year) are recorded; in this case Weissman's method is not applicable, only the classical one.

Both methods lose much information with respect to the (underlying) full sample of $n = km$ *; the efficiencies of both methods with respect to the full sample tend to zero as* $k \to \infty$ *since the procedures are based on very special subsamples*.

Then the parameters $(\lambda + \delta \log k, \delta)$ of the $y_j, (j = 1, \dots, m)$ have the usual maximum likelihood estimators $(\hat{\lambda} + \hat{\delta} \log k, \hat{\delta})$.

From the asymptotic $(m \to \infty)$ variance-covariance matrix of $(\hat{\lambda} + \hat{\delta} \log k, \hat{\delta})$, given in Chapter 5, we get for the asymptotic variance-covariance matrix $\widehat{\sum}$ of $(\hat{\lambda}, \hat{\delta})$,

$$\widehat{\sum} = \frac{\delta^2}{m} \begin{bmatrix} 1 + \frac{6}{\pi^2}(1 - \gamma - \log k)^2 & -6(1 - \gamma - \log k)/\pi^2 \\ -6(1 - \gamma - \log k)/\pi^2 & 6/\pi^2 \end{bmatrix}$$

and, for $k > 1$,

$$\rho(\hat{\lambda}, \hat{\delta}) = -(1 + \frac{\pi^2}{6(1 - \gamma - \log k)^2})^{-1/2} \to -1, \text{as } k \to +\infty,$$

but slowly.

Consider now estimation based on the m largest values $x_1'' > x_2'' > \cdots > x_m''$ of the i.i.d. sample of $\{x_{ij}\}$ $(i = 1, \dots, k;\ j = 1, \dots, m)$.

For fixed k and m we get the usual maximum likelihood estimators (λ, δ) (with a slight change of notations) from the censored sample of m largest observations of a (possible) sample of $n = km$. We have, as $k \to +\infty$,

$$\lambda^* = x_m'' - \delta^* \log k + O_p(k^{-1})$$

$$\delta^* = \bar{x}_m'' - x_m'' + O_p(\frac{\log k}{k})$$

with $\bar{x}_m'' = \frac{1}{m} \sum_1^m x_i''$.

The asymptotic variance-covariance matrix of (λ^*, δ^*) is

$$\sum{}^* \sim \delta^2 \begin{bmatrix} \sigma_m^2 + \frac{m-1}{m^2}(\log m)^2 & \frac{m-1}{m^2}\log m \\ \frac{m-1}{m^2}\log m & \frac{m-1}{m^2} \end{bmatrix}$$

with

$$\rho(\lambda^*, \delta^*) = (1 + m^2\sigma_m^2/(m-1)\ (\log m)^2)^{-1/2} \qquad \text{for } m \geq 2$$

where $\sigma_m^2 = \frac{\pi^2}{6} - S_m'$, with $S_m' = \sum_1^{m-1} j^{-2}$, $S_m' = 0$

Clearly $\rho(\lambda^*, \delta^*) \to 1$, as $k \to \infty$, also slowly. To compare these two procedures we will use the well-known Cramér efficiency. It is defined as

$$\text{eff}((\lambda^*, \delta^*)/(\hat{\lambda}, \hat{\delta})) = \det(\hat{\sum})/\det(\sum^*)$$

$$= \frac{6}{\pi^2} \frac{1}{(m-1)\sigma_m^2} \downarrow \frac{6}{\pi^2} = .60793$$

as $m \to \infty$.

We could also study the efficiency as defined in Tiago de Oliveira (1982), defined as the worst possible for the estimation of all quantiles. It was shown that this efficiency is the smallest root of $\det(\hat{\sum} - \mu \sum^*) = 0$. But as $\mu_{max} \cdot \mu_{min} = \text{eff}(\lambda^*, \delta^*)/(\hat{\lambda}, \hat{\delta}) \approx \frac{\pi^2}{6}$, we see that μ_{min} is (approximately) smaller than $\sqrt{\pi^2/6} = .77970$ for large m. It was shown, in the basic paper referred to, that $(\hat{\lambda}, \hat{\delta})$ — the yearly maxima method — is better than (λ^*, δ^*) — the largest maxima method — if the quantile corresponds to a probability larger than $p = .9$, for $m \geq 1$ using the asymptotic approximation, and in practice for every p if $m \geq 15$ which is the general case.

This shows a definite superiority of the yearly maxima method over some wide-spread thinking, when both methods can he applied. Evidently a symbiosis of both methods can be useful in some cases.

These results are based on the exact assumption that the $\{x_{ij}\}$ have the Gumbel distribution Λ. But these results can hold even with the assumption that the distribution $\{x_{ij}\}$ of belongs to the domain of attraction of Λ with a sufficiently fast rate of convergence of F^n to Λ. This will certainly be true if F is sufficiently close to Λ in an intuitive sense.

One might think, if the Gumbel distribution is valid, that the largest maxima method can always be applied for small $m's$. But the assumption of the independence of the x_{ij}'s is rarely satisfied in applications. The crucial point for this method is mainly the local dependence of the x_{ij}'s . Often the largest values occur in clusters, which prevents the application of the largest maxima method; but the yearly maxima method k can still be applied.

We can mention that these results may useful for the design of experiments: when only a small amount of information is available, i.e. if only m instead of $n = km$ observations can be recorded, a choice of design must be made, and if we are interested in large quantiles the block method should be used, chiefly because it guarantees the practical independence of the blocks.

An analysis similar to the one given here was made by Reiss (1987) for the Fréchet distribution, with special reference to the Hill estimator.

9.4 Estimation using two or three quantiles:

It is important, at this point, to recall that $\chi_p^* \xrightarrow{P} \chi_p, \sqrt{n}\, f(\chi_p) \frac{\chi_p' \chi_p}{\sqrt{p(1-p)}}$ is asymptotically standard normal if $0 < f(\chi_p) < \infty$, that (χ_p^*, χ_q^*), where $0 < p < q < 1$ conveniently reduced is an asymptotically binormal pair with standard margins and correlation coefficient $\rho = \sqrt{\frac{p}{q}\,\frac{1-q}{1-p}} > 0$. Its extension to more sample quantiles is immediate. See in Chapter 1 "A note on the asymptotic behaviour of sample quantiles".

Consider now the Gumbel distribution $\Lambda((x-\lambda)/\delta)$. Then as $\chi_p = \lambda + \delta(-\log(-\log p))$, the estimation equations are

$$Q_p(n) = \chi_p^* = \lambda^* - \delta^* \log(-\log p)$$

$$Q_q(n) = \chi_q^* = \lambda^* - \delta^* \log(-\log q),$$

and so the estimators are

$$\lambda^* = [\log(-\log p) Q_q - \log(-\log q) Q_p] / \log\left(\frac{\log p}{\log q}\right)$$

$$\delta^* = (Q_q - Q_p) / \log\left(\frac{\log p}{\log q}\right).$$

As the variance-covariance matrix of (Q_q, Q_p) is

$$V \sim \frac{1}{n} \begin{bmatrix} \frac{p(1-p)}{f^2(\chi_p)} & \frac{p(1-q)}{f(\chi_p)\, f(\chi_q)} \\ \frac{p(1-q)}{f(\chi_p)\, f(\chi_q)} & \frac{q(1-q)}{f^2(\chi_q)} \end{bmatrix} = \frac{\delta^2}{n} \begin{bmatrix} \frac{1-p}{p(\log p)^2} & \frac{1-q}{q \log p \log q} \\ \frac{1-q}{q \log p \log q} & \frac{1-p}{q(\log q)^2} \end{bmatrix}$$

with $\det(V) \sim \frac{\delta^4}{n^2} \frac{(1-q)(q-p)}{pq^2 (\log p)^2 (\log q)^2}$, the variance-covariance matrix of (λ^*, δ^*) as

$$[Q_q, Q_p]^T = \begin{bmatrix} 1 & -\log(-\log p) \\ 1 & -\log(-\log q) \end{bmatrix} (\lambda^*, \delta^*)^T,$$

is

$$V^* = \begin{bmatrix} 1 & -\log(-\log p) \\ 1 & -\log(-\log q) \end{bmatrix} - 1 \;\; V \begin{bmatrix} 1 & -\log(-\log p) \\ 1 & -\log(-\log q) \end{bmatrix} - 1^T$$

with $\det(V^*) = \det(V) / \det^2 \begin{bmatrix} 1 & -\log(-\log p) \\ 1 & -\log(-\log q) \end{bmatrix}$.

As the Cramér-Rao bound for the estimators of (λ, δ) — corresponding to the ML estimators $(\hat{\lambda}, \hat{\delta})$ — is

$$\hat{V} = \frac{\delta^2}{n}\begin{bmatrix} 1 + \frac{6(1-\gamma)^2}{\pi^2} & \frac{6(1-\gamma)}{\pi^2} \\ \frac{6(1-\gamma)}{\pi^2} & \frac{6}{\pi^2} \end{bmatrix}$$

the Cramér efficiency is

$$\det(\hat{V})/\det(V^*) = \frac{6}{\pi^2}\,\frac{(\log(\frac{\log p}{\log q}))^2\; p\, q^2 (\log p)^2\; (\log q)^2}{(1-q)^2\; (q-p)^2}.$$

The maximum efficiency is obtained for $p = .07$ and $q = .76$ and its value is 40.8%.

One could also compute the efficiency as defined in Tiago de Oliveira (1982), connected with the study of quantile estimators that follows.

The quantile estimator of the ξ-quantile χ_ξ is

$$\chi_\xi^*(p, q) = c_1\, Q_p + c_2\, Q_q$$

with $\quad c_1 + c_2 = 1$

and $\quad c_1\, \chi_p + c_2\, \chi_q = \chi_\xi$ or $(-\log p)^{c_1}\, (-\log q)^{1-c_1} = -\log \xi$

to be quasi-linearly invariant.

It is asymptotically normal with mean value χ_ξ and asymptotic variance

$$V(\chi_\xi^*(p, q)) \sim \frac{1}{n}\{c_1^2\, \frac{p(1-p)}{f^2(\chi_p)} + 2c_1(1-c_1)\frac{p(1-q)}{f(\chi_p)\;\; f(\chi_q)} + (1-c_1)^2\, \frac{q(1-q)}{f^2(\chi_q)}\} =$$

$$\frac{\delta^2}{n}\{c_1^2\, \frac{1-p}{p(\log p)^2} + 2c_1(1-c_1)\frac{1-q}{q \log p\;\; \log q} + (1-c_1)^2\, \frac{1-q}{q(\log q)^2}\}.$$

We can compute its asymptotic efficiency with respect to the Cramér-Rao bound, i.e., with respect to the ML estimator $\hat{\chi}_\xi = \hat{\lambda} + \xi\, \hat{\delta}$.

As $\quad V(\hat{\chi}_\xi) \sim \frac{\delta^2}{n}\{1 + (1 - \gamma + \chi_\xi)^2\} = \frac{\delta^2}{n}\{1 + (1 - \gamma - \log(-\log \xi))^2\}$, the asymptotic efficiency is

$$\lim_{n\to\infty} \frac{V(\hat{\chi}_\xi)}{V(\chi_\xi^*\,(p, q))}$$

and it can be maximized in (p, q) $(0 < p < q < 1)$.

Neves (1986) has made the calculations and the optimal results are as follows with p, q, and c_1 (recall that $c_2 = 1 - c_1$) as functions of ξ see Table 9.1.

Table 9.1

ξ	p	q	c_1	eff (%)
0.01	0.066	0.933	1.143720	65
0.02	0.072	0.943	1.104300	65
0.03	0.008	0.051	0.339046	67
0.05	0.014	0.086	0.360599	75
0.10	0.030	0.180	0.412024	82
0.20	0.070	0.350	0.459707	82
0.30	0.130	0.510	0.524218	81
0.40	0.210	0.640	0.574626	82
0.50	0.310	0.740	0.613838	83
0.60	0.430	0.820	0.653152	84
0.70	0.560	0.880	0.678641	83
0.80	0.710	0.930	0.723894	79
0.85	0.777	0.948	0.716725	73
0.89	0.010	0.755	-0.314795	68
0.90	0.010	0.759	-0.341731	68
0.95	0.013	0.776	-0.562649	68
0.99	0.021	0.793	-1.115900	67

Notice that we do not always have $p < \xi < q$, as might be expected from a naive analysis. This is connected to the fact that the graph of efficiency has two relative maxima and the absolute maximum changes with ξ. For more details and the graphs see Neves (1986). The efficiency is thus reasonable for the usual quantiles.

Having thus sketched the way of analysing the use of quantiles, we will in the next cases/distributions concentrate only on the efficiency for quantiles estimators.

The Fréchet distribution has the form $\Phi_\alpha((x-\lambda)/\delta) = 0$ if $x \le \lambda, \Phi_\alpha((x-\lambda)/\delta) = \exp\{-((x-\lambda)/\delta)^\alpha\}$ if $x \ge \lambda$. Supposing λ known ($\lambda = \lambda_0 = 0$) for convenience we know that the increasing transformation $Y = \log X$ leads

to a Gumbel distribution $\Lambda(\frac{y-\log\delta}{1/\alpha})$. The empirical quantiles of the $\{y_i\}$ are the corresponding ones of the $\{x_i\}$, and so the estimator of the ξ-quantile χ_p is given by $\log\chi_\xi^* = c_1 \log Q_p + (1-c_1)\log Q_q, \text{i.e.}, \chi_\xi^* = Q_p^{c_1}\, Q_q^{1-c_1}$; recall that this is in correspondence with the second condition which, here, takes the form $c_1 \log\chi_p + (1-c_1)\ \log\chi_q = \log\chi_\xi$ or $\chi_\xi = \chi_p^{c_1}\, \chi_q^{1-c_1}$, defining $c_1(p,q,\xi)$.

χ_ξ^* is homogeneous, i.e., if the $\{x_i\}$, are multiplied by $c(>0)$ the same happens to χ_ξ^*; as the asymptotic variance of the χ_ξ^* is the same as for χ_ξ^* for the Gumbel distribution above, the δ-method shows that $\chi_\xi^*(p,q)$ is also asymptotically normal with mean value and asymptotic variance $\frac{\delta^2}{(-\log\xi)^{2/\alpha}\,\alpha^2 n}\{1+(1-\gamma-\log(-\log\xi))^2\}$; the efficiency is the same as before. Notice that the variance depends on δ and α and not only on α as could be expected $(1/\alpha$ is dispersion parameter of the transformed variable $y)$. This is, evidently, connected with the homogeneity but the not quasi-linearity of the estimator.

If for the Fréchet distribution we have (λ,δ) unknown, but $\alpha=\alpha_0$ is known, the situation is completely analogous to the location-dispersion situation above. The ξ-quantile is $\chi_\xi = \lambda + (-\log\xi)^{-1/\alpha_0}\cdot\delta$ and $c_1 = c_1(p,q,\xi)$ is defined by $c_1\,\chi_p + (1-c_1)\,\chi_q = \chi_\xi$ or $c_1(-\log p)^{-1/\alpha_0} + (1-c_1)\,(-\log q)^{-1/\alpha_0} = (-\log\xi)^{-1/\alpha_0}$. The estimator χ_ξ^* of the ξ-quantile is

$$\chi_\xi^* = c_1 Q_p + (1-c_1)Q_q$$

which is asymptotically normal with mean-value χ_ξ and asymptotic variance

$$V(\chi_\xi^*) \sim \frac{1}{n}\{c_1^2\,\frac{p(1-p)}{f^2(\chi_p)} + 2c_1(1-c_1)\frac{p(1-q)}{f(\chi_p)\ \ f(\chi_q)} + (1-c_1)^2\,\frac{q(1-q)}{f^2(\chi_q)^2}\}$$

$$= \frac{\delta^2}{\alpha_0^2 n}\{c_1^2\,\frac{1-p}{p(-\log p)^{2+2/\alpha_0}} + 2c_1(1-c_1)\,\frac{1-q}{q\,(\log p\cdot\log q)^{1+1/\alpha_0}} + (1-c_1)^2\,\frac{1-q}{q(-\log q)^{2+2/\alpha_0}}\},$$

and the ML estimator $\widehat{\chi_\xi} = \hat{\lambda} + (-\log\xi)$, associated with the Cramér-Rao bound, has mean value χ_ξ and asymptotic variance

$$V(\widehat{\chi_\xi}) \sim \frac{\delta^2}{n}\{\frac{(-\log\xi)^{-2/\alpha_0}}{\alpha_0^2} + \frac{(1-\Gamma(2+1/\alpha_0)\ \ (-\log\xi)^{-1/\alpha_0})^2}{(\alpha_0+1)^2\ (\Gamma(1+2/\alpha_0)-\Gamma^2(1+1/\alpha_0))}\}.$$

The asymptotic efficiency can then be computed as before. For $\alpha_0 = 3$ we have

Table 9.2 ($\alpha_0 = 3$)

ξ	p	q	c_1	eff (%)
0.010	0.0399	0.8632	1.062470	64
0.015	0.0500	0.8500	1.064880	63
0.018	0.0045	0.0320	0.379776	69
0.020	0.0050	0.0356	0.362109	71
0.050	0.0132	0.0924	0.407834	80
0.100	0.0300	0.1890	0.465436	82
0.200	0.0770	0.3720	0.550790	82
0.300	0.1500	0.5180	0.613496	84
0.400	0.2450	0.6280	0.655708	86
0.500	0.3570	0.7150	0.688727	87
0.600	0.4780	0.7870	0.713217	84
0.700	0.6050	0.8490	0.732972	78
0.740	0.6500	0.9000	0.788330	73
0.750	0.0043	0.5952	-0.399820	73
0.800	0.0050	0.6090	-0.558671	73
0.900	0.0070	0.6320	-1.155680	72
0.950	0.0082	0.6448	-1.901690	72
0.990	0.0095	0.6591	-4.454210	71

We can make comments similar to those for the Gumbel distribution.

Let us recall that, as seen before, $\Phi_\alpha(1+\frac{x-\lambda}{\alpha\,\delta}) = \Phi_\alpha(\frac{x-(\lambda-\alpha\,\delta)}{\alpha\,\delta}) \to \Lambda(\frac{x-\lambda}{\delta})$ as $\alpha \to +\infty$. Then $\lambda - \alpha\,\delta$ takes the place of λ and $\alpha\,\delta$ that of δ in the previous formulation, and the ξ-quantile is $\tilde{\chi}_\xi = \lambda - \alpha(1-(-\log\xi)^{-1/\alpha})\delta$ and for fixed α_0 the MLestimator of $\tilde{\chi}_\xi$ is $\tilde{\tilde{\chi}}_\xi = \tilde{\lambda} - \alpha_0(1-(-\log\xi)^{-1/\alpha_0})\tilde{\delta}$ where $\tilde{\lambda}$ and $\tilde{\delta}$ are the ML estimators of the *new* parameters (λ, δ). $\tilde{\tilde{\chi}}_\xi$ is also asymptotically normal with mean value $\tilde{\chi}_\xi$ and variance asymptotic to

$$\frac{\alpha_0^2\,\delta^2}{n}\{\frac{(-\log\xi)^{-2/\alpha_0}}{\alpha_0^2} + \frac{(1-\Gamma(2+1/\alpha_0)\ (-\log\xi)^{-1/\alpha_0})^2}{(\alpha_0+1)^2\ (\Gamma(1+2/\alpha_0)-\Gamma^2(1+1/\alpha_0))}\};$$

when $\alpha_0 \to +\infty$ we see that $\tilde{\chi}_\xi \to \lambda - \log(-\log\xi)\,\delta$ and the variance asymptotic is to $\frac{\delta^2}{n}\{1+\frac{6}{\pi^2}(1-\gamma-\log(-\log\xi))^2\}$ which are the corresponding values for the Gumbel distribution $\Lambda((x-\lambda)/\delta)$, as could be expected from

$$\Phi_{\alpha_0}(1+\tfrac{x-\lambda}{\alpha_0}) \to \Lambda(\tfrac{x-\lambda}{\delta}) \text{ as } \alpha_0 \to \infty.$$

Consider now the three-parameter case for the Fréchet distribution.

It is not possible to obtain linear combinations of three quantiles (Q_p, Q_q, Q_r), with constant coefficients depending on (p, q, r), to estimate parameters; they should always depend also on α being estimated, and this could be presumed because it appears in a weak form in the estimation of the ξ-quantile when $\lambda = \lambda_0$ is known. The parameter α can be estimated by the location-dispersion-free ratio.

$$\frac{Q_r - Q_q}{Q_q - Q_p} \xrightarrow{P} g(\alpha) = \frac{(-\log r)^{-1/\alpha} - (-\log q)^{-1/\alpha}}{(-\log q)^{-1/\alpha} - (-\log p)^{-1/\alpha}}.$$

Thus the estimator α^* can be given by $g(\alpha^*) = \frac{Q_r - Q_q}{Q_q - Q_p}$ which, by the δ-method, can be shown to be asymptotically normal with mean value α and variance $O(n^{-1})$; the triple $(\lambda^*, \delta^*, \alpha^*)$ is also asymptotically trinormal with mean value $(\lambda, \delta, \alpha)$ and variance-covariance matrix $O(n^{-1})$.

A simple solution is to take $\frac{\log p}{\log q} = c > 1$ and $\frac{\log r}{\log q} = c^{-1} < 1$; we get

$$\frac{Q_r - Q_q}{Q_q - Q_p} = c^{1/\alpha} \quad \text{and} \quad p = q^c, q = r^c;$$

any choice of c gives a system for the estimation of α.

The best choice of (p, q, r) for each ξ, or even the choice of (p, q, r) that maximizes the asymptotic efficiency, has not yet been studied and should not be expected to be very efficient.

The method only seems useful, at present, to obtain the first estimates of $(\lambda, \delta, \alpha)$ from the equations

$$Q_p = \chi_p^* = \lambda^* + (-\log p)^{-1/\alpha^*}\delta^*$$

$$Q_q = \chi_q^* = \lambda^* + (-\log q)^{-1/\alpha^*}\delta^*$$

$$Q_r = \chi_r^* = \lambda^* + (-\log r)^{-1/\alpha^*}\delta^*$$

to be used to seed the solution of the ML equations and then to estimate the quantiles; using these estimators to estimate the quantiles their asymptotically normal behaviour

can be obtained, through simple but lengthy computations, by the use of the δ-method.

Recall that for the Fréchet distribution we must always have $\lambda^* \le X_1'$, which imposes a new condition on the (X_1', Q_p, Q_q) or (X_1', Q_p, Q_q, Q_r) to allow the use of the method; note that $\text{prob}\{\lambda^* \le X_1'\} \to 1$.

Consider finally the Weibull distribution (for minima) $W_\alpha((x-\lambda)/\delta)$ and suppose that $\lambda = \lambda_0$ is known ($\lambda_0 = 0$ for convenience). Then by the transformation $Y = -\log X$ we get a Gumbel random variable with parameters $(-\log\delta, 1/\alpha)$. As the transformation is decreasing we have $\chi_{1-\xi}(\text{Gumbel}) = -\log\chi_\xi$ (Weibull), and so the estimator of the ξ-quantile is

$$\chi_\xi^* = Q_p^{d_1}\, Q_q^{1-d_1}$$

where $d_1(p,q,\xi) = c_1(1-q, 1-p, 1-\xi)$; χ_ξ^* is also homogeneous; as the asymptotic variance of $-\log\chi_\xi^*$ is the same as that for $\chi_{1-\xi}^*$ in the Gumbel distribution, the δ-method shows that $\chi_\xi^*(p,q)$ is also asymptotically normal with mean value χ_ξ and asymptotic variance $V(\chi_\xi^*(p,q)) \sim \frac{\delta^2}{(-\log(1-\xi))^{2/\alpha}\,\alpha^2\,n}\{1 + (1-\gamma-\log(-\log(1-\xi)))^2\}$, the efficiency being the same as before with the exchange of (ξ, p, q) by $(1-\xi, 1-p, 1-q)$.

The same could be obtained by noting that if X has the distribution $W_\alpha(x/\delta)$ then $1/X$ has the distribution $\Phi_\alpha(x/(1/\delta))$.

Suppose now that we have the Weibull distribution (for minima) with $\alpha = \alpha_0 > 2$ known, i.e., the distribution is $W_{\alpha_0}((x-\lambda)/\delta)$ whose quantiles are $\chi_\xi = \lambda + (-\log(1-\xi))^{1/\alpha_0}\,\delta$. Consider the estimator $\chi_\xi^* = \chi_\xi^*(p,q) = c_1\,Q_p + c_2\,Q_q$ with $c_1 + c_2 = 1$ and $c_1(-\log(1-p))^{1/\alpha_0} + c_2(-\log(1-q))^{1/\alpha_0} = (-\log(1-\xi))^{1/\alpha_0}$. Evidently χ_ξ^* is asymptotically normal with mean value χ_ξ and asymptotic variance

$$V(\chi_\xi^*) \sim \frac{\delta^2}{\alpha_0^2\,n}\{c_1^2\,\frac{p}{(1-p)(-\log(1-p))^{2-2/\alpha_0}} + 2\,c_1(1-c_1)$$

$$\frac{p}{(1-p)\,(\log p\,\log q)^{1-1/\alpha_0}} + (1-c_1)^2\,\frac{q}{(1-q)(-\log(1-q))^{2-2/\alpha_0}}\}.$$

The ML estimator, for $\alpha_0 > 2$, associated with the Cramér-Rao bound $\hat\chi_\xi = \hat\lambda + (-\log(1-\xi))^{1/\alpha_0}\,\hat\delta$ has the mean value χ_ξ and asymptotic variance

$$V(\hat\chi_\xi) \sim \frac{\delta^2}{n}\{\frac{(-\log(1-\xi))^{2/\alpha_0}}{\alpha_0^2} + \frac{(1-\Gamma(2-1/\alpha_0)(-\log(1-\xi))^{-1/\alpha_0})^2}{(\alpha_0-1)^2\,\{\Gamma(1-2/\alpha_0)-\Gamma^2(1-1/\alpha_0)\}}\}.$$

Note the similarity of this result (for $\alpha_0 > 2$) with the corresponding one for the Fréchet distribution already given.

Let us obtain an asymptotic result analogous to the one connected with the convergence $\Phi_\alpha(\lambda + \frac{x-\lambda}{\alpha\ \delta}) \to \Lambda(\frac{x-\lambda}{\delta})$.

In our case we have $1 - W_\alpha(1 - \frac{x-\lambda}{\delta}) = 1 - W_\alpha(-\frac{x-(\lambda+\alpha\ \delta)}{\alpha\ \delta}) \to \Lambda\,(\frac{x-\lambda}{\delta})$, $\alpha \to \infty$, and the ξ-quantile is $\tilde{\chi}_\xi = \lambda + \alpha(1 - (-\log\xi)^{1/\alpha})\ \delta$ corresponding to the substitution of λ by $\lambda + \alpha\ \delta$, of δ by $\alpha\ \delta$ and ξ by $1 - \xi$ as we are, for W_α, dealing with minima and for Λ dealing with maxima.

$\tilde{\tilde{\chi}}_\xi = \tilde{\lambda} + \alpha_0(1 - (-\log\xi)^{1/\alpha_0})\ \tilde{\delta}$ is the ML estimator of $\tilde{\chi}_\xi$, for $\alpha_0 > 2$, $(\tilde{\lambda}, \tilde{\delta})$ being the ML estimators of the new (λ, δ).

Thus, as the transformations are linear, we know that $\tilde{\chi}_\xi$ is asymptotically normal with mean value $\tilde{\chi}_\xi$ and variance

$$V(\tilde{\tilde{\chi}}_\xi) \sim \frac{\alpha_0^2\ \delta^2}{n}\{\frac{(-\log\xi)^{2/\alpha_0}}{\alpha_0^2} + \frac{(1 - \Gamma(2 - 1/\alpha_0)\ \ (-\log\xi)^{1/\alpha_0})^2}{(\alpha_0 - 1)^2\ (\Gamma(1 - 2/\alpha_0) - \Gamma^2(1 - 1/\alpha_0))}\}.$$

Letting $\alpha_0 \to +\infty$, and corresponding to the fact that $1 - W_{\alpha_0}(1 - \frac{x-\lambda}{\alpha_0}) \to \Lambda(\frac{x-\lambda}{\delta})$, we see that $\tilde{\chi}_\xi \to \lambda - \log(-\log\xi))\delta$ and the variance is asymptotic to $\frac{\delta^2}{n}\{1 + \frac{6}{\pi^2}(1 - \gamma - \log(-\log\xi))^2\}$ which are the corresponding values for the Gumbel distribution, as happened for the Fréchet distribution before.

Consider now the three-parameter case for the Weibull distribution (of minima). The estimation is completely analogous to the previous one for the Fréchet distribution (for minima) with the added difficulty of not having regular ML estimators if $\alpha \leq 2$. The equations are analogous to the ones for the Fréchet distribution and quantile estimation is also analogous.

We must recall that any estimator λ^* should be such that $\lambda^* \leq X_1'$, which imposes, conditions on (X_1', Q_p, Q_q) or (X_1', Q_p, Q_q, Q_r) to allow the use of the method; although we know that $\text{Prob}\{\lambda^* \leq X_1'\} \to 1$.

See also the exercises concerning the use of X_1' and of one or two empirical quantiles to estimate the parameters of the Fréchet distribution or the Weibull distribution (for minima) and their ξ-quantiles.

9.5 Estimation of the parameters using block partitions of the sample:

As a matter of convenience, suppose we have the i.i.d. sample $(x_1, \dots, x_n)$, the ordered sample $(x_1^{'} \le x_2^{'} \le \cdots \le x_n^{'})$, we choose probability levels (p, q) with $0 < p < q < 1$, and we split the sample into three blocks:

$x_1^{'} \le \cdots \le x_{[np]}^{'}$ with average $\bar{x}_1 = \sum_1^{[np]} x_i^{'}/[np]$,

$x_{[np]+1}^{'} \le \cdots \le x_{[nq]}^{'}$ with average $\bar{x}_2 = \sum_{[np]+1}^{[nq]} x_i^{'}/([nq] - [np])$,

$x_{[nq]+1}^{'} \le \cdots \le x_n^{'}$ with average $\bar{x}_3 = \sum_{[np]+1}^{n} x_i^{'}/(n - [nq])$;

we are supposing n sufficiently large such that $1 < [np] < [nq] < n$ or $p > 1/n$ and $q - p > 1/n$.

As can be expected, the triple $(\bar{x}_1, \bar{x}_2, \bar{x}_n)$ has an asymptotic trinormal distribution. Their mean values are $(\lambda + \delta\mu_1, \lambda + \delta\mu_2, \lambda + \delta\mu_3)$ and the variance-covariance matrix is

$$\frac{\delta^2}{n}\,\Omega^{-1} \sim \frac{\delta^2}{n}\begin{bmatrix} \sigma_1^2/p & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2/(q-p) & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2/(1-q) \end{bmatrix}$$

where $F(\frac{x-\lambda}{\delta})$ is the form of the distribution function of the x_i. We have, for the mean values,

$$\mu_1 = \frac{1}{p}\int_0^p F^{-1}(w)\ dw,$$

$\mu_2 = \frac{1}{q-p}\int_p^q F^{-1}(w)\ dw$, and

$$\mu_3 = \frac{1}{1-q}\int_q^1 F^{-1}(w)\ dw,$$

and for the terms of the variance-covariance matrix :

$$\sigma_1^2 = \frac{1}{p}\int_0^p (F^{-1}(w))^2\ dw - \mu_1^2 + (\chi_p - \mu_1)^2\,(1-p),$$

$$\sigma_2^2 = \frac{1}{q-p}\int_p^q (F^{-1}(w))^2\ dw - \mu_2^2 + \frac{1}{q-p}\{p(1-p)(\mu_2 - F^{-1}(p))^2 +$$

$$+ q(1-q)\big(F^{-1}(q) - \mu_2\big)^2 + 2\,p(1-q)\big(\mu_2 - F^{-1}(p)\big)\big(F^{-1}(q) - \mu_2\big)\},$$

$\sigma_3^2 = \frac{1}{1-q}\int_q^1 (F^{-1}(w))^2\ dw - \mu_3^2 + q(\mu_3 - F^{-1}(q))^2$,

$\sigma_{12} = (F^{-1}(p) - \mu_1)\,[\mu_2 + \frac{1-q}{1-p}\,F^{-1}(q) - \frac{1-p}{q-p}\,F^{-1}(p)]$,

$\sigma_{13} = (\mu_3 - F^{-1}(q))\,(F^{-1}(p) - \mu_1)$, and

$\sigma_{23} = (\mu_3 - F^{-1}(q))(-\mu_2 + \frac{q}{q-p}F^{-1}(q) - \frac{p}{q-p}\,F^{-1}(p))$.

Notice that these values are expressed in the reduced distribution function.

Our purpose is to determine, first, coefficients (l_1, l_2, l_3) and (d_1, d_2, d_3) such that

$\lambda^* = \sum_1^3 l_i\ \bar{x}_i$ and $\delta^* = \sum_1^3 d_i\ \bar{x}_i$

are the least-squares estimators of (λ, δ). Denoting by $[\mu]^T = (\mu_1, \mu_2, \mu_3)$, $[1]^T = (1,1,1)$, $[\bar{x}]^T = (\bar{x}_1, \bar{x}_2, \bar{x}_3)$, and by Ω , as written before, the inverse of the variance-covariance matrix, with $\Gamma = \Omega([1]\,[\mu]^T - [\mu]\,[1]^T)\,\Omega/\Delta$, where $\Delta = [1]^T\,\Omega\,([1]\,.\,[\mu]^T\,\Omega\,[\mu] - [1]^T\Omega\,[\mu]^2)$, we have

$\lambda^* = -\,[\mu]^T\,\Gamma[\bar{x}]$,

$\delta^* = -\,[1]^T\,\Gamma[\bar{x}]$,

the coefficients of the linear combinations being, in obvious notation,

$[l]^T = -[\mu]^T\,\Gamma$ and $[d]^T = [1]^T\,\Gamma$.

The pair (λ^*, δ^*) is asymptotically binormal with mean values (λ, δ) and variance-covariance matrix

$$\frac{\delta^2}{n\,\Delta}\begin{bmatrix} [\mu]^T\,\Omega\,[\mu] & [1]^T\Omega\,[\mu] \\ [1]^T\,\Omega\,[\mu] & [1]^T\Omega\,[1] \end{bmatrix}$$

and the asymptotic efficiency relative to the Cramér-Rao bounds (corresponding to the ML estimators) is

$$\text{eff} = \Delta/\delta^4 \times \{M(\frac{\partial \log f}{\partial \lambda})^2\ .\ M(\frac{\partial \log f}{\partial \delta})^2 - M(\frac{\partial \log f}{\partial \lambda}\ \frac{\partial \log f}{\partial \delta})^2\}$$

where $f(\frac{x-\lambda}{\delta}) = \frac{d\,F\,((x-\lambda)/\delta)}{d\,x} = \frac{1}{\delta}\,F'\,(\frac{x-\lambda}{\delta})$.

Notice that the coefficients, when a shape parameter exists, should be written more correctly as $l_i(\alpha)$ and $d_i(\alpha)$.

The study of the maxima efficiency for the Fréchet distribution can be summarized in the following short table (the μ_i exist only for $\alpha_0 > 2$):

Table 9.3

	max. eff (%)	p	q
$\alpha_0 = 3$	84.3	.10	.82
$\alpha_0 = 6$	88.7	.11	.72
$\alpha_0 = 8$	90.9	.07	.45
$\alpha_0 = 9$	91.9	.07	.44
$\alpha_0 = 10$	92.6	.06	.41

Recall that the result for $\alpha_0 = 10$ is very similar to the one corresponding to Gumbel distribution, as known.

We can also, in the same way, study right, left and doubly censored samples. It corresponds, in the first and second case, to choosing three probabilities $0 < p < q < r < 1$ and discarding the right-block average $\bar{x}_4$ or the left-block average $\bar{x}_1$ (obvious notations), which is equivalent to taking $l_4 = d_4 = 0$ or $l_1 = d_1 = 0$. Clearly the triples $(\bar{x}_1, \bar{x}_2, \bar{x}_3)$ and $(\bar{x}_2, \bar{x}_3, \bar{x}_4)$ are asymptotically trinormal with a variance-covariance matrix which is the obvious sub-matrix of that for a 4-block partition, analogous to the one given. By the technique above we can obtain, when possible, the asymptotic relative efficiency of estimators (λ^*, δ^*) which are linear combinations of the averages, quasi-linearity invariant, i.e., for $\alpha_0 > 2$ in both the Fréchet and Weibull (for minima) cases and in Gumbel case, and optimize for $0 < p < q < r < 1$, and then compute the best coefficients that, as before, depend on α_0. Note that the best efficiency using one of the three averages of the simply censored sample can be better than the use of three averages in a non-censored sample; this seems a paradox but is easily explained. When using a simply censored sample we are partitioning it into four blocks and discarding the right or the left one: so comparison should be made with the efficiency of a 4-block partition (not a 3-block one!) and then the efficiency of the uncensored sample is better.

A doubly censored sample corresponds to choosing four probabilities $0 < p < q < r < s < 1$, and to discharging the extreme averages $\bar{x}_1$ and $\bar{x}_5$. The remaining triple $(\bar{x}_2, \bar{x}_3, \bar{x}_4)$ is asymptotically trinormal with a variance-covariance matrix which is the obvious sub-matrix of the one corresponding to the 5-block partition. The best choice of the coefficients of the linear combination can be made, as before, by seeking the probabilities that maximize efficiency and then computing the best coefficients, depending as always on α_0.

Let us, finally, consider the case when we have the three-parameters $(\lambda, \delta, \alpha > 2)$ unknown the Gumbel case is contained in the model for $\alpha = +\infty$ or, practically speaking, α large) and a non-censored sample with a 3-block partition.

We will not make all the computations but only sketch the method, along the lines of what was done previously.

Supposing α known, we obtained the best estimators of (λ, δ)

$$\lambda^* = \sum_1^3 l_i(\alpha)\, \bar{x}_i$$

$$\delta^* = \sum_1^3 d_i(\alpha)\, \bar{x}_i$$

in the beginning of the section.

The ratio statistic $\frac{\bar{x}_3 - \bar{x}_2}{\bar{x}_2 - \bar{x}_1} \to g(\alpha) = \frac{\mu_3 - \mu_2}{\mu_2 - \mu_1}$ is location-dispersion-free.

Then, by the δ-method, the ratio statistic is asymptotically normal with mean value $g(\alpha)$ and thus, by the same method, α^*, given by $g(\alpha^*) = \frac{\bar{x}_3 - \bar{x}_2}{\bar{x}_2 - \bar{x}_1}$, is asymptotically normal with variance $O(n^{-1})$. Also the triple $(\lambda^*, \delta^*, \alpha^*)$ is asymptotically normal with mean value $(\lambda, \delta, \alpha)$ and variance-covariance matrix $O(n^{-1})$. But, in fact, $\lambda^* = \sum_1^3 l_i(\alpha)\, \bar{x}_i$ and $\delta^* = \sum_1^3 d_i(\alpha)\, \bar{x}_i$, written above, depend on α and we should use $\lambda^{**} = \sum_1^3 l_i(\alpha^*) \bar{x}_i$, $\delta^{**} = \sum_1^3 d_i(\alpha^*)\, \bar{x}_i$, the triple $(\lambda^{**}, \delta^{**}, \alpha^*)$ being, also by the δ-method, asymptotically trinormal with mean value $(\lambda, \delta, \alpha)$ and variance-covariance matrix of $O(n^{-1})$. The new coefficients $l_i(\alpha^*), d_i(\alpha^*)$ are not the best ($\alpha > 2$ supposed known), although they can be expected to be close to the best ones as α^* should be close to $(\alpha > 2)$. But the asymptotic efficiency of the method drops down to values around 0.1 and 0.2 and so the method is poor, as could be expected. We could seek functions $\bar{l}_i(\alpha)$, $\bar{d}_i(\alpha)$ such that $\sum_1^3 \bar{l}_i((\alpha^*)\, \bar{x}_i, \sum_1^3 \bar{d}_i(\alpha^*)\, \bar{x}_i, \alpha^*)$ have maximum asymptotic efficiency. This is an open and important problem which is very practical and allowing a quick decision; recall that as we want quasi-linear invariance we must have $\sum_1^3 \bar{l}_i(\alpha) = 1$ and $\sum_1^3 \bar{d}_i(\alpha) = 0$. The extension of the previous reasonings to censored samples is immediate as well as to the unsolved problem.

9.6 Estimation using exceedances over thresholds

This section contains two different approaches, the peaks-over-threshold (P.O.T.) method and the all-excesses-over-threshold (A.E.O.T.) method.

The P.O.T. method can be connected with the notion of the return period, already given, the papers by Gumbel (1955) on calculated risk, Tiago de Oliveira (1984) on large earthquakes, and Rosbjerg and Knudsen (1984) on extreme sea states.

Let us fix a threshold u and consider in the sequence of random variables $\{X_i\}$ the *exceedances* over u, i.e., the random variables $X_j > u$. If the X_i has the distribution function $F(x)$, the exceedance $X_j(> u)$ has the distribution function $F(x|u) = \frac{F(x) - F(u)}{1 - F(u)} (u \leq x \leq \bar{\omega}_F)$; we will suppose $F(x)$ continuous.

In Chapter 2 we considered a sequence of thresholds (there called levels) such that $n(1 - F(u_n)) \to \tau$ as $n \to +\infty$, and sketched a brief way of obtaining the asymptotic distribution of maxima. In the same way, see Leadbetter, Lindgren and Rootzén (1983), we can prove that if $n(1 - F(u_n)) \to \tau$ as $n \to +\infty$ then

$$\text{Prob}\{\#\{X_i > u_n\} \leq k\} \to e^{-\tau} \sum_{s=0}^{k} \frac{\tau^s}{s!},$$

showing asymptotically the Poissonian character of excesses of large thresholds in i.i.d. case. The exceedance may be considered a large storm, large earthquake, large flood, etc.

We will *assume* that the large exceedances over a level x are i.i.d. with a Poisson distribution with mean value $\beta(1 - F(x|u))$ per cycle (e.g., 1 year), and so the mean value of exceedances of u per cycle is $v = \beta$ as $F(u|u) = 0$ and in t cycles βt. Thus the level x_T such that the mean value of exceedances in T cycles is 1 is given by $1 = \beta T(1 - F(x_T|u))$ or $F(x_T|u) = 1 - 1/\beta T$; x_T is evidently the *design value* for a return period of T (for the excess variables).

Also

$$\text{Prob}\{\max X_i \text{ in } t \text{ cycles} \leq x\}$$

$$= \text{Prob}\{\#\{X_i > x\} \text{ in } t \text{ cycles equal to } 0\} = e^{-\beta t(1 - F(x|u))}.$$

Let $x(L, R)$ be the level to be exceeded with probability R in L cycles. In the same way we have

$$1 - R = e^{-\beta L(1 - F(x(L,R)|u))}$$

or $F(x(L, R)|u) = 1 + \frac{\log(1-R)}{\beta L}$ which with $F(x_T|u) = 1 - \frac{1}{\beta T}$ leads to the non-parametric relation

$$1 - R = e^{-L/T}$$

which connects the design value x_T for one exceedance on average in T cycles to the level $x(L, R)$ that can be exceeded in cycles with probability R (the risk). Thus the two levels are connected and x_T can be used for design, with a new interpretation.

Notice that the hypothesis that we were dealing with exceedances over large thresholds was necessary to justify the use of the Poisson approximation.

In fact exceedances cluster in general, and this Poisson distribution applies to sequenced approximately independent clusters (with convenient β).

Let us now consider the peaks of each cluster and *assume*, as an example that has been useful, that $F(x)$ is the exponential distribution and so $F(x|u) = 1 - e^{-(x-u)/\delta}, x \geq u, \delta > 0$, independent of cluster distribution. Then we have $x_T = u + \delta \log(\beta T)$.

Thus if we have, in t_1 cycles, n_1 clusters above u we get

$$\hat{\beta} = n_1/t_1$$

and if peaks of the clusters are x_i we get

$$\hat{\delta} = \frac{1}{n_1} \sum_1^{n_1} (x_i - u) = \frac{1}{n_1} \sum_1^{n_1} x_i - u = \bar{x} - u\,.$$

Once more by the δ-method, as $\hat{\beta}$ is asymptotically normal with mean value β and variance β/t_1, $\hat{\delta}$ is asymptotically normal with mean value δ and variance δ^2/n_1 and zero covariance (independence), we get $\hat{x}_T = u + \hat{\delta} \log(\hat{\beta} T)$ asymptotically normal with mean value x_T and asymptotic variance $V(\hat{x}_T) \sim \frac{\delta^2}{\beta\, t_1} (1 + (\log(\beta\, T)))^2$.

Evidently the corresponding results can be obtained when using Gumbel and Fréchet distributions or distributions attracted to them.

We will finally deal with the A.E.O.T. method. We will use the papers by Smith (1984) and Davison (1984).

Let us assume an i.i.d. sample $\{X_i\}$ with distribution function $F(x)$, fix a threshold u, and select only the exceedances $X_i > u$. To a certain extent we are choosing, on average, the $n(1 - F(u))$ largest values, and so the results are analogous to those concerning $m = n(1 - F(u))$ largest values, dealt with in second section of this chapter. What is important, now, is that we are assuming that the largest values come from an i.i.d. sample which, even forgetting the seasonality, is not necessarily valid because large values tend to cluster and independence for large values is necessarily false. Note that in the P.O.T. method, what was assumed was the (practical) independence of clusters plus a stable distribution of the peaks of the

exceedances, which is much more reasonable. Because of this, the treatment will necessarily be sketchy.

Here we will use some preasymptotic form, the generalized Pareto distribution; note that the θ here corresponds to the θ of the von Mises-Jenkinson from $G(z|\theta)$ and is symmetric to the θ of the usual generalized Pareto distribution. The generalized Pareto distribution has the form

$$G(y/\delta|\theta) = 1 - (1 + \theta\, y/\delta)_+^{-1/\theta} \quad \text{if} \quad \theta \neq 0$$

$$= 1 - e^{-y+/\delta} \quad \text{if} \quad \theta = 0$$

where $\delta = \delta(u) > 0$ and $-\infty < \theta < +\infty$; also $0 \leq y < +\infty$ for $\theta \geq 0$ and $0 \leq y \leq -\delta/\theta$ for $\theta < 0$.

Let $\bar{F}(\Delta|u) = F(u + \Delta|u)$ for $0 \leq \Delta \leq \bar{w}_F - u$ be the distribution of the *excess* $X - u$ over the threshold u. Pickands (1975) proved that $F(.)$ is attracted to $G(.|\theta)$ iff

$$\lim_{u \to \bar{w}_F} \inf_{0<\delta<+\infty} \sup_{\Delta} |\bar{F}(\Delta|u) - G(\Delta/\delta(u)|\theta)| = 0,$$

and so the generalized Pareto distribution $G(\Delta/\delta(u)|\theta)$ is an approximation to $\bar{F}(\Delta|u)$, the distribution of the excess.

Then *assuming* all the excesses $X - u$ to be independent and that $\theta > -1/2$, we can obtain the ML estimators $(\hat{\delta}, \hat{\theta})$ of (δ, θ) which are asymptotically normal with mean values (δ, θ) and the variance-covariance matrix asymptotically

$$\frac{1}{m}\begin{bmatrix} 2\,\delta^2(1+\theta) & \delta(1+\theta) \\ \delta(1+\theta) & (1+\theta)^2 \end{bmatrix}$$

where m is the number of exceedances over u. We skip the results for $\theta \leq -1/2$.

Thus, once a high threshold u is fixed, *assuming* independence, the m exceedances have a binomial distribution with exceedance (or survival) probability $S(u) = 1 - F(u)$. Then with the observed excesses Δ assumed independent and with a generalized Pareto distribution we can conditionally on m, obtain the ML estimators $(\hat{\delta}, \hat{\theta})$. An approximation to the design value x_T is then given by $F(x_T)$ as $F(x_T) = F\big(u + (x_T - u)\big) - F(u) + F(u) \approx S(u)\, G(\frac{x_T - u}{\delta(u)}|\theta) + F(u) = 1 - 1/T$, as in the first part of the section, or equivalently with $x_T = u + \Delta_T$, with Δ_T is given by

$$G(\Delta_T/\delta(u)|\theta) = 1 - \frac{1}{S(u)\,T}$$

or $\Delta_T = \frac{(S(u)T)^\theta - 1}{\theta}\delta(u)$ and thus $x_T = u + \frac{(S(u)T)^\theta - 1}{\theta}\delta(u)$ if $\theta \neq 0$ and $x_T = u + \log(S(u)\,T)\,\delta(u)$ if $\theta = 0$; clearly $\hat{x}_T$ is asymptotically standard normal with mean value x_T and variance of $O(n^{-1})$. The last formula can be compared with $x_T = u + \log(\beta T)$ given previously for the exponential distribution, which corresponds to the Gumbel limiting distribution ($\theta = 0$); for large u, β will be in general smaller than 1 and thus comparable with $S(u)$.

This second part of the section needs a thorough revision owing to the hypothesis made.

For more details connected to tail estimation see Annex 4.

References

Boos, D. D., 1983. Using extreme value theory to estimate large percentiles. Techn. Rep., Dept. Statist., North Carolina State University.

David, H. A., 1981. *Order Statistics* (2nd ed), Wiley, New York.

Davison, Anthony C., 1984. Modelling excesses over high thresholds, with an application. *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 461-482, D. Reidel, Dordrecht.

Diaz, Jean –Pierre., 1985. Estimation et prédiction de valeurs extrêmes dans le domaine d'atraction de la loi de Fréchet. *C. R. Acad. Sc. Paris*, t. 301, ser. I, 541-544.

Gomes, M. I., 1981. An i-dimensional limiting distribution of largest values and its relevance to the statistical theory of extremes. *Statistical Distributions in Scientific Work*, 6, C. Taillie *et al* eds., 389-410, D. Reidel, Dordrecht.

Gomes, M. I., 1984. Concomitants in a multidimensional extreme model. *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 353-354, D.Reidel, Dordrecht.

Gumbel, E. J., 1955. The calculated risk in flood control. *Appl. Sc. Research* (Holland), 5, A, 273-280.

Gumbel, E. J., 1958. *Statistics of Extremes*, Columbia University Press, New York.

Hall, P., 1982. On estimating the end -point of a distribution. *Ann. Statist*., 10, 556-568.

Hill, B. M., 1975. A simple general approach to inference about the tail of a distribution. *Ann. Statist*., 3, 1163-1174.

Hüsler, J. and Schüpbach, M., 1986. On simple block estimators for the parameters of the extreme -value distribution. *Commun. Statist.-Simul. Comput*., 15, 61-76.

Hüsler, J. and Tiago de Oliveira, J., 1988. The usage of largest observations far parameter and quantile estimation for Gumbel distribution; an efficiency analysis. *Publ Inst. Statist. Univ. Paris*, 33 (1), 41-56.

Kubat, P. and Epstein, B., 1980. Estimation of quantiles of location scale distribution based on 2 or 3 order statistics. *Trab. Estad. Inv. Oper*., 22, 575-581.

Kubat, P., 1982. Simple large sample estimators of scale and location parameters based on blocks of order statistics. *Trab. Estad. Inv. Oper*., 33, 86 118.

Leadbetter, M. R., Lindgren, G. and Rootzen, H., 1983. Extremes and Related Properties. *Random Sequences and Processes*, Springer-Verlag, Heidelberg.

Neves, M., 1986. Estimacäo de quantis das distribuicöes de Gumbel e Fréchet com parâmetros de escala e de localizacäo desconhecidos, baseados em duas estatísticas ordinais. *Centro de Estatística e Aplicacöes*, (25/86), Lisboa.

Neves, M., 1988. Estimacäo por blocos dos parâmetros da distribuicäo de Fréchet. Universidade Nova de Lisboa, Tese de Doutoramento.

Pickands, J. III, 1975. Statistical inference using extreme order statistics. *Ann. Statis*t., 3, 119-131.

Reiss, R. D., 1987. Estimating the tail index of a claim size distribution. *Bläter Deutschen Gesllsch. Versicher. Math*., 18, 21-25.

Rosbjerg, D. and Knudsen, J., 1984. POT-estimation of extreme sea states and the benefit of using wind data. *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 611-620, D. Reidel, Dordrecht.

Smith, R. L. and Weismann, I., 1985. Maximum likelihood estimation of the lower tail of a probability distribution. *J. Roy. Statist. Soc.*, 47, 285-298.

Smith, R. L., 1984. Threshold methods for sample extremes. *Statistical Extremes Applications*, J. Tiago de Oliveira ed., 621-638, D. Reidel, Dordrecht.

Smith, R. L., 1986. Extreme value theory based on the R largest annual events. *J. Hydrology*, 86, 27-43.

Tiago de Oliveira, J, 1982. Efficient estimation for quantiles of Weibull distributions. *Rev. Beige Statist. Rech. Oper*., 22, 3-10.

Tiago de Oliveira, J, 1984. Weibull distributions and large earthquake modeling. *Probabilistic Methods in the Mechanics of Solids and Structures* (TUTAM Symposium in honour of Dr. W. Weibull), 81-89, S. Eggwertz and N.C. Lind eds., Springer –Verlag, Heidelberg.

Tiago de Oliveira, J., 1972. Statistics for Gumbel and Fréchet distributions. *Structural Safety and Reliability*, A. M. Freudenthal ed., 91-105.

Weissman, I., 1978. Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc*., 73, 812-815.

Weissman, I., 1981. Confidence intervals for the threshold parameter-II: unknown shape parameters. *Commun. Statist.- Theor. Meth.*, 11, 2461-2467.

Weissman, I., 1981b. Confidence intervals for the threshold parameter, *Commun. Statist.- Theor. Meth.*, A10, 549-557.

Weissman, I., 1984. Statistical estimation in extreme value theory. *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 109-115, D. Reidel, Dordrecht.

*****