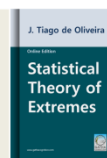




# Statistical Theory of Extremes

Homepage: <http://www.gathacognition.com/book/gcb14>  
<http://dx.doi.org/10.21523/gcb14>



## Part 2

Statistics for Univariate Extremes

## Annex 4

# On Extremes Values and Tail Estimation

**J. Tiago de Oliveira**

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

### Editor(s)

J.C. Tiago de Oliveira

### Published Online

23 June 2017

## 1 Introduction

The estimation of the tail of a distribution, using extreme order statistics, sometimes referred as a possibility, has been approached in different ways. We will describe the essential features of these, with some comments and a comparison between them as well as some remarks about some open questions; we have drawn heavily on the review paper by [Tiago de Oliveira \(1991\)](#).

As there is complete duality between the maxima and the minima, and so between the right and left tail, we will only analyse the right tail estimation using the largest order statistics.

Under this umbrella two different problems can be dealt with:

- a) Estimation of the large quantiles (for probabilities close to 1) and
- b) Estimation of the probability of overpassing some given level.

Originally published in 'Statistical Analysis of Extremes', 1997, 2016

<http://dx.doi.org/10.21523/gcb1.1715>

© 2017 GATHA COGNITION® All rights reserved.

As was shown by Einmahl (1990), the problem of right tail estimation is only interesting when we are beyond (or near) the largest value of the sample, thus constituting a *statistical extrapolation* and, from this perspective, some hypothesis needs to be made. For statistical extrapolation we must suppose that the distribution function of the i.i.d. sample lies in the domain of attraction of some asymptotic distribution of maxima. On the other hand, if we are interested only in the behaviour of the distribution in the interior of the support of the sample, i.e., the closed interval [minimum, maximum], the utilization of the usual sample distribution function is completely sufficient and for this *statistical interpolation* we do not need to add any hypothesis except the usual (and convenient one) of continuity; in fact, it is a classical non-parametric approach.

Let us recall that a distribution function  $F(z)$  is attracted, for maxima, to  $G(z|\theta)$ , noted  $F(.) \in \mathcal{D}(G(.|\theta))$ , if there exist attraction coefficients  $\{\lambda_n, \delta_n(> 0)\}$ , not uniquely defined, such that for an i.i.d. sample  $(x_1, \dots, x_n)$ , with the distribution function  $F(x)$ , we have:

$$P\{(\max(x_1, \dots, x_n) - \lambda_n)/\delta_n \leq z\} = P\{\max_{1 \leq i \leq n} (x_i) \leq \lambda_n + \delta_n z\} = F^n(\lambda_n + \delta_n z) \rightarrow G(z|\theta)$$

as  $n \rightarrow \infty$ ; as the  $G(z|\theta)$  are continuous the convergence is uniform. It should be noticed that if  $F(.) \in \mathcal{D}(\Phi_\alpha)$  we have  $\bar{w} = +\infty$  and we can take  $\lambda_n = 0$  and if  $F \in \mathcal{D}(\Psi_\alpha)$  we have  $\bar{w} < +\infty$  and we can take  $\lambda_n = \bar{w}$ .

The necessary and sufficient conditions for attraction to one of the asymptotic distributions of maxima have been given before, the most practical being the one to be recalled in the description of the third approach although some others will appear in the other two approaches.

For some time the limiting relation  $F^n(\lambda_n + \delta_n z) \rightarrow G(z|\theta)$  as  $n \rightarrow \infty$  led to the approximation  $F^n(\lambda_n + \delta_n z) \approx G(z|\theta)$  or  $F^n(y) \approx G((x - \lambda)/\delta|\theta)$  where the index  $n$  was omitted because  $\theta, \lambda$  and  $\delta$  were to be estimated from a sample of maxima. This way gave an approximation to  $F^n(y)$ , and sometimes  $G^{1/n}((y - \lambda)/\delta|\theta)$ , equal to  $G(y - \lambda')/\delta'|\theta)$  by stability, was used as a (very) rough approximation to  $F(y)$ . Thus the idea of using the largest observations

of a sample to estimate  $F(y)$  came naturally. As a side remark, it should be noted that, for instance, if  $F^n(\lambda_n + \delta_n z) \rightarrow \Lambda(z)$  as  $n \rightarrow \infty$ , for  $n$  not sufficiently large, a better approximation to  $F^n(y)$  can be obtained using, instead of  $G((y - \lambda)/\delta|0)) = \Lambda((y - \lambda)/\delta)$ , a penultimate distribution function  $G((x - \lambda)/\delta|\theta)$  [with  $\lambda$ ,  $\delta$  and  $\theta$  to be estimated from the sample of i.i.d. maxima]. Such is the case for the normal distribution when the criterion used for comparison of the asymptotic approximation is the maximum absolute error in the probability evaluation, i.e., when we try to minimize  $\sup_y |N^n(\lambda_n + \delta_n y) - G(y|\theta)|$  where we have a sequence of  $\theta = \theta_n \uparrow 0$ . The global approach used here integrates this penultimate approximation naturally by *estimating*, from the sample of i.i.d. maxima, the  $\theta$ ,  $\lambda$  and  $\delta$  that fits the data better.

Before continuing an important remark must be made. When we are dealing with maxima *only*, in many cases, we can split the sequence of observations (the concrete time-series) into natural blocks for which the maxima are surely independent, thus having a sample of independent maxima; this was the case for various examples in this Part 2. These maxima, as was said for stationary sequences, for markovian sequences, for  $m$ -dependent sequences, for strongly mixing sequences and even for some evolutionary sequences, as will be shown in Part 4, have, under some more strong or mild conditions on association (many times, on the correlation), the same asymptotic distribution for maxima; the same happens for analogous continuous time stochastic processes for which we sometimes have records, many times discretized for the statistical analysis. The natural block used for hydrological data is the natural hydrological year, but for oceanographical and meteorological data the block used, in general, is the civil (!) year. This is the well-known “yearly method”. But if we use the largest values of the time-series it is obvious that the observations are not independent, not even stationary, because — even for an additive model — they have underlying the effects of trend and of cyclic variation, which have to be removed giving many times the largest values at *other* time points. Even when assumed to be stationary sequences, large (and also small) values occur by clusters and so in general the 2nd, 3rd, ... maximum occur at time points close to that of the (1st) maximum, as the association (correlation) is very strong; this drawback can be avoided by “cleaning” the time series from the cluster effect, which means

that once the 1st maximum at instant  $t$  is obtained, data should be “cleaned” from observations obtained in the instants (time steps)  $t - m$  to  $t + m$ , and the 2nd maximum sought in the remaining data, etc.; *this is the natural procedure if we are dealing with a stationary  $m$ -dependent sequence*. For other situations something similar must be found. Another difficulty appears when we are using the largest values in a time-series split in the “wrong” blocks in which some number of largest values are to be used; if the block is, for instance, a civil year (and not a natural one like the hydrologic year), it may happen that a storm occurs by the end/beginning of the year and the maxima registered for the two successive years correspond to observations made at 9 p.m. on 31st December and to 3 a.m. on the following 1st January, which clearly, belong to the same cluster against the “cleaning” procedure sketched above, thus giving rise to errors in planning and design.

The “cleaning” procedure is thus essential to avoid errors; a good example of the “independentization” of observations in a time series was done by [Dijk and de Haan \(1990\)](#); the behaviour of the auto-correlation coefficient is sometimes a good tool to steer the procedure of choosing  $m$ .

## 2 The three approaches to tail estimation

We will now consider a large sample of i.i.d. observations, which is supposed to have a distribution function  $F(\cdot)$  such that  $F(\cdot) \in \mathcal{D}(G(\cdot|\theta))$ . This means, *for practical applications*, that the removal of the trend and cyclic variation and the cluster “cleaning” have been made. So far, it seems that only two situations exist where independence can be assumed directly: largest ages of death for men, for women or, for both, in some country if, in the years considered, no war, epidemic, national disaster or famine struck the region under study; and large fires data (with compensation for inflation) in some region if we discard fires with “epidemics” like forest fires, fires from bombing, and large fires in built-up areas. In these cases the “pooling” of data is natural, giving rise to a (very) large sample from which a relatively large set of largest values can be used to estimate tail properties. Possibly analogous cases exist but the conditioning to *assume* independence is very clear. Once the estimator of  $F(\cdot)$  is obtained, we can estimate either the desired *exceedance probability* or the *quantile*.

The first approach is due to [Pickands \(1975\)](#) and has been used with variations in [Smith \(1987\)](#), [Smith and Weismann \(1985\)](#), [Smith \(1989\)](#) and [Pickands \(1989\)](#). Basically it consists of the approximation of the asymptotics of the distribution to the *excesses over a threshold* by a generalized Pareto distribution which was obtained in the important [Pickands \(1975\)](#) paper. The method, then, corresponds to the P.O.T. (peaks over threshold) method sometimes used by engineers, against the yearly method.

The second approach is based on [Weismann \(1978\)](#) where the joint asymptotic distribution of a fixed number of the largest order statistics is used.

Both approaches have been used *as if* the *asymptotic or preasymptotic* distribution was the *actual* one, evidently raising the already classical criticism to the usage of  $G((y - \lambda)/\delta|\theta)$ , with  $\lambda$ ,  $\delta$  and  $\theta$ , estimated, for the actual distribution  $F^n(y)$  of the maximum of  $n$  observations when, in fact, both methods are approximations; note that the use of penultimate distributions referred to, is one way to compensate by using an approximation to the distribution of independent maxima. A guide to this approach was given by [Tiago de Oliveira \(1981\)](#) and [\(1984\)](#) where the theory of statistical choice for  $\theta < 0$  (Weibull),  $\theta = 0$  (Gumbel) or  $\theta > 0$  (Fréchet) was developed; this statistical choice can give a better way of approaching the description of the statistical behaviour of independent maxima, even when generated by a dependence process, as happens in the yearly method.

A third and new approach has been developed essentially by [Dekkers and de Haan \(1989\)](#), [Dekkers, Einmahl and de Haan \(1989\)](#) and [Dijk and de Haan \(1990\)](#); it is connected with some previous results for the estimation of the shape parameter for Fréchet distribution, such as the Hill estimator.

This approach, although steered by the asymptotic distribution of maxima, *does not* use it as the *actual* distribution, in contrast to the previous ones that, at some point, substitute the actual (unknown) distribution by its asymptotic approximation.

Notice that the initial assumption that all observations are i.i.d. and their distribution function is attracted to some asymptotic distribution of extremes  $G(z|\theta)$  leads to the use of upper order statistics — sometimes called extremal, sometimes intermediate, see [Reiss \(1989\)](#) — which are evidently

dependent; independence thus is at the beginning but *not* in the observations chosen for the statistical analysis.

The asymptotic normality of [Pickands' \(1975\)](#) estimator of  $\theta$  has been proved in various papers, such as that of [Dekkers and de Haan \(1989\)](#).

Complementary results can be found in [Tiago de Oliveira ed., \(1984\)](#), especially the papers by [Weismann \(1984\)](#), [Smith \(1984\)](#) and [Davison \(1984\)](#), and also in [Davison and Smith \(1990\)](#).

This Annex applies results of [Chapters 3 and 9](#) for the estimation of large quantiles and small exceedance probabilities; but it must be remarked that optimal procedures for the choice of the number and the orders of large quantiles, the efficiency evaluation, the speed of convergence, etc., are yet open problems.

### 3 The first approach

Let  $(x_1, \dots, x_n)$  be a sample of i.i.d. random variables (observations) with distribution function  $F(x)$ , attracted to some  $G(z|\theta)$  and let  $\bar{w} = \inf \{x|F(x) = 1\}$  be the right-end point of  $F(\cdot)$ . As we are going to use the upper order statistics, let us recall that  $x_k''$  denotes the  $k$ -th maximum, i.e., the ascending order statistics  $x'_{n+1-k} = x_{n+1-k:n}$ ; the maximum is obviously  $x_1''$  and the minimum is  $x_n''$ .

Recall that [Pickands \(1975\)](#) gave the name of standard *generalized Pareto distribution* to the distribution function  $P(x|\theta) = 1 - (1 + \theta x)_+^{-1/\theta}$  where  $P(x|0)$ , defined as  $P(x|0^+) = P(x|0^-) = 1 - e^{-x}$ , is the standard exponential distribution.

If  $\theta \geq 0$  we have for  $P(x|\theta)$  the right-end point  $\bar{w} = +\infty$  but for  $\theta < 0$  the right-end point is  $\bar{w} = -1/\theta$ ; in general  $\bar{w} = 1/(-\theta)_+$  for the standard generalized Pareto distribution.

Evidently if  $X$  has the distribution function  $F(x)$ , the conditional survival function of the excess  $X - u$ , when  $X \geq u$ , is given by

$$S(x|u) = P(X > u + x | X > u) = \frac{1 - F(u + x)}{1 - F(u)} \quad \text{for } x \geq 0$$

and the conditional distribution function of the excess is evidently

$$F(x|u) = 1 - S(x|u) = \frac{F(x+u) - F(u)}{1 - F(u)} \quad \text{for } x \geq 0.$$

Pickands (1975) proved the following result:

$$\lim_{u \uparrow w} |F(x|u) - P(x|\delta(u)|\theta)| = \lim_{u \uparrow w} |S(x|u) - (1 + \theta x|\delta(u))_+^{-1/\theta}| = 0$$

iff  $F(\cdot) \in \mathcal{D}(G(\cdot|\theta))$ .

Notice that the scale parameter  $\delta(u)$  depends on  $u$ .

Also in the same paper it was proved that:

If  $m_n$  is an integer such that  $1 \leq 4m_n \leq n$  and  $m_n \rightarrow \infty$  but  $m_n/n \rightarrow 0$

$$\text{then } \theta^* = \log\left(\frac{x_{m_n}'' - x_{2m_n}''}{x_{2m_n}'' - x_{4m_n}''}\right) / \log 2$$

and  $\delta^* = \theta^*(x_{2m_n}'' - x_{4m_n}'') / (2^{\theta^*} - 1)$  are such that

$$\sup_x |F(x|x_{4m_n}'') - P(x|\delta^*|\theta^*)| \xrightarrow{a.s.} 0 \quad \text{and also that } \theta^* \xrightarrow{P} \theta.$$

This theorem shows that for the (large) random threshold  $x_{4m_n}''$  the conditional distribution of the random excess  $x - x_{4m_n}''$  is asymptotically approximated by a Pareto generalized distribution.

A rule for the choice of  $m_n$  is also given.

Later Dekkers and de Haan (1989) proved the asymptotically normality of  $\theta^*$  (see the third approach, first result, of this Annex).

Smith (1987) (with a change of sign in  $\theta$ ) used Pickands (1975) result with either a moving or a random threshold  $u_n$ . Evidently, conditional on a given  $u$ , the excesses  $x_i - u$  are independent random variables and its number (the number of exceedances of  $u$ ) is random (a binomial variate if  $u$  is not random). Thus  $u$  should not be neither too large to have few exceedances or too small to have almost all the sample giving rise to exceedances. Then the

excesses are taken as having, as *actual* distribution, a generalized Pareto one, and estimation is made by the method of maximum likelihood which is only regular if  $\theta > -1/2$ , with an extension if  $\theta \leq -1/2$ . Let us denote by  $(\hat{\theta}, \hat{\delta})$  the maximum likelihood estimators. Then [Smith \(1987\)](#), assuming that the excesses have an *actual* generalized Pareto distribution, shows, with  $N$  denoting the random number of excesses, that:

$\sqrt{N}(\hat{\delta}|\delta - 1, \hat{\theta} - \theta)$  is asymptotically binormal with zero mean values and variance-covariance matrix  $V = (1 + \theta) \begin{bmatrix} 2 & 1 \\ 1 & 1 + \theta \end{bmatrix}$  which is the Cramér-Rao bound.

It is then shown that if the excesses do not have the actual generalized Pareto distribution but

$$F(x|u) - P(x/\delta(u)|\theta) = O(1/\sqrt{N}) \quad \text{then}$$

$\sqrt{N}(\hat{\delta}|\delta - 1, \hat{\theta} - \theta)$  is asymptotically binormal with mean values not necessarily zero but with the same variance-covariance matrix  $V$ .

The non-zero mean values allow for the existence of a bias in the approximation; but knowledge of this bias depends on the structure and the form of the tail  $1 - F(x)$ . The study is made for each form of the asymptotic distribution (Weibull, Gumbel or Fréchet) in particular, with additional conditions.

The distribution function for values  $y = u + x, x \geq 0$ , naturally using  $1 - \hat{F}(u) = N/n$ , is then estimated by  $1 - \hat{F}(u + x) = \frac{N}{n} (1 + \hat{\theta} x/\hat{\delta})_+^{-1/\hat{\theta}}$  or  $\hat{F}(y) = 1 - \frac{N}{n} (1 + ((y - u)/\hat{\delta})_+^{-1/\hat{\theta}})$ , for  $y \geq u$  which is asymptotically normal by the use of the  $\delta$ -method; see [Tiago de Oliveira \(1982\)](#).

The quantile for probability  $p = 1 - \varepsilon$  is estimated by solving back the equation  $\hat{F}(y) = 1 - \varepsilon$  and so is given by  $\hat{y}(1 - \varepsilon) = u + \hat{\delta}[(\frac{n\varepsilon}{N})^{-\hat{\theta}} - 1]$  which is also asymptotically normal by the  $\delta$ -method.



The choice of the threshold  $u_n$  (random or moving) and the way that  $u_n \uparrow \bar{w}$  is partially discussed, distribution by distribution, using some additional assumptions on the type of variation of the tail  $1 - F(y)$ .

Here are some examples:

If  $F \in \mathcal{D}(\Phi_\alpha)$ , that is *iff*  $\bar{w} = +\infty$  and  $L(x) = x^\alpha(1 - F(x))$  is of *slow variation* (i.e.,  $\frac{L(xt)}{L(x)} \rightarrow 1$  when  $x \rightarrow \infty$ ) such that  $L(tx)/L(x) = 1 + k(t)\phi(x) + o(\phi(x))$ , where  $k(t) = \frac{c}{\rho}(t^\rho - 1)$  ( $\rho < 0$ ), then the previous asymptotic result is valid with the mean values  $(\frac{\mu(1+\theta)(1-2\theta\rho)}{1+\theta-\theta\rho}, \frac{-\mu(1+\theta)\theta(1+\rho)}{1+\theta-\theta\rho})$  and variance-covariance matrix  $V$  given; here we denote by  $\mu = p \lim \frac{\sqrt{N} C \Phi(uN)}{\alpha - \rho}$ . Recall that  $\theta = 1/\alpha$  and  $\delta(u) = u/\alpha$ .

The Hill estimator for  $\alpha$  is given by

$$\alpha_n^* = \frac{N}{\sum_1^N \log x_i'' - N \log u}$$

and the asymptotic ratio of  $\frac{MSE(\hat{\alpha}_n)}{MSE(\alpha_n^*)}$  is

$$(\alpha + 1)^2 \left( \frac{1+\rho}{\alpha+1-\rho} \right)^{2\alpha/(\alpha-2\rho)},$$

which shows, that according to the value of  $\rho$ , one or other of the estimators is better.

The choice of a deterministic sequence  $\{u_n\}$  must be such that  $n(1 - F(u_n)) \rightarrow \infty, \sqrt{n(1 - F(u_n))} C \Phi(u_n) \rightarrow \mu(\alpha - \rho)$ , a result of which is of doubtful usefulness because  $F(x)$  is generally unknown, as well as  $\mu$  and  $\alpha$ ; in the case where  $u$  is a random threshold we choose a sequence of  $N_n$  such that  $N_n \rightarrow \infty, N_n/n \xrightarrow{P} 0$  and use  $u_n = x_{N_n+1}''$ .

The case for  $\Psi_\alpha$ , when  $\alpha > 2$ , has some similarities with the previous one. The necessary and sufficient condition to be attracted to  $\Psi_\alpha(x)$  is that  $\bar{w} < +\infty$  and  $L(x) = x^\alpha(1 - F(\bar{w} - 1/x))$  is also of slow variation.

Assuming the same condition as before, with  $\rho < 0$ , and defining  $\mu$  by  $\mu = \lim \sqrt{N} C(\phi((\bar{w} - u_n)^{-1})/(\alpha - \rho)$ , we see that it is also valid with the mean values

$$\left( \frac{\mu(1 + \theta)(1 + 2\theta\rho)}{1 + \theta + \theta\rho}, \frac{-\mu\theta(1 + \theta)(1 - \rho)}{1 + \theta + \theta\rho} \right)$$

and also the same variance-covariance matrix  $V$ . We have, then,  $\theta = -1/\alpha$  and  $\delta = (\bar{w} - u)/\alpha$ . This result is essentially due to [Smith and Weismann \(1985\)](#).

The estimator of the right-end point  $\bar{w}$  is given by  $\hat{w} = u_n + \hat{\alpha} \hat{\delta} = u_n - \hat{\delta}/\hat{\theta}$  which is also asymptotically normal. The random threshold  $u_n = x_{N_n+1}''$  must be such that  $\bar{w} - u_n = O_p((N/n)^{1/\alpha}) = O_p((n/N)^\theta)$ , but  $\theta$  is unknown except for the bounds  $-1/2 < \theta < 0$ .

The cases for  $\alpha \in ]0, 2]$  are analogous but more complex.

Now consider the asymptotic distribution  $\Lambda$ . Although three approaches are presented in [Smith \(1987\)](#), we will describe only the second one because it corresponds to the P.O.T. methods used by engineers, which directly assume  $\theta = 0$ . The estimator of  $\delta$  is then  $\hat{\delta} = \frac{1}{N} \sum_1^N \log x_i'' - u_n$  where the  $x_i''$  are the order statistics larger than  $u_n$ . Then for the estimator of the distribution function

$$\hat{F}(y) = 1 - \frac{N}{n} \exp\left(\frac{y - u_n}{\hat{\delta}}\right)$$

it can be shown that

$$\sqrt{N} \left( \frac{1 - \hat{F}(u_n + z \phi(u_n))}{1 - F(u_n + z \phi(u_n))} - 1 \right)$$

is asymptotically normal with mean value  $\mu z - \frac{\mu z^2}{2}$  and variance  $1 + z^2$ , where we define  $\mu$  by  $\mu = P \lim \sqrt{N} \phi(u_n)/n$  and  $\phi(t)$  is such that

$$1 - F(x) = c(x) \exp\left(-\int_{-\infty}^x \frac{dt}{\phi(t)}\right), x < \bar{w}$$

with  $c(x) \rightarrow 1$  as  $x \rightarrow \bar{w}$ ,  $\phi$ , is a positive differentiable function, and  $\phi'(x) \rightarrow 0$  as  $x \rightarrow \bar{w}$ .

Notice that, in this result, we have an asymptotic evaluation of the relative error of the approximation by the exponential distribution. The fact that the result, once more, depends on knowledge (at least global) of the initial  $F$  leads to difficulties.

#### 4 The second approach

This approach stems essentially from [Weissmann's \(1978\)](#) paper and assumes that the number of upper order statistics used is fixed, although the sample size increases indefinitely. The sample  $(x_1, \dots, x_n)$  is assumed, once more, to be i.i.d. with a distribution function which is in the domain of attraction of one of the three asymptotic distributions supposed known, analogously to what happened before.

Let us then suppose that  $F \in \mathcal{D}(\Lambda)$ , i.e., that  $F^n(\lambda_n + \delta_n x) \rightarrow \Lambda(x)$ . Denoting by  $y_{i,n} = (x_i'' - \lambda_n)/\delta_n = y_i$  for short, it can be shown that:

*The vector  $(y_1, \dots, y_m)$ , if  $F \in \mathcal{D}(\Lambda)$  with attraction coefficients  $\lambda_n$  and  $\delta_n (> 0)$ , has the limiting density  $\exp(-e^{-y^m} - \sum_1^m y_i)$  with  $y_1 \geq y_2 \geq \dots \geq y_m$ .*

Thus we can approximate the actual distribution of the  $x_i''$  (or  $y_i$ ) ( $i = 1, 2, \dots, m$ ) by the asymptotic one, then having the likelihood

$$\frac{1}{\delta^m} \exp\left(e^{-\frac{x_m'' - \lambda}{\delta}} - \sum_1^m \frac{x_i'' - \lambda}{\delta}\right).$$

The random pair  $(\sum_1^m x_i'', x_m'')$  is a sufficient statistic for  $(\lambda, \delta)$  and the maximum likelihood estimators  $(\hat{\lambda}, \hat{\delta})$  and minimum variance unbiased estimators  $(\lambda^*, \delta^*)$  of  $(\lambda, \delta)$  are

$$\hat{\delta} = \frac{1}{m} \sum_1^m x_i'' - x_m'', \hat{\lambda} = x_m'' - \log m \cdot \hat{\delta}$$

and

$$\delta^* = \frac{1}{m-1} \sum_1^{m-1} x_i'' - x_m'' = \frac{m}{m-1} \delta^*, \lambda^* = x_m'' + (S_m - \gamma) \delta^*$$

with  $S_m = \sum_1^{m-1} \frac{1}{i}$ . We have  $V(\delta^*) = \frac{\delta^2}{m-1}$  and

$$V(\delta^*) = \left[ \frac{(S_m - \gamma)^2}{m-1} + S_m' \right] \delta^2 \text{ where } S_m' = \frac{\pi^2}{6} - \sum_1^{m-1} \frac{1}{i^2}.$$

By a result from [Mejzler \(1949\)](#), also given by [de Haan \(1971\)](#), we have, for the quantile  $F^{-1}(1 - c/n)$ , the basic relation:  $(F^{-1}(1 - c/n) - \lambda_n)/\delta_n \rightarrow \log c$  which shows that for  $\lambda_n$  we can take  $F^{-1}(1 - 1/n)$ , and for  $\delta_n = F^{-1}(1 - e/n) - F^{-1}(1 - 1/n)$ , as known since [Gnedenko's \(1943\)](#) paper. Thus

*The quantile  $F^{-1}(1 - c/n)$ , if  $F \in \mathcal{D}(\Lambda)$ , can be estimated by  $\hat{F}^{-1}(1 - c/n) = \hat{\lambda}_n + \hat{\delta}_n \log c$  or by  $F^{*-1}(1 - c/n) = \lambda_n^* + \delta_n^* \log c$ .*

In the case where  $F \in \mathcal{D}(\Phi_\alpha)$ , as previously stated  $\bar{w} = +\infty$  and we can take  $\lambda_n = 0$  and as  $\Phi_\alpha(z) = \Lambda(\alpha \log z)$  ( $z > 0$ ),  $\log X$  ( $X > 0$ ) is attracted to  $\Lambda$ . Reducing to the previous case and afterwards returning to the observed values we have

*If  $F \in \mathcal{D}(\Phi_\alpha)$ , ( $\alpha$  unknown) we have*

$$\hat{\alpha} = \left( \frac{1}{m} \sum_1^m \log x_i'' - \log x_m'' \right)^{-1} \text{ and } \hat{F}^{-1}(1 - c/n) = (m/c)^{1/\hat{\alpha}} x_m''.$$

For the case  $F \in \mathcal{D}(\Psi_\alpha)$  we know that  $\bar{w} < +\infty$  and  $\Psi_\alpha(z) = \Lambda(-\alpha \log(-z))$  for  $z < 0$  and then

*If  $\bar{w}$  is known and  $F \in \mathcal{D}(\Psi_\alpha)$ , then we have*

$$\hat{\alpha} = \left( \log(\bar{w} - x_m'') - \frac{1}{m} \sum_1^m (\log \bar{w} - x_i'') \right)^{-1} \text{ and}$$

$$\hat{F}^{-1}(1 - c/n) = \bar{w} - \left(\frac{c}{n}\right)^{1/\hat{\alpha}} x_m''(\bar{w} - x_m'').$$

The case where  $\bar{w}$  is unknown was not considered.

## 5 The third approach

As stated previously, the third approach is different from the previous ones, in that it does not use the approximation given by an asymptotic or preasymptotic distribution as the *actual* one, as was done in the first and second approaches in one way or another.

The essential result for large quantile estimation, is based in the attraction conditions using the quantile function  $Q(v) = F^{-1}(v)$ , the inverse function  $F^{-1}(v) = \inf\{x | F(x) \geq v\}$ , given by [Mejzler \(1949\)](#) and [de Hann \(1970\)](#). We have the basic result:

$$F(.) \in \mathcal{D}(G(.|\theta)) \text{ iff } \frac{Q(1-tx)-Q(t)}{Q(1-ty)-Q(t)} \rightarrow \frac{x^{-\theta}-1}{y^{-\theta}-1}$$

$$\left(\frac{\log x}{\log y} \text{ if } \theta = 0\right) \text{ locally uniformly when } t \rightarrow 0.$$

Let us denote by  $F_n(x)$  the sample distribution function and by  $Q_n(v) = F_n^{-1}(v)$  the sample quantile function, which are both in correspondence with  $F(x)$  and  $Q(v)$ . It is immediate that  $Q_n(1 - m_n/n) = x_{m_n}''$ . Thus if we are seeking the  $p$ -quantile of  $F(x)$ , i.e., the solution of  $F(x) = p = 1 - \varepsilon$ , this is equivalent to seeking  $Q(1 - \varepsilon)$ . Evidently if we want to design a dam with a probability  $\varepsilon'$  of being overpassed for  $n'$  units of time under the i.i.d. hypothesis, we have to solve the equation  $F^{n'}(x) = 1 - \varepsilon'$  or  $F(x) = (1 - \varepsilon')^{1/n'}$ , that is use  $Q(1 - \varepsilon)$  where  $\varepsilon = 1 - (1 - \varepsilon')^{1/n'}$  is much closer to zero as  $\varepsilon'$  is small and  $n'$  large.

Our purpose is thus to estimate  $Q(1 - \varepsilon)$  from a sample of  $n$  i.i.d. observations whose distribution function is attracted to  $G(.|\theta)$  using the some largest order statistics.

As we can write

$$Q(1 - \varepsilon) = \frac{Q(1-\varepsilon) - Q(1 - \frac{m_n}{n})}{Q(1 - \frac{m_n}{n}) - Q(1 - \frac{2m_n}{n})} [Q(1 - \frac{m_n}{n}) - Q(1 - 2\frac{m_n}{n})] + Q(1 - \frac{m_n}{n})$$

replacing  $Q$  by  $Q_n$  where convenient, as  $\frac{m_n}{n} \rightarrow 0$  by, hypothesis, and, substituting the first quotient by its limit given by the attraction conditions, as  $\varepsilon = \frac{n}{m_n} \frac{m_n \varepsilon}{n}$ ,

$$Q_n^*(1 - \varepsilon) = \frac{(m_n/\varepsilon n)^{\theta_n^*} - 1}{1 - 2^{-\theta_n^*}} [x_{m_n}'' - x_{2m_n}''] + x_{m_n}''$$

where  $\theta_n^*$  is the estimate of  $\theta$  given by [Pickands \(1975\)](#).

$$Q_n^* = \log \frac{x_{m_n}'' - x_{2m_n}''}{x_{2m_n}'' - x_{m_n}''} / \log 2.$$

Only three largest order statistics (intermediate as  $m_n \rightarrow \infty$  also, but at a slower rate) are sufficient to deal directly with the estimation of the quantile of probability  $1 - \varepsilon$ .

Then [Dekkers and de Haan \(1989\)](#) show the following results:

*If  $Q(t)$  has a positive derivative and a positive function  $b(t)$  exists such that*

$$t^{1+\theta} \frac{x^{1+\theta} Q(1-tx) - Q(1-t)}{b(t)} \rightarrow \pm \log x \text{ as } t \rightarrow 0$$

*then  $F(\cdot) \in \mathcal{D}(G(\cdot|\theta))$  and  $\sqrt{m_n}(Q_n^* - \theta)$  is asymptotically normal with mean value zero and variance  $\theta^2(2^{2\theta+1} + 1)/(2(2^\theta - 1) \log 2)^2$  for sequences  $m_n \rightarrow \infty$  such that  $m_n = o(n/g^{-1}(1/n))$  with  $g(t) = t^{1+2\theta} Q'(1-t)^2/b^2(t)$  (when  $\theta = 0$  the asymptotic variance is  $\frac{3}{4(\log 2)^4}$  as follows from (Hôpital's rule).*

This allows us to conclude the consistency of  $\theta_n^*$ . As we are dealing with intermediate order statistics that are asymptotically normal — see [Reiss \(1989\)](#) —  $Q_n^*(1 - \varepsilon)$  is expressed in them and also in  $\theta_n^*$ , all asymptotically

normal, by the  $\delta$ -method we can expect that  $Q_n^*(1 - \varepsilon)$  is asymptotically normal. In fact we have

*If  $Q'(t)$  exists and  $t^{1+\theta} Q'(1 - t)$  is of slow variation when  $t \downarrow 0$  then*

$$\sqrt{2m_n} \frac{x_{m_n}'' - Q(1 - \varepsilon_n)}{x_{m_n}'' - x_{2m_n}''}$$

*is asymptotically normal with mean value zero and variance  $\theta^2 2^{2\theta+1}/(2^\theta - 1)^2$  if  $\varepsilon_n \rightarrow 0$  and  $m_n = [n \varepsilon_n]$  (when  $\theta = 0$  we have for the variance  $2/(\log 2)^2$ ).*

This results allows us as before to obtain confidence intervals for the quantile of probability  $1 - \varepsilon_n$ .

In Dekkers, Einmahl and de Haan (1989) there was introduced a general estimator of  $\theta$  also asymptotically normal that can be used for all values of  $\theta$ , an estimator which is a modification of Hill's estimator only applicable for  $\theta > 0$ . Denoting by

$$M_n = \frac{1}{m_n} \sum_1^{m_n} (\log x_i'' - \log x_{m_n}'')^2,$$

it is shown that

*If  $Q'(t)$  exists and  $t^{1+\theta} Q'(1 - t)$  is of slow variation when  $t \downarrow 0$  then*

$$\sqrt{m_n} \frac{x_{m_n}'' - Q(1 - \varepsilon_n)}{x_{m_n}'' \cdot M_n}$$

*is asymptotically normal, with mean value zero and variance  $(1 - \min(0, \theta))^2$  if  $\varepsilon_n \rightarrow 0, n \varepsilon_n \rightarrow \infty$  and  $m_n = [n \varepsilon_n]$ .*

Let us finally study the estimator of the exceedance probabilities.

In Dijk and de Haan (1990) the problem is approached in a somewhat similar way to the generalized Pareto distribution. Consider a level  $y_n$  such that, with an i.i.d. sample from a population with distribution function  $F(\cdot) \in \mathcal{D}(G(\cdot) | \theta)$ , we have  $n(1 - F(y_n)) \rightarrow 0$ . As is well known, the

attraction of  $F(\cdot)$  to  $G(\cdot|\theta)$  is equivalent to  $n(1 - F(\lambda_n + \delta_n z)) \rightarrow (1 + \theta z)_+^{-1/\theta}$ . With  $m_n \rightarrow \infty$ ,  $\frac{m_n}{n} \rightarrow 0$ , we have also  $\frac{n}{m_n}(1 - F(\lambda_{n/m_n} + \delta_{n/m_n} z)) \rightarrow (1 + \theta z)_+^{-1/\theta}$ , and so

$$1 - F(y_n) \approx \frac{m_n}{n} (1 + \theta \frac{y_n - \lambda_{n/m_n}}{\delta_{n/m_n}})_+^{-1/\theta}$$

which suggests the estimator

$$F_n^*(y_n) = 1 - \frac{m_n}{n} (1 + \theta_n^* \frac{y_n - \lambda_{n/m_n}}{\delta_{n/m_n}})_+^{-1/\theta_n^*}$$

when we assume  $y_n = \lambda_{n/m_n} + \delta_{n/m_n} \frac{c_n^\theta - 1}{\theta}$

(note that  $c_n \sim \frac{m_n}{n(1-F(y_n))}$ ).

Then the asymptotic normality of  $F_n^*(y_n)$  is proved.

Let us put  $q\theta(x) = \int_1^x s^{\theta-1} \log s \, ds [\sim (x^\theta \log x)/\theta \text{ if } \theta > 0, \sim (\log x)^2/2 \text{ if } \theta = 0 \text{ and } \sim 1/\theta^2 \text{ if } \theta < 0]$ ,  $c_n^* = \frac{m_n}{n(1-F_n^*(y_n))}$  and denote by  $\theta_n^*$  the estimator of  $\theta$  given in Dekkers, Einmahl and de Haan (1989). Then we can prove that:

$$\text{If } m_n \rightarrow \infty, m_n/n \rightarrow \infty, \frac{q\theta(c_n)}{c_n^\theta \sqrt{m_n}} \rightarrow 0,$$

$$\frac{\sqrt{m_n}}{q\theta(c_n)} (\frac{c_n^\theta - 1}{\theta})^2 \sup_{u \geq \lambda_{n/m_n}} |1 + \theta + \frac{(1 - F(u)) F''(u)}{F'(u)^2}| \rightarrow 0$$

as  $n \rightarrow \infty$ , and the conditions that allow the estimation of  $\theta$  by  $\theta_n^*$  are valid then

$$\frac{\sqrt{m_n} c_n^{*\theta_n^*}}{q_{\theta_n^*}(c_n^*)} \left[ \frac{1 - F(y_n)}{(1 - F_n^*(y_n))} - 1 \right]$$

is asymptotically normal with mean value zero and variance  $1 + \theta^2$  if  $\theta \geq 0$ ,



$$\frac{(2-3\theta)^2}{1-2\theta} + \frac{(1-2\theta)^2(5-11\theta)}{(1-3\theta)(1-4\theta)} - \frac{4(2-3\theta)(1-2\theta)}{1-3\theta} \text{ if } \theta < 0.$$

The choice of  $m_n$  is strongly dependent on the initial distribution function  $F$ ; for instance for a Cauchy distribution we must have  $m_n = O(n^{4/5})$  but for the normal distribution we must have  $m_n = O((\log n)^2)$ .

## 6 Some final comments

Although the results contained in the third approach may be difficult to utilize with regard to the estimation of the exceedance probability, this one seems to be the most appropriate procedure as it only assumes the *existence* of one asymptotic distribution of maxima to the *actual distribution, all observations supposed i.i.d.*, instead of *using* the asymptotics directly and afterwards, when possible, correcting for bias. Although for this procedure there is a study of optimal  $m_n$  by Dekkers and de Haan (1991), the results are so different that there does not seem to exist a general rule to be used universally. To compare the first and the third method, one way could be the comparison of the asymptotic MSE's. In fact as the MSE for the first method is of the order of  $\frac{A}{N}$  or, as  $N \sim n(1 - F(u_n))$ , of the order of  $\frac{A}{n(1-F(u_n))}$ , and for the third method the MSE (for the first  $\theta^*$  the MSE equals the variance) is of the order of  $\frac{B}{m_n}$  (the exact values of the constants  $A$  and  $B$  are given in the text), we have  $\frac{MSE_1}{MSE_3} = O(\frac{m_n}{n(1-F(u_n))})$ . But, in fact, comparison is difficult as the optimal (or approximately optimal)  $u_n$  in the first method depends on the *assumed* or *supposed* behaviour of  $1 - F(x)$ , as well as  $A$  depending on the constants of slow variation but  $m_n$  in the third method depends on the behaviour of  $Q(v)$  through  $g(t)$ .

If we consider  $F(x) = \Phi_\alpha(x)$  we see  $m_n \sim B n^{2/3}$  and  $n(1 - \Phi_\alpha(v_n)) \sim A n^{2/3}$ ,  $m_n$  being optimal for the second estimator in the third approach and the methods are of the same order.

We can also devise a statistical choice procedure to choose between  $\theta < 0$ ,  $\theta = 0$  and  $\theta > 0$ , analogous to the one developed by Tiago de Oliveira (1981); see Tiago de Oliveira (1991b) for details.

## References

- Davison, A. C. and Smith, R. L., 1990. Models for exceedance over high thresholds, (with discussion). *J. Roy. Statist. Soc.*, 8 (52), 393-442.
- Davison, A. C., 1984. Modelling excesses over high thresholds, with an application, *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 461-482, D. Reidel, Dordrecht.
- Dekkers, A.L.M. and de Haan, L., 1989. On the estimation of the extreme value index and large quantile estimation. *Ann. Statist.*, 17, 1795-1832.
- Dekkers, A.L.M. and de Haan, L., 1991. Optimal choice of the sample fraction in extreme-value estimation. Centre for Mathematics and Computer Science, Amsterdam.
- Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L., 1989. A moment estimator for the index of an extreme value distribution. *Ann. Statist.*, 17, 1883-1855.
- Dijk, V. and de Haan, L., 1990. On the estimation of the exceedance probability of a high level, Report 9051/A, Econometric Institute, Erasmus University, Rotterdam.
- Einmahl, 1990. The empirical distribution function as a tail estimator. *Statist. Neerland.*, 44, 79-82.
- Gnedenko, B. V., 1943. Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.*, 44, 423-453.
- de Haan, L., 1970. On regular variation and its applications to weak convergence of sample extremes. *Math. Centre Tracts*, 32, Amsterdam.
- de Haan, L., 1971. A form of regular variation and its applications to the domain of attraction of the double exponential distribution, *Z. Wahrsch. verw. Gebiete*, 17, 241-258.
- Mejzler, D., 1949. On a theorem of B. V. Gnedenko. *Sb. Trudov Inst. Math. Akad. Nauk. SSR*, 12, 31-35 (in russian).
- Pickands, J. III, 1975. Statistical inference using extreme order statistics. *Ann. Statist.*, 3, 119-131.
- Pickands, J. III, 1989. *Consistent estimation of extreme quantiles*, Preprint, University Pennsylvania.
- Reiss, R. - D., 1989. *Approximate Distribution of Order Statistics*, Springer-Verlag.
- Smith, R. L. and Weismann, I., 1985. Maximum likelihood estimation of the lower tail of a probability distribution. *J. Roy. Statist. Soc.*, B, 47, 285-298.
- Smith, R. L., 1987. Estimating tails of probability distribution. *Ann. Statist.*, 15, 1174-1207.
- Smith, R. L., 1989. Approximations in extreme value theory. Preprint, Centre for Stochastic Processes, University of North Carolina, at Chappel Hill.
- Smith, R. L., 1984. Threshold methods for sample extremes. *Statistical Extremes Applications*, J. Tiago de Oliveira ed., 621-638, D. Reidel, Dordrecht.
- 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., 1981. Statistical choice of univariate extreme models. *Statistical Distributions in Scientific work*, C. Taillie et al. eds., 6, 367-387, D. Reidel, Dordrecht.
- Tiago de Oliveira, J., 1982. The  $\delta$ -method for obtention of asymptotic distributions; applications. *Publ. Inst. Statist. Univ. Paris*, 27, 49-70.
- Tiago de Oliveira, J., 1984b. Initiation of statistical decision for Weibull distribution. *Cuad. Bioestad. Aplic. Informat.*, 2, 495-499.
- Tiago de Oliveira, J., 1991. Perspectivas sobre a estatística de extremos e problemas em aberto. *Rev. Real Acad. Ciencias Exactas, Físicas y Naturales*, Madrid.
- Tiago de Oliveira, J., 1991b. Tail estimation and extremes. Home Ahmed A. Sarhan, Alexandria, Egypt.
- Weissman, I., 1978. Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.*, 73, 812-815.
- Weissman, I., 1984. Statistical estimation in extreme value theory. *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 109-115, D. Reidel, Dordrecht.

\*\*\*\*\*