

# Extrapolation into the unknown: modelling tails, extremes, and bounds

M. A. MAES

*University of Calgary  
Department of Civil Engineering  
Calgary, Alberta, Canada, T2N 1N4  
mamaes@ucalgary.ca*

The high reliability associated with civil engineering systems requires us to model carefully the tails of the probability distributions associated with various system uncertainties. Unlike other areas of engineering statistics, reliability-based design and optimisation is concerned with very small probability events. Typically, we have little interest in the central behaviour of random variables, but we do want to make sure that we model correctly whatever goes on in the lower tails and/or the upper tails. Dealing with this problem of extrapolating into the unknown, as we may irreverently call this challenge, requires the use of a number of specific techniques based largely on extreme value statistics, risk-based optimisation, and tail-sensitivity analysis. We emphasize how tails of random variables should be scaled depending on the type of decision we wish to make. Basic extreme value results are given and extended to illustrate the use of weighted tail models. This leads to the definition and the application of the tail heaviness index. An appropriate modelling of tails in limit state random variables is shown to be related to tail sensitivity measures. We also discuss bounded variables, both from the viewpoint of suspected upper or lower bounds, as well as confirmed but unknown upper or lower bounds. Finally the important problem of extreme ratios, and extremes of random variables subjected to one or more constraints is briefly discussed.

*Key words: tails, extremes, decision-based extrapolation, tail identification, tail heaviness, bounds.*

## 1. Introduction

### 1.1. Scope

Risk analysis and reliability-based decision making in structural engineering revolves fundamentally around the simultaneous consideration of (very) small probabilities and (very) large consequences. The former is related to the problem of modelling tails, while the latter is related to the problem of

extremes. Both are in fact different faces of the same coin, as small probabilities usually involve extreme events and vice versa. There is fundamentally no difference between the study of extremes values and the analysis of tails of distributions.

Figure 1 shows a univariate probability density function (pdf) with its tail areas and a large central portion. Clearly, our efforts should focus on the very left and the very right-hand side of this pdf. Unfortunately, a large majority of statistical modeling techniques is applicable to the central portions of random variables, and hence we often find ourselves poorly equipped to deal with tails and extremes. The objective of this paper is to identify the important issues and techniques that can be used in tail modeling, so as to be consistent with the basic principles of decision analysis and reliability-based design.

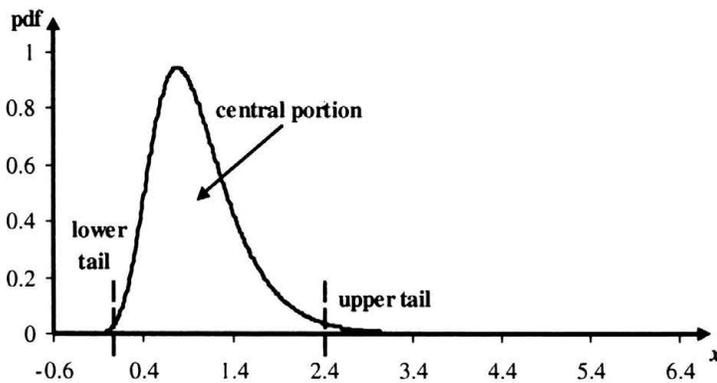


FIGURE 1. Central and tail portions of a random variable.

## 1.2. Tails in risk analysis and decision making

Although recent worldwide tragic events, ranging from environmental catastrophes and accidents caused by inadequate engineering to politically and socially motivated acts of violence, gave a new and grim meaning to the phrase “extreme events”, its study is not new in risk and decision analysis. The occurrence of extreme events in financial risk management has become a major focus of study in relevant years as financial instruments have become increasingly complex and volatile. What is relevant today is not only financial shocks but also events from other sources such as political, social and environmental events. Political leaders warn us of more terrorist attacks that can affect economic stability. Climatologists predict more turbulent weather conditions as a result of global warming that can affect agricultural production. Health experts predict different forms of diseases that can affect productiv-

ity and industrial output. These events manifest into a series of shocks that affect the economy continuously.

No industry has been more affected than the insurance industry that not only witness an increase in frequency of losses but also in magnitude. The claim amounts for recent catastrophic events have been huge and are indeed staggering. Tentative assessments of the damage from the attacks on September 11<sup>th</sup> range from \$30 billion to \$70 billion, approximately half the record damage claims from the Kobe earthquake in 1995 (Flynn, 2002).

Similarly, shocks to the financial sector appear to also increase in frequency and size. There is sufficient evidence to indicate that inadequate risk management tools can be blamed partially in the high incidence of frauds and the inability to forecast catastrophic mishaps. The calls for sophisticated risk management tools to counter the increasingly complex financial instruments appear not to be in place, resulting in such extreme events.

Several obstacles hamper the proper inclusion of extreme events in risk management tools. Risk management tools often involve a trade-off between the quality of the 'central' versus the 'tail' properties. It is not surprising that in the early phases, more emphasis was placed on the former rather than on the latter. Inferences about the tail of a distribution are usually much harder to make, since only a few observations enter the tail region. Moreover, the inferences are very sensitive to the largest observed losses and the introduction of new extreme losses to a dataset may have a substantial impact.

Further, since extreme events are rare by definition, most managers will not be confronted with them during their time on a specific job. Few organizations have incentives to stimulate the performance of future managers at the expense of the current generation. Finally, for many users the psychological perception of the risk of an extreme event is more determined by how vividly they remember the last instance than by the statistical probability of re-occurrence.

### 1.3. Structural design

It goes without saying that structural design is very sensitive to design specifications extracted both from the upper tail areas of load effect variables, as well as from the lower tail portions of resistance variables. Rather than discussing this fact at length, let us illustrate the background thinking used by industry to specify or select extreme design criteria. Consider the following example, reported in Maes and Gu (1994). In 1994, five high level joint industry participants using the same NESS (Grant et al., 1993; Günther and Rosenthal, 1983) offshore database, independently, were asked to

provide their best estimate of the 100 year return significant wave height at a given grid point in the North Sea. All of the participants had expertise in statistics; engineering, and met-ocean modeling. The selection of this design wave height, *HS-100*, is of critical importance to the design, construction and operation of an offshore platform.

We could not help being pleasantly surprised with the astonishing array of techniques and approaches used by the participants: all submissions attest to the fact that the contributors had an expert understanding of the NESS statistics and the extreme value methods needed to formulate engineering design criteria. Our second impression was equally compelling: notwithstanding the diversity of selected EV methods and the variety of subsequently applied “adjustment/corrections,” it was interesting to observe that the recommended *HS-100* values ended up lying very close to one another. The process involved therefore several steps: data base interpretation, selection of a “company-standard” tail model, a statistical uncertainty analysis, and, last but not least, negotiation at the engineering as well as at the manager’s level.

Table 1 summarizes the final results. The first row lists the *HS-100* obtained from the companies own EV analysis of the NESS data: all values submitted can essentially be rounded off to the same number: 11.0 m. But the second row reflects the inclusion of statistical uncertainty, as well as a number of corrections, some applied with, some applied without further justification. The last row lists the final numbers, as they emerged from the final meeting with the decision makers.

TABLE 1. Summary of recommended *HS-100* [m].

	A	B	C	D	E
Value based exclusively on NESS data	10.8	11.0	11.3	11.0	10.5
Recommended value including all corrections / uncertainties	12.6	12.4	12.2	12.0	12.0
Final recommendation	14.3	12.4	12.0	12.0	13.6

Each submission contained a fair number of steps that require the use of good judgement and subjective reasoning. Clearly, several issues were simply not amenable to quantitative evaluation. For instance, the reason for selecting a particular approach may have been that it is a given group’s standard way of dealing with extreme value problems, or it may have been an approach strongly favoured by one or more people, or it may be a series of procedures developed over the years, which enjoyed a history of frequent and successful use. At the same time, each group was attempting to derive a result that

would in all likelihood be acceptable to the outside world (management, designers, regulatory agencies, etc.)

Consequently, there were several aspects of the submissions that were difficult to interpret. Keeping these limitations in mind, it seemed reasonable to identify the following basic criteria to assess the quality of a particular approach:

1. How practical and clear is the suggested approach? A convincing tail model/analysis procedure must be logical and simple to use.
2. Is the method theoretically sound and does it lead to accurate results? Is it based on recognized statistical techniques and proven results from extreme value theory?
3. Can the method be generalized easily to other gridpoints and locations or is it very dependent on a particular data structure? How wide is its range of applicability?
4. How sensitive is the method to assumptions regarding data, distribution types? Is the method robust? Can confidence intervals easily be constructed? Is parameter/statistical uncertainty taken into account?
5. How explicit is any non-analytical input? To which extent is it justified and how streamlined is the process by which this information is implemented?

## 2. Tails: reality or fiction?

It is common practice in the fields of risk analysis and reliability-based design, to devote very little time and effort to checking whether an assumed distribution of an input variable  $X$  is indeed a fair and risk-consistent representation of reality. This is typically the case for variables of which little is known, particularly model uncertainties or other non-physical uncertainties.

But the choice of the type of pdf of a random variable based only on central data can have an enormous impact on the tails (Ditlevsen, 1993; Maes and Breitung, 1993). This is illustrated in Fig. 2 where two random variables are contrasted. Both have mean 1 and standard deviation 0.1, but one is normal and one is lognormal. The central portions of these two random variables are nearly identical but there are huge differences in the tail, as shown by some of the very small and very large quantiles in Fig. 2.

In a typical reliability analysis, risk is often critically dependent on the upper and/or lower tail behaviour of one or a few basic uncertainties. If, for instance, the tail model of one of these variables is altered slightly, it may very well be that the risk level associated with the model changes by an order of magnitude, even though the uncertainty modeling itself may be

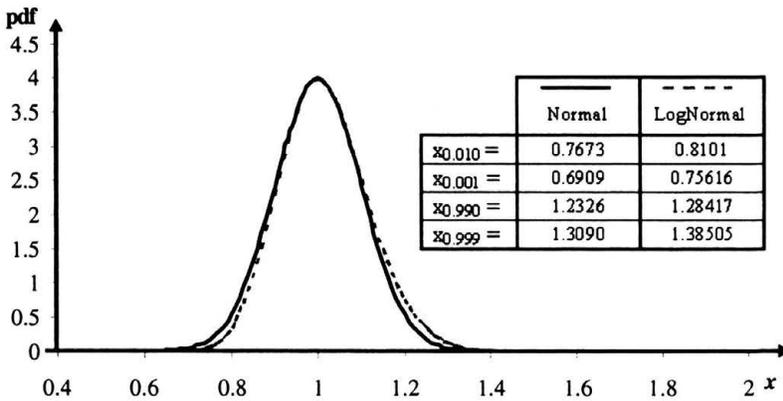


FIGURE 2. Illustration of the difference in tails between two random variables having nearly the same central portion.

satisfactory from a statistical point of view. This situation is in sharp contrast with the advertised high level of accuracy now associated with (commercial) analysis tools available to solve risk analysis problems given probabilistic assumptions regarding its basic uncertainties. This situation is, of course, highly undesirable. Professionals involved in quantitative risk analysis (QRA) are not well served by it, because it undermines their claim to “correctness” or usability of their risk calculations. We are aware of quite a few cases in industry (even in a regulatory context) where consultants, unhappy with the results provided by their analysis, simply decided to replace some lognormal-type by normal-type uncertainties (without any measurable and certainly not alarming loss of statistical goodness of fit) until the risk level became acceptable.

There is then a major difference between “central models” for which the use of classical statistical inference tools is appropriate, and “tail models” which are applicable to risk and reliability problems where the interest lies in the occurrence of rather exceptional events.

A fundamental issue must first be addressed: how can we compare tails or distinguish tails from one another? What are the characteristics of tails that make them into what they are?

### 3. Tail equivalence

A key concept in characterizing tail behaviour is that of tail equivalence (Maes, 1995). It provides a basic principle for deciding if two random variables  $X$  and  $Y$  with cumulative distribution functions (cdf)  $F_X$  and  $F_Y$  have the same ultimate tail behaviour; for the upper tail (in what follows right tails

will be assumed),  $F_X$  and  $F_Y$ , are tail equivalent if

$$\lim_{t \rightarrow \omega} \frac{1 - F_X(t)}{1 - F_Y(t)} = 1 \tag{3.1}$$

where  $\omega$  is the maximum value attained by both random variables, that is,

$$\omega(F) = \sup \{t | F(t) < 1\} \tag{3.2}$$

for both distributions  $F_X$  and  $F_Y$ . The value of  $\omega$  can be infinite.

Equation (3.1) provides a criterion for the quality of approximation of a distribution function  $F_X$  by another distribution  $F_Y$  in the upper tail region.

In the remainder of this section we will preview some of the results that will be derived in later sections. This will improve assimilation of the subsequent flow of information.

In extreme value theory one typically looks for probability distributions which arise under general conditions imposed on the maximum of a data set. An example is the generalized extreme value distribution (GEV). When a random sample of size  $n$  from a distribution  $F$  is considered, and if the following condition on the maximum or highest order statistic of the ordered data  $X_1^* \leq X_2^* \leq \dots \leq X_n^*$  is imposed:

$$\Pr \left( \frac{X_n^* - a_n}{b_n} \leq x \right) \text{ exists as } n \rightarrow \infty, \tag{3.3}$$

then for some sequence of constants  $a_n$  and  $b_n$  and for all  $x$ , the limiting distribution is the GEV (Fisher, and Tippet, 1928) which can be written as

$$F_{\text{GEV}}(x) = \begin{cases} \exp(-1 + \gamma x)^{-1/\gamma}, & \text{if } \gamma \neq 0, \\ \exp(-\exp(-x)), & \text{if } \gamma = 0. \end{cases} \tag{3.4}$$

This result will be discussed further in Secs. 4-6. In engineering, it is also common to work with excesses over thresholds rather than maxima, in which case the generalized Pareto distribution (GPD) arises as the limiting distribution, as shown in Sec. 9:

$$F_{\text{GPD}}(x) = 1 - (1 + \frac{\gamma}{\sigma}(x - u))^{-1/\gamma}, \quad u < x < \omega, \tag{3.5}$$

where

$$\omega = \infty, \gamma \geq 0 \quad \text{or} \quad \omega = u - \frac{\sigma}{\gamma}, \gamma \geq 0. \tag{3.6}$$

Here  $u$  is a high threshold,  $\sigma$  is a positive scaling factor and  $\gamma$  is the so-called extreme value index which links the GEV with the GPD. As shown by

Pickands (1975), the GPD arises as the limiting distribution of the excesses  $X - u$  of a random variable  $X$  over a high threshold  $u$ .

Thanks to the tail equivalence principle (3.1), all we need to do is to examine if the conditional distribution of  $Y = X - u$ , given that  $X$  exceeds the threshold  $u$ , is tail equivalent with the GPD with a specific set of parameters  $\gamma$ ,  $\sigma$  and  $u$ . Moreover, if one considers the scaled random variable  $Y = (X - u)/\sigma$ , tail equivalence can be established with the standardized GPD:

$$F_{\text{GPD}}(y) = 1 - (1 + \gamma y)^{-1/\gamma},$$

$$0 < y < -\frac{1}{\gamma}, \quad \gamma < 0, \quad (3.7)$$

$$0 < y < \infty, \quad \gamma \geq 0.$$

As can be seen in equations (3.4), (3.5) and (3.7), the extreme value index (EVI)  $\gamma$  plays a key role in assessing the weight of the tail. This will lead to the definition of the tail heaviness index (THI) in Sec. 7.

#### 4. Tails of the exponential type

Consider the tail of an arbitrary unbounded pdf  $f_X(x)$  in Fig. 3. The shaded area under the tail is equal to  $1 - F_X(x)$ , and it becomes smaller and smaller as  $x$  grows. At the same time, however, the ordinate  $f_X(x)$  decreases. Generally speaking, the ratio  $f_X(x)/(1 - F_X(x))$  can be determinate or indeterminate, and it can increase, decrease, or remain constant as  $x$  grows. Gumbel (1958) refers to this ratio as the “extremal intensity”. It corresponds

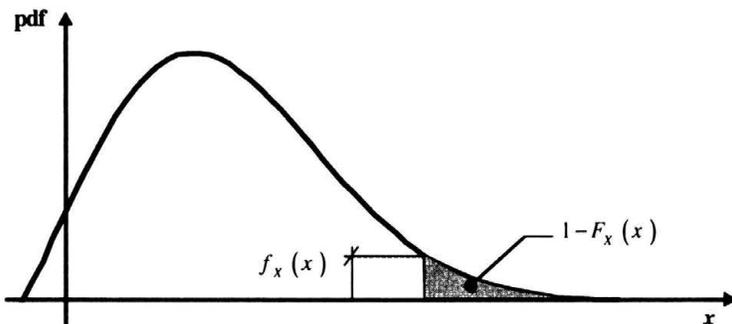


FIGURE 3. Study of the tail of an arbitrary pdf.

in fact to the following conditional probability,

$$\begin{aligned} \Pr(x < X < x + dx | x < X) &= \frac{f_X(x)dx}{1 - F_X(x)} \\ &= -d(\ln(1 - F_X(x))) = d(L_X(x)). \end{aligned} \tag{4.1}$$

It can also be seen that this ratio is equal to  $d(L(x))$  where  $L_X(x)$  is the minus log-exceedance function  $L_X(x) = -\ln(1 - F_X(x))$  which will be discussed in detail in Sec. 8.

The exponential tail-type (ETT) family of distributions includes all the pdfs for which the preceding ratio is indeterminate, i.e.  $0/0$ , as  $x \rightarrow \infty$ . If this is the case, de l'Hôpital's rule (H) may be applied repeatedly as follows:

$$\lim_{x \rightarrow \infty} \left( \frac{f}{1 - F} \right) \stackrel{H}{=} \lim_{x \rightarrow \infty} \left( \frac{-f'}{f} \right) \stackrel{H}{=} \dots \stackrel{H}{=} \lim_{x \rightarrow \infty} \left( -\frac{f^{(i-1)}}{f^{(i)}} \right), \quad i = 1, 2, \dots \tag{4.2}$$

This results in an alternative, but more commonly used definition of ETT distributions: random variables  $X$  belong to the ETT family, if and only if, the so-called extremal quotient  $Q(x)$  given by:

$$Q(x) = \frac{-f_X^2(x)}{f_X'(x)[1 - F_X(x)]} \xrightarrow{x \rightarrow \infty} 1, \tag{4.3}$$

approaches 1 in the limit as  $x \rightarrow \infty$ . This equation follows from the first identity appearing in (4.2).

Just prior to reaching the value of 1 at  $x \rightarrow \infty$ , the value of  $Q(x)$  must either be larger than, equal to, or smaller than one; this fact distinguishes

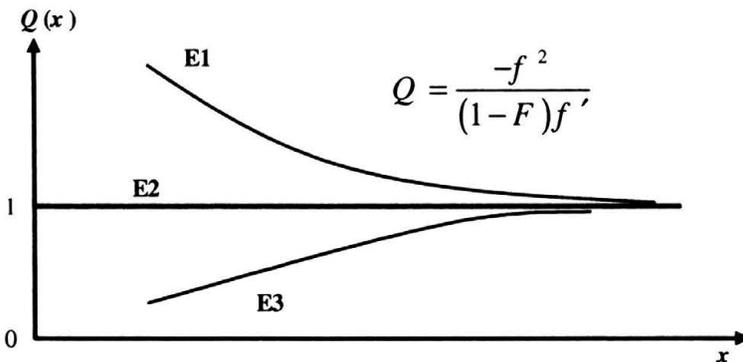


FIGURE 4. Definition of the three classes of ETT distributions.

three classes among the ETT family of distributions, as shown in Fig. 4, namely:

$$\begin{cases} E1 : Q(x) = 1 + \varepsilon(x), \\ E2 : Q(x) = 1, \\ E3 : Q(x) = 1 - \varepsilon(x). \end{cases} \quad (4.4)$$

With  $\varepsilon(x) \geq 0$  and  $\lim_{x \rightarrow \infty} \varepsilon(x) = 0$ .

The generic name of the ETT family originates from the fact that the tail of the pdf declines in almost the same way as the tail of a simple exponential pdf. Indeed, a pdf of the type  $f_X(x) = \lambda e^{-\lambda x}$  has  $Q(x) = 1$  for all  $x$ , and it belongs, therefore, to the class E2 of ETT distributions. It can be checked that all of the unbounded pdfs used in engineering practice are of the exponential tail type, by expanding the expression  $Q(x)$  in (4.3) and checking that its limit is equal to one. ETT includes the normal, lognormal, Weibull, gamma distributions as well as all valid distribution functions based on  $F(x) = 1 - e^{-h(x)}$  where  $h(x)$  is a positive increasing function, as well as their mixtures.

As mentioned in the previous paragraph, there are a few distributions that are not a member of the ETT family. First, there is the Pareto type  $F_X(x) = 1 - x^{-k}$  ( $1 \leq x < \infty$ ), for which  $Q(x)$  is a constant and equal to  $\frac{k}{k+1} \neq 1$ . This pdf has no mode and no moments of order  $\geq k$  and it is marked by a tail with a very slow, non-exponential decline. Secondly, distributions of the Cauchy type which have no moments at all,  $F_X(x) = 1/2 + \arctan x/\pi$ , or  $f_X(x) = (\pi(1+x^2))^{-1}$  are not of the exponential tail-type since the ratio of the shaded area to ordinate in Fig. 3 does not become indeterminate, but actually grows with increasing  $x$ .

Suppose now that a sample of size  $n$  is generated from an ETT parent distribution. In order to have a quantity that grows with  $n$  at approximately the same rate as the maximum value in the sample, the characteristic extreme  $a_n$  of size  $n$  can be defined as the value of the random quantity  $X$  corresponding to a mean recurrence interval equal to  $n$ :

$$F_X(a_n) = 1 - \frac{1}{n}, \quad \text{or} \quad n = \frac{1}{1 - F_X(a_n)}, \quad n = 2, 3, \dots \quad (4.5)$$

This definition follows directly from the notion of the return period. In fact, if  $X_i$  were random quantities occurring during a unit of time, for example 1 year, then the return period of the value  $a_n$  would be equal to  $n$ . In general, however, the definition of the characteristic extreme is not linked to the concept of time scaling; it merely states that in a sample of size  $n$ , the expected number of values that exceed  $a_n$  is exactly equal to one. As an

example,  $a_n$  for the simple exponential pdf  $f_X = \lambda e^{-\lambda x}$  is equal to  $\frac{\ln n}{\lambda}$ , using (4.5).

A second extremal parameter, namely the dispersion factor  $b_n$ , can now be defined as the value of the area-ordinate ratio in Fig. 3, evaluated at  $x = a_n$ , or as the inverse of Gumbel’s “extremal intensity” (4.1):

$$b_n = \left( \frac{1 - F_X}{f_X} \right)_{x=a_n}, \tag{4.6}$$

or, substituting  $1 - F_X$  by the value indicated in (4.5),

$$b_n = \frac{1}{n f_X(a_n)}. \tag{4.7}$$

Note that  $b_n$  is always positive and that it has the same dimension as  $x$  and  $a_n$ . It is also referred to as the extremal scale factor. The following identity can also be established by differentiating (4.6) with respect to  $n$ , assumed to be a continuous variable:

$$\frac{da_n}{dn} = \frac{dF_X}{dn} / \frac{dF_X}{da_n} = \frac{1}{n^2} \frac{1}{f_X(a_n)} = \frac{b_n}{n}. \tag{4.8}$$

In a sample of size  $n$ , we can also determine the “most probable largest” sample value,  $\tilde{z}$  as the mode of the maximum  $Z$  of the random sample

$$Z = \max_{i=1, \dots, n} X_i. \tag{4.9}$$

This random variable has a cdf given by

$$F_Z(z) = Pr(Z < z) = Pr \left( \bigcap_{i=1, \dots, n} X_i < z \right) = [F_X(z)]^n, \tag{4.10}$$

and a pdf, shown in Fig. 5, which is equal to

$$f_z(z) = n F_X(z)^{n-1} f_X(z), \tag{4.11}$$

which has a mode  $\tilde{z}$  that can be found by setting the derivative function equal to zero (Fig. 5):

$$f'_Z(\tilde{z}) = n(n - 1)F_X(\tilde{z})^{n-2} f_X^2(\tilde{z}) + nF_X(\tilde{z})^{n-1} f'_X(\tilde{z}) = 0. \tag{4.12}$$

This results in the following expression for the most probable largest  $\tilde{z}$  in a sample of size  $n$ :

$$- \left( \frac{f'_X F_X}{f_X^2} \right)_{x=\tilde{z}} = n - 1. \tag{4.13}$$

So far, three extremal parameters have been defined: the most probable largest  $\tilde{z}$  (4.13), the characteristic extreme  $a_n$  (4.6) and the dispersion factor  $b_n$  (4.7). How do they interrelate and what is their behaviour with increasing  $n$ ? Equivalently, what is the importance of the three ETT classes?

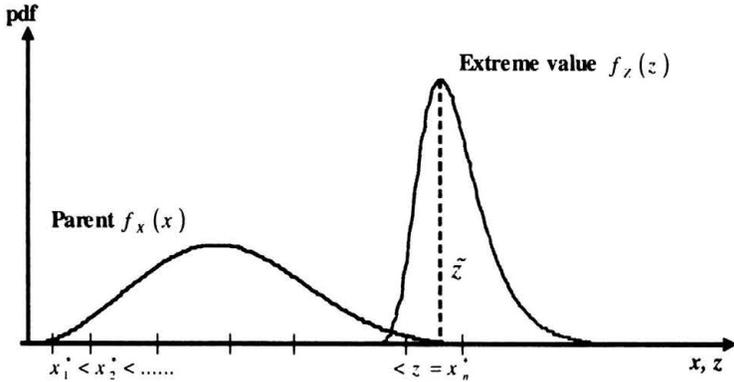


FIGURE 5. Most probable largest  $\tilde{z}_n$  in a random sample size  $n$  from an arbitrary parent pdf  $f_X$ .

(a) **Relation between  $a_n$ ,  $\tilde{z}_n$  and class:** By substitution of the expression (4.4) into the constitutive equation (4.13) for the most probable largest value, we obtain:

$$F_X(\tilde{z}) = 1 - \frac{1}{Q(\tilde{z})(n-1) + 1}, \tag{4.14}$$

and, with  $Q(z) = 1 \pm \varepsilon(z)$  as in (4.5), the following expression is obtained:

$$F_X(\tilde{z}) = 1 - \frac{1}{n \pm (n-1)\varepsilon(\tilde{z})}, \tag{4.15}$$

and must be compared with the corresponding equation for the characteristic extreme, i.e. (4.5). Clearly, since  $F_X$  is always a non-decreasing function and since, by the definition of ETT,  $\varepsilon \rightarrow 0$  for large values of the  $x$ , we may conclude that:

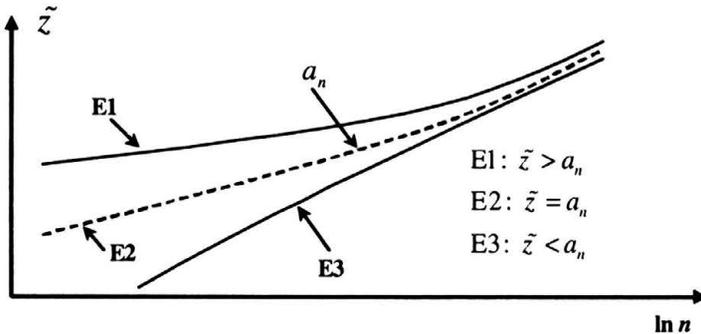


FIGURE 6. Comparison of the most probable largest  $\tilde{z}$  with the characteristic extreme  $a_n$ .

1. The most probable largest  $\tilde{z}_n$  converges to the characteristic extreme  $a_n$  as  $x \rightarrow \infty$ .
2. Prior to reaching  $n$ , the class of the parent pdf determines whether  $\tilde{z}$  is larger than, equal to, or smaller than  $a_n$ . This is represented in Fig. 6.

**(b) Behaviour of  $b_n$  with increasing  $n$ :** Assuming for large samples  $n$  to be continuous, the derivative of  $b_n$  (4.7) with respect to  $n$  may be studied:

$$\frac{db_n}{dn} = \frac{d}{dn} \left( \frac{1}{nf_X(a_n)} \right) = -\frac{1}{n^2 f_X(a_n)} - \left( \frac{f'}{nf^2} \right)_{a_n} \frac{da_n}{dn}. \quad (4.16)$$

Introducing (4.5) and (4.8), and introducing the expression (4.3) for  $Q(x)$ , results in:

$$\frac{db_n}{dn} = -\frac{1}{n^2 f_X(a_n)} + \frac{1}{Q(a_n)} \left( \frac{1}{n^2 f_X(a_n)} \right) = \frac{1}{n^2 f_X(a_n)} \left\{ \frac{1}{Q} - 1 \right\} a_n. \quad (4.17)$$

The first ratio and  $a_n$  in (4.17) are always positive, but the expression in brackets is zero, positive or negative, for ETT class E1, E2, E3 respectively. Therefore, the tails of distributions belonging to class E1, E2 and E3 parent pdfs are characterized by dispersion factors that decrease, remain constant, and increase, respectively, with growing sample size. In the limit ( $n \rightarrow \infty$ ), the dispersion becomes constant (i.e.,  $\frac{db_n}{dn} = 0$ ). This behaviour is shown in Fig. 7.

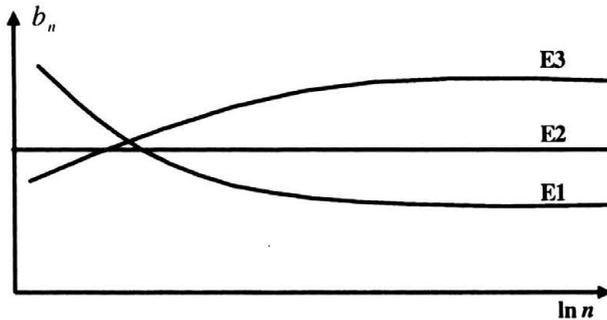


FIGURE 7. The dispersion factor  $b_n$  as a function of sample size  $n$ .

**(c) Behaviour of  $a_n$  with  $\ln n$ :** We noted earlier that the characteristic extreme for a simple exponential pdf, is equal to  $\ln n$ ; consequently,  $\ln n$  can serve as an appropriate benchmark for the growth of  $a_n$ :

$$\frac{da_n}{d(\ln n)} = n \frac{da_n}{dn} = b_n > 0, \quad (4.18)$$

which implies, as expected, that  $a_n$  is always increasing. But more relevantly the second derivative of  $a_n$  with respect to  $\ln n$  is equal to  $n \frac{db_n}{dn}$  in (4.17), resulting in a concave, convex or linear growth of  $a_n$  with  $\ln n$ , depending on the ETT class of the tail of the random variable. Figure 8 also shows that the absolute value of  $a_n$  increases slower (E1), faster (E3), or at the same rate as  $\ln n$  (E2).

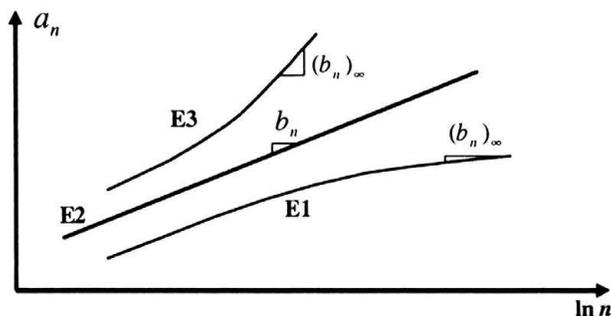


FIGURE 8. Increase of the characteristic extreme  $a_n$  as a function of  $\ln n$ .

At this point, the most famous result in extreme value theory may be introduced. We owe the original derivation to von Mises in 1936, almost a decade after the double exponential distribution  $\Lambda_3$  was obtained as a limiting distribution by Fisher and Tipett. Gnedenko (1943) proved that the conditions were necessary and sufficient, and Gumbel (1958) simplified the proof considerably. The theorem states that all probability distributions of the ETT family have  $\Lambda_3(z|a_n, b_n)$  as their extreme value limiting distribution, where  $\Lambda_3$  is the double exponential distribution,  $a_n$  is the characteristic extreme and  $b_n$  is the dispersion factor, both parameters being related to the tail properties of the parent distribution.

This result can easily be derived from the fact that the condition (4.3) that  $Q(x) \rightarrow 1$ , is equivalent with imposing  $\lim_{x \rightarrow \infty} L^{(i)}(x) = 0$  for the derivatives of order 2 and higher, where  $L(x)$  is the minus-log-exceedance function  $L = -\ln(1 - F)$ . It will also be seen in Sec. 8 that this amounts to a linear asymptote in an  $(L, x)$  plot. Based on the definitions of  $a_n$  and  $b_n$ , we have

$$L(a_n) = \ln n, \quad (4.19)$$

$$b_n = \frac{1}{L'(a_n)}. \quad (4.20)$$

As a result, the tail of the parent distribution can be written as

$$F(x) = 1 - \exp(-L(x)) = 1 - \frac{1}{n} \exp(-(L(x) - L(a_n))). \quad (4.21)$$

The Taylor expansion of  $L(x)$  around  $L(a_n)$  results in a simple linear relationship because of the above condition that all derivations of  $L(x)$  of order 2 and higher must be zero for the ETT classes:

$$\begin{aligned} L(x) - L(a_n) &= (x - a_n)L'(a_n) + 0 \\ &= \left(\frac{x - a_n}{b_n}\right). \end{aligned} \tag{4.22}$$

Introducing this result in (4.21) and raising it to the power  $n$  in order to obtain the cdf of the maximum  $Z = \max_n X_i$  shows that, for larger  $n$ ,

$$\begin{aligned} \ln F_Z(z) &= n \ln F_X(z) = n \ln(1 - (1 - F_X(z))) \\ &\cong -n(1 - F_X(z)) \\ &\cong -\exp\left(-\left(\frac{z - a_n}{b_n}\right)\right), \end{aligned} \tag{4.23}$$

or

$$F_Z(z) \rightarrow \Lambda_3(z|a_n, b_n) = \exp\left(-\exp\left(-\left(\frac{z - a_n}{b_n}\right)\right)\right). \tag{4.24}$$

It can also be seen that for exceedances of a high threshold  $u$ , all members of the ETT class converge towards the linear  $L$ -expression (4.22) (except for a linear shift), in other words, tails exceeding a high threshold have an exponential distribution for  $X > u$ . We will refer to this fact when examining the role of the Generalized Pareto Distribution (GPD) in Sec. 10.

### 5. Tails of the Pareto Type (PT)

As discussed in the previous section, ETT tails are characterized by the condition that  $Q(x) \rightarrow 1$  as  $x$  increases. This was seen to result in tails that become nearly exponential for a large enough threshold, since for the exponential distribution  $Q(x) \rightarrow 1$  for all  $x$ . Consider now the case that these excesses  $Y$ , when properly scaled, have a Pareto-type polynomial distribution:

$$F_Y(y) = 1 - (1 + \xi y)^{-1/\xi}, \quad \xi > 0, \quad 0 < y < \infty. \tag{5.1}$$

It can easily be seen that for all  $y$ , the extremal quotient  $Q(y)$  given by (4.3) is positive and equal to

$$Q_Y(y) = \frac{1}{1 + \xi}. \tag{5.2}$$

When  $\xi$  approaches 0,  $Q$  approaches 1 as for the ETT tails, but otherwise the extremal quotient takes on small values between 0 and 1. Frechet and Gnedenko showed that the, properly scaled, corresponding asymptotic extreme

value distribution for  $Z = \max_n X_i$  is equal to the so-called Fréchet extreme value distribution:

$$F_z(z) \rightarrow \exp\left(- (1 + \xi z)^{-1/\xi}\right), \quad \xi > 0, \quad 0 < z < \infty. \quad (5.3)$$

As can be seen from the expression for  $Q(y)$  in (4.3) the PT tails are long and heavy. They include the tails of the Pareto, Cauchy and Fréchet distributions.

## 6. Beta-type tails: bounded distributions

Short tails or the suspicion of short tails usually denote the existence of a finite upper bound (for simplicity we will focus on right-hand tails). One would expect the extremal quotient  $Q$  given by (4.3) to be (very) large for such tails as the density close to the upper bound is still large when  $1 - F(x)$  approaches 0 close to the upper bound  $\omega$ . If we consider a threshold  $u$  close to the upper bound  $\omega$ , then the following right-leaning beta distribution with positive exponent  $-\frac{1}{\xi}$  (hence  $\xi < 0$ ) can be used to model the scaled excesses  $y = \frac{(x-u)}{\sigma}$  on the tail between  $u$  and  $\omega$ :

$$F_Y(y) = 1 - (1 + \xi y)^{-1/\xi}, \quad 0 < y < -\frac{1}{\xi}, \quad \xi < 0. \quad (6.1)$$

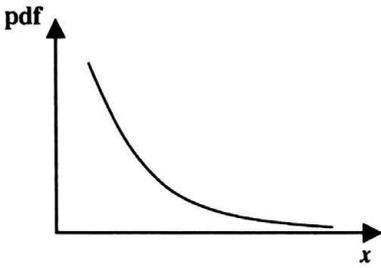
Clearly, in this case the value of  $\xi$  is directly related to the nature of the upper bound  $\omega$ . It can be seen that the upper bound  $\omega$  is equal to  $-\frac{1}{\xi}$ , so that we can rewrite the distribution and its associated pdf as:

$$F_Y(y) = 1 - \left(1 - \frac{y}{\omega}\right)^{-1/\xi}, \quad 0 < y < \omega, \quad (6.2)$$

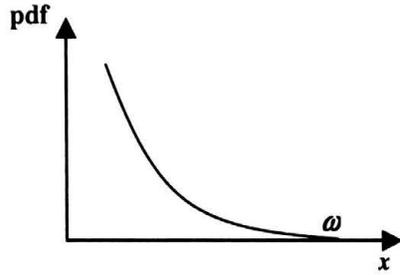
$$f_Y(y) = \left(1 - \frac{y}{\omega}\right)^{-1/\xi - 1}, \quad 0 < y < \omega. \quad (6.3)$$

These equations show that as  $Y$  attains its upper bound  $\omega$ , the value of  $\xi$  controls in which way  $\omega$  is approached. This is illustrated in Fig. 9. The larger values of  $\xi$  (those close to zero) generate distributions where the bound is reached very slowly: at  $\omega$ , at least the first  $[\text{int}(1/|\xi|) - 2]$  derivatives of the pdf are zero. When  $\xi = -\frac{1}{2}$  the slope of the density function at  $\omega$  is nonzero but finite, whereas  $\xi = -1$  corresponds to the case of truncation. Even more negative values of  $\xi$  lead to infinite densities at the end-point  $\omega$  as shown in Fig. 9.

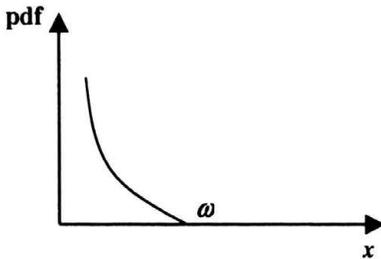
The classification of tails according to the  $\xi$  is useful for a variety of reasons. For instance, it allows one to investigate if a data set has been truncated, by testing whether the conditional empirical distribution is tail



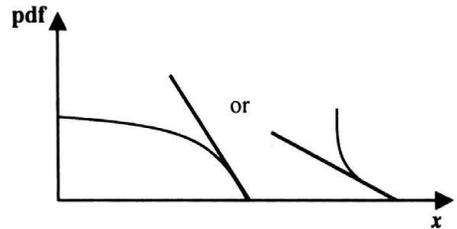
(a) no upper bound;  $Q = 1, 0 \leq \xi$  (ETT)



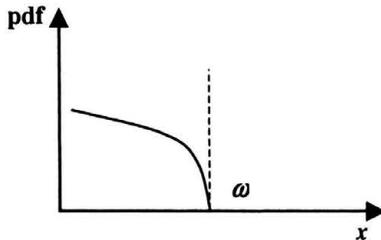
(b) finite end-point  $\omega$ , zero density;  $1 < Q < \infty, -1 < \xi < 0$



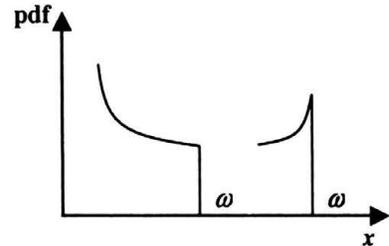
(c) zero density at  $\omega$ , first  $(k - 2)$  derivatives of pdf are zero;  $\xi = -1/k$  ( $k$  integer,  $k \leq 3$ ),  $Q = (k + 1)/k$



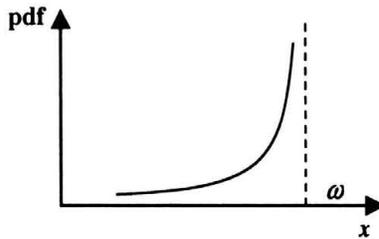
(d) zero density, finite first derivative at  $\omega$ ;  $\xi = -1/2, Q = 2$



(e) zero density at end-point  $\omega$ , infinite first derivative;  $2 < Q < \infty, -1 < \xi < -1/2$



(f) non-zero finite density at  $\omega$ ;  $Q = \infty, \xi = -1$ , and  $Q < 0, \xi < -1$



(g) non-zero finite density at  $\omega$ ;  $-\infty < Q < 0, \xi < -1$

FIGURE 9. Different types of tails and upper bounds.

equivalent with (6.2) having  $\xi = -1$ . Or, one can examine if a data set possess a tail behaviour which is significantly different from that of the central part of the distribution. And, as will be seen later, general results in safety and reliability analysis can be expressed as a function of  $\xi$  (Maes and Huyse, 1995; Maes, 1995).

Since the tails close to upper bound  $\omega$  are in this case modeled by the beta distribution (6.1), we refer to this type of tail as the beta type (BT). Similar to (5.2) the extremal quotient for the BT tail is constant and larger either than one or less than zero depending on the value of  $\xi$ :

$$Q_Y(y) = \frac{1}{1 + \xi} \begin{cases} > 1 & -1 \leq \xi < 0, \\ \rightarrow \infty & \text{for } \xi \rightarrow -1, \\ < 0 & -\infty < \xi < -1. \end{cases} \quad (6.4)$$

All of these tail  $Q$  values imply light tails and they are consistent with the above discussions of the various types of bounded tails shown in Fig. 9. Weibull and Gnedenko showed that the (bounded and properly scaled) maximum  $Z = \max_n X_i$  has a so-called Weibull asymptotic extreme value distribution:

$$F_z(z) \rightarrow \exp\left(- (1 + \xi z)^{-1/\xi}\right), \quad (6.5)$$

$$\xi < 0, \quad 0 < z < \omega, \quad \omega = -\frac{1}{\xi}.$$

The three extreme value distributions given by (4.24), (5.3) and (6.5) for the tails of the ETT type, the PT type and the BT, respectively, can be united in the generalized extreme value distribution given earlier in Eq. (3.4). All of these distributions have the important property of being tail equivalent (Sec. 3) with the tails of their parent distributions.

## 7. The tail heaviness index (THI)

The tail heaviness index (THI) was introduced by Breiman et al. (1979) and first used by Boos (1984) in a comparative study of techniques used to estimate large quantiles. The idea is to benchmark heaviness against that of the exponential tail which is assigned a value of zero; it is negative for lighter than exponential tails (subexponential) and positive for heavier than exponential tails (superexponential).

The index is, generally speaking, a function of the position on the tail, i.e. left- versus right-handed tail, as well as the exceedance probability  $q$  and

the corresponding quantile  $x_q$ . It is defined as

$$H(x_q) = -q \left[ \frac{f'}{f^2} \right]_{x=x_q} - 1 = - \left( \frac{(1-F)f}{f^2} \right)_{x=x_q} - 1, \tag{7.1}$$

where  $f$  is the density function of  $X$ . Alternatively (7.1) may be expressed in terms of the loglikelihood function  $l(x) = \ln f(x)$ :

$$H(x_q) = -q \left[ l' e^{-l} \right]_{x=x_q} - 1. \tag{7.2}$$

Since  $q = 1 - F(x_q)$ , it can easily be checked that the minus-log-exceedance function

$$L(x) = -\ln(1 - F(x)) \tag{7.3}$$

where  $F(x)$  is the cumulative distribution function, can also be used in the definition of  $H(x_q)$ :

$$H(x_q) = - \left[ \frac{L''}{L'^2} \right]_{x=x_q}. \tag{7.4}$$

It should be noted that if  $q$  is a probability per unit time (e.g. annual exceedance probability), then  $L(x)$  represents the log-return-period function.

The tail heaviness index is, in fact, closely related to previously defined in Eq. (4.3) extremal quotient  $Q(x)$  and it therefore inherits similar properties:

$$H(x) = \frac{1}{Q(x)} - 1. \tag{7.5}$$

### 8. Detecting heaviness in the $(L, x)$ plot

The optimal way of representing and investigating tail behaviour is the  $(L, x)$  plot, where the minus-log-exceedance function  $L$  defined in (7.3) is plotted as a function of  $x$ . In what follows, we will focus on right-hand tails only, i.e.,  $x$  approaches  $\infty$  or some upper bound  $\omega$ . In an  $(L, x)$  plot, the exponential tail is represented by a straight line, because  $Q(x) = 1$  and hence  $H = 0$  and  $L'' = 0$  according to the above equations.

It is clear from the identity (7.4) that the tail heaviness index is proportional to the negative of the curvature  $L''(x)$  at any point  $x$  in the graph, as shown in Fig. 10. Consequently, a convex plot (“dogtail”) has a negative  $H$ -value; this points to a light tail of the beta-type, and, it suggests the likely existence of some upper bound on  $x$ . Conversely, a positive tail heaviness index corresponds to a concave  $(L, x)$  plot of a heavy tail of the Pareto-type, as shown in Fig. 10.

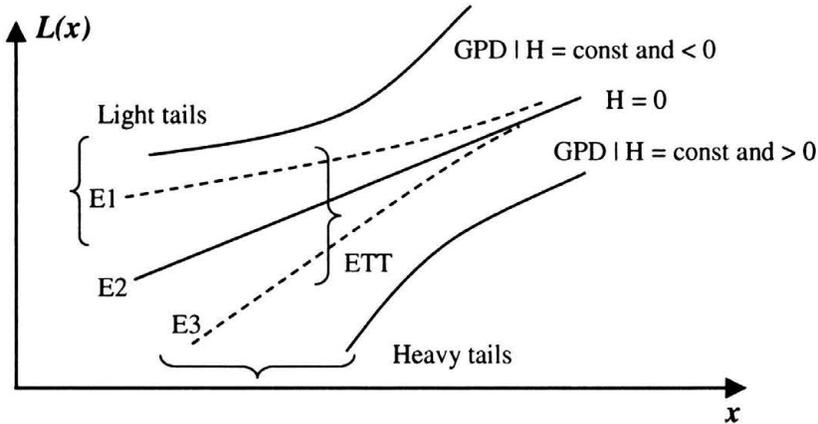


FIGURE 10.  $(L, x)$  plot showing various tail characteristics.

It should be noted that the use of a Gumbel plot (Castillo, 1992) rather than an  $(L, x)$  plot is equally effective. In this plot of the Gumbel ordinate  $\eta(x) = -\ln[-\ln F(x)]$  versus  $x$ , a double exponential distribution is represented by a straight line. But for small exceedance probabilities  $[1 - F(x)] \rightarrow 0$ , and, consequently, for tail values,  $\eta(x)$  is virtually equal to  $L(x)$ , which means that the  $(L, x)$  plot and the Gumbel plot are tail equivalent (see Sec. 3).

Unlike other probability-paper methods,  $(L, x)$  plots should be used only to assess tail behaviour, i.e., as a tool for decision-making with respect to small exceedance probabilities, large quantiles, tail extrapolation, or other applications in reliability.

### 9. Tails exceeding a high threshold

We referred earlier to three types of tails with negative, zero, and positive tail heaviness indices. The concept of tail equivalence allows us to identify the tail behaviour using the excesses  $X - u$  of the random variable  $X$  over a high threshold  $u$ . Pickands (1975) showed that the generalized Pareto distribution (GPD) arises as the limiting distribution of the excesses provided the tail belongs to the domain of attraction of one of the extreme value distributions. The GPD summarizes the three previously discussed tail-over threshold distributions (4.22), (5.1), (6.1), so that (3.5) can also be written as:

$$F_{GPD}(x) = 1 - \left[ 1 + \frac{\xi(x - u)}{\sigma} \right]_+^{-1/\xi}, \quad x > u, \quad (9.1)$$

where  $u$  is the high threshold, and  $\sigma$  is a positive constant. As before, the case  $\xi = 0$  is interpreted as the limit  $\xi \rightarrow 0$ , which results in  $F(x) = 1 - \exp\left(-\frac{(x-u)}{\sigma}\right)$ , i.e., the excess  $X - u$  over the threshold  $u$  is an exponential random variable with mean  $\sigma$ . If  $\xi > 0$ , (9.1) represents one of several forms of the (unbounded) Pareto tail type, whereas the case  $\xi < 0$  restricts the range of the excess to the interval  $0 < X - u < -\sigma/\xi$ , so that the GPD (9.1) becomes a (bounded) beta distribution, representing the Beta tail type.

The GPD enjoys widespread use in areas such as hydrology and oceanography. In fact, it forms a key component of peak-over-threshold methods of analysis. It can easily be seen that a Poisson process of exceedances of a high level having excesses with a generalized Pareto distribution results in maxima which have a generalized extreme value distribution (6.5).

Consequently, GEV and GPD are tail equivalent according to criterion given in Eq. (3.1) and the three types of extreme value distribution correspond directly to the three tail-over-threshold types. Moreover, it can be seen from the derivatives of the minus-log-exceedance function:

$$L_{\text{GPD}}(x|u, \varepsilon, \sigma) = \frac{1}{\xi} \ln \left[ 1 + \frac{\xi(x-u)}{\sigma} \right]_+ \quad \text{for } x > u, \quad (9.2)$$

and also from the previous expressions for  $Q(x)$ , that the tail heaviness index of the GPD is constant over the entire range of  $x$ , and equal to:

$$H_{\text{GPD}} =_+ \xi \quad \text{for all } x > u. \quad (9.3)$$

The GPD represents the only family of distributions with *constant* heaviness; this provides justification for the labels in Fig. 10.

Figure 10 shows the three types of GPD distributions with their respective constant tail heaviness indices. Also included in this plot are tails belonging to the three ETT classes. Now these tails have  $H(x)$  values that are not constant but instead are negative and slowly increase to zero as  $x \rightarrow \infty$  for class E3, or else, are positive and slowly decrease to zero as  $x \rightarrow \infty$ , for class E1. The advantage of the  $(L, x)$  plot is that the GPD never fails to provide a reasonable fit to the curvature associated with the empirical distribution. This is discussed in the next two paragraphs and it also illustrated in Fig. 11 which is based on random samples generated from various pdfs.

Examples of light tails (i.e. negative heaviness indices) include the normal, the gamma (with  $\alpha > 1$ ), the logistic, the uniform ( $H = -1$ ), the beta, and most distributions with a finite upper bound. Whereas some of the above eventually reach zero tail heaviness,  $\lim_{x \rightarrow \infty} H(x) = 0$  because they belong to the ETT class of distributions, their behaviour in the practical

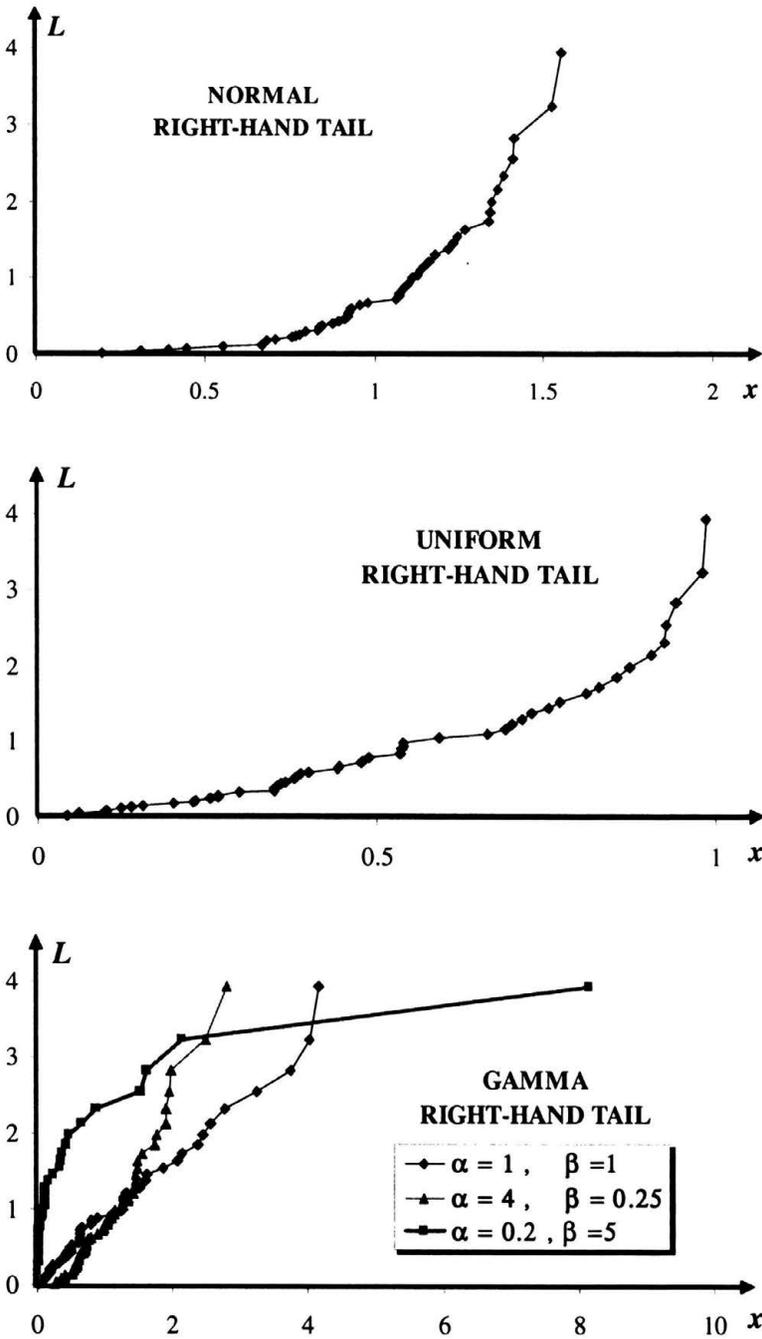


FIGURE 11.  $(L, x)$  plot of random samples from various distributions.

pre-asymptotic range is that of subexponential (BT class) tails (Figs 10-11). Since the tail heaviness becomes more negative as  $x$  becomes smaller on the tail, this effect is more noticeable in applications with smaller sample sizes.

Positive  $H$ -values characterizing long, heavy (PT type) tails include the gamma (with  $\alpha < 1$ ), the lognormal, the Pareto, the Cauchy and the Fréchet. Similar to the gamma or the  $t$  distribution, the Weibull distribution as a member of the ETT family  $F(x) = 1 - \exp\left(-\left(\frac{x}{u}\right)^k\right)$  has a tail heaviness index that can be either positive (class E1), zero (class E2), or negative (class E3) before reaching zero at  $x \rightarrow \infty$ ; see also Figs. 10 and 11:

$$H_{\text{Weibull}}(x) = -\left(1 - \frac{1}{k}\right) \left(\frac{u}{x}\right)^k. \quad (9.4)$$

Thus, for exponents  $k > 1$ , the Weibull tail is light, for  $k = 1$  it is exponential, and for  $k < 1$  the tail is heavy, but since the right-hand tail of the Weibull belongs to the ETT,  $\lim_{x \rightarrow \infty} H(x) = 0$ .

## 10. GPD estimation

Let the empirical distribution function of  $X$  in an  $(L, x)$  plot be described by the  $n$  pairs  $(L_i, x_i)$  of the ordered sample  $x_1 \leq x_2 \leq \dots \leq x_n$ , where  $L_i = -\ln\left(1 - \frac{i}{n+1}\right)$ . In the upper tail area  $\mathbf{T}$ , we are now interested in fitting a GPD distribution (9.1) which is tail-equivalent to this empirical distribution.

The first problem is to define the extent of the upper tail area  $\mathbf{T}$ , which amounts to determining the threshold  $u$ , the lower cut-off of  $\mathbf{T}$ . This is a somewhat subjective task, although a quick visual inspection of an  $(L, x)$  plot often suffices to identify the best threshold.

A useful tool is based on the equivalent of the mean residual life (MRL) diagram for lifetime modeling. This is a plot of the conditional mean of the excesses  $E(X - u | X > u)$  as a function of  $u$ . The only stable part of this graph is located in the tail, and it often provides for a straightforward threshold selection. The slope and the intercept of the best MRL straight line fit can subsequently be related to the parameters  $\xi$  and  $\sigma$  of the GPD in the tail area. A more rigorous approach to the problem of how to separate the central portion from the tail portion is discussed in Sec. 12.

Other methods for estimating the GPD parameters (Davison and Smith, 1989) include maximum likelihood estimation and the Hosking and Wallis' (1987) method based on a simple extension of the MRL plot.

In reality, however, tail estimation is essentially a decision-making problem: no single tail fit is satisfactory for all purposes and for all possible

different types of future usage. There is no single “good-for-all” tail fitting criterion. Rather, we wish to conjecture that the criteria selected for achieving a suitable tail fit are critically dependent on what it is one wants to achieve; they should be consistent with the measure of risk associated with the application affected by the tail behaviour. Hence, a tail model used by decision maker A may be different from the model used by B even though identical tail data is used. This is a clear case of the ends justifying the means.

The most flexible option for achieving this objective is weighted least squares on the  $(L_i, x_i)$  data in the tail region. The weights  $w_i$  should depend on the quantity (or, property) of interest so that, in general, the GPD parameters can be found from the minimization with respect to  $\xi$  and  $\sigma$  of the sum of weighted square errors SWSE:

$$\text{SWSE} = \sum_{i \in T} w_i [L_i - L(x_i|u, \xi, \sigma)]^2 \quad (10.1)$$

where  $L$  is given by (9.2). This implies that errors  $\Delta L = L_i - L$  are iid with mean 0 and variance proportional to  $1/w_i$ . By treating this heteroscedastic model as a likelihood problem, similar to the Maes and Breitung (1994) analysis for the GEVD, probability intervals may be determined to describe model and parameter uncertainty.

Let us list some examples of appropriate expressions (10.1) consistent with a few specific tail evaluation objectives:

**Case A:** The objective is to determine small exceedance probabilities with the smallest absolute error.

Very rarely does one encounter engineering applications in which risk can be measured in terms of absolute errors  $\Delta q_i$  on small probabilities of exceedance  $q_i$ . The relative error, or the error on log (probability), forms a much more frequent (and appropriate) criterion for tail evaluation. But in order to minimize absolute errors on probability in the tail area  $-\Delta q_i = \Delta F_i = (1 - F_i) \Delta L_i$ , it is clear that we need to use weights  $w_i = (1 - F_i)^2$  in (10.1). As expected, these weights reduce the importance of the upper tail, since the  $(L, x)$  plot amplifies precisely this region.

**Case B:** The objective is to determine small exceedance probabilities with the smallest relative error.

Since  $-\Delta q_i/q_i = -\Delta(1 - F_i)/(1 - F_i) = -\Delta \ln(1 - F_i) = \Delta L_i$ , unit weights  $w_i = 1$  are required in (10.1). This reduces (10.1) to regular least squares together with straightforward parameter uncertainty analysis.

**Case C:** The objective is to determine return periods with the smallest absolute error.

If the random variable  $X$  applies to a unit period of time, then  $\frac{1}{1-F_i}$  represents a return period. If we wish to consider errors on this function, then we have  $\Delta\left(\frac{1}{1-F_i}\right) = \frac{\Delta L_i}{(1-F_i)}$ . Accordingly, we require weights  $w_i = \frac{1}{(1-F_i)^2}$  in (10.1) which increase with distance on the tail. The large tail weights indicate that, for this purpose, the  $(L, x)$  plot does not "amplify" the tail area sufficiently.

**Case D:** The objective is to determine return period with the smallest relative error.

More frequently, it is the relative error on the return period, which serves as a true risk indicator. Because  $\frac{\Delta\frac{1}{1-F_i}}{\frac{1}{1-F_i}} = \frac{\Delta F_i}{1-F_i} = \Delta L_i$ , this case is equivalent to the base case  $B$  with unit weights.

**Case E:** The objective is to determine large tail quantiles  $x_q$ .

If the decision criterion involves a risk measure, which can be expressed as a linear function of the error of the unknown quantile (a nonlinear function requires linearization but the treatment is otherwise the same), then the weights required in (10.1) are equal to  $\frac{1}{L'_i} = \frac{1}{\sigma + \xi(x_i - \mu)}$ , since the error on the quantile can be written as  $\Delta x_{qi} = \Delta L_i / L'_i$ .

**Case F:** The objective is to respect the principle of tail equivalence in general.

If one wishes to construct or extrapolate a distribution tail to a (usually limited) set of tail data, and no clear-cut risk-based decision criterion is available for the goodness of fit, then the basic principle of tail equivalence should apply. With  $(F + \Delta F)$  equal to the empirical distribution function, (3.1) becomes  $\lim_{x \rightarrow \infty} \left[ \frac{1-(F+\Delta F)}{1-F} - 1 \right] = 0$ , which shows that errors  $\frac{\Delta F_i}{1-F_i} = \Delta L$  need to be minimized. As a result, the best tail fit is also achieved with  $w_i = 1$ , similar to case  $B$ .

**Case G:** The objective is to model tails so that we are consistent with their subsequent use in a structural reliability analysis.

This subject is discussed in more detail in the following section. It will be shown that for GPD tail fits on independent basic random variables, the case  $B$  with unit weights in (10.1) is the most appropriate choice.

## 11. Tail sensitivity in structural reliability applications

Consider a set of basic random variables  $X = \{X_1, \dots, X_m\}$  and a failure set  $F$  described by  $\{X|g(X) < 0\}$ . When the failure probability  $P(F) = \Pr(g(X) < 0)$  is small, then the tail behaviour of (at least) some of these variables will have a critical effect on  $P(F)$ . The question addressed

here is exactly how sensitive  $P(F)$  is to errors on the tail of one the basic variables, say variable  $X_j$  having a minus-log-exceedance function  $L_j(x)$  and a loglikelihood function  $l_j(x)$ .

As shown in Maes (1991) as well as from the previous discussion of tail evaluation criteria, errors on  $[-\ln P(F)]$  often represent an appropriate measure of risk than errors on  $P(F)$  itself. Consequently, we will focus on the effect of  $\Delta(-\ln P(F))$ . According to Breitung and Faravelli's (1994) asymptotic result describing the sensitivity of  $P(F)$  to  $x_j$ ; we have

$$\frac{\partial P(F)}{\partial x_j} \sim P(F) \frac{\partial l(\mathbf{x}^*)}{\partial x_j} \quad (11.1)$$

where  $l(\mathbf{x}^*)$  is the loglikelihood function evaluated at the point of maximum likelihood (PML)  $\mathbf{x}^*$ . This point is the unique point (if it exists) which maximizes  $l(\mathbf{x})$  subject to  $g(\mathbf{x}) < 0$ . If the basic random variables are independent, then it follows that:

$$\frac{\partial \ln P(F)}{\partial x_j} \sim l'_j(x_j^*) = \left( \frac{f'_j}{f_j} \right)_{x_j^*} . \quad (11.2)$$

The error on  $\ln P(F)$  due to a tail modelling error  $\Delta L_j$  on the  $j$ -th variable can consequently be written as:

$$\Delta(-\ln P(F)) = \left[ - \left( \frac{\partial \ln P(F)}{\partial x_j} \right) \left( \frac{\partial L_j}{\partial x_j} \right)^{-1} \Delta L_j \right]_{x_j^*} , \quad (11.3)$$

to which the results (4.8) and (3.3) can be applied, to yield:

$$\Delta(-\ln P(F)) \sim \left[ - \frac{f'_j}{f_j} \cdot \frac{1 - F_j}{f_j} \Delta L_j \right]_{x_j^*} \quad (11.4)$$

or,

$$\Delta(-\ln P(F)) \sim \frac{1}{Q_j(x_j^*)} \Delta L_j \sim [1 + H_j(x_j^*)] \Delta L_j . \quad (11.5)$$

Two conclusions may be drawn from this result. First, expression (11.5) shows that if the tail behaviour of variable  $X_j$  is modelled using the GPD, then because of (9.3), the factor  $[1 + H_j]$  is constant over the entire tail region, and the effect of an error  $\Delta L_j$  anywhere on the tail is homogeneous and directly proportional to the error on  $\ln P(F)$ . Consequently, no additional weights ( $w_i = 1$ ) are required in the SWSE criterion (10.1), that is, the base case  $B$  of the previous section is applicable.

The second conclusion concerns the matter of tail sensitivity. If the upper or lower tail regions of  $k$  independent random variables contribute to  $P(F)$ , then the uncertainty on  $\ln P(F)$  as a result of the tail modelling uncertainties  $\sigma_{L_j}^2$  can be expressed as:

$$\sigma_{\ln P(F)}^2 \cong \sum_{j=1}^k (1 + H_j)^2 \sigma_{L_j}^2 . \quad (11.6)$$

As discussed above, there is no need to specify that  $H_j$  be evaluated at the PML if GPD tails are used. If we exclude the exceptional cases for which  $H_j \leq -1$ , this results shows that  $\ln P(F)$  is considerably more sensitive to heavy tails, where  $H_j > 0$ , than to light tails, where  $H_j < 0$ . For instance, in the case of a limit state function using both normal and lognormal variables, the modelling of the tails of the latter variables has a much more critical effect on  $P(F)$  from the point of view of tail sensitivities. Similar to importance factors ( $\alpha$  factors) in FORM, we can also define the relative tail sensitivity of tail  $j$  as the factor:

$$(1 + H_j) / \left[ \sum_{j=1}^k (1 + H_j)^2 \right]^{\frac{1}{2}} . \quad (11.7)$$

## 12. Where is the tail?

We have seen that characterization of tails of random variables involves several challenging tasks. First, a probabilistic model had to be selected for large (or small) values of the random variable. Then, appropriate statistics and uncertainty modeling had to be used. But we avoided one important problem: we must decide where exactly the upper and lower tail region of the variable are located. In other words: how do we decide what is “tail”, and what is “central”? In this respect, a compromise must be struck between using too many data to define the tail, which introduces bias towards central values, and using too few which causes excessive scatter for the final estimates.

This raises questions with regard to the extent of the tail: how many data should be included in each tail? We will apply a minimum mean square error-criterion (MSE) to a specific parameter such as the tail heaviness index, an endpoint, an extreme quantile, a reliability index or any other decision-based measure of risk. In order to evaluate the MSE needed to determine the optimal tail range and to establish confidence intervals on tail-related parameters, a simple variation of the non-parametric bootstrap method is used.

The vast majority of statistical techniques are based on estimating the central part of the distribution model of a random variable. The distributional form can be parametric or non-parametric. In both cases some measure of distance (likelihood,  $x^2$ , moments) is used to define the misfit between the data and the theoretical (or expected) model. These measures are generally based on the central limit theorem.

When extrapolation beyond the highest or lowest data is needed, central methods often fail to produce accurate results. Moreover, in many practical cases, the tail behaviour of the variable can be considerably different from the bulk or central part of the data. In that case one wants to extrapolate from the empirical tail using only those data which yield relevant information for the actual or true tail behaviour. The extrapolation is based on the assumption of the continuity of the tail beyond the data; an assumption which-from a strict mathematical point of view-cannot be justified. The important question that is raised in this section is to find the value or threshold above (below), which the tail actually "starts", i.e., which part of the data can be labelled "tail" and which part "central".

In the following we concentrate on upper tails but the methodology easily carries over to lower tails. To distinguish between central and tail data, many practitioners use visual methods based on quantile plotting or probability paper. The standard procedure consists of drawing the empirical distribution or quantiles versus the expected ones and looking for steady trends in the extreme part of the plot (Gumbel, 1958; Castillo, 1988). In some cases, there exists a distinct threshold above which there is consistent behaviour. Due to the lack of data in others, the choice of a clear threshold is prone to considerable uncertainty or variation, and the picking of such levels requires creative and - one would hope - educated guessing.

The most powerful quantile plot is the so-called generalized Pareto-quantile plot (GP-quantile plot) as described in Caers and Rombouts (1996) and Beirlant et al. (1996b). This is the scatter plot of points

$$\left(-\log \frac{j}{n}, \log UH_{j,n}\right), \quad j = 1, \dots, n-1 \quad \text{for right tails,} \quad (12.1)$$

$$\left(\log \frac{j}{n}, \log UH_{n-j,n}\right), \quad j = 1, \dots, n-1 \quad \text{for left tails,} \quad (12.2)$$

where

$$UH_{k,n} = X_{n-k}^* H_{k,n} \quad (12.3)$$

and

$$H_{k,n} = \frac{1}{k} \sum_{j=1}^k (\log X_{n-j+1}^* - \log X_{n-k}^*). \quad (12.4)$$

Here  $j$  is the index that defines the scatter of points on the quantile plot, while  $k$  denotes the number of highest order statistics used to calculate the statistics  $UH$  and  $H$ . The justification for the selection of the  $UH$  statistics is given in for instance Smith (1987) and Beirlant et al. (1996a, 1996b). At this stage, it is sufficient to point out that the plot matches the extreme behaviour associated with Eq. (9.1).

The GP-quantile plot ultimately becomes linear as one takes higher thresholds (Caers, 1996; Beirlant et al., 1996b). Therefore, the main property of the plot is that the slope of a line fitted using a weighted mean square error criterion to the  $k$  highest data above that threshold converges in probability to the extreme value index  $\xi$ . In the case of Gumbel domain

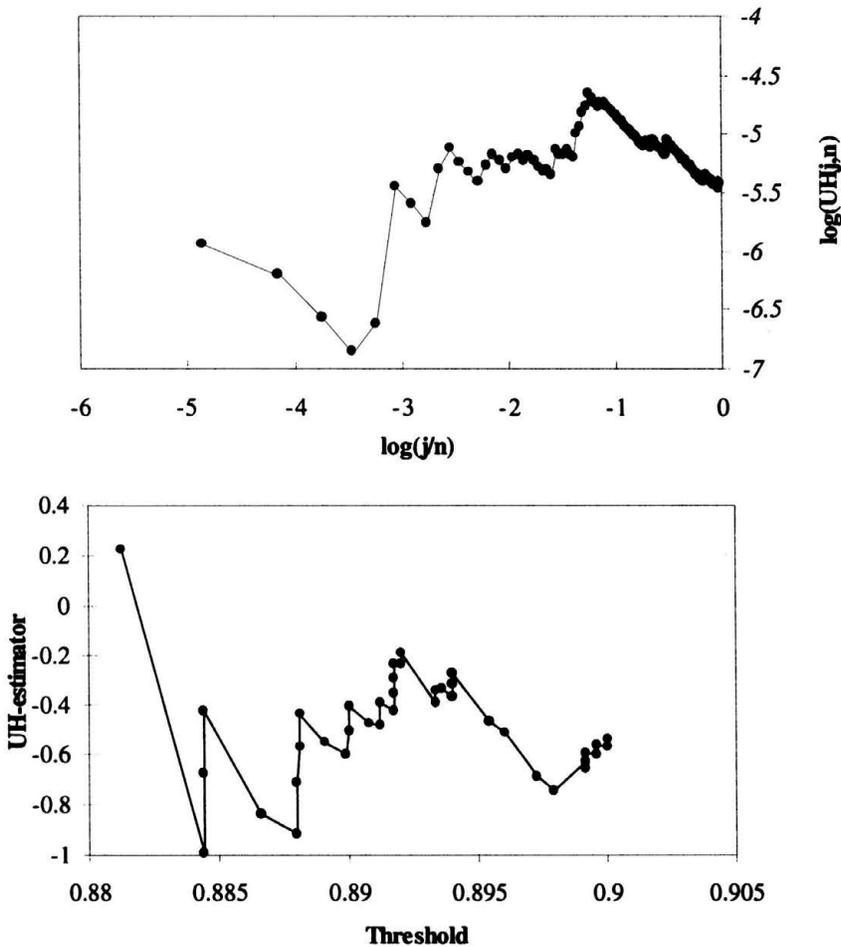


FIGURE 12. Wall thickness data: GP-quantile plot (top), UH-estimator (bottom).

attraction, one expects the GP-quantile plot to level to a horizontal line, in the case of Pareto-type tails and bounded tails, to a positive and negative slope, respectively.

Even so, the use of GP-quantile plot does not eliminate the problem of identifying a clear threshold between the central part or the upper and lower tail. To illustrate some aspects of the resulting "guestimation", consider a set of wall thickness data as measured on casing tubes exiting the steel mill. Typically, these data have some form of lower bound because of the nature of the manufacturing process. Here we consider the ratio of the actual wall thickness to the nominal wall thickness. Figure 12 shows the GP-quantile plot and estimated slopes for the lower tail of the wall thickness data.

The plot of the estimated slope versus the threshold  $u$  (Fig. 12(b)) reveals that as the threshold moves down (and  $k$  decreases), the slope seems to vary around a value of  $\xi \simeq -0.5$  to  $-0.8$ , which means the data are bounded. However the overall variability increases:  $\xi = -0.994$  at  $u = 0.884$ ,  $\xi = +0.221$  at  $u = 0.881$ . The same observation applies to the GP-quantile plot in Fig. 12(a). At first, the central or bulk part of the data shows a positive slope, in contrast to the negative slope in the extreme part. Then as one looks more and more to the left of the plot, the variability increases. Therefore, the choice of the threshold has important repercussions on the estimated value of  $\xi$ , even in the downward trend of the quantile plot. The amount of extremes above a threshold has its own sampling distribution. Parametric methods such as the maximum likelihood method (Smith, 1987, 1989) take a fixed threshold and a fixed amount of extremes above that threshold.

In a sense, the GP quantile plot is similar to the aforementioned  $(L, x)$  plot and the Gumbel plot (Gumbel, 1958). This latter is a plot of the ordered data on the  $x$ -axis versus the Gumbel quantiles on the  $y$ -axis

$$\left( X_{n-j+1}^*, -\log \left( -\log \frac{j}{n+1} \right) \right), \quad j = 1, \dots, n \quad \text{for right tails,} \quad (12.5)$$

$$\left( X_j^*, -\log \left( -\log \frac{n+1-j}{n+1} \right) \right), \quad j = 1, \dots, n \quad \text{for left tails.} \quad (12.6)$$

Whereas in the GP-case one always expects a *linear* behaviour, the Gumbel plot is either linear ( $\xi = 0$ ), convex ( $\xi > 0$ ) or concave ( $\xi < 0$ ), but always monotonically ascending. A similar method of finding regression estimators is used by Castillo (1988) on the Gumbel plot. He suggests to look for the ratio of the mean slope in two neighbouring zones, the quotient of this slope being a measure of curvature. This quotient is then used as a statistic to decide on the tail behaviour. In this respect, the GP-plot seems more powerful since it is easier to detect linearity than to detect convexity or concavity, certainly when

only few data samples are available. In Caers and Maes (1998) it is shown that the GP-quantile plot leads to a simple estimator which is efficient in estimating negative values of  $\xi$ .

However, the challenge remains: how does one choose a "good" threshold  $u$ ? The threshold cannot be too low, since in that case, too many central values will disturb the estimation in the sense that it introduces bias. On the other hand, as less data points are available above higher thresholds, the variance increases considerably. In response to this question, Boos (1984) recommends that the ratio of  $k$  (number of tail data) over  $n$  (total number of data) should be  $k/n = 0.02$  ( $50 < n < 500$ ) and  $k/n = 0.1$  for  $500 < n < 1000$ . In a recent discussion, Hasofer (1996) suggests to use  $k \simeq 1.5\sqrt{n}$ .

Due to the trade-off between the conflicting trends of bias and variance, we propose to use a finite sample mean square error (MSE) as a criterion for estimating the threshold. The idea is to select as an optimal threshold the one which minimizes this MSE. Suppose that one is interested in estimating some extremal property,  $\theta$ , which can be the THI, an upper bound  $\omega$ , a high quantile, or a safety index, or a failure probability. It is important to realize that the choice of the estimator is entirely dictated by the scope of decision we ultimately need to make. The estimator of the (fixed but unknown)  $\theta$  is denoted as the random variable  $\hat{\theta}$ . Then the MSE can be applied to  $\hat{\theta}$  as follows

$$\begin{aligned} \text{MSE}(\hat{\theta}) &= E \left[ (\hat{\theta} - \theta)^2 \right] = \left( E(\hat{\theta} - \theta) \right)^2 + E \left[ \left( (\hat{\theta} - \theta) - E(\hat{\theta} - \theta) \right)^2 \right] \\ &= \left( E[\hat{\theta}] - \theta \right)^2 + E \left[ \left( \hat{\theta} - E[\hat{\theta}] \right)^2 \right] = \text{bias}^2(\hat{\theta}) + \text{var}(\hat{\theta}). \end{aligned} \tag{12.7}$$

The MSE can now be evaluated for any threshold  $u$ , based on the estimate of  $\hat{\theta} = \hat{\theta}(u)$  using the data above the threshold  $u$ . The first term in (12.7) usually increases when the threshold is lowered, while the second one increases when the threshold is increased and the plot becomes more scattered. The dependence on  $u$  of the MSE becomes apparent. Minimizing the MSE can be contrasted with recent work in extreme value statistics which, instead, concentrates on optimizing the asymptotic mean square error (ASME), (Beirlant et al. 1996a,b) in order to estimate  $\xi$ . Expressions for bias and variance under the asymptotic conditions of some estimators of  $\xi$ , for example, exist in literature (Beirlant and Teugels, 1986; Dekkers and de Haan, 1989; Csörgo et al., 1985).

It is important to note that the proposed MSE-criterion can be applied to any extremal property, such as an extreme quantile, an exceedance probability, an end-point or a reliability index and that it does not require the

asymptotic conditions needed for AMSE which, in any case, seem unrealistic for small samples.

Expressions for the finite sample MSE as in (12.7) are generally not readily available for most estimators and the question remains how to estimate MSE. A natural choice is to use the bootstrap as an estimate for finite sample bias and variance at each threshold. Details of this procedure and the subsequent question of finite sample confidence intervals are discussed in Caers and Maes, (1998).

### 13. Conclusions

Tail heaviness plays an important role in all of risk analysis and reliability-based design. Following the fundamental principle of tail equivalence, any tail can be represented by a GPD with selected tail heaviness coefficient. Risk-based tail estimation criteria are seen to be dependent on the objectives of the analysis. In most of the cases discussed here, least squares in an  $(L, x)$  plot is shown to be consistent with these objectives. Tail sensitivity in structural reliability is formulated asymptotically in terms of the individual tail heaviness indices. The use of an MSE-based criterion on the tail/extreme parameters that will ultimately influence the decision we need to make is suggested to decide on the optimal extent of the tail to be used in a GPD tail analysis.

### References

1. BEIRLANT, J., and TEUGELS, J.L. (1989), Asymptotic normality of Hill's estimator, *Lecture Notes in Statistics*, No.51, Extreme value theory, Proceedings of the Oberwolfach, 1987, New York: Springer-Verlag.
2. BEIRLANT, J., VYNCKIER, P., and TEUGELS, J.L. (1996), Excess functions and estimation of the extreme value index, *Bernoulli*, No.2, pp.293-318.
3. BEIRLANT, J., VYNCKIER, P., and TEUGELS, J.L. (1996), Tail index estimation, Pareto-quantile plots and regression diagnostics, *Journal of the American Statistics Association*, 91, pp.1659-1667.
4. BOOS, D.D. (1984), Using extreme value theory to estimate large percentiles, *Technometrics*, Vol.2, No.1, pp.33-39.
5. BREIMAN, L., STONE, C.J., and GINS, J.D. (1979), New methods for estimating tail probabilities and extreme value distributions, Technology Service Corporation-PD-A226-1, Santa Monica, California.
6. BREITUNG, K., and FARAVELLI, L. (1994), Loglikelihood maximization and response surface reliability assessment, *International Journal of Nonlinear Dynamics*, Vol.5, No.3, pp.273-286.

7. CAERS, J., and MAES, M.A. (1998), Identifying tails, bounds, and end-points of random variables structural safety, *Structural Safety*, No.20, pp.1-23.
8. CASTILLO, E. (1992), Engineering analysis of extreme value data: selection of models, *Journal of Waterway, Port, Coastal and Ocean Engineering*, ASCE, Vol.118, No.2, pp.129-146.
9. CASTILLO, E. (1988), *Extreme Value Theory in Engineering*, Academic Press, San Diego, California.
10. CSÖRGO, S., DEHEUVELS, P., and MASON, D. (1985), Kernel estimates of the tail index of a distribution, *Annual Statistics*, No.13, pp.1050-1077.
11. DAVID, H.A. (1981), *Order Statistics*, (2<sup>nd</sup> ed.), John Wiley & Sons, New York.
12. DAVISON, A.C., SMITH, R.L. (1989), Models for exceedances over high thresholds, *Journal of Statistics*, Soc. B, Vol.52, No.2.
13. DEKKERS, L.M., DE HAAN, L. (1989), On the estimation of the extreme-value index and large quantile estimation, *Annual of Statistics*, No.17, pp.1795-1832.
14. DITLEVSEN, O. (1993), Distribution arbitrariness in structural reliability, *Proceedings of the ICOSSAR 1993*, Innsbruck, Vol.2, pp.1241-1247.
15. FISHER, R.A., TIPPET, L.H.C. (1928), Limiting forms of the frequency distribution in the largest particle size and smallest member of a sample, *Proceedings of the Cam. Phil. Society*, Vol.24, pp.180-190.
16. FLYNN, P. (2002), Financial bailout of september 11: rapid response, *Challenge*, Vol.45, No.1, pp.104-116.
17. FRECHET, M. (1927), Sur la loi de probabilité de L'Écart maximum, *Annales de la Soc. Polonaise de Mathematics*, Cracow, Vol.6, 93.
18. GNEDENKO, B. (1943), Sur la distribution limite du terme maximum d'une série aléatoire, *Annales de la Mathematics*, Vol.44, No.3, July.
19. GRANT, C.K., HEIDEMAN, J.C., PETERS, D.J., SHAW, C.J. and SZABO, D. (1993), Modelling the North Sea through the North European Storm Study, *Proceedings OTC 93*, Houston, TX.
20. GUMBEL, E.J. (1958), *Statistics of Extremes*, Columbia University Press, New York.
21. GÜNTHER, H., and ROSENTHAL, W. (1983), Shallow water surf: wave model based on the Texel-Marsen-Arsloe (TMA) Wave Spectrum, *Proceedings of the 20<sup>th</sup> Congress of the International Association of Hydraulic Research (IAHR)*, Moscow/Soviet Union.
22. HASOFER, A.M. (1996), Parametric estimation of failures probabilities, In: *Mathematical Models for Structural Reliability Analysis*, eds. Casciati and Roberts, Boca Raton, Florida, CRC Press.
23. HOSKING, J., and WALLIS, J.R. (1987), Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, Vol.29, pp.339-349.
24. MAES, M.A. (1991), Codification of design load criteria subject to modeling uncertainty, *Journal of Structural Engineering Division*, ASCE, Vol.117, No.10, pp.2988-3007.

25. MAES, M.A., and BREITUNG, K. (1994), Reliability-based tail estimation, *Probabilistic Structural Mechanics: Advances in Structural Reliability*, eds. Spanos, P.D. and Wu, Y.T., Springer Verlag, New York, pp.335-346.
26. MAES, M.A. (1995), Tail heaviness in structural reliability, *Applications of Statistics and Probability*, Proceedings of the ICASP7 Conference, Paris, France, July 10-3.
27. MAES, M.A., and GU, G. (1994), Techniques used to determine extreme wave heights from the NESS data set, *Journal of Research of the National Institute of Standards and Technology*, Vol.99, No.4, July-Aug, pp.441-442.
28. PICKANDS, J. (1975), Statistical inference using extreme order statistics, *Annals of Statistics*, Vol.3, pp.119-131.
29. VON MISES, R. (1936), La distribution de la plus grande de  $n$  valeurs, *Rev. Math. de l'Union Interbalkanique*, Vol.1, No.1.
30. WEIBULL, W. (1951), A statistical distribution function of wide applicability, *Journal of Applied Mechanics*, No.18.

