Arvid Naess

Applied Extreme Value Statistics

With a Special Focus on the ACER Method

Springer

Applied Extreme Value Statistics

Arvid Naess

Applied Extreme Value Statistics

With a Special Focus on the ACER Method

Springer

Arvid Naess
Faculty of Information Technology,
Mathematics and Electrical Engineering
Norwegian University of Science and
Technology
Trondheim, Norway

ISBN 978-3-031-60768-4 ISBN 978-3-031-60769-1 (eBook)
https://doi.org/10.1007/978-3-031-60769-1

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To Dorothea and Vemund

Preface

This book grew out of many years of involvement with the practical applications of extreme value analysis to measured or simulated data. This is a fascinating area of research because of the fundamental dichotomy inherent in this problem area. On the one hand, you have beautiful mathematical results for asymptotic extreme value distributions. On the other hand, you have real-life data, which are hardly asymptotic. So, the unavoidable question becomes: To what extent can you use the asymptotic distributions to analyze real-life data? Personally, I have always felt uncomfortable with the use of the parametric classes of asymptotic extreme value distributions in applications. This was largely due to the fact that the justification for applying them generally seemed dubious, and amazingly enough, the problem of justification is rarely discussed at all in papers using asymptotic distributions on real-life data. The problems of justification and other issues related to the fundamental dichotomy are discussed in Chap. 1.

Of course, I was not the only one who disliked asymptotics for use on real-life data. A consequence of this situation was that alternative procedures for extreme value analysis were developed in several engineering disciplines. Some of these alternative procedures were based on ideas similar to those developed in Chap. 4. This chapter contains what was largely my world view on applied extreme value statistics for quite some time, and to some extent, it still is. However, the development of the ACER method, which is a central theme in this book, cf. Chap. 5, allows for a much wider perspective. Its use in practice basically involves two separate steps. The first step is based exclusively on the data and ends up with a nonparametric representation of the extreme value distribution inherent in the data. This is the crucial element of the ACER method. The second step consists of an optimization procedure to fit a parametric function to the nonparametric distribution. This step is necessary in order to be able to predict extremes larger than those contained in the data, which is typically demanded by applications. To develop a rational method of optimized fitting, the asymptotic extreme value theory by necessity becomes an essential ingredient which guides the construction of the parametric functions used in the optimization procedure. It is therefore necessary to identify the correct asymptotic extreme value distribution, since the fitted extreme value distribution by necessity must approach the relevant asymptotic form in the limit.

Even if my own work in developing methods for use in applied extreme value analysis more or less avoided direct use of the asymptotic extreme value distributions, I have always clearly understood their importance as an unavoidable foundation. The publication in 1983 of the important book *Extremes and Related Properties of Random Sequences and Processes* by Leadbetter, Lindgren, and Rootzén happened when my own interest in extreme value analysis more or less started. I, therefore, read this book very carefully, and it gave me a very good grip on the asymptotic extreme value theory. Of course, I also read parts of the seminal book by E. J. Gumbel, published in 1958, which also has a focus on asymptotic results. However, by the mid-1980s, that book, which was written in a pre-computer era, appeared as more or less obsolete when compared to the book by Leadbetter et al.

I have written this book, not because I want to convince people to abandon the classical asymptotic approaches, which, unfortunately, too often in practice are reduced to blindfolded curve fitting exercises to asymptotic parametric distributions with no real analysis to back it up. No, I have written the book because I would also like to show that it is now possible to make a more rationally based extreme value analysis of observed data. I want to show that the ACER method very often provides a unique practical diagnostic tool for a rational extreme value analysis. If, as a result, asymptotic distributions turn out to be more or less acceptable, then their use would at least have a reasonable justification.

It is also important to emphasize that the book is not a comprehensive treatment of methods for applied extreme value analyses, but to a large extent a collection of methods that I have personally worked with on and off over a period of three decades, and which I have found to be relevant and useful. I have made an effort to write the book as much as possible like an introduction to extreme value statistics with emphasis on applications. Therefore, the book also contains introductory chapters to the classical asymptotic theories and the threshold exceedance models, as well as many illustrative examples. The mathematical level is elementary, and detailed mathematical proofs have been avoided in favor of heuristic arguments to increase readability. Hopefully, this makes the book useful and appealing to a large audience of people representing a wide range of diverse applications.

Since the topic of this book is applied extreme value statistics, an inevitable component to go along with it, is access to computer programs for carrying out the analysis of available data. For the methods based on the asymptotic results described in Chaps. 2 and 3, there are several excellent programs easily available. Specific recommendations are not given here. Whichever program is chosen, good results can be obtained within the framework of asymptotic distributions. On the other hand, the ACER method has not yet attained a comparable level of software development. References to computer programs for univariate and bivariate analyses by the ACER method have therefore been given in this book. These programs can be freely downloaded.

Writing on the technical level necessary for this book requires a lot of attention to details. It is in practice impossible to avoid errors and mistakes, poorly formulated explanations, or misprints in initial versions of such a book. Fortunately, I have some very good friends and colleagues who have helped me identify and correct many such shortcomings, and for this, I am forever thankful. Any mistakes, which may still remain, are entirely my own responsibility. The first group of people that I would like to mention for their important contributions to improving the book are Professors Bernt J. Leira, Bo H. Lindqvist, and Sverre Haver and Dr. Karl W. Breitung. Previous collaborators and PhD students that have been important in helping me in various ways are Professors Torgeir Moan, Oleg Gaidai, Sjur Westgaard, Marc Maes, Nilanjan Saha, Wei Chai, and Arild Brandrud Næss; Drs. Oleksandr Batsevich, Oleh Karpa, Ali Cetin, Hans K. Karlsen, and Kai Erik Dahlen; and Morten Skjong. I am also very grateful to my many good students that I have had the pleasure of working with over the years, who have also inevitably been part of my own never-ending education as a researcher.

Trondheim, Norway
March, 2023

Arvid Naess

Contents

Chapter 1
Challenges of Applied Extreme Value Statistics

1.1 Introduction

This book provides an introduction to the calculation of extreme value statistics for measured or simulated data. "Extreme" here means "the largest", interpreted in a way that follows from the context. As opposed to books on asymptotic extreme value statistics, the focus is also on methods specifically developed to work for real-life data. A consequence of this, is that the book contains much less theoretical issues about the asymptotic properties of extreme value statistics than is usual. However, the most important elements from the asymptotic extreme value statistics will be discussed, since they are still widely used in practical applications.

Although two of the asymptotic methods described in this book have been used extensively over several decades for prediction of the extreme value statistics of many natural phenomena, the prerequisites for their application are often not satisfied, and in some cases, not even approximately. Under such circumstances, there would appear to be a problem. It is this situation that will be highlighted in this chapter.

1.2 A Brief Summary of Status, Problems and Challenges

Statistical distributions of the extreme values of large samples of data were derived almost one hundred years ago by Fisher and Tippett (1928), cf. also Fréchet (1927); Gnedenko (1943); de Haan (1970). The main prerequisite for the existence of the derived results were that the data could be considered as outcomes of independent and identically distributed random variables. As it turned out, in non-degenerate cases there are only three possible types of limiting extreme value distributions with increasing amounts of data. It means that these results are asymptotic, as the technical term goes. On the positive side, the fact that we know explicitly what the

A. Naess, *Applied Extreme Value Statistics*,
https://doi.org/10.1007/978-3-031-60769-1_1

possible distributions look like, even if only in the limit of large samples increasing indefinitely, is very satisfactory. And there are criteria that can tell us which type of distribution applies if the underlying distribution of the data is known (Leadbetter et al. 1983). However, on the negative side, it is not possible to know to what extent one of the three types of limiting distributions actually applies to a real-life case with only a limited amount of data, even though there may be reasons to expect that the true extreme value distribution should not deviate too much from one of the limiting forms. Unfortunately, there are no useful convergence results that are precise enough to really help us decide quantitatively on this issue. Still, the common practice has been to assume an appropriate limiting form as the extreme value distribution to use. This can easily be understood from the simple fact that the limiting distributions are known explicitly, while the exact extreme value distributions inherent in the data, are largely unknown. The procedure to identify the appropriate limiting distribution is to optimize the fit of the extreme values derived from the observed data to the asymptotic forms. Typically, the extreme values from the data are taken as the maxima of specified blocks of data, e.g. annual maxima.

The three asymptotic types of extreme value distributions are essentially characterized by the value of one parameter, γ say, called the shape parameter. As will be seen later, the most important case for us in this book is when $\gamma = 0$. This is called a Type I, or Gumbel, distribution. For positive values of γ, Type II, or Fréchet, distributions are obtained, while for negative values of γ, the distributions are of Type III, or Weibull (for maxima). As it turns out, all three distribution types may be expressed in terms of one parametric form called the generalized extreme value (GEV) distribution. A standard recommendation is then to use the GEV parametric form for the sake of optimized fitting of the obtained extreme value sample. There is, however, one serious flaw with this procedure. The extreme value sample being extracted from limited amounts of data, are hardly a sample from an asymptotic distribution. Hence, one cannot expect that the estimated parameters will point to the correct asymptotic distribution. This is an issue of importance for extrapolation to out-of-sample long return period levels. For example, a practical task may be to say something about a 100 year return period level on the basis of 25 years of measured data. Then the correct asymptotic distribution is of paramount importance because the different types of extreme value distribution may lead to quite different extrapolation results. An additional issue is, of course, that with limited amounts of data follows considerable uncertainty on the estimated quantities. It may, in fact, happen that the estimated value of γ is slightly negative, pointing to a Type III distribution, but with the confidence interval accounted for, also $\gamma = 0$, or even $\gamma > 0$ are possible candidates for the value of γ. Hence, all three types of extreme value distributions seems to be possible alternatives in such a case. Since these asymptotic distributions have very different behaviour when extrapolated to high quantile values, the previous comments on the importance of this aspect, would often necessitate a more careful analysis of the situation to decide which asymptotic distribution to apply.

The peaks-over-threshold (POT) method for extreme value analysis will be discussed to some extent in this book. This method is also based on asymptotics.

The data extracted for its use, are the exceedances above high thresholds. Asymptotically, these data are assumed to follow a generalized Pareto (GP) distribution, which is then the equivalent of the GEV for the block maxima method. The POT method also has three classes of distributions, again characterized by the γ parameter. For example, the singular case $\gamma = 0$ corresponds to the exponential distribution. It is a rather popular method, mainly because it uses more of the data for inference. Unfortunately, it has certain deficiencies, which will be highlighted in this book.

There is an important and interesting observation to be made at this initial stage of our exposition of extreme value statistics. As already been stated, for all negative values of the shape parameter γ, the Type III class of extreme value distributions apply, while for all positive values of γ, it is the Type II class of extreme value distributions that is obtained. This would seem to indicate that there are two huge classes of extreme value distributions that would tend to make the singular case $\gamma = 0$ a rather special and maybe uninteresting case. The fact of the matter is quite the opposite. For almost all environmental processes that will be dealt with in this book, it is the Gumbel distribution that has prevailed as the correct asymptotic extreme value distribution. There has over the years been some suggestions to the other types as well, but these have almost all been finally rejected in the face of overwhelming evidence for Type I distributions. Of course, it is impossible to fully answer the fundamental question: To what extent do our statistical models apply to real-life data? But so far, it seems that these statistical methods work rather well on such data, but being overconfident in these methods is perhaps an unwarranted position to take.

One important reason that the singular asymptotic Gumbel case is so important in practice, is that from the perspective of a sub-asymptotic world, the picture of the size of the extreme value distribution classes looks very different. When only limited amounts of data are available, the asymptotic limiting distributions, strictly speaking, do not apply, except in very special cases. Hence, we are in a sub-asymptotic situation. As will be seen in large parts of this book, there is a huge class of extreme value distributions that apply to a range of different problems, which all end up at the Gumbel distribution asymptotically. So, the apparent singularity of the Gumbel case is an artefact of the asymptotic limiting process, and does not reflect the situation in the sub-asymptotic regime.

In an effort to resolve the inconsistency between real-life data and asymptotic distributions, a new method has been developed that is based on the concept of the average conditional exceedance rate (ACER). The method proceeds by establishing a cascade of empirical, non-parametric distribution functions that converge to the extreme value distribution inherent in the data. The advantages of the method is that no assumptions about independent and stationary data have to be made. For example, seasonal variations of the data do not require special modelling. The method also has a unique diagnostic feature in how it displays the effect of dependence between the data on the extreme value distribution. This may be of significance for the choice of which data can be included in the analysis. The ACER method will be discussed in detail in this book.

Whatever method of extreme value statistics is chosen for the analysis of the available data, the goal is almost always to predict extreme values with return periods larger, and often much larger, than the period of data collection. This inevitably requires extrapolation techniques to be used. The seemingly stochastic mechanism generating the sampled data is often sufficiently well understood to support the assumption of the validity of extrapolation. Unfortunately, this may not always be the situation. Ideally, in such cases, the predicted extreme values obtained by extrapolation should then be accompanied by a cautionary note. However, this is rarely done, simply because more credible alternatives for the prediction process are not available.

The extrapolation procedure is, in general, based on obtaining estimates of the parameters that determine the extreme value distribution type adopted for the data at hand. If an asymptotic approach is used, the GEV distribution is often preferred for parameter estimation in the case of the block maxima method, or the GP distribution for the POT method. Since these are parametrized forms covering all three types of asymptotic extreme value distributions, it is often recommended to use these forms, allowing the data to determine which type of extreme value distribution to use. As already mentioned, such a procedure may not always be a good idea. Also for the ACER method, a parametrized family of functions is proposed for the purpose of extrapolation, which is tailored to reflect the sub-asymptotic character of the data.

The parameter estimates calculated for the examples in this book, are based on either the method of moments or the maximum likelihood method in the case of the GEV or the GP distributions. For the ACER method, the optimized fitting is obtained by using a Levenberg-Marquardt method on an objective function expressed as a weighted mean square deviation measure between the empirical and the proposed parametric ACER functions on the log level. Uncertainty quantification is also a very important aspect of any statistical inference. In this book, the use of bootstrapping will serve to illustrate this issue, since it has some attractive properties.

Chapter 2
Classical Extreme Value Theory

2.1 Introduction

Classical extreme value statistics is concerned with the distributional properties of the maximum of a number of independent and identically distributed (iid) random variables when the number of variables becomes large. A partial result was obtained by Fréchet (1927), while Fisher and Tippett (1928) discovered that there are three types of possible limiting or asymptotic distributions, which are now contained in the Extremal Types Theorem, which is discussed in the next section. These three asymptotic distributions are typically referred to as the Gumbel, Fréchet, and Weibull distributions. It is also common practice to refer to them as Type I, Type II, and Type III, in the same order. Important contributions to this theory were later made by Gnedenko (1943), Gumbel (1958), and de Haan (1970).

2.2 The Asymptotic Limits of Extreme Value Distributions

The classical extreme value theory starts by looking at a sequence of independent and identically distributed (iid) random variables $X_1, X_2, \dots$ with common distribution function $F_X(x)$. The extreme value of a finite number $X_1, \dots, X_n$ is then $M_n = \max\{X_1, \dots, X_n\}$. The distribution of M_n can be easily derived as

$$\begin{aligned} F_{M_n}(x) &= \text{Prob}(M_n \le x) = \text{Prob}(X_1 \le x, \dots, X_n \le x) \\ &= \text{Prob}(X_1 \le x) \cdot \ldots \cdot \text{Prob}(X_n \le x) = \big(F_X(x)\big)^n . \end{aligned} \tag{2.1}$$

This relation is not very helpful in practice, because in most cases the distribution function $F_X(x)$ is not known exactly. Therefore, it would have to be estimated from recorded data. However, small discrepancies in the estimates of $F_X(x)$ can lead

A. Naess, *Applied Extreme Value Statistics*,
https://doi.org/10.1007/978-3-031-60769-1_2

to substantial discrepancies, in a relative sense, in the values of $\left(F_X(x)\right)^n$ for large values of n. In classical extreme value theory, one proceeds by studying the behavior of $\left(F_X(x)\right)^n$ as $n \to \infty$, but with a twist. Obviously, for any x such that $F_X(x) < 1$, $\left(F_X(x)\right)^n \to 0$ as $n \to \infty$. This necessitates a rescaling. Specifically, instead of studying M_n, one introduces a renormalized version of M_n:

$$M_n^* = \frac{M_n - b_n}{a_n} \tag{2.2}$$

for suitable sequences of constants $a_n > 0$ and b_n that are chosen to stabilize the location and scale of M_n^* as $n \to \infty$. It is then proven that there are, in fact, only three types of limiting distributions for this renormalized M_n^*. This is the famous Extremal Types Theorem (Leadbetter et al. 1983), which can be expressed as follows.

If there exist sequences of constants $a_n > 0$ and b_n such that

$$\text{Prob}\left(\frac{M_n - b_n}{a_n} \le x\right) \to G(x)\,, \quad n \to \infty, \tag{2.3}$$

where $G(x)$ is a nondegenerate distribution function, then $G(x)$ belongs to one of the following three families:

$$\text{I} \qquad G(x) = \exp\left\{-\exp\left[-\left(\frac{x-b}{a}\right)\right]\right\}\,, \quad -\infty < x < \infty\,; \tag{2.4}$$

$$\text{II} \qquad G(x) = \begin{cases} 0 & , \quad x \le b\,, \\ \exp\left\{-\left(\frac{x-b}{a}\right)^{-c}\right\} & , \quad x > b\,; \end{cases} \tag{2.5}$$

$$\text{III} \qquad G(x) = \begin{cases} \exp\left\{-\left(\frac{b-x}{a}\right)^{c}\right\} & , \quad x < b\,, \\ 1 & , \quad x \ge b\,; \end{cases} \tag{2.6}$$

for parameters $a > 0$, b and for families II and III, $c > 0$.

These three types of extreme value distributions are also commonly referred to as Gumbel, Fréchet, and Weibull, respectively. Note that the Weibull distribution given here is not the same as the commonly known Weibull distribution, which corresponds to the type III extreme value distribution for minima. Also, carefully note that even if the Weibull distribution is the only type of extreme value distribution with a finite upper limit on its values, this does not mean that extremes of limited data must follow this distribution. For such data, it may very well happen that the rescaling constant $a_n \to 0$ as n increases. Hence, even the Gumbel distribution may be the appropriate asymptotic limit for the extreme values of bounded data.

It may be verified that it is, in fact, possible to express all three types of extreme value distributions in a common form, which is known as the generalized extreme value (GEV) distribution. This is achieved as follows:

$$G(x) = G(x; \mu, \sigma, \gamma) = \exp\left\{ -\left[1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\gamma} \right\}, \tag{2.7}$$

defined on the set $\{x : 1 + \gamma\big((x-\mu)/\sigma\big) > 0\}$, where the parameters satisfy $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \gamma < \infty$. This distribution has three parameters: a location parameter μ, a scale parameter σ, and a shape parameter γ. The type II distributions correspond to $\gamma > 0$, while type III corresponds to $\gamma < 0$. The case $\gamma = 0$ must be interpreted as a limiting case when $\gamma \to 0$, which leads to the Gumbel distribution:

$$G(x) = \exp\left\{ -\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\}, \quad -\infty < x < \infty. \tag{2.8}$$

The statistical moments of the GEV distributions can now be calculated based on the explicit formulas of Eqs. (2.7) and (2.8). Denoting the random variable determined by a GEV distribution by M, its first two moments are

$$\mathbb{E}(M) = \mu + (e_1 - 1)\frac{\sigma}{\gamma}, \quad \gamma \neq 0, \gamma < 1 \tag{2.9}$$

and

$$\mathrm{Var}(M) = \left(e_2 - e_1^2\right)\frac{\sigma^2}{\gamma^2}, \quad \gamma \neq 0, \gamma < 1/2, \tag{2.10}$$

where $e_k = \Gamma(1 - k\gamma)$, $k = 1, 2$, and $\Gamma(\cdot)$ is the gamma function. For $\gamma \geq 1$, $\mathbb{E}(M) = \infty$, while $\mathbb{E}(M) = \mu + \lambda_E\, \sigma$ when $\gamma = 0$, that is, for the Gumbel case. Here, $\lambda_E = 0.5772...$ denotes Euler's constant. For $\gamma \geq 1/2$, $\mathrm{Var}(M) = \infty$, while $\mathrm{Var}(M) = \sigma^2\pi^2/6$, when $\gamma = 0$.

For statistical inference on experimental data, the unified form expressed by Eq. (2.7) has the advantage that the data themselves determine which type of distribution is appropriate, thereby avoiding a prior subjective judgment about any specific tail behavior. The uncertainty in the estimated value of γ is also a reflection of the uncertainty about the correct distribution for the data. Unfortunately, in practice, it may very well happen that the uncertainty in γ may cover all three types of extreme value distribution, which would necessitate a more careful analysis of the data. Also note that the data used for estimation purposes are never truly asymptotic, thereby introducing additional uncertainty when trying to identify the correct asymptotic distribution. Since the results of extrapolation to determine long return period design values may depend very much on the asymptotic extreme value distribution used, identifying the correct one is clearly important in such cases.

2.3 The Block Maxima Method

In practical application of the GEV distributions to a long time series of observed data, it is assumed that the maximum observation of a reasonably large chunk of the time series follows a GEV distribution. This is recognized by observing that from (2.3) we would assume that for large n,

$$\mathrm{Prob}\Big(\frac{M_n - b_n}{a_n} \le x\Big) \approx G(x)\,. \tag{2.11}$$

But this may be rewritten as ($y = a_n x + b_n$)

$$\mathrm{Prob}\big(M_n \le y\big) \approx G\Big(\frac{y - b_n}{a_n}\Big) = G^*(y)\,, \tag{2.12}$$

where G^* is also a member of the GEV family of distributions. Hence, if the main theorem applies, that is, by (2.3), $M_n^* = (M_n - b_n)/a_n$ approximately follows a GEV distribution, then M_n itself will approximately follow a GEV distribution, but with different parameters. Anyway, in practice, it is the parameters of G^* that would be of most interest.

This leads to the following approach, which is often referred to as the block maxima method. Assume that a sequence of independent observations $x_1, x_2, \ldots$ from a stationary time series is long enough to allow segmenting it into blocks of data of length n, for some large value of n, generating a series of m block maxima, $M_{n,1}, \ldots, M_{n,m}$, say, to which a GEV distribution is tentatively fitted. A typical application of the block maxima method would be to yearly extreme value observations of an environmental parameter, e.g., wind speed. In such a case, it is also often referred to as an annual maxima method. There is a practical argument behind extracting the maximum over the period of 1 year, because by choosing shorter periods, the assumption that the sampled maxima are outcomes of a common distribution would easily be violated due to seasonal variations. Still, of course, the underlying assumption that the block maxima are extracted from a set of iid random variables is clearly violated. Fortunately, by experience, this does not seem to pose a serious obstacle to the practical use of the block maxima method.

A quantity of specific interest in applications is the return period level x_p, where $G(x_p) = 1 - p$. For the annual maxima method, x_p has a return period of $1/p$ years. That is, x_p would be exceeded on the average every $1/p$ years. Inverting (2.7), it is found that for $\gamma \neq 0$,

$$x_p = \mu - (\sigma/\gamma)\big[1 - \big(-\log(1-p)\big)^{-\gamma}\big]\,, \tag{2.13}$$

while for $\gamma = 0$,

$$x_p = \mu - \sigma \log\big(-\log(1-p)\big)\,. \tag{2.14}$$

Coles (2001) discusses how to estimate confidence intervals on x_p using profile likelihood methods, which seem to provide reasonable accuracy. In this book the focus is on the bootstrap method, cf. Sect. 2.8.

2.4 Outline Proof of the Extremal Types Theorem

The proof of the Extremal Types Theorem is not a very complicated proof, but it is rather lengthy and technical (Leadbetter et al. 1983). Since it is not central to the focus of this book, only a sketch will be given here to illustrate the main ingredients. The concept of max-stability is needed. It is defined as follows:

A distribution G is called max-stable if, for every $m = 2, 3, \ldots$, there are constants $\alpha_m > 0$ and β_m such that

$$G^m(\alpha_m x + \beta_m) = G(x)\,. \tag{2.15}$$

G^m is the distribution function of $M_m = \max\{Z_1, \ldots, Z_m\}$, where the Z_i are iid random variables with distribution function G. Therefore, max-stability is a property satisfied by distributions that are invariant under the operation of taking sample maxima, except for a change of scale and location. The following result brings forward the connection between max-stability and extreme value distributions (Leadbetter et al. 1983),

A distribution is max-stable if, and only if, it is a GEV distribution.

To check that a GEV distribution is max-stable is a straightforward exercise in algebra. The converse is much harder. Anyway, this result can now be used to prove the Extremal Types Theorem. Consider first $M_{nk} = \max\{X_1, \ldots, X_{nk}\}$ of a sample of nk iid random variables X_i, for some large value of n. This large sample can be divided into k subsamples of n variables in each. Hence, there will be k iid random variables like $M_n = \max\{X_1, \ldots, X_n\}$. n is chosen large enough to claim that

$$\text{Prob}\Big(\frac{M_n - b_n}{a_n} \le x\Big) \approx G(x)\,, \tag{2.16}$$

for suitable constants a_n and b_n and for the limiting distribution G. Hence, for any integer $k \ge 2$, since $nk > n$,

$$\text{Prob}\Big(\frac{M_{nk} - b_{nk}}{a_{nk}} \le x\Big) \approx G(x)\,, \tag{2.17}$$

Eq. (2.16) leads to $\text{Prob}\big(M_n \le z\big) \approx G\big((z - b_n)/a_n\big)$, while Eq. (2.17) gives $\text{Prob}\big(M_{nk} \le z\big) \approx G\big((z - b_{nk})/a_{nk}\big)$. However, M_{nk} is obviously the maximum of k variables having the same distribution as M_n. But then,

$$\text{Prob}\big(M_{nk} \le z\big) = \Big[\text{Prob}\big(M_n \le z\big)\Big]^k\,. \tag{2.18}$$

From this, it is deduced that (in the limit)

$$G\Big(\frac{z-b_{nk}}{a_{nk}}\Big) = G^k\Big(\frac{z-b_n}{a_n}\Big). \tag{2.19}$$

From this it follows that G and G^k are identical apart from location and scale parameters. Hence, G is max-stable, and by the result above, it is a member of the GEV family of distributions.

2.5 Domains of Attraction for the Extreme Value Distributions

In practice, the exact statistical distribution of the data being analyzed is rarely known. However, in many cases there may be rather strong evidence as to what type of distribution to expect. For instance, average wind speeds over periods of 10 minutes in northern Europe have been found to follow a Weibull type distribution. Then it would be useful to know what kind of extreme value distribution to expect for such data. The answer to such questions is the subject of the theory of domains of attraction for extreme value distributions. It is beyond the scope of our treatment of this topic here to go into much detail, but some useful results seem worthwhile presenting. A more thorough discussion is given by Gnedenko (1943) and Leadbetter et al. (1983).

A time series $X_1, X_2, \ldots$ of iid random variables with distribution function F, and with a density function f, is considered. x_F is defined to be the right endpoint of F by $x_F = \sup\{x;\, F(x) < 1\} \quad (x_F \le \infty)$. Then the following sufficient conditions due to von Mises apply:

Suppose that F is absolutely continuous with density f. Then sufficient conditions for F belonging to each of the three possible domains of attraction are as follows:

Type I: f has a negative derivative f' for all x in some interval (x_0, x_F), $(x_F \le \infty)$, and

$$\lim_{x \nearrow x_F} \frac{f'(x)(1-F(x))}{f^2(x)} = -1.$$

Type II: $f(x) > 0$ for $x \ge x_0$ finite, and for some constant $\alpha > 0$,

$$\lim_{x\to\infty} \frac{xf(x)}{1-F(x)} = \alpha.$$

Type III: $f(x) > 0$ for all x in some finite interval (x_0, x_F), $f(x) = 0$ for $x > x_F$, and for some constant $\alpha > 0$,

$$\lim_{x \nearrow x_F} \frac{(x_F - x) f(x)}{1 - F(x)} = \alpha.$$

Using these results, it is straightforward to verify that the following list of distributions belongs to the domain of attraction of the Type I (Gumbel) case, just to mention a few well-known cases: normal, lognormal, exponential, Weibull, gamma, and, of course, the Gumbel distribution itself.

Distributions belonging to the domain of attraction of Type II are, e.g., the Pareto, the generalized Pareto for positive shape parameter, and the Type II extreme value distribution itself. For Type III may be mentioned, e.g., the uniform distributions, distributions truncated on the upper side (provided a smooth density function), and the Type III extreme value distribution itself.

2.6 Parameter Estimation for the GEV Distributions

The practical application of the block maxima method involves the need to decide on how to divide the observed data into blocks. Obviously, there will be two conflicting issues that have to be dealt with. The desire to have large blocks so that the distribution of the block maxima will approximate a GEV distribution may easily lead to a sample of few block maxima. Statistical inference on small samples may entail large uncertainties. On the other hand, increasing the sample of block maxima by choosing smaller blocks may violate the asymptotic approximation by assuming a GEV distribution for the block maxima. These issues may be further complicated by the issues of independence and stationarity, which were discussed in Sect. 2.3. While establishing general rules for the choice of block size relative to the amount of data available is hardly feasible, for some practical cases the accumulated experience has led to what may be called a consensus. For example, in wind engineering, the choice of 1 year as a block size has become very close to a standard procedure. An important consideration for this choice is that the data may then reasonably be assumed to belong to the same population since seasonal effects have effectively been removed.

When the sample of block maxima has been determined, the next step would be to estimate the parameters of the GEV model, or one of the three types, if that can be ascertained a priori. In this book the focus is on two rather popular estimation methods, which is the method of moments (primarily for the Gumbel model) and the maximum likelihood method. The probability weighted moment method has also been used to some extent. For this method, please cf. Hosking et al. (1985).

To simplify notation, the block maxima are denoted by $Z_1, \ldots, Z_k$, assuming k blocks. These random variables are assumed to be iid with a common GEV distribution, the parameters of which are to be estimated from the outcomes of the block maxima, that is, the observed data.

2.6.1 Estimation by the Method of Moments

The exposition of the method of moments for parameter estimation is limited to the Gumbel model. It is for this case that it seems to be most popular, maybe due to its simplicity in this case. The general Gumbel model has two parameters. Since the first two statistical moments $m_1 = \mathbb{E}(Z)$ and $m_2 = \mathbb{E}(Z^2)$ of a Gumbel distributed variable Z can be expressed in terms of these two parameters, for estimation the following two empirical moments are calculated:

$$\hat{m}_j = (1/k)\sum_{i=1}^{k} z_i^j \, , j = 1, 2, \tag{2.20}$$

where $z_1, \ldots, z_k$ are the observed data.

Assuming that Z has the general Gumbel distribution $G(z) = \exp\{-\exp[-(z-\mu)/\sigma]\}$, then $m_1 = \mathbb{E}(Z) = \mu + 0.5772\sigma$ and $m_2 = \mathbb{E}(Z^2) = m_1^2 + \pi^2\sigma^2/6$, cf. Sect. 2.2. Denote by μ_k and σ_k the estimated values of the parameters based on the k observations of block maxima. It is then obtained that

$$\mu_k = \hat{m}_1 - 0.5772\sigma_k \tag{2.21}$$

and

$$\sigma_k = (\sqrt{6}/\pi)\sqrt{\hat{m}_2 - \hat{m}_1^2}. \tag{2.22}$$

2.6.2 Maximum Likelihood Estimation

The maximum likelihood (ML) method is very popular and has widespread use in almost every branch of statistics. It turns out that the application of the ML methods for estimation on GEV models requires some caution. Fortunately, it seems that for the applications relevant for this book, the restrictions that need to be observed are rarely an issue. Specifically, for values of the shape parameter $\gamma > -0.5$, the ML estimators behave regularly. The only thing to note is that there are some small sample issues related to the use of ML estimators also for GEV models, cf. Coles and Dixon (1999).

Based on the assumption that $Z_1, \ldots, Z_k$ are iid random variables having a common GEV distribution, then the log-likelihood function for the GEV parameters when $\gamma \neq 0$ has the following expression:

$$\ell(\mu, \sigma, \gamma) = -k \log \sigma - (1 + 1/\gamma) \sum_{i=1}^{k} \log \left[1 + \gamma \left(\frac{z_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^{k} \left[1 + \gamma \left(\frac{z_i - \mu}{\sigma} \right) \right]^{-1/\gamma}, \qquad (2.23)$$

provided that

$$1 + \gamma \left(\frac{z_i - \mu}{\sigma} \right) > 0, \quad \text{for} \quad i = 1, \ldots, k. \qquad (2.24)$$

If the last condition is violated, the likelihood becomes zero and the log-likelihood therefore $-\infty$. The case $\gamma = 0$ needs to be considered separately, using the Gumbel model. In this case the log-likelihood becomes

$$\ell(\mu, \sigma) = -k \log \sigma - \sum_{i=1}^{k} \left(\frac{z_i - \mu}{\sigma} \right) - \sum_{i=1}^{k} \exp \left\{ - \left(\frac{z_i - \mu}{\sigma} \right) \right\}. \qquad (2.25)$$

To obtain the numerical maximum likelihood estimates from the observed data by using (2.23) and (2.25), standard numerical optimization programs may be utilized. If (2.23) is used, care must be exercised to avoid numerical problems in cases where the optimization algorithms tend to parameter estimates in the close vicinity of $\gamma = 0$. Then it is strongly advisable to use (2.25).

Confidence intervals on the estimated parameter values can be calculated exploiting that the approximate distribution of the estimators $(\hat{\mu}, \hat{\sigma}, \hat{\gamma})$ is multivariate normal with mean value (μ, σ, γ). This is discussed by Coles (2001).

2.7 Model Validation

As is well known from basic courses in statistics, the use of probability (or PP) plots and quantile (or QQ) plots may reveal very useful information about the extent of agreement between an assumed or estimated probability distribution and the empirical distribution of the data. These are also highly useful tools for a visual check of fitted GEV models in particular cases. For a thorough discussion of the use of these plots, cf. Beirlant et al. (2004).

A probability or a PP plot is a direct comparison of the fitted distribution model to the empirical distribution. Assume that the sample of block maxima has been ordered by increasing value: $z_{(1)} \leq z_{(2)} \leq \ldots \leq z_{(k)}$. The empirical distribution function, $\tilde{G}$ say, evaluated at $z_{(i)}$ is given by

$$\tilde{G}(z_{(i)}) = i/(k+1). \tag{2.26}$$

The proposed GEV model distribution is obtained by substituting the parameter estimates into (2.7)

$$\hat{G}(z_{(i)}) = \exp\left\{-\left[1+\hat{\gamma}\left(\frac{z_{(i)}-\hat{\mu}}{\hat{\sigma}}\right)\right]^{-1/\hat{\gamma}}\right\}, \tag{2.27}$$

provided $\hat{\gamma} \neq 0$. If $\hat{\gamma} = 0$, the plot is constructed using the Gumbel distribution. If the GEV model is a good approximation, then

$$\hat{G}(z_{(i)}) \approx \tilde{G}(z_{(i)}) \tag{2.28}$$

for each index i, so that the PP plot consisting of the points

$$\left(\hat{G}(z_{(i)}), \tilde{G}(z_{(i)})\right) \;\; i = 1, \ldots, k \tag{2.29}$$

should follow approximately the unit diagonal.

For the case of extreme value distributions, a quantile or QQ plot is usually considered to be more informative than a PP plot because it shows more clearly the agreement at high values of the observed data, which is of primary concern when fitting extreme value models. Assuming again that $\hat{\gamma} \neq 0$, the QQ plot is traced out by the point graph

$$\left(\hat{G}^{-1}(i/(k+1)), z_{(i)}\right), \;\; i = 1, \ldots, k, \tag{2.30}$$

where

$$\hat{G}^{-1}(i/(k+1)) = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\gamma}}\left[1 - \left\{-\log\left(i/(k+1)\right)\right\}^{-\hat{\gamma}}\right]. \tag{2.31}$$

This graph should also approximately follow a straight line. These procedures are discussed at greater length in Chap. 9.

2.8 Estimating Confidence Intervals by Bootstrapping

The bootstrapping method is a statistical technique of fairly recent origin that can be used for estimating confidence intervals on quantities derived from a statistical distribution on the basis of a limited sample generated by that same distribution (Efron and Tibshirani 1993; Davison and Hinkley 1997). It is based on resampling from a distribution determined by the available sample of data. Despite the fact that the name of the method alludes to lifting oneself up by the bootstraps (Baron von Munchausen), the method appears to be reasonably effective for the specific purpose of estimating confidence bands. For convenience, a brief discussion of some basic features of the bootstrapping method is provided here.

Assume that $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ is a sample or a vector consisting of n independent observations of a random variable Z and that this is the only empirical information available about Z. Confidence intervals for a statistical quantity require the estimation of quantiles from the distribution of a relevant estimator. There are in principle two available options for obtaining bootstrap estimates of such quantiles. One is the nonparametric approach, where a purely empirical distribution function is established for Z on the basis of the observed data by allocating a probability of $1/n$ to each of the observed data points. The other is the parametric bootstrap, which is obtained by assuming that Z has a specified distribution function $F_Z(z; \theta) = \text{Prob}(Z \leq z)$, where θ denotes a vector of unknown parameters, which determine the distribution. These parameters are then estimated from the observed data $\mathbf{z}$, giving $\hat{\theta}$, and $F_Z(z; \hat{\theta})$ is adopted as the distribution of Z.

In this section on the block maxima method using GEV models, only the parametric bootstrap is used. The goal is to estimate some statistical quantity V, e.g., a high quantile like $100(1-\alpha)\%$ $(0 < \alpha << 1)$, given by the unknown distribution. Let $\hat{V}$ denote the estimate of V obtained from the fitted model distribution $F_Z(z; \hat{\theta})$, which is a GEV distribution. The parametric bootstrapping technique for estimating confidence intervals on V is based on resampling from the GEV model obtained.

This is done as follows: Let Z^* denote the random variable with distribution function $F_Z(z; \hat{\theta})$. ℓ bootstrap samples $\mathbf{z}_j^*$, $j = 1, \ldots, \ell$, with n independent observations of Z^* in each sample are now generated. Each sample $\mathbf{z}_j^*$ is used to fit a new GEV model from which an estimate V_j^* of V is obtained.

Simple estimates for confidence intervals on V are derived by calculating the sample standard deviation s_V^*:

$$s_V^* = \sqrt{\frac{1}{\ell - 1} \sum_{j=1}^{\ell} (V_j^* - \bar{V}^*)^2}, \tag{2.32}$$