# The Renext package

## Computing details

2022-11-25

Yves Deville

Yves Deville Statistical consultant, 569 rue Nicolas Parent 73000 Chambéry FR
Mail deville.yves@alpestat.com

# Contents

# Introduction

## POT and Renext

The present document provides details about the computations implemented in the R package **Renext** Deville and IRSN (2019). This package was designed by IRSN[1] as a tool allowing the use within the R environment R Development Core Team (2010) of the statistical framework called *méthode du renouvellement* also known as Peaks Over Threshold (POT) method. The classical POT context relies on a marked process; the assumptions are described in the *Renext Guide*[2] as well as in many classical sources: see e.g. Davison and Smith (1990) or the book Embrechts, Klüppelberg, and Mikosch (1996).

## Complete and partial observations

In the usual context of application of POT, the process is completely observed, meaning that the available data contains the events $T_k$ and the marks $X_k$ over some time period with known duration. However, it is not infrequent to have *partial observations* of the marked process completing the observations $[T_k,\, X_k]$. Such partial observations typically arise from *historical information*.

The following vocabulary has been used in **Renext**.

**Blocks** A (time) block is time period with known duration $w$.

**Main sample** or OT data. It is a set $n$ observations $X_k$ and a main threshold $u$.

**MAX blocks** For each MAX block $b$, we are given the $r_b$ largest observations, all assumed to be $> u$. They can be assumed to be in decreasing order.

$$Z_{b,1} \geqslant Z_{b,2} \geqslant \cdots \geqslant Z_{b,r_b} > u.$$

The $r_b$ r.vs $Z_{b,i}$ are the Largest Order Statistics of an unknown number of $X_k$: those falling in the block $b$.

**OTS blocks** For each OTS block $b$, we are given a threshold $u_b \geqslant u$ as well as the $r_b$ observations $X_{b,i}$ that exceeded $u_b$. Their number $r_b$ is potentially zero, in which case one may say that $u_b$ is an *unobserved level.*

[1] Institut de Radioprotection et de Sûreté Nucléaire.

[2] A pdf document shipped with the package.

It will be assumed that the time blocks used in the analysis do not overlap; thus the observations for $B$ different blocks are $B$ mutually independent random vectors. However, the random variables $Z_{b,i}$ within a MAX block $b$ are dependent. It will be assumed that the number $r_b$ of the observations in a MAX or OTS block is fixed or that it is random but independent from the marks $X_{b,i}$.

The partial observation contexts correspond to a *temporal aggregation* of the underlying marked process. Note that we do not assume to be given the time stamps $T_k$ for the observations provided in a MAX or an OTS block, nor do we assume to be given the total number $N_b$ of observations in a such block.

MAX blocks can be compared to the blocks used in the classical Extreme Value methods of block maxima and $r$ largest (Coles 2001, chap. 3). Although these classical methods rely on asymptotic considerations and imply the the use of Generalised Extreme Value (GEV) distribution, they can be compared to the POT context where only partial observations for MAX blocks are available. The comparison between the two approaches is detailed in the third chapter of this report.

## Distributions for the excesses

While classical POT normally relies on the use of the Generalised Pareto Distribution (GPD) for the excesses, some sources Miquel (1981) use the *méthode du renouvellement* with different distributions e.g. with a Weibull distribution. **Renext** was designed to allow Maximum Likelihood (ML) estimation in these contexts.

**Renext** allows the use of a quite general distribution for the excesses, used in a black-box fashion. This possibility mainly relies on the ML estimation. Actually, the likelihood of a POT model can be evaluated in a fairly general context, with an arbitrary distribution for the excesses and a combination of ordinary observations and partial observations. The likelihood function can then be maximised using the `optim` function.

Such a black-box approach suffers from a number of well-known problems concerning e.g. initial values, possible non-convergence or difficulties in the numerical evaluation of the hessian. Therefore, a number of distributions has received more attention in **Renext**, and specific algorithms or initialisation procedures have been proposed and implemented for these, with the following goals.

- Give a general expression for the likelihood in the general context described above.
- Concentrate the likelihood with respect to some of the parameters when possible.
- Derive initial values for the parameters.
- Use *analytical* (rather than numerical) derivatives when possible. Analytical derivatives can in some cases overcome the difficulties in the evaluation of the hessian. In order to use analytical derivatives when partial observations are available, the derivatives with respect to the parameters must be provided for the log-density of the marks and for the survival function.

The present document give detailed information about these goals. It also intends to clarify the link between Extreme Value Theory (EVT) contexts: the POT framework, and those relying on a temporal aggregation: block maxima and $r$-largest.

## Content

This document is organised as follows.

- The first chapter gives the log-likelihood and its derivatives in the general context of partial observations described above. This general likelihood can be maximised as such, or it can be used with the Poisson rate $\lambda$ concentrated out.
- The second chapter is devoted to the "classical" ML estimation, where Independent and Identically Distributed (i.i.d.) observations are used, as is the case when only complete observations of the marks $X_k$ are used. It focuses on some specific distributions of excesses, for which a concentration of the likelihood can be used and/or initial values can be found for the likelihood maximisation. The Lomax and maxlo[3] receive more attention because they are special cases of the GPD and also because some useful properties or features of their likelihood given here seem not to be well-known.
- The third chapter focuses on the temporal aggregation contexts. It provides details on the relation between POT and the classical block maxima and $r$ largest. It also studies the question of the initial estimates when only partial observations are available.
- The fourth chapter is devoted to the computational problems met in the determination of the plotting positions used in several diagnostic plots and especially in the Return Level plot.
- The fifth chapter describes some statistical tests for the distribution of the excesses.

Some technical computations are given in appendices. The notations vary across chapters, e.g. excesses can be denoted by $X$ or by $Y$ depending on the context.

[3]Non-official name.

# Chapter 1

# Likelihood with partial observations

## 1.1 Likelihood

### 1.1.1 The context

As exposed in the introduction, we consider the ML estimation of a marked process using heterogeneous data of the following types.

**Main sample** A set $n$ observations $X_i$ and a main threshold $u$. The $X_i$ represent all marks over a time period with known duration $w$.

**MAX blocks** For each block $b$ with known duration $w_b$, we are given the $r_b$ largest observations $Z_{b,i}$, all assumed to be $> u$.

**OTS blocks** For each block $b$ with known duration $w_b$, we are given a threshold $u_b \geqslant u$ as well as the $r_b$ observations $X_{b,i}$ that exceeded $u_b$. Their number $r_b$ is potentially zero.

The main sample can be regarded as a block "OT" block (for Over the Threshold). Assuming that the different blocks do not overlap, hence that the observations are independent across blocks, the log-likelihood is the sum of block contributions. We give here an expression of the log-likelihood of a block in each of the three cases.

**Remark.** In **Renext** versions `<= 2.1-0`, it was necessary to provide a main sample from which ML estimate for the distribution of marks can be found. This is no longer necessary for later versions. ■

The rate of the Homogeneous Poisson Process is denoted by $\lambda$. The distribution of the marks depends of a parameter vector $\boldsymbol{\theta}$ and has density $f(x; \boldsymbol{\theta})$ or simply $f(x)$. The distribution and survival functions are $F(x; \boldsymbol{\theta})$ and $S(x; \boldsymbol{\theta})$ or simply $F(x)$ and $S(x)$.

### 1.1.2 Likelihood for main sample

**Theorem 1.1.** *Consider the main sample with $n$ observed values $X_i$ for $i = 1, 2, \ldots, n$ and total duration $w$. The log-likelihood $\ell := \log L$ is*

$$\ell = n \log(\lambda w) - \log(n!) - \lambda w + \sum_{i=1}^{n} \log f(X_i) \tag{1.1}$$

*Proof.* The likelihood is

$$L = \Pr\{N = n\} \times \prod_{i=1}^{n} f(X_i) = \frac{(\lambda w)^n}{n!} e^{-\lambda w} \times \prod_{i=1}^{n} f(X_i)$$

$\square$

### 1.1.3 Likelihood for MAX data

**Theorem 1.2.** *Consider an historical MAX block $b$ with duration $w_b$ and with given $r_b$ largest order statistics $Z_{b,i}$ for $i = 1, 2, \ldots, r_b$. The corresponding log-likelihood $\ell_b = \log L_b$ writes as*

$$\ell_b = r_b \log(\lambda w_b) - \lambda w_b \, S(Z_b^\star;\, \boldsymbol{\theta}) + \sum_{i=1}^{r_b} \log f(Z_{b,i};\, \boldsymbol{\theta}) \tag{1.2}$$

*where $S(x;\, \boldsymbol{\theta}) = 1 - F(x;\, \boldsymbol{\theta})$ is the survival and $Z_b^\star = \min_i Z_{b,i}$*

*Proof.* To simplify let us temporarily omit the index $b$, thus using $w$ or $r$ in place of $w_b$ and $r_b$. The density and distribution functions are denoted $f(x)$ and $F(x)$ with their dependence to the parameter $\boldsymbol{\theta}$ omitted. The $Z_i$ are assumed in decreasing order $Z_1 \geqslant Z_2 \geqslant \cdots \geqslant Z_r$. The index $i$ no longer refers to the order of the events but to that of the observations. The number of events in the block is unknown but must be greater or equal to $r$. For $n \geqslant r$, the probability of observing the $Z_i$ conditional on $\{N = n\}$ is

$$\Pr[Z_1, Z_2, \ldots, Z_r \,|\, N = n] = \frac{n!}{(n-r)!} F(Z_r)^{n-r} \prod_{i=1}^{r} f(Z_i)$$

Indeed $n - r$ observations among the $n$ must be less or equal to $Z_r$ while the $r$ remaining ones must be equal to the observed values. Using the total probability formula we get

$$L_b = \sum_{n=r}^{\infty} \Pr\{N = n\} \times \frac{n!}{(n-r)!} F(Z_r)^{n-r} \prod_{i=1}^{r} f(Z_i)$$

For an Homogeneous Poisson Process with rate $\lambda$ we have $\Pr\{N = n\} = (\lambda w)^n \, e^{-\lambda w}/n!$, thus

$$\begin{aligned} L_b &= \sum_{n=r}^{\infty} \frac{(\lambda w)^n}{n!} \, e^{-\lambda w} \frac{n!}{(n-r)!} F(Z_r)^{n-r} \prod_{i=1}^{r} f(Z_i) \\ &= (\lambda w)^r \, e^{-\lambda w} \left[\prod_{i=1}^{r} f(Z_i)\right] \sum_{n=r}^{\infty} \frac{(\lambda w)^{n-r}}{(n-r)!} F(Z_r)^{n-r} \end{aligned}$$

The change of index $k = n - r$ in the sum at the right side gives

$$\sum_{n=r}^{\infty} \quad = \sum_{k=0}^{\infty} \frac{(\lambda w)^k}{k!} F(Z_r)^k = \exp\{\lambda w \, F(Z_r)\}$$

Finally, the likelihood of the block is

$$L_b = (\lambda w)^r \left[\prod_{i=1}^{r} f(Z_i)\right] \exp\{-\lambda w \, [1 - F(Z_r)]\}$$

which after taking the log is as claimed.

$\square$

### 1.1.4 Likelihood for OTS data

**Theorem 1.3.** *Consider an historical OTS block $b$ with threshold $u_b$, duration $w_b$ and with given $r_b$ observed levels $X_{b,i}$ for $i = 1, 2, \ldots, r_b$. Up to an additive constant, the corresponding log-likelihood $\ell_b = \log L_b$ writes as*

$$\ell_b = r_b \log(\lambda w_b) - \lambda w_b \, S(u_b; \boldsymbol{\theta}) + \sum_{i=1}^{r_b} \log f(X_{b,i}; \boldsymbol{\theta}) \tag{1.3}$$

*where $S(x; \boldsymbol{\theta}) = 1 - F(x; \boldsymbol{\theta})$ is the survival.*

*Proof.* The events with levels $X > u_b$ form a HPP with rate $\lambda \times S(u_b; \boldsymbol{\theta})$, and their number $R_b$ follows a Poisson distribution. Conditional on $\{R_b = r_b\}$, the $r_b$ observed levels are independent with density $f(x)/S(u_b)$ for $x \geqslant u_b$. Hence

$$\begin{aligned} L_b &= \Pr\{R_b = r\} \times \prod_{i=1}^{r_b} \frac{f(X_{b,i}; \boldsymbol{\theta})}{S(u_b; \boldsymbol{\theta})} \\ &= \frac{[\lambda w_b \, S(u_b; \boldsymbol{\theta})]^{r_b}}{r_b!} \, e^{-\lambda w_b \, S(u_b; \boldsymbol{\theta})} \times \prod_{i=1}^{r_b} \frac{f(X_{b,i}; \boldsymbol{\theta})}{S(u_b; \boldsymbol{\theta})} \end{aligned}$$

and the result follows by taking the log. □

### 1.1.5 Global log-likelihood

The global log-likelihood writes (up to an additive constant)

$$\begin{aligned} \ell(\lambda, \boldsymbol{\theta}) = {} & n \log(\lambda w) - \lambda w + \sum_{i=1}^{n} \log f(X_i; \boldsymbol{\theta}) \\ & + \sum_{b \in \text{MAX}} \left\{ r_b \log(\lambda w_b) - \lambda w_b \, S(Z_b^\star; \boldsymbol{\theta}) + \sum_{i=1}^{r_b} \log f(Z_{b,i}; \boldsymbol{\theta}) \right\} \\ & + \sum_{b \in \text{OTS}} \left\{ r_b \log(\lambda w_b) - \lambda w_b \, S(u_b; \boldsymbol{\theta}) + \sum_{i=1}^{r_b} \log f(X_{b,i}; \boldsymbol{\theta}) \right\} \end{aligned}$$

Note that the value $\widehat{\lambda}(\boldsymbol{\theta})$ of $\lambda$ maximising $\log L$ for a given value of $\boldsymbol{\theta}$ is given by the simple formula

$$\widehat{\lambda}(\boldsymbol{\theta}) = \frac{n + \sum_{b \in \text{MAX}} r_b + \sum_{b \in \text{OTS}} r_b}{w + \sum_{b \in \text{MAX}} w_b \, S(Z_b^\star; \boldsymbol{\theta}) + \sum_{b \in \text{OTS}} w_b \, S(u_b; \boldsymbol{\theta})} \tag{1.4}$$

where empty sums are to be replaced by zero. Moreover, if each block duration is replaced by a discounted duration defined as the product of the duration by the survival $S$ at the smallest observable value according to $w \leftarrow w \times S$, then the estimate for the rate can be written in standard form

$$\widehat{\lambda}(\boldsymbol{\theta}) = \frac{\text{total number of observed events}}{\text{total (discounted) duration}}$$

where the denominator depends on $\boldsymbol{\theta}$.

The concentrated log-likelihood (with respect to $\lambda$) can be expressed by replacing the products of $\widehat{\lambda} \times$ (discounted duration) by the corresponding number of events

$$\ell_{\mathsf{c}}(\boldsymbol{\theta}) = [n + r_{\text{MAX}} + r_{\text{OTS}}] \times \log(\widehat{\lambda}) + \sum_{i=1}^{n} \log f(X_i;\, \boldsymbol{\theta})$$
$$+ \sum_{b \in \text{MAX}} \sum_{i=1}^{r_b} \log f(X_{b,i};\, \boldsymbol{\theta}) + \sum_{b \in \text{OTS}} \sum_{i=1}^{r_b} \log f(X_{b,i};\, \boldsymbol{\theta}) + C$$

where $r_{\text{MAX}}$ and $r_{\text{OTS}}$ are the total numbers of observations for the two categories. Here $C$ denotes an additive constant that can be ignored as far as the goal is maximising the likelihood. However the value of $C$ can be required for instance once the maximum is found. This is

$$C = -\sum_b r_b[1 - \log w_b]$$

where $b$ describes all blocks an $r_b := n$ for the main sample considered as a block and similarly $w_b := w$ then.

## 1.2 Log-likelihood derivatives

### 1.2.1 Goals

The expression of the log-likelihood $\ell$ given in section 1.1.5 can be differentiated w.r.t. the parameters, thus allowing the computation of the observed information matrix in the general framework where MAX and OTS data are available.

### 1.2.2 Main sample

For the main sample in 1.1.2, the derivatives of the log-likelihood $\ell$ of (1.1) are

$$\frac{\partial \ell}{\partial \lambda} = \frac{n}{\lambda} - w,$$
$$\frac{\partial \ell}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{r} \frac{\partial}{\partial \boldsymbol{\theta}} \log f(X_i).$$

At the second order

$$\frac{\partial^2 \ell}{\partial \lambda^2} = -\frac{n}{\lambda^2},$$
$$\frac{\partial^2 \ell}{\partial \lambda \partial \boldsymbol{\theta}} = 0,$$
$$\frac{\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \sum_{i=1}^{n} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \log f(X_i).$$

### 1.2.3 MAX data

Given an historical block as in 1.1.3, the derivatives of the log-likelihood $\ell_b$ of (1.2) are

$$\begin{aligned}
\frac{\partial \ell_b}{\partial \lambda} &= \frac{r_b}{\lambda} - w_b\, S,\\
\frac{\partial \ell_b}{\partial \boldsymbol{\theta}} &= -\lambda w_b \frac{\partial S}{\partial \boldsymbol{\theta}} + \sum_{i=1}^{r_b} \frac{\partial}{\partial \boldsymbol{\theta}} \log f(Z_{b,i}),
\end{aligned}$$

where the values of $S$ and its derivatives are taken at the minimum $Z_b^\star$. At the second order

$$\begin{aligned}
\frac{\partial^2 \ell_b}{\partial \lambda^2} &= -\frac{r_b}{\lambda^2},\\
\frac{\partial^2 \ell_b}{\partial \lambda \partial \boldsymbol{\theta}} &= -w_b \frac{\partial S}{\partial \boldsymbol{\theta}},\\
\frac{\partial^2 \ell_b}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} &= -\lambda w_b \frac{\partial^2 S}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} + \sum_{i=1}^{r_b} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \log f(Z_{b,i}).
\end{aligned}$$

### 1.2.4 OTS data

Given an historical block as in 1.1.4, the derivatives of the log-likelihood $\ell_b$ of (1.3) are

$$\begin{aligned}
\frac{\partial \ell_b}{\partial \lambda} &= \frac{r_b}{\lambda} - w_b\, S,\\
\frac{\partial \ell_b}{\partial \boldsymbol{\theta}} &= -\lambda w_b \frac{\partial S}{\partial \boldsymbol{\theta}} + \sum_{i=1}^{r_b} \frac{\partial}{\partial \boldsymbol{\theta}} \log f(X_{b,i}),
\end{aligned}$$

where the values of $S$ and its derivatives are taken at the minimum $u_b$. At the second order

$$\begin{aligned}
\frac{\partial^2 \ell_b}{\partial \lambda^2} &= -\frac{r_b}{\lambda^2},\\
\frac{\partial^2 \ell_b}{\partial \lambda \partial \boldsymbol{\theta}} &= -w_b \frac{\partial S}{\partial \boldsymbol{\theta}},\\
\frac{\partial^2 \ell_b}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} &= -\lambda w_b \frac{\partial^2 S}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} + \sum_{i=1}^{r_b} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \log f(X_{b,i}).
\end{aligned}$$

### 1.2.5 Computation

For some distributions, working with the survival $S(x) = 1 - F(x)$ and the cumulative hazard $H(x) = -\log S(x)$ is more convenient. For instance in the Weibull case we have for $x > 0$ the simple expressions

$$S(x) = \exp\left\{-(x/\beta)^\alpha\right\}, \qquad H(x) = (x/\beta)^\alpha.$$

The derivatives of $F(x)$ and those of $S(x)$ are related according to

$$\frac{\partial F}{\partial \boldsymbol{\theta}} = -\frac{\partial S}{\partial \boldsymbol{\theta}} \qquad \frac{\partial^2 F}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = -\frac{\partial^2 S}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}.$$

Since $S = \exp(-H)$ we have the two relations

$$\frac{\partial S}{\partial \boldsymbol{\theta}} = -\frac{\partial H}{\partial \boldsymbol{\theta}} \times S \qquad \frac{\partial^2 S}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = -\frac{\partial^2 H}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \times S + \left[\frac{\partial H}{\partial \boldsymbol{\theta}}\right] \left[\frac{\partial H}{\partial \boldsymbol{\theta}}\right]^\top \times S \tag{1.5}$$

which might be convenient in many cases.

For some distributions such as the mixture of exponentials, the derivatives of $f$ are easier to compute than those of $\log f$. These can be computed then as

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log f = \frac{1}{f} \frac{\partial f}{\partial \boldsymbol{\theta}} \qquad \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \log f = -\frac{1}{f^2} \left[\frac{\partial f}{\partial \boldsymbol{\theta}}\right] \left[\frac{\partial f}{\partial \boldsymbol{\theta}}\right]^\top + \frac{1}{f} \frac{\partial^2 f}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}$$

The previous derivatives (or some of them) are provide for some classical distributions used by **Renext** in section A.2 page 74.

# Chapter 2

# Classical maximum likelihood

## 2.1 Scope

In this chapter, we review the ML estimation for several distributions used in **Renext** when an "ordinary sample" of size $n$ is used, that is: a collection of $n$ i.i.d. r.vs $X_1$, $X_2$, $\dots$, $X_n$ following the target distribution.

In most cases, a likelihood concentration is used, allowing a one-dimensional optimisation rather than a two-dimensional one. The derivatives of the concentrated log-likelihood may be given in closed form and used to compute an information matrix, either *observed* or *expected*.

The main application in **Renext** is for POT estimation; then the r.vs $X_i$ are the excesses over the threshold $u$. Their distribution typically involves two parameters, a shape parameter and a scale parameter. This may be, as for the GPD, a special case of a three-parameter distribution with a location parameter taken as the threshold $u$.

The section 2.8 devoted to the Negative Binomial Lévy process is quite different from the others and does not concern the marks of the marked process. It describes the ML estimation of this process from partial observations $N_b$ representing event counts.

## 2.2 Classical Weibull

### 2.2.1 The distribution

In this section, we consider the ML estimation and inference for the two-parameter Weibull distribution. This distribution depends on a shape parameter $\alpha > 0$ and a scale parameter $\beta > 0$; the density is

$$f(x) = \frac{\alpha}{\beta^\alpha}\, x^{\alpha-1} \exp\left\{ - \left[\frac{x}{\beta}\right]^\alpha \right\}, \qquad x > 0. \tag{2.1}$$

The parameter vector is $\boldsymbol{\theta} := [\alpha,\, \beta]^\top$. We reparameterise of the model by using $\delta = \beta^\alpha$ in place of $\beta$. Thus, the density writes

$$f(x) = \frac{\alpha}{\delta}\, x^{\alpha-1} \exp\left\{ -\frac{x^\alpha}{\delta} \right\}, \qquad x > 0 \tag{2.2}$$

and the new parameter vector is $\boldsymbol{\eta} := [\alpha,\, \delta]^\top$.

### 2.2.2 The likelihood

Let $X_1, X_2, \ldots, X_n$ be a sample of the Weibull distribution above. The log-likelihood $\ell = \log L(\boldsymbol{\eta})$ for the parameterisation (2.2) using $\boldsymbol{\eta} = [\alpha,\, \delta]^\top$ is

$$\ell = \sum_{i=1}^{n} \left\{ \log\alpha - \log\delta + (\alpha - 1)\log X_i - \frac{1}{\delta} X_i^\alpha \right\}$$

Its derivatives with respect to the two parameters are

$$\begin{cases} \partial_\alpha \ell &= \dfrac{n}{\alpha} + \sum_i \log X_i - \dfrac{1}{\delta} \sum_i X_i^\alpha \log X_i \\ \partial_\delta \ell &= -\dfrac{n}{\delta} + \dfrac{1}{\delta^2} \sum_i X_i^\alpha \end{cases}$$

The derivative of $\ell$ with respect to $\delta$ is zero iff $\delta = \widehat{\delta}(\alpha)$ given by

$$\widehat{\delta}(\alpha) = \frac{1}{n} \sum_i X_i^\alpha$$

Replacing $\delta$ by this value gives the concentrated log-likelihood (with respect to $\delta$), i.e. $\ell_{\mathrm{c}}(\alpha) = \ell[\alpha,\, \widehat{\delta}(\alpha)]$. The derivative of this concentrated log-likelihood cancels iff

$$\frac{1}{\alpha} + \frac{1}{n} \sum_i \log X_i - \frac{\sum_i X_i^\alpha \log X_i}{\sum_i X_i^\alpha} = 0$$

which can be rewritten as

$$g(\alpha) - g(0) = \frac{1}{\alpha}, \qquad \text{with} \qquad g(\alpha) := \frac{\sum_i X_i^\alpha \log X_i}{\sum_i X_i^\alpha}$$

Deriving $g(\alpha)$ with respect to $\alpha$

$$g'(\alpha) = \frac{\left\{\sum X_i^\alpha \log^2 X_i\right\}\left\{\sum X_i^\alpha\right\} - \left\{\sum X_i^\alpha \log X_i\right\}^2}{\left\{\sum X_i^\alpha\right\}^2}$$

and thus $g'(\alpha) \geqslant 0$ using Cauchy-Schwarz inequality, so $g(\alpha)$ is increasing. Writing $L_i := \log X_i$ it easy to see that

$$\lim_{\alpha \to 0} g(\alpha) = \bar{L}, \qquad \lim_{\alpha \to \infty} g(\alpha) = \max_i L_i.$$

In the general case where the $X_i$ are not all equal, $g(\alpha) - g(0)$ increases strictly from 0 to the positive value $\max_i L_i - \bar{L}$ and thus must be equal to $1/\alpha$ for a unique value of $\alpha$. Therefore a unique vector $\boldsymbol{\eta} = [\alpha,\, \delta]^\top$ maximises the likelihood.

### 2.2.3 Estimation algorithm

The maximising $\alpha$ can be found using a zero finding algorithm. A possibility is to use a few steps of a Newton-Raphson which computes $\alpha_k$ for $k \geqslant 1$ from

$$\alpha_k = \alpha_{k-1} - h\left(\alpha_{k-1}\right) / h'\left(\alpha_{k-1}\right)$$

with $h(\alpha) := g(\alpha) - g(0) - 1/\alpha$. and a suitable initial value $\alpha_0$. At each step, it might be useful to compute the three values $R_k(\alpha)$ for $k = 0,\, 1,\, 2$ where

$$R_k(\alpha) = \frac{1}{n} \sum_i X_i^\alpha \log^k X_i \tag{2.3}$$

since these are used in the computations and are useful once the optimum when reached.

The initial value for $\alpha$ can be obtained using the following classical results. When $X$ has density $f(x)$ the r.v. $Z = X^\alpha/\beta^\alpha$ follows the exponential distribution with mean 1, and $-\log Z$ has the standard Gumbel distribution with variance $\pi^2/6$. Since $\log Z = \alpha L - \alpha \log \beta$ with $L := \log X$, we can make use of the following estimator

$$\widetilde{\alpha} = \frac{\pi}{\sqrt{6}} \times \frac{1}{S_L} \approx \frac{1.2825}{S_L}$$

where $S_L^2$ is the sample variance of the sample $L_i = \log X_i$.

### 2.2.4 Information matrix and inference

The second derivatives of $\ell = \log L$ can be written using the $R_k(\alpha)$ of (2.3) above

$$\begin{cases} \partial^2_{\alpha,\alpha}\ell & = -\dfrac{n}{\alpha^2} - \dfrac{n}{\delta}\, R_2(\alpha) \\ \partial^2_{\delta,\alpha}\ell & = \dfrac{n}{\delta^2}\, R_1(\alpha) \\ \partial^2_{\delta,\delta}\ell & = \dfrac{n}{\delta^2} - \dfrac{2n}{\delta^3}\, R_0(\alpha) \end{cases}$$

These formulas give the (opposite of the) observed information matrix $\mathbf{J}(\boldsymbol{\eta})$.

It is even possible to compute the *expected* information matrix $\mathbf{I} = -\mathbb{E}\left\{\partial^2\ell/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top\right\}$. The expectations of $R_k$ are

$$\mathbb{E}\left[R_k(\alpha)\right] = \int_0^\infty x^\alpha \log^k x\, f(x)\, \mathrm{d}x = \frac{\delta}{\alpha^k} \int_0^\infty u\,(\log\delta + \log u)^k\, e^{-u}\, \mathrm{d}u$$

the rightmost integral being obtained through the change of variable $x^\alpha = \delta u$. Let

$$\lambda_k = \int_0^\infty u \log^k u\, e^{-u}\, \mathrm{d}u$$

Using classical computations or values of the digamma and trigamma functions

$$\lambda_0 = 1 \qquad \lambda_1 = 1 - \gamma \qquad \lambda_2 = \frac{\pi^2}{6} + \gamma^2 - 2\gamma$$

where $\gamma$ is Euler's constant ($\gamma \approx 0.577216$). The theoretical expected (or Fisher) information is

$$\mathbf{I}(\boldsymbol{\eta}) = n \begin{bmatrix} \dfrac{1}{\alpha^2}\left(1 + \log^2\delta + 2\lambda_1 \log\delta + \lambda_2\right) & \dfrac{-1}{\alpha\delta}\,(\lambda_1 + \log\delta) \\ \dfrac{-1}{\alpha\delta}\,(\lambda_1 + \log\delta) & \dfrac{1}{\delta^2} \end{bmatrix}$$

The asymptotic variance $\mathrm{AVar}(\widehat{\boldsymbol{\eta}})$ for the estimator $\widehat{\boldsymbol{\eta}}$ of the vector $\boldsymbol{\eta} = [\alpha,\, \delta]^\top$ is obtained as the inverse of the information matrix at the estimate.

The ML estimation is easily reverted to the original parameterisation. The estimate $\widehat{\boldsymbol{\theta}}_{\mathtt{ML}}$ is computed using the transformation $\beta = \delta^{1/\alpha}$ at the found maximum, and the asymptotic covariance matrix is computed as

$$\mathrm{AVar}(\widehat{\boldsymbol{\theta}}) = \mathbf{F}\,\mathrm{AVar}(\widehat{\boldsymbol{\eta}})\,\mathbf{F}^\top$$

where $\mathbf{F}(\boldsymbol{\eta}) := \partial\boldsymbol{\theta}/\partial\boldsymbol{\eta}$ is the jacobian matrix

$$\mathbf{F}(\boldsymbol{\eta}) = \begin{bmatrix} \partial_\alpha \alpha & \partial_\delta \alpha \\ \partial_\alpha \beta & \partial_\delta \beta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\delta^{1/\alpha}\log(\delta)/\alpha^2 & \delta^{(1/\alpha)-1}/\alpha \end{bmatrix}$$

evaluated at $\widehat{\boldsymbol{\eta}}$.

The ML estimation of the two-parameter Weibull distribution using the likelihood concentration described above is available in the `fweibull` function of **Renext**.

## 2.3 Classical gamma

### 2.3.1 The likelihood

Let $X_1,\, X_2,\, \dots,\, X_n$ be a sample of the classical gamma distribution with shape parameter $\alpha > 0$, scale parameter $\beta > 0$ with density

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\, x^{\alpha-1} \exp\left\{-\frac{x}{\beta}\right\} \qquad x > 0$$

Let $\boldsymbol{\theta} = [\alpha,\, \beta]^\top$. The log-likelihood $\ell = \log L(\boldsymbol{\theta})$ is

$$\ell = \sum_{i=1}^{n} \Big\{ -\log\Gamma(\alpha) - \alpha\log\beta + (\alpha-1)\log X_i - X_i/\beta \Big\}$$

Its derivatives with respect to the two parameters are

$$\begin{cases} \partial_\alpha \ell &= -n\,\psi(\alpha) - n\log\beta + \sum_i \log X_i \\ \partial_\beta \ell &= -n\,\dfrac{\alpha}{\beta} + \dfrac{1}{\beta^2}\sum_i X_i \end{cases}$$

where $\psi(x) = \mathrm{d}\log\Gamma(x)/\mathrm{d}x$ is the digamma function. The second order derivatives are

$$\begin{cases} \partial^2_{\alpha,\alpha}\ell &= -n\,\psi_1(\alpha) \\ \partial^2_{\alpha,\beta}\ell &= -\dfrac{n}{\beta} \\ \partial^2_{\beta,\beta}\ell &= n\,\dfrac{\alpha}{\beta^2} - \dfrac{2}{\beta^3}\sum_i X_i \end{cases}$$

where $\psi_1(x) = \mathrm{d}\psi(x)/\mathrm{d}x$ is the trigamma function.

### 2.3.2 Information matrix and inference

The information matrix and its inverse approximating the estimation variance are

$$\mathbf{I} = n \begin{bmatrix} \psi_1(\alpha) & 1/\beta \\ 1/\beta & \alpha/\beta^2 \end{bmatrix} \qquad \mathrm{Var}[\widehat{\boldsymbol{\theta}}] \approx \frac{1}{n\ [\alpha\psi_1(\alpha) - 1]} \begin{bmatrix} \alpha & -\beta \\ -\beta & \psi_1(\alpha) \end{bmatrix}.$$

Note that there is no practical difference here between the observed and the expected information matrices.

### 2.3.3 Concentration of the likelihood

The likelihood can be concentrated with respect to $\beta$ using $\widehat{\beta}(\alpha) = \bar{X}/\alpha$. The concentrated log-likelihood with respect to $\beta$ writes

$$\ell_{\mathrm{c}}(\alpha) = -n \log \Gamma(\alpha) - n\alpha\, \log(\bar{X}/\alpha) + (\alpha - 1) \sum_i \log X_i - n\,\alpha$$

The first two derivatives are

$$\frac{\partial \ell_{\mathrm{c}}}{\partial \alpha} = -n\,\psi(\alpha) + n\, \log(\alpha) - n\, \log \bar{X} + n\, \overline{\log X}, \qquad \frac{\partial^2 \ell_{\mathrm{c}}}{\partial \alpha^2} = -n\, [\psi_1(\alpha) - 1/\alpha]$$

The interesting point is that the second derivative is negative and thus the concentrated likelihood is a concave function of $\alpha$ This results from the inequality $\psi_1(x) \geqslant 1/x$ for any $x > 0$, easily shown using a series representation of $\psi_1(x)$

$$\psi_1(x) = \sum_{k=0}^{\infty} \frac{1}{(x+k)^2}$$

For any $K \geqslant 0$

$$\psi_1(x) \geqslant \sum_{k=0}^{K} \frac{1}{(x+k)^2} \geqslant \sum_{k=0}^{K} \frac{1}{(x+k)(x+k+1)}$$

the rightmost sum is

$$\sum_{k=0}^{K} \left\{ \frac{1}{x+k} - \frac{1}{x+k+1} \right\} = \frac{1}{x} - \frac{1}{x+K+1}$$

since $K$ can be arbitrarily large we have $\psi_1(x) \geqslant 1/x$.

Another interesting property is that the difference $\psi_1(x) - 1/x$ is large for small values of $x$. Hence the concentrated log-likelihood takes a strong curvature for small values of $\alpha > 0$. Since estimation is often related as more difficult for small values of $\alpha$, the transformation should improve things.

The ML estimation of the two-parameter gamma using the likelihood concentration described above is available in the `fgamma` function of **Renext**.

## 2.4 Lomax

### 2.4.1 The distribution

The Lomax distribution is a continuous univariate distribution with density given by

$$f(x) = \frac{\alpha}{\beta}\left[1 + \frac{x}{\beta}\right]^{-(\alpha+1)} \qquad x > 0.$$

the support is $[0, +\infty)$. This distribution is a reparameterisation of the GPD with positive shape $\xi > 0$. The Lomax scale and shape parameters are $\beta$ and $\alpha$ and must both be positive.

This distribution is also known as *Pareto Type II distribution* and is described in Wikipedia's page for the Lomax distribution or in Johnson, Kotz, and Balakrishnan (1994). Giles, Feng, and Godwin (2013) provide many details about the ML estimation.

### 2.4.2 The likelihood

Let $X_1, X_2, \ldots, X_n$ be a sample of the Lomax distribution with shape parameter $\alpha$ and scale $\beta$, and Let $\boldsymbol{\theta} = [\alpha,\, \beta]^\top$ the vector of parameters. The log-likelihood $\ell = \log L(\boldsymbol{\theta})$ is

$$\ell = \sum_{i=1}^{n}\Big\{\log\alpha - \log\beta - (\alpha+1)\log[1 + X_i/\beta]\Big\}, \tag{2.4}$$

and its derivatives with respect to the two parameters are

$$\begin{cases} \partial_\alpha \ell &= \dfrac{n}{\alpha} - \sum_i \log[1 + X_i/\beta], \\ \partial_\beta \ell &= -\dfrac{n}{\beta} + \dfrac{(\alpha+1)}{\beta}\sum_i \dfrac{X_i/\beta}{1 + X_i/\beta}. \end{cases}$$

### 2.4.3 Concentrated log-likelihood

Solving $\partial_\alpha \ell = 0$ in $\alpha$ leads to the value $\widehat{\alpha}$

$$\widehat{\alpha}(\beta) = 1/A(\beta),$$

where $A(\beta)$ is defined as

$$A(\beta) := \frac{1}{n}\sum_i \log[1 + X_i/\beta]. \tag{2.5}$$

Note that the function $A(\beta)$ is clearly convex and decreasing and tending to 0 for $\beta \to +\infty$.

The concentrated log-likelihood (with respect to $\alpha$) is obtained by replacing $\alpha$ by $\widehat{\alpha}(\beta)$ in (2.4). Together with its first derivative, it is given by

$$\ell_{\mathrm{c}}(\beta) = -n\left[\log A + \log\beta + A + 1\right], \qquad \ell_{\mathrm{c}}'(\beta) = -n\left[\frac{A+1}{A}A' + \frac{1}{\beta}\right]. \tag{2.6}$$

The derivative $\ell_{\mathrm{c}}'(\beta)$ is not always decreasing – in other words: the concentrated likelihood is not always log-concave. The behaviour of $A(\beta)$ hence that of $\ell_{\mathrm{c}}(\beta)$ for large $\beta$ can be made clear by using an asymptotic expansion of $A(\beta)$ and $A'(\beta)$ when $\beta \to +\infty$. For instance, we have from the definition (2.5)

$$A(\beta) = \frac{1}{n} \sum_i \left\{ X_i\, \beta^{-1} - \frac{1}{2}\, X_i^2\, \beta^{-2} + o(\beta^{-2}) \right\} = \bar{X}\, \beta^{-1} - \frac{1}{2}\, M_2\, \beta^{-2} + o(\beta^{-2})$$

where $M_2$ is the non-central empirical moment $M_2 := (1/n) \sum_i X_i^2$. Similarly, $A'(\beta)$ – which is a rational function in $\beta$ – has an asymptotic development involving the moments $\bar{X}$ and $M_2$. With some algebra, we find

$$\ell_c(\beta) \underset{\beta \to \infty}{\sim} \ell_c(\infty) := -n \log \bar{X} - n \tag{2.7}$$

and

$$\ell_c'(\beta) \underset{\beta \to +\infty}{\sim} \frac{n}{2}\, \bar{X}\, \left[1 - \mathrm{CV}^2\right]\, \beta^{-2} \tag{2.8}$$

where CV is the coefficient of variation of the sample with a denominator $n$ chosen for the variance, i.e, $\mathrm{CV}^2 := [M_2 - \bar{X}^2]/\bar{X}^2$.

When the coefficient of variation is less than 1, the derivative $\ell_c'(\beta)$ turns out to be positive for large $\beta$. It might then be the case that no zero exist for $\ell_c'(\beta)$, meaning that the ML estimator $\widehat{\beta}$ does not exist, morally being $\widehat{\beta} = +\infty$. The likelihood is not log-concave in general. It is worth noting that the limit of $\ell_c(\beta)$ when $\beta \to \infty$ i.e, the right hand side of (2.7) is nothing but the maximal likelihood of the sample under the exponential distribution. Thus, if a finite maximum likelihood estimate $\widehat{\beta}_{\mathtt{ML}} < \infty$ exists, then $\ell_c(\widehat{\beta}_{\mathtt{ML}}) \geqslant -n \log \bar{X} - n$.

### 2.4.4 Fisher information and inference

#### Observed information

The hessian matrix is given by

$$\begin{cases} \partial^2_{\alpha,\alpha} \ell & = -\dfrac{n}{\alpha^2} \\[2mm] \partial^2_{\alpha,\beta} \ell & = \dfrac{1}{\beta} \sum_i \dfrac{X_i/\beta}{1 + X_i/\beta} \\[2mm] \partial^2_{\beta,\beta} \ell & = \dfrac{n}{\beta^2} - \dfrac{(\alpha+1)}{\beta^2} \sum_i \dfrac{X_i/\beta}{1 + X_i/\beta} - \dfrac{(\alpha+1)}{\beta^2} \sum_i \dfrac{X_i/\beta}{[1 + X_i/\beta]^2}. \end{cases}$$

The observed information $\mathbf{J}(\boldsymbol{\theta})$ results by inversion.

#### Expected information

With some simple algebra the hessian matrix can be rewritten as

$$\begin{cases} \partial^2_{\alpha,\alpha} \ell & = -\dfrac{n}{\alpha^2} \\[2mm] \partial^2_{\alpha,\beta} \ell & = \dfrac{n}{\beta}\, [R_1(\alpha) - 1] \\[2mm] \partial^2_{\beta,\beta} \ell & = \dfrac{n}{\beta^2}\, [-\alpha + (\alpha+1)\, R_2(\alpha)] \end{cases}$$

where

$$R_1(\alpha) := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1+X_i/\beta}, \qquad R_2(\alpha) := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{[1+X_i/\beta]^2}.$$

It is easy to prove that the r.v. $V := 1/[1+X/\beta]$ has density $p(v) = \alpha\, v^{\alpha-1}$ on the interval $(0,\, 1)$, leading to

$$\begin{cases} \mathbb{E}R_1(\alpha) &= \dfrac{\alpha}{\alpha+1} \\ \mathbb{E}R_2(\alpha) &= \dfrac{\alpha}{\alpha+2} \end{cases}$$

Thus the expected information matrix is

$$\mathbf{I}(\boldsymbol{\theta}) = n \begin{bmatrix} \dfrac{1}{\alpha^2} & -\dfrac{1}{\beta}\,\dfrac{1}{\alpha+1} \\ -\dfrac{1}{\beta}\,\dfrac{1}{\alpha+1} & \dfrac{1}{\beta^2}\,\dfrac{\alpha}{\alpha+2} \end{bmatrix}$$

and its inverse can be used as an approximation of (co)variance for the estimation

$$\mathbf{I}^{-1}(\boldsymbol{\theta}) = \frac{1}{n} \begin{bmatrix} \alpha^2(\alpha+1)^2 & \alpha(\alpha+1)(\alpha+2)\,\beta \\ \alpha(\alpha+1)(\alpha+2)\,\beta & (\alpha+1)^2(\alpha+2)\,\beta^2/\alpha \end{bmatrix},$$

and $\mathrm{Var}(\widehat{\boldsymbol{\theta}}) \approx \mathbf{I}^{-1}(\widehat{\boldsymbol{\theta}})$ for $n$ large enough.

### 2.4.5 Bounds for $\widehat{\beta}$

As proved in appendix A.5, it can be shown that the ML estimate $\widehat{\beta}$ verifies $\beta_{\mathtt{L}} \leqslant \widehat{\beta} \leqslant \beta_{\mathtt{U}}$ where

$$\beta_{\mathtt{L}} := 0.15 \times \min\{X_i\}, \qquad \beta_{\mathtt{U}} := \max\left\{ \max\{X_i\},\, \beta_2 \right\}$$

and $\beta_2$ is the largest root of the equation in $\beta$

$$[M_1^2 - M_2/2]\,\beta^2 + [M_3 - M_1M_2]\,\beta + M_1M_3 = 0.$$

with $M_k$ standing for the empirical (non central) moment of order $k$.

### 2.4.6 CV close to 1

When the empirical coefficient of variation CV is close to 1 (but still greater than 1), the value of $\widehat{\beta}_{\mathtt{ML}}$ is large. Its determination may be difficult because the concentrated log-likelihood is very flat near the optimum. Using (2.6), the derivative can be shown to have the following asymptotic expansion for large $\beta$

$$\ell'_{\mathtt{c}}(\beta) = -n\left\{a_1\beta^{-2} + a_2\beta^{-3} + o(\beta^{-3})\right\}$$

where

$$a_1 = \frac{M_2 - 2M_1^2}{2M_1}, \qquad a_2 = -\frac{8M_1M_3 - 3M_2^2 - 12M_1^2M_2}{12M_1^2}.$$

Then $\widehat{\beta}_{\mathtt{ML}}$ can be computed as $-a_2/a_1$. However, this value may fail to be positive.

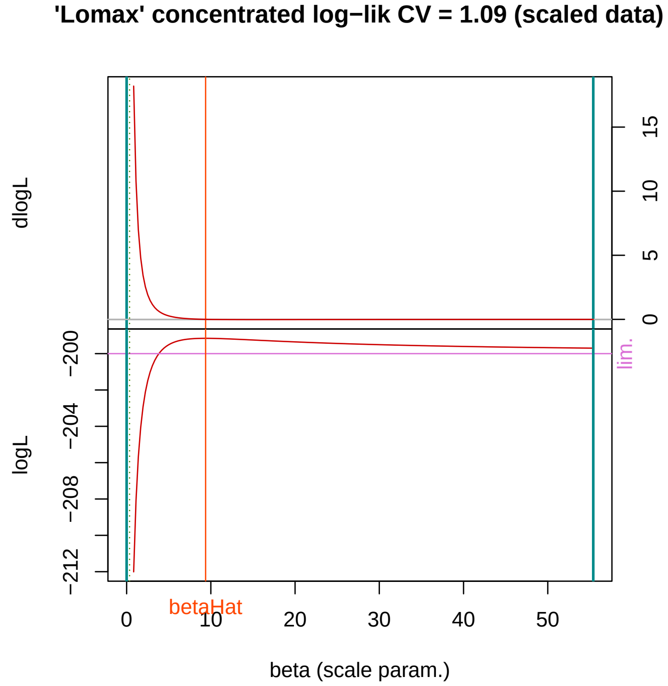



Figure 2.1: Concentrated log-likelihood (bottom) and its derivative (top) for a Lomax sample.

### 2.4.7 Example

The following code is taken from the help `?flomax`. It produces the plot shown on Figure 2.1. It can be seen from the figure that the concentrated log-likelihood $\ell_{\mathsf{c}}(\beta)$ decreases to a finite limit for large $\beta$. The limit materialised with an horizontal line is nothing but the exponential log-likelihood. Note that the data shown on the figure have been scaled for the sake of numerical precision: $X_i$ is replaced by $X_i/\bar{X}$. Thus the shown maximum has to be transformed back to give the estimate $\widehat{\beta}$

```
R> set.seed(1234)
R> n <- 200; alpha <- 2 + rexp(1); beta <- 1 + rexp(1)
R> x <- rlomax(n, scale = beta, shape = alpha)
R> res <- flomax(x, plot = TRUE)
R> res$estimate

    shape     scale
10.370332  3.295373

R> res$estimate["scale"] / mean(x)

   scale
9.377393
```

## 2.5 Maxlo

### 2.5.1 The distribution

The maxlo distribution is a continuous univariate distribution with density given by

$$f(x) = \frac{\alpha}{\beta}\left[1 - \frac{x}{\beta}\right]^{\alpha-1} \qquad 0 < x < \beta.$$

The shape parameter is $\alpha > 0$ and scale parameter is $\beta > 0$; the support is $[0, \beta]$. This distribution is a reparameterisation of the GPD with negative shape $\xi < 0$.

### 2.5.2 The likelihood

Let $X_1, X_2, \ldots, X_n$ be a sample of the maxlo distribution with shape parameter $\alpha > 0$ and scale $\beta > 0$, and let $\boldsymbol{\theta} := [\alpha, \beta]^\top$. The likelihood is defined only when $\beta \geqslant \max\{X_i\}$. The log-likelihood $\ell = \log L(\boldsymbol{\theta})$ is

$$\ell = \sum_{i=1}^{n} \Big\{ \log\alpha - \log\beta + (\alpha - 1)\log[1 - X_i/\beta] \Big\}.$$

Let $X_{(n)} := \max\{X_i\}$; note that for small $\beta$ i.e. $\beta \approx X_{(n)}$

$$\lim_{\beta \to X_{(n)}} \ell(\alpha, \beta) = \begin{cases} \infty & 0 < \alpha < 1, \\ -n\log X_{(n)} & \alpha = 1, \\ -\infty & \alpha > 1 \end{cases}.$$

Thus whatever be the $X_i$, the log-likelihood is maximised for $\beta = X_{(n)}$ and $0 < \alpha < 1$. Strictly speaking, we can say that there exists an infinity of ML estimates. Note that $X_{(n)}$ estimates $\beta$ with a positive bias.

This problem of an infinite likelihood for an infinity of parameter values is generally hidden in numerical computations, because the values of $\log(\epsilon)$ that can be computed for $\epsilon \approx 0$ remain quite small, typically near $-36.7$ for $\epsilon \approx 10^{-16}$. The problem disappears as soon as a constraint $\alpha \geqslant \alpha_{\text{L}}$ with $\alpha_{\text{L}} > 1$ is imposed to $\alpha$. In practice, such a constraint will rarely be active because the numerical evaluations can hardly detect a value $\ell$ greater than the local maximum when it exists.

The derivatives of $\ell$ with respect to the two parameters are

$$\begin{cases} \partial_\alpha \ell & = \dfrac{n}{\alpha} + \sum_i \log[1 - X_i/\beta], \\ \partial_\beta \ell & = -\dfrac{n}{\beta} + \dfrac{(\alpha - 1)}{\beta} \sum_i \dfrac{X_i/\beta}{1 - X_i/\beta}. \end{cases}$$

A likelihood concentration is possible for the unconstrained case and also, with some care, for the constrained case.

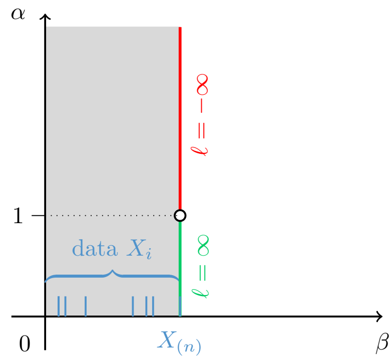



Figure 2.2: Whatever be the $X_i$, the log-likelihood $\ell$ for the maxlo distribution takes infinite values on the border line $\beta = X_{(n)}$.

### 2.5.3 Concentrated log-likelihood

**Unconstrained**

Solving $\partial_\alpha \ell = 0$ in $\alpha$ leads to the value $\widehat{\alpha}$

$$\widehat{\alpha}(\beta) = \frac{1}{A(\beta)}, \qquad A(\beta) := -\frac{1}{n}\sum_i \log[1 - X_i/\beta]. \tag{2.9}$$

The concentrated log-likelihood (with respect to $\alpha$) and its first derivative is given by

$$\ell_{\mathrm{c}}(\beta) = -n\,[\log A + \log\beta - A + 1], \qquad \ell_{\mathrm{c}}'(\beta) = -n\left[\frac{1-A}{A}A' + \frac{1}{\beta}\right]. \tag{2.10}$$

Since $A(\beta)$ tends to 0 when $\beta \to +\infty$, we have $A(\beta) \leqslant 1$ for $\beta$ large enough and the concentrated log-likelihood $\ell_{\mathrm{c}}(\beta)$ is decreasing for large $\beta$. The function $A(\beta)$ is clearly convex and decreasing. However, the derivative $\ell_{\mathrm{c}}'(\beta)$ is not always decreasing – in other words: the concentrated likelihood is not always log-concave. Using an asymptotic expansion of $A(\beta)$ and $A'(\beta)$ for $\beta \to +\infty$, we have

$$\ell_{\mathrm{c}}(\beta) \underset{\beta\to\infty}{\sim} -n\log\bar{X} - n \tag{2.11}$$

and

$$\ell_{\mathrm{c}}'(\beta) \underset{\beta\to+\infty}{\sim} -\frac{n}{2}\,\bar{X}\,\left[1 - \mathrm{CV}^2\right]\,\frac{1}{\beta^2}$$

where CV is the coefficient of variation of the sample, computed using $n$ as the denominator in the variance, i.e. $\mathrm{CV}^2 := [M_2 - \bar{X}^2]/\bar{X}^2$.

When the coefficient of variation is greater than 1, the derivative $\ell_{\mathrm{c}}'(\beta)$ is positive for large $\beta$. It might then be the case that no zero exist for $\ell_{\mathrm{c}}'(\beta)$, meaning that the ML estimator does not exist. The likelihood is not log-concave in general. As for the Lomax distribution, the limit of $\ell_{\mathrm{c}}(\beta)$ when $\beta \to \infty$ i.e, the right hand side of (2.11) is the maximal likelihood of the sample under the exponential distribution. Thus if a maximum likelihood estimate $\widehat{\beta} < \infty$ exists then $\ell_{\mathrm{c}}(\widehat{\beta}) \geqslant -n\log\bar{X} - n$.

### Constrained

Assume that we impose the constraint $\alpha \geqslant \alpha_{\mathtt{L}}$ with a fixed value $\alpha_{\mathtt{L}} > 1$. Let $\ell^\star_{\mathtt{c}}(\beta)$ denote the concentrated log-likelihood with the constraint, i.e.

$$\ell^\star_{\mathtt{c}}(\beta) := \max_{\alpha \geqslant \alpha_{\mathtt{L}}} \ell(\beta,\, \alpha).$$

We still can find the value $\widehat{\alpha}(\beta)$ of $\alpha$ such that $\ell(\beta,\, \alpha)$ is maximal for given $\beta$

$$\widehat{\alpha}(\beta) = \begin{cases} \alpha_{\mathtt{L}} & \text{if } X_{(n)} < \beta < \beta^\star, \\ 1/A(\beta) & \text{if } \beta \geqslant \beta^\star, \end{cases}$$

where $\beta^\star$ is the unique solution of $A(\beta) = 1/\alpha_{\mathtt{L}}$ over $(X_{(n)},\, \infty)$. No closed expression can be given for $\beta^\star$; however since $A(\beta)$ is decreasing, the condition $\beta < \beta^\star$ is evaluated thanks to the equivalence

$$\beta < \beta^\star \quad \text{iif} \quad A(\beta) > 1/\alpha_{\mathtt{L}}. \tag{2.12}$$

Note that $\ell^\star_{\mathtt{c}}(\beta)$ and its derivative remain given by the expression in (2.10) only for $\beta \geqslant \beta^\star$, and

$$\ell^\star_{\mathtt{c}}(\beta) = \begin{cases} n\left\{\log \alpha_{\mathtt{L}} - \log \beta - (\alpha_{\mathtt{L}} - 1)A(\beta)\right\} & \text{if } X_{(n)} < \beta < \beta^\star, \\ \ell_{\mathtt{c}}(\beta) & \text{if } \beta \geqslant \beta^\star. \end{cases} \tag{2.13}$$

Compared to the unconstrained case, the first expression leads to a different behaviour for $\beta \approx X_{(n)}$ since $\ell^\star_{\mathtt{c}}(\beta)$ tends to $-\infty$ there, while $\ell_{\mathtt{c}}(\beta)$ tends to $\infty$. One can prove that $\ell^\star_{\mathtt{c}}(\beta)$ is increasing for $X_{(n)} < \beta \leqslant \beta_{\mathtt{L}}$ where $\beta_{\mathtt{L}}$ is given by

$$\beta_{\mathtt{L}} := \min\left\{\beta^\star,\, \lambda X_{(n)}\right\}, \qquad \lambda := \frac{1}{1 - (\alpha_{\mathtt{L}} - 1)/(n\alpha_{\mathtt{L}})}. \tag{2.14}$$

with $\lambda > 1$, see appendix A.6. Provided that $\mathtt{CV} < 1$, the concentrated log-likelihood is increasing for $\beta \approx X_{(n)}$ and is decreasing for $\beta$ large enough. Then, there exists at least one global maximum over $(X_{(n)},\, \infty)$. Moreover, the constrained ML estimate is easily found by using a one-dimensional maximisation of $\ell^\star_{\mathtt{c}}(\beta)$. It might happen that the estimate $\widehat{\beta}$ is such that $\widehat{\beta} \leqslant \beta^\star$, in which case the corresponding $\widehat{\alpha}$ is located at the bound $\alpha_{\mathtt{L}}$. In such circumstance, the information matrix must not be used since the likelihood may not be differentiable at the estimate.

Note that (2.14) has a limited practical value inasmuch the value $\beta^\star$ is not known explicitly. To determine which of the two candidates in (2.14) is the min, we can compute $A(\beta)$ with $\beta = \lambda X_{(n)}$, and then use the equivalence (2.12) rather than a zero-finding for $\beta^\star$.

## 2.5.4 Information matrix and inference

### Observed information

The hessian matrix is given by

$$\begin{cases} \partial^2_{\alpha,\alpha}\ell &= -\dfrac{n}{\alpha^2} \\ \partial^2_{\alpha,\beta}\ell &= \dfrac{1}{\beta}\sum_i \dfrac{X_i/\beta}{1 - X_i/\beta} \\ \partial^2_{\beta,\beta}\ell &= \dfrac{n}{\beta^2} - \dfrac{(\alpha-1)}{\beta^2}\sum_i \dfrac{X_i/\beta}{1 - X_i/\beta} - \dfrac{(\alpha-1)}{\beta^2}\sum_i \dfrac{X_i/\beta}{[1 - X_i/\beta]^2} \end{cases}$$

from which the observed information matrix $\mathbf{J}(\boldsymbol{\theta})$ is obtained by inversion.

**Expected information**

Using some simple algebra, the hessian above can be rewritten as

$$\begin{cases} \partial^2_{\alpha,\alpha}\ell & = -\dfrac{n}{\alpha^2} \\ \partial^2_{\alpha,\beta}\ell & = \dfrac{n}{\beta}\,[R_1(\alpha) - 1] \\ \partial^2_{\beta,\beta}\ell & = \dfrac{n}{\beta^2}\,[\alpha - (\alpha - 1)\,R_2(\alpha)] \end{cases}$$

where

$$R_1(\alpha) := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1 - X_i/\beta}, \qquad R_2(\alpha) := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{[1 - X_i/\beta]^2}.$$

It is easy to prove that the r.v. $V := 1/[1 - X/\beta]$ is Pareto distributed with survival $S_V(v) = (1/v)^\alpha$ on the interval $(1, +\infty)$. It has a finite expectation for $\alpha > 1$ and a finite second moment for $\alpha > 2$ corresponding to

$$\begin{cases} \mathbb{E}R_1(\alpha) & = \dfrac{\alpha}{\alpha - 1} \\ \mathbb{E}R_2(\alpha) & = \dfrac{\alpha}{\alpha - 2} \end{cases}$$

Thus the expected information matrix exists for $\alpha > 2$ and is given by

$$\mathbf{I}(\boldsymbol{\theta}) = n \begin{bmatrix} \dfrac{1}{\alpha^2} & -\dfrac{1}{\beta}\dfrac{1}{\alpha - 1} \\ -\dfrac{1}{\beta}\dfrac{1}{\alpha - 1} & \dfrac{1}{\beta^2}\dfrac{\alpha}{\alpha - 2} \end{bmatrix}$$

and its inverse can be used as an approximation of (co)variance for the estimation

$$\mathbf{I}^{-1}(\boldsymbol{\theta}) = \frac{1}{n}\begin{bmatrix} \alpha^2(\alpha - 1)^2 & \alpha(\alpha - 1)(\alpha - 2)\,\beta \\ \alpha(\alpha - 1)(\alpha - 2)\,\beta & (\alpha - 1)^2(\alpha - 2)\,\beta^2/\alpha \end{bmatrix},$$

and $\text{Var}(\widehat{\boldsymbol{\theta}}) \approx \mathbf{I}^{-1}(\widehat{\boldsymbol{\theta}})$ for $n$ large enough.

**Remark.** The information matrix matrix can only be computed when $\alpha > 2$. When $\widehat{\alpha}_{\text{ML}} \leqslant 2$, no covariance matrix or even standard deviations can be given for the ML estimate.

### 2.5.5 Bounds for $\widehat{\beta}$

As proved in appendix A.6, it can be shown that the constrained ML estimate $\widehat{\beta}$ verifies

$$\beta_{\text{L}} \leqslant \widehat{\beta} \leqslant \beta_{\text{U}}$$

where $\beta_{\text{L}}$ is given by (2.14) above, and $\beta_{\text{U}}$ defined as

$$\beta_{\text{U}} := \max\left\{2 \times \max\{X_i\},\, \beta_2\right\},$$

with $\beta_2$ being the largest root of the equation

$$3\left[M_2 - 2M_1^2\right]\beta^2 + 2\left[5M_3 - 3M_1M_2\right]\beta - 12M_1M_3 = 0,$$

and $M_k$ standing for the empirical (non central) moment of order $k$.

### 2.5.6 CV close to 1

When the empirical coefficient of variation CV is close to 1 (but still smaller than 1), the value of $\widehat{\beta}_{\texttt{ML}}$ is large. Its determination may be difficult because the concentrated log-likelihood is very flat near the optimum. Using (2.10), the derivative can be shown to have the following asymptotic expansion for large $\beta$

$$\ell'_{\texttt{c}}(\beta) = -n\left\{a_1\beta^{-2} + a_2\beta^{-3} + o(\beta^{-3})\right\}$$

where

$$a_1 = \frac{M_2 - 2M_1^2}{2M_1}, \qquad a_2 = \frac{8M_1M_3 - 3M_2^2 - 12M_1^2M_2}{12M_1^2}.$$

Then $\widehat{\beta}_{\texttt{ML}}$ can be computed as $-a_2/a_1$. However, this value may fail to be positive.

### 2.5.7 Example

Mirroring that of the Lomax example given in section 2.4.7, the following code is taken from the help `?fmaxlo`. It produces the plot shown on Figure 2.3. It can be guessed from the figure that the concentrated log-likelihood $\ell_{\texttt{c}}(\beta)$ decreases to a finite limit for large $\beta$. The limit materialised with an horizontal line is nothing but the exponential log-likelihood. As in 2.4.7, the data shown have been scaled: $X_i$ is replaced by $X_i/\bar{X}$, and the shown maximum has to be transformed back to give the estimate $\widehat{\beta}$

```
R> ## generate sample
R> set.seed(1234)
R> n <- 200; alpha <- 2 + rexp(1); beta <- 1 + rexp(1)
R> x <- rmaxlo(n, scale = beta, shape = alpha)
R> res <- fmaxlo(x, plot = TRUE)
R> res$estimate

   shape    scale
4.204632 1.175853

R> res$estimate["scale"] / mean(x)

   scale
5.209352
```

## 2.6 Generalised Pareto Distribution (GPD)

### 2.6.1 The distribution

- For $\xi < 0$ the distribution is the maxlo distribution with $\alpha = -1/\xi$ and $\beta = -\sigma/\xi$.
- For $\xi > 0$ the distribution is Lomax with $\alpha = 1/\xi$ and $\beta = \sigma/\xi$.

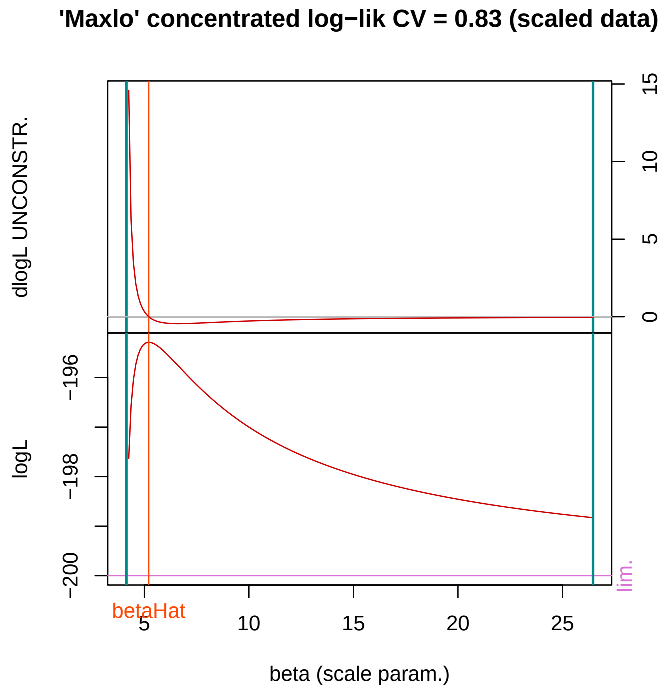



Figure 2.3: Concentrated log-likelihood (bottom) and its derivative (top) for a Maxlo sample.

### 2.6.2 ML estimation

The likelihood can be evaluated when $1 + \xi X_i/\sigma > 0$ for all $i$. For $\xi < 0$, these conditions boils down to $X_{(n)} < -\sigma/\xi$.

As remarked before in sections 2.4 and 2.5, and as it was also noted by del Castillo and Daoudi (2009), the sign of the ML estimate $\widehat{\xi}$ is that of $\texttt{CV} - 1$ where `CV` is the sample coefficient of variation. Having chosen a small technical parameter $\delta > 0$, we can proceed as follows.

- if $\texttt{CV} < 1 - \delta$, fit the maxlo distribution by ML as was explained in section 2.5.
- If $\texttt{CV} > 1 + \delta$, fit the maxlo distribution by ML as was explained in section 2.4.
- If $1 - \delta \leqslant \texttt{CV} \leqslant 1 + \delta$, then $\widehat{\xi} \approx 0$ and $\widehat{\sigma} \approx \bar{X}$.

This approach is essentially similar to that proposed by Grimshaw (1993). It is implemented in the `fGPD` function of **Renext**.

**Remark.** Without imposing any constraints on the shape parameter $\xi$, the ML estimate of $\boldsymbol{\theta}$ is never unique and whatever be the $X_i$, the likelihood is maximal for $\xi < -1$ and $\sigma = -\xi\, X_{(n)}$. For such a value, the likelihood is infinite: $L(\boldsymbol{\theta}) = \infty$.

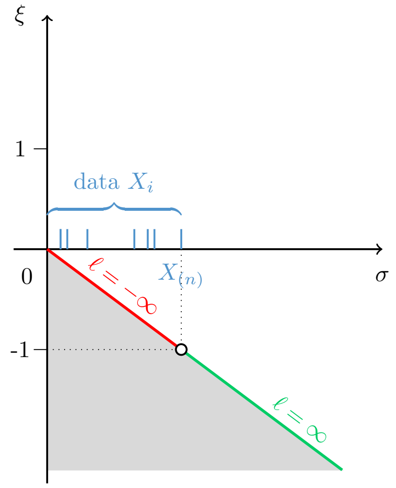



Figure 2.4: Whatever be the $X_i$, the log-likelihood $\ell$ for the GPD takes infinite values on the border line $\sigma + \xi X_{(n)} = 0$.

### 2.6.3 Information matrix and inference

#### Observed information

When the coefficient of variation is far enough from 1, a Lomax or a maxlo distribution is fitted and the expected information is obtained. The expected information for the GPD results by using the jacobian of the transformation into a GPD ("Lomax to GPD" or "maxlo to GPD").

When the coefficient of variation is close to 1, the ML estimates are $\widehat{\sigma} = \bar{X}$ and $\widehat{\xi} = 0$. By using a development when $\xi \approx 0$ for each of the quantities $A$, $B$ and $B^2$ defined in appendix A.2.3, the observed information matrix and its inverse are found to be

$$\mathbf{J}(\boldsymbol{\theta}) = n \begin{bmatrix} 1/\sigma^2 & 1/\sigma \\ 1/\sigma & \frac{2}{3}[M_3^\star - 3] \end{bmatrix}, \qquad \mathbf{J}^{-1}(\boldsymbol{\theta}) = \frac{1}{n\,\left[\frac{2}{3}M_3^\star - 3\right]} \begin{bmatrix} 2\sigma^2\,[M_3^\star/3 - 1] & -\sigma \\ -\sigma & 1 \end{bmatrix}.$$

where $M_3^\star = M_3/M_1^3$ is the standardised third moment of the $X_i$. It can be remarked that $\mathbf{J}$ is positive definite when $M_3^\star > 9/2$, while a direct application of the Cauchy-Schwarz inequality with $M_2 = 2M_1$ only gives $M_3^\star > 4$. For an exponentially distributed sample with $n$ large, we have $M_3^\star \approx 6$.

#### Expected information

By inverting the expected information matrix

$$\mathbf{I}^{-1}(\boldsymbol{\theta}) = \frac{\xi + 1}{n} \begin{bmatrix} 2\sigma^2 & -\sigma \\ -\sigma & \xi + 1 \end{bmatrix},$$

and $\mathrm{Var}(\widehat{\boldsymbol{\theta}}) \approx \mathbf{I}^{-1}(\widehat{\boldsymbol{\theta}})$ for $n$ large enough.

**Remark.** It can be shown that $\xi$ and $\nu := \log[(1+\xi)\sigma]$ are orthogonal parameters, which can be useful for some applications. The GPD can be reparameterised using $[\xi,\, \nu]^\top$.

## 2.7 Shifted Weibull ⋆

As in the classical Weibull case, let us write $\eta := \beta^\alpha$.

$$f(y) = \frac{\alpha}{\eta}\,(y+\delta)^{\alpha-1}\,\exp\left\{-\frac{(y+\delta)^\alpha}{\eta} + \frac{\delta^\alpha}{\eta}\right\} \qquad y > 0$$

$$\ell = \sum_i \left\{\log\alpha - \log\eta + (\alpha-1)\log(Y_i+\delta) - \frac{1}{\eta}\,(Y_i+\delta)^\alpha + \frac{1}{\eta}\,\delta^\alpha\right\}$$

Its derivatives with respect to the three parameters are

$$\begin{cases} \dfrac{\partial\ell}{\partial\alpha} &= \dfrac{n}{\alpha} + \sum_i \log(Y_i+\delta) - \dfrac{1}{\eta}\sum_i \left[(Y_i+\delta)^\alpha \log(Y_i+\delta) - \delta^\alpha \log\delta\right] \\ \dfrac{\partial\ell}{\partial\eta} &= -\dfrac{n}{\eta} + \dfrac{1}{\eta^2}\sum_i \left[(Y_i+\delta)^\alpha - \delta^\alpha\right] \\ \dfrac{\partial\ell}{\partial\delta} &= (\alpha-1)\sum_i \dfrac{1}{Y_i+\delta} - \dfrac{\alpha}{\eta}\sum_i \left[(Y_i+\delta)^{\alpha-1} - \delta^{\alpha-1}\right] \end{cases}$$

The value of $\eta$ that maximises $\ell$ for given $\alpha$ and $\delta$ is

$$\widehat{\eta}(\alpha,\,\delta) = \frac{1}{n}\sum_i \left[(Y_i+\delta)^\alpha - \delta^\alpha\right]$$

and the concentrated log-likelihood is

$$\ell^{\mathrm{c}} = n\log\alpha - n\log\widehat{\eta} + (\alpha-1)\sum_i \log(Y_i+\delta) - n$$

## 2.8 Negative binomial Lévy process

### 2.8.1 Estimation

Let $N_1$, $N_2$, ..., $N_n$ be $n$ independent r.vs having negative binomial distribution with the same probability parameter $p$. The size parameter for $N_k$ is assumed to be $r = \gamma w_k$ where $w_k > 0$ is the duration of the $k$-th observation period. Following an "R-like" notation

$$N_k \sim \texttt{nbinom(size =}\gamma w_k\texttt{, prob =}p\texttt{)}$$

The parameters to be estimated are $\gamma > 0$ and $p$. Note that the model is a special case of Generalized Linear Model (GLM) with one covariate $w$ and no constant. The classical negative binomial GLM uses $\ell(\mu) = \log(\mu)$ as link function with a constant size parameter $r$, hence a variable $p$. Our model specifies an identity link $\ell(\mu)$ and a fixed (unknown) $p$ which is very different.

When the $w_k$ are all equal, classical estimators can be obtained from the (sample) dispersion coefficient. For the standard case: $w_k = 1$ for all $k$, the estimators $\widetilde{\gamma}$ and $\widetilde{p}$ are given by

$$\widetilde{p} = \frac{\overline{N}}{S^2} \qquad \widetilde{\gamma} = \frac{\overline{N}\,\widetilde{p}}{1-\widetilde{p}}$$

$\overline{N}$ and $S^2$ being the sample mean and variance of the $N_k$.

We now turn to the ML estimation. The likelihood for observation $k$ is

$$\frac{\Gamma(\gamma w_k + N_k)}{\Gamma(\gamma w_k)\Gamma(N_k+1)}\, p^{\gamma w_k} q^{N_k}$$

and the sample log-likelihood is

$$\log L = \sum_{k=1}^{a} \log\Gamma(\gamma w_k + N_k) - \log\Gamma(\gamma w_k) - \log\Gamma(N_k+1) + \gamma w_k \log p + N_k \log q$$

The derivatives are (sums are from $k = 1$ to $n$)

$$\begin{cases} \partial_\gamma \log L & = \sum_k w_k \left[\psi(\gamma w_k + N_k) - \psi(\gamma w_k) + \log p\right] \\ \partial_p \log L & = \sum_k \gamma w_k/p - N_k/(1-p) \end{cases} \tag{2.15}$$

where $\psi = \Gamma'/\Gamma$ is the digamma function. The first order condition $\partial \log L/\partial\gamma = 0$ writes

$$\frac{p}{1-p} = \gamma \times \frac{\sum_k w_k}{\sum_k N_k} = \frac{\gamma}{R} \tag{2.16}$$

where $R$ is the event rate over the observation period

$$R = \frac{\sum_k N_k}{\sum_k w_k}$$

Equation (2.16) gives the walue $\widehat{p}(\gamma)$ of $p$ maximising $\log L$ for a fixed $\gamma$

$$\widehat{p}(\gamma) = \frac{\gamma}{\gamma + R}$$

One can therefore use the *concentrated log-likelihood* with respect to $p$, i.e. the function $L_{\mathsf{c}}(\gamma) = L\left[\gamma,\, \widehat{p}(\gamma)\right]$ depending on the sole parameter $\gamma$.

The second order derivatives of $\log L$ are

$$\begin{cases} \partial^2_{\gamma,\gamma} \log L &= \sum_k w_k^2 \left[\psi_1(\gamma w_k + N_k) - \psi_1(\gamma w_k)\right] \\ \partial^2_{\gamma\, p} \log L &= \sum_k w_k / p \\ \partial^2_{p,p} \log L &= -\sum_k \gamma w_k/p^2 - N_k/(1-p)^2 \end{cases} \tag{2.17}$$

where $\psi_1 = \psi'$ is the trigamma function (second order derivative of $\log \Gamma$). The trigamma function is strictly decreasing and the right side of the first equation is negative as expected.

As will be justified later, the log-likelihood is a *concave* function of $[\gamma,\, p]^\top$ and so is the concentrated log-likelihood $\log L_{\mathsf{c}}(\gamma)$. In other words $\log L_{\mathsf{c}}(\gamma)$ has a decreasing first derivative. The ML estimator of $\gamma$ can therefore be found in a safe way. One can either find the unique root of the derivative of $\log L_{\mathsf{c}}(\gamma)$, or maximise this concave function of a single variable. The zero-finding begins with the dertermination of an interval containing the root or with the determination of a good initial value. A few steps of an iterative method like Newton-Raphson will then be enough to reach the maximum.

Once the ML estimators $\widehat{\gamma}$ and $\widehat{p}$ have been obtained, the observed information matrix can be computed using (2.17). Note that the expected and observed information matrices are identical here.

### 2.8.2 Log-concavity

The log-likelihood can be written as

$$\log L = C + \left\{ \sum_k \log \Gamma(\gamma w_k + N_k) - \log \Gamma(\gamma w_k) \right\} + \left\{ \sum_k \gamma w_k \log p + N_k \log(1-p) \right\}$$

where $C$ does not depend on the parameters. We claim that in both sums on right side *all terms are concave functions of their arguments*. Since the sum of concave functions is also concave, $\log L$ will be concave.

The concavity of any term in the second sum is easily checked from its second derivative matrix (hessian). For the first sum (which depends only on $\gamma$)

$$\frac{\mathrm{d}^2}{\mathrm{d}\gamma^2}\left[\log \Gamma(\gamma w_k + N_k) - \log \Gamma(\gamma w_k)\right] = w_k^2 \left[\psi_1(\gamma w_k + N_k) - \psi_1(\gamma w_k)\right] \leqslant 0$$

since the trigamma function $\psi_1(x)$ is decreasing.

When the log likelihood is concave and regular, it can be shown that the concentrated log-likelihood (with respect to some of the parameters) is also concave. Therefore our $\log L_{\mathsf{c}}(\gamma)$ is a concave function of $\gamma$.

# Chapter 3

# GPD-POT and GEV block maxima

This chapter is devoted to the relation between POT models and Block Maxima (BM). We only consider the classical frameworks where POT excesses are assumed to be GPD, and where the BM follow a GEV distribution. We also study the relation between POT and the classical $r$ largest statistics context.

Although the BM and $r$ largest are usually derived from asymptotic considerations, it is quite well-known that the same models result from a temporal aggregation of the Marked Process $[T_k,\, X_k]$ as used in POT.

> "The distribution of the maximum of a Poisson number of i.i.d. excesses over a high threshold is GEV". *(Embrechts* et al. *1996, chap. 3).*

However, a number of details are most generally omitted. These relate to the possibility that a block has no observations, in which case the maximum does not exist.

## 3.1 GPD-POT to GEV-max

### 3.1.1 Temporal aggregation of the marked process

In this section we consider the case of a partial observation (or temporal aggregation) for the marked process $[T_k,\, X_k]$. We assume that we have $B$ disjoints periods or *blocks* with known durations $w_b$ for $b = 1,\, 2,\, \dots,\, B$. We consider the following r.vs representing the number of events and the maximum of the marks for the block $b$

$$N_b := \#\{k;\, T_k \in \text{block } b\}, \qquad M_b := \max_{T_k \in \text{ block } b} X_k.$$

Observe that $M_b$ is defined only when the block $b$ contains at least one observation, a condition which is fulfilled with the probability $\Pr\{N_b \neq 0\} = 1 - \exp\{-\lambda w_b\}$. When a block duration is small relative to $1/\lambda$, this probability is not that close to one; e.g. for $\lambda w_b = 1$ we find $\Pr\{N_b \neq 0\} \approx 0.63$. Moreover for $B = 10$ blocks with $\lambda w_b = 1$, there is about one chance in a hundred that all blocks have an observation. When $B$ is large enough, systematically missing observations $M_b$ due to empty blocks will ineluctably occur, unless $\lambda$ is large enough. This will be discussed later.

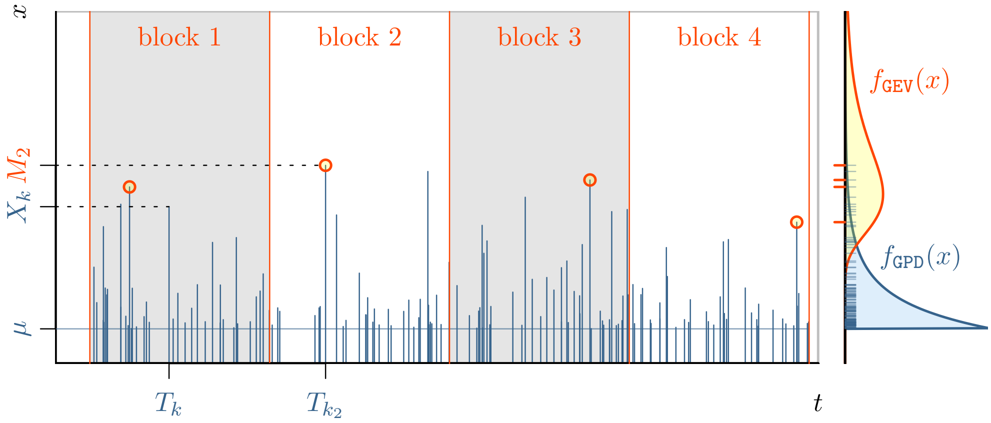



Figure 3.1: Temporal aggregation of the marked process with constant block durations $w_b \equiv w$. The distribution of the marks $X_k$ is $\texttt{GPD}(\mu,\, \sigma,\, \xi)$ while block maxima $M_b$ have distribution $\texttt{GEV}(\mu^\star,\, \sigma^\star,\, \xi^\star)$.

Since $N_b$ and the observations $X_k$ are independent and since $M_b$ is the maximum of $N_b$ independent r.vs we know that

$$\Pr\left[M_b \leqslant x \,|\, N_b = k\right] = F_X(x)^k \qquad \text{for } k \geqslant 1, \tag{3.1}$$

which allows the determination of the distribution of $M_b$ in the following theorem. Since the maxima $M_b$ corresponding to disjoint blocks are independent, the joint distribution of the maxima results. When $N_b = 0$ we can set $M_b := -\infty$ and the joint distribution is of mixed type with a positive probability mass on vectors with some $M_b$ equal to $-\infty$.

It will be convenient to denote by $\mathcal{I}_{\texttt{GPD}}(\mu,\, \sigma,\, \xi)$ the support of the GPD with parameters $\mu$, $\sigma$ and $\xi$. Similar notations will be used for the GEV distribution. The support of the block maxima $M_b$ is the same as that of the marks $X_k$.

**Theorem 3.1.** *The r.vs $M_b$ corresponding to disjoint blocks are independent. The marginal distribution of $M_b$ is given by*

$$\Pr\{N_b \neq 0,\, M_b \leqslant x\} = \exp\{-\lambda w_b S_X(x)\} - \exp\{-\lambda w_b\}. \tag{3.2}$$

*If the marks are GPD with $X_k \sim \texttt{GPD}(\mu,\, \sigma,\, \xi)$, then for all $x \in \mathcal{I}_{\texttt{GPD}}(\mu,\, \sigma,\, \xi)$, we have*

$$\Pr\left\{N_b \neq 0,\, x \leqslant M_b \leqslant x + \mathrm{d}x\right\} = f_{\texttt{GEV}}(x;\, \mu_b^\star,\, \sigma_b^\star,\, \xi_b^\star)\,\mathrm{d}x \tag{3.3}$$

*where the GEV parameters are given by*

$$\mu_b^\star = \mu + \frac{(\lambda w_b)^\xi - 1}{\xi}\,\sigma, \quad \sigma_b^\star = (\lambda w_b)^\xi\,\sigma, \quad \xi_b^\star = \xi, \qquad \textit{for } \xi \neq 0, \tag{3.4}$$

*or by*

$$\mu_b^\star = \mu + \log(\lambda w_b)\,\sigma, \quad \sigma_b^\star = \sigma, \quad \xi_b^\star = \xi = 0, \qquad \textit{for } \xi = 0, \tag{3.5}$$

*depending on the value of $\xi$. In both cases, the likelihood of an observation $M_b$ is computed as if $M_b$ comes from a sample of $\texttt{GEV}(\mu_b^\star,\, \sigma_b^\star,\, \xi_b^\star)$.*

*When the blocks have the same duration $w_b \equiv w$, the maxima $M_b$ form an i.i.d. sample of the GEV distribution $\texttt{GEV}(\mu^\star,\, \sigma^\star,\, \xi^\star)$.*

*Proof.* In the proof we will omit the index $b$. Consider $x$ in the support of the distribution of the marks $X_k$. We have

$$\Pr\left[M \leqslant x \,|\, N \neq 0\right] = \sum_{k=1}^{\infty} \Pr\left[M \leqslant x \,|\, N = k\right] \times \Pr\left[N = k \,|\, N \neq 0\right] = \sum_{k=1}^{\infty} F_X(x)^k \,\Pr\left[N = k \,|\, N \neq 0\right]$$

where (3.1) was used in the second equality. Now, multiplying by $\Pr\{N \neq 0\}$

$$\begin{aligned}
\Pr\{N \neq 0,\, M \leqslant x\} &= \sum_{k=1}^{\infty} F_X(x)^k \times \Pr\{N = k\} = \sum_{k=1}^{\infty} F_X(x)^k \times e^{-\lambda w}[\lambda w]^k/k! \\
&= e^{-\lambda w} \sum_{k=1}^{\infty} [\lambda w\, F_X(x)]^k/k! = e^{-\lambda w}\left[\exp\{\lambda w\, F_X(x)\} - 1\right] \\
&= \exp\{-\lambda w\, S_X(x)\} - \exp\{-\lambda w\}
\end{aligned}$$

which is (3.3). By derivation with respect to $x$, we get

$$\Pr\left\{N \neq 0,\, x \leqslant M \leqslant x + \mathrm{d}x\right\} = \frac{\mathrm{d}}{\mathrm{d}x}\left\{e^{-\lambda w\, S_X(x)}\right\} \times \mathrm{d}x. \tag{3.6}$$

From now on let us assume that the marks are $\texttt{GPD}(\mu,\, \sigma,\, \xi)$ and that $\mu^\star$, $\sigma^\star$ and $\xi^\star$ are given by (3.4) for $\xi \neq 0$ or (3.5) for $\xi = 0$. First, since $\mu - \sigma/\xi = \mu^\star - \sigma^\star/\xi^\star$ for $\xi \neq 0$, it is easy to see that for any vector $[\mu,\, \sigma,\, \xi]^\top$ we have

$$\mathcal{I}_{\texttt{GPD}}(\mu,\, \sigma,\, \xi) = \mathcal{I}_{\texttt{GPD}}(\mu^\star,\, \sigma^\star,\, \xi^\star) \subset \mathcal{I}_{\texttt{GEV}}(\mu^\star,\, \sigma^\star,\, \xi^\star)$$

see Figure 3.2. Moreover, we have for any $x \in \mathcal{I}_{\texttt{GPD}}(\mu,\, \sigma,\, \xi)$

$$F_{\texttt{GEV}}(x;\, \mu^\star,\, \sigma^\star,\, \xi^\star) = \exp\left\{-S_{\texttt{GPD}}(x;\, \mu^\star,\, \sigma^\star,\, \xi^\star)\right\}$$

which can be checked from the closed form expressions. It will thus be enough to prove that for any $x \in \mathcal{I}_{\texttt{GPD}}(\mu,\, \sigma,\, \xi)$ we have

$$\lambda w\; S_{\texttt{GPD}}(x;\, \mu,\, \sigma,\, \xi) = S_{\texttt{GPD}}(x;\, \mu^\star,\, \sigma^\star,\, \xi^\star) \tag{3.7}$$

indeed, the derivative at the right hand side of (3.6) will thus be the GEV density $f_{\texttt{GEV}}(x;\, \mu^\star,\, \sigma^\star,\, \xi^\star)$. Separating the two cases $\xi \neq 0$ and $\xi = 0$, the verification of (3.7) is simple algebra. □

**Remark.** Note that (3.3) only holds when $x$ is in the support $\mathcal{I}_{\texttt{GPD}}(\mu,\, \sigma,\, \xi)$ and it does not hold over the full superset $\mathcal{I}_{\texttt{GEV}}(\mu^\star,\, \sigma^\star,\, \xi^\star)$. It is easy to check that by integrating (3.3) with respect to $x \in \mathcal{I}_{\texttt{GPD}}(\mu,\, \sigma,\, \xi)$ that we get $\Pr\{N_b \neq 0\} = 1 - \exp\{-\lambda w_b\}$. ■

Note that while $\lambda$ is related to a time scale (it expresses as an inverse time), no time unit is found in the GEV parameters. The reason is that the time unit is hidden in the block duration $w$ which is needed to compute the return level curve in time units (typically years).

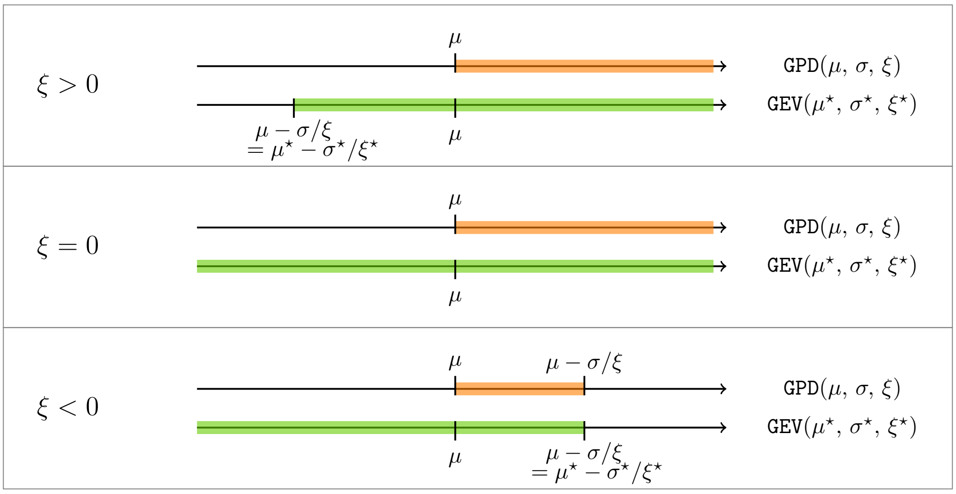



Figure 3.2: Supports $\mathcal{I}_{\texttt{GPD}}(\mu,\, \sigma,\, \xi)$ and $\mathcal{I}_{\texttt{GEV}}(\mu^\star,\, \sigma^\star,\, \xi^\star)$.

### 3.1.2 Links with Extreme Value regression

In the general case where the block duration $w_b$ is not constant, the distribution of $M_b$ depends on $w_b$. Ignoring the previous derivation, one could have used $w_b$ as a covariate in an extreme value regression. It is very unlikely that by proceeding in this way we would find the exact relations to the covariate as given in (3.4). In the exponential case $\xi = 0$, the exact form of dependence is quite usual

$$\mu_b^\star = \beta_0 + \beta_1 \log w_b, \qquad \sigma_b^\star = \sigma^\star \text{ (constant)},$$

with parameters $\beta_0$, $\beta_1$, $\sigma^\star$. Thus the location parameter is related to the log duration of the blocks, which may seem natural. However, when $\xi \neq 0$ the true relationship would require the links

$$\mu_b^\star = \beta_0 + \beta_1\, w_b^{\xi^\star}, \qquad \sigma_b^\star = \gamma_1\, w_b^{\xi^\star}, \qquad \xi^\star = \xi \text{ (constant)},$$

and the parameters $\beta_0$, $\beta_1$, $\gamma_1$ and $\xi^\star$. These equations do not fit in the standard framework where each of the three parameters is connected to the covariates through its own link function (Coles 2001, chap. 5). Thus, the very simple situation of a temporal aggregation does not lead to a simple extreme value regression. Note also that as a function of $w$, the variations of the term $w^{\xi^\star}$ are large for small values of $w$, since in practice $|\xi^\star| \leqslant 1$.

## 3.2 GEV-max to GPD-POT

### 3.2.1 Disaggregation for constant block duration

#### Problem

We now assume to be given a sequence $M_b$ corresponding to disjoint blocks $b = 1,\, 2,\, \dots,\, B$ with the *same duration* $w_b \equiv w$, so that the $M_b$ form a sample of a GEV distribution. In other words, we have partial observations of the marked process. We may then estimate the parameters of the underlying marked process and infer on them. However, the marked process embeds *four* parameters $\lambda$, $\mu$, $\sigma$ and $\xi$, while the GEV distribution only involves *three* parameters $\mu^\star$, $\sigma^\star$ and $\xi^\star$. Given the vector $\boldsymbol{\theta}^\star := [\mu^\star,\, \sigma^\star,\, \xi^\star]^\top$, there is an infinity of vectors $\boldsymbol{\theta} = [\lambda,\, \mu,\, \sigma,\, \xi]^\top$ satisfying the relations (3.4). The marked process model that generated the observations $M_b$ can correspond to any vector provided that all the $M_b$ lie in the interior of the support $\mathcal{I}_{\texttt{GPD}}(\mu,\, \sigma,\, \xi)$. There is an infinity of vectors $\boldsymbol{\theta} = [\lambda,\, \mu,\, \sigma,\, \xi]^\top$ satisfying these conditions that have the same log-likelihood. The corresponding marked process models can be said *observationally equivalent* with respect to the given sequence of block maxima $M_b$.

**Remark.** The observational equivalence is tightly related to the "POT-stability" property of the GPD. By increasing the threshold $u$ and lowering the rate $\lambda$ it is possible to maintain the same return level curve.

A natural idea to overcome the problem of identifiability is to fix one of the four POT parameters. The following two strategies can be considered

- Choose the rate $\lambda > 0$, and then compute or estimate the GPD parameters $\mu$, $\sigma$ and $\xi$.
- Choose the GPD location $\mu$, and then compute or estimate the rate $\lambda$ as well as $\sigma$ and $\xi$.

In the first case, any positive rate $\lambda > 0$ can be chosen and we simply have a re-parameterisation of the GEV distribution. Taking $\lambda = 1/w$, the three GPD parameters of the renewal process become identical to their GEV correspondent, that is: $\mu = \mu^\star$, $\sigma = \sigma^\star$ and $\xi = \xi^\star$.

The second approach is very attractive when the model must be fitted using observations $M_b$, since it boils down to a POT estimation from aggregated data as discussed now.

### Fixing $\mu$: a "GEV to POT" function

The relations (3.4) or (3.5) give the BM parameter vector $\boldsymbol{\theta}^\star$ as a function of the POT parameter $\boldsymbol{\theta}$; we aim to clarify here a possible inverse relation, i.e. the determination of $\boldsymbol{\theta}$ from $\boldsymbol{\theta}^\star$ for a fixed value of $\mu$.

For the GEV context we will denote by $\Theta^\star = \{[\mu^\star,\, \sigma^\star,\, \xi^\star]^\top; \sigma^\star > 0\}$ the domain of admissible parameters. The notations $\boldsymbol{\theta}_{(-\lambda)}$ and $\boldsymbol{\theta}_{(-\mu)}$ are for the vectors obtained by omitting $\lambda$ or $\mu$ in the vector $\boldsymbol{\theta} = [\lambda,\, \mu,\, \sigma,\, \xi]^\top$. The relations (3.4) giving $\boldsymbol{\theta}^\star$ as a function of $\boldsymbol{\theta}$ and $\mu$ can be written as

$$\boldsymbol{\theta}^\star = \boldsymbol{\psi}^\star(\boldsymbol{\theta}_{(-\mu)};\, \mu) \tag{3.8}$$

which can be called a "POT to GEV" transformation. The Jacobian of this transformation is easily computed, see A.3 page 78. The same notations can be used for the Gumbel context and the relations (3.5), provided it is understood then that $\boldsymbol{\theta}^\star = [\mu^\star,\, \sigma^\star]^\top$ and $\boldsymbol{\theta} = [\lambda,\, \mu,\, \sigma]^\top$.

**Theorem 3.2.** *Let $\boldsymbol{\theta}^\star = [\mu^\star,\, \sigma^\star,\, \xi^\star]^\top$ be a vector of GEV parameters with $\sigma^\star > 0$. A solution $\boldsymbol{\theta}_{(-\mu)} = [\lambda,\, \sigma,\, \xi]^\top$ of (3.4) exists if and only if $\mu$ is an interior point of the support $\mathcal{I}_{\texttt{GEV}}(\boldsymbol{\theta}^\star)$. Then the solution $\boldsymbol{\theta}_{(-\mu)}$ is unique and we may write it as a function of $\boldsymbol{\theta}^\star$ and $\mu$, i.e. as $\boldsymbol{\theta}_{(-\mu)} = \boldsymbol{\psi}(\boldsymbol{\theta}^\star;\, \mu)$.*

*For a vector of Gumbel parameters $\boldsymbol{\theta}^\star = [\mu^\star,\, \sigma^\star]^\top$ with $\sigma^\star > 0$, a unique solution $\boldsymbol{\theta}_{(-\mu)}$ of (3.5) exists.*

*Proof.* Consider the GEV case. From (3.4) we have by simple algebra

$$[\lambda w]^{-\xi^\star} = 1 + \xi^\star\,(\mu - \mu^\star)/\sigma^\star. \tag{3.9}$$

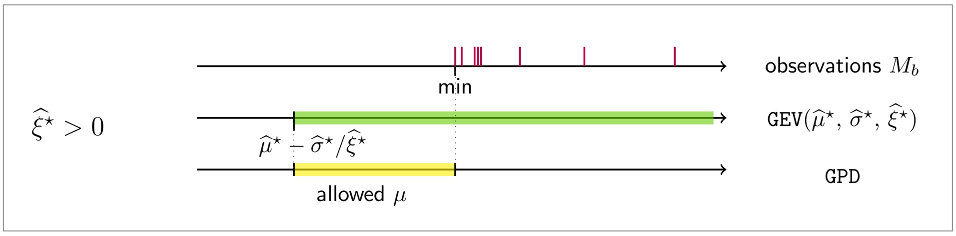



Figure 3.3: The fixed parameter $\mu$ must lie in the interior of the support $\mathcal{I}_{\texttt{GEV}}(\widehat{\boldsymbol{\theta}}^\star)$. When $\widehat{\xi}^\star > 0$, we must have $\mu > \widehat{\mu}^\star - \widehat{\sigma}^\star/\widehat{\xi}^\star$.

We have $\lambda > 0$ if and only the right hand side is positive, i.e. if $\mu$ is located in the interior of the support $\mathcal{I}_{\texttt{GEV}}(\boldsymbol{\theta}^\star)$. We may then take the power $-1/\xi^\star$ of each side, leading to $\lambda w = -\log F_{\texttt{GEV}}(\mu;\, \boldsymbol{\theta}^\star)$. We then easily find $\sigma$ and $\xi$. To summarise

$$\lambda = -\frac{1}{w}\log F_{\texttt{GEV}}(\mu;\, \boldsymbol{\theta}^\star), \qquad \sigma = (\lambda w)^{-\xi^\star}\sigma^\star, \quad \xi = \xi^\star. \tag{3.10}$$

The proof is straightforward for the Gumbel case.

□

#### Fitting BM from POT

Given $B$ block maxima $M_b$ we now consider the estimation of a GEV distribution by using a POT model with $\mu$ fixed. Using the notations of the previous section, we can estimate $\boldsymbol{\theta}_{(-\mu)}$ rather than the vector $\boldsymbol{\theta}^\star$ of GEV parameters and get this later using the POT to BM transform described in the previous section. More precisely, we can maximise with respect to $\boldsymbol{\theta}_{(-\mu)}$ the POT likelihood $L_{\texttt{POT}}(\boldsymbol{\theta}_{(-\mu)};\, \mu)$ where the second argument is meant to recall that $\mu$ is used as the threshold required in POT.

Not all values of the fixed parameter $\mu$ can be chosen. The fixed value of $\mu$ must obviously be such that $M_b > \mu$ for every block $b$, and it must also lie in the interior of the support $\mathcal{I}_{\texttt{GEV}}(\widehat{\boldsymbol{\theta}}^\star)$, see Figure 3.3. Assume that we are given a subset $\Theta_0^\star$ of the parameter space $\Theta^\star$ containing all the parameters $\boldsymbol{\theta}^\star$ that could have generated the observations, and that $\mu$ is an interior point of $\mathcal{I}_{\texttt{GEV}}(\boldsymbol{\theta}^\star)$ for all $\boldsymbol{\theta}^\star \in \Theta_0^\star$. Then

$$L_{\texttt{GEV}}(\boldsymbol{\theta}^\star) = L_{\texttt{POT}}(\boldsymbol{\theta}_{(-\mu)};\, \mu), \quad \text{with} \quad \boldsymbol{\theta}_{(-\mu)} = \boldsymbol{\psi}(\boldsymbol{\theta}^\star;\, \mu)$$

holds for all $\boldsymbol{\theta}^\star \in \Theta_0^\star$. It is thus clear that maximising the POT likelihood with respect to $\boldsymbol{\theta}_{(-\mu)}$ with $\mu$ fixed and transforming it with the function $\boldsymbol{\psi}^\star$ will lead to the same solution $\widehat{\boldsymbol{\theta}}^\star$ as fitting a GEV by ML. In other words

$$\widehat{\boldsymbol{\theta}}_{(-\mu)} = \boldsymbol{\psi}(\widehat{\boldsymbol{\theta}}^\star,\, \mu) \tag{3.11}$$

and the estimated joint distribution for the observations $M_b$ will be the same in the two cases.

An advantage of the POT approach lies in the possibility of likelihood concentration seen in chapter 1 (section 1.1.5 page 9). We can fit the model using a two-parameter optimisation involving $\sigma$ and $\xi$ while $\lambda$ is concentrated out through

$$\widehat{\lambda} = \frac{B}{\sum_b w_b\, S_{\texttt{GPD}}(M_b;\, \mu,\, \widehat{\sigma},\, \widehat{\xi})} \tag{3.12}$$

which is a special case of the formula (1.4) page 9. So the optimisation uses a parameter less than the classical GEV ML. Note that we will invariably find $\widehat{\lambda} > 1/w$ as is clear from the last formula.

A simple and interesting choice is $\mu := \min_b M_b$. This choice ensures that $\mu$ is an interior point of the support for any GEV distribution that could have generated the observations $M_b$, as required before. In practice, we will take $\mu := \min_b M_b - \epsilon$ for a small $\epsilon > 0$, smaller than the observation error[1]. This value of $\mu$ can be regarded as the maximal allowed value, and it corresponds to the minimal value of the rate $\lambda$. Hence it is not possible to choose a value of $\mu$ such that $\{\widehat{\lambda} < 1\}$ has a positive probability. More information on the estimation of $\lambda$ can be found using the following result.

**Theorem 3.3.** *Assume that the rv.s $M_1, M_2, \ldots, M_B$ are i.i.d. with distribution* $\texttt{GEV}(\mu^\star, \sigma^\star, \xi^\star)$ *with $\xi^\star \neq 0$, $\xi^\star > -1/2$, and that $w > 0$ is fixed. Let $\widetilde{\mu} := \min_b M_b$ and let $\widehat{\lambda}$ be the ML estimate of the rate $\lambda$ obtained by using $\widetilde{\mu}$ as threshold and an (aggregated) POT ML estimation.*

*Then for large $B$ the distribution of $\widehat{\lambda}w$ is that of the maximum of $B$ i.i.d random variables with standard exponential distribution. Thus*

$$w\,\mathbb{E}[\widehat{\lambda}] \approx \log B + \gamma, \qquad w^2\,\mathrm{Var}[\widehat{\lambda}] \approx \frac{\pi^2}{6}$$

*where $\gamma \approx 0.577$ is the Euler-Mascheroni constant. In other words, $\widehat{\lambda}w - \log B$ converges in distribution to the standard Gumbel* $\texttt{Gum}(0, 1)$.

*Proof.* See Appendix A.8 page 87. □

In practice, the number of blocks will typically be between 30 and 100, and an order of magnitude for $\widehat{\lambda}$ is from 4 to 5.

Note that the variance of $\widehat{\lambda}$ does not tend to zero, which may suggest a non-consistent estimator. However, $\widehat{\lambda}$ is not really an estimator because we do not have a fixed known threshold $u$ as would be required in POT, hence we do not have a true $\lambda$ that has to be estimated. We rather use a *random* threshold $u = \widetilde{\mu}$. This method allows us to find a consistent estimator of the GEV parameters, and thus of $\mu^\star$. Yet none of the two random variables $\widetilde{\mu}$ and $\widehat{\lambda}$ tends to a (non-random) constant for large $B$, as it is clear from the previous result and the first relation in (3.4). By using a larger number of blocks $B$, we will have to use a lower threshold and hence a larger rate.

### Return levels

A given parameter vector $\boldsymbol{\theta}^\star = [\mu^\star, \sigma^\star, \xi^\star]^\top$ for block maxima can be translated into a vector $\boldsymbol{\theta} = [\lambda, \mu, \sigma, \xi]^\top$ for POT, among an infinity of them. All possible POT models share *exactly* the same return level curve. Indeed, recall that in the POT context, the return level for period $T$ is given by

$$x_{\texttt{POT}}(T) = \mu + \sigma\,\phi(\lambda T, \xi) \tag{3.13}$$

where $\phi(z, \rho)$ is the Box-Cox transformation

$$\phi(z, \rho) = \begin{cases} \dfrac{z^\rho - 1}{\rho} & \rho \neq 0, \\ \log z & \rho = 0. \end{cases}$$

[1]Excesses exactly equal to zero are impossible in POT fitting.

Using (3.4) and simple algebra, the return level $x_{\texttt{POT}}(T)$ in (3.13) expresses as a function of the GEV parameters

$$x_{\texttt{POT}}(T) = \mu^\star + \sigma^\star \phi(T/w,\, \xi^\star). \tag{3.14}$$

So all POT models verifying (3.4) have exactly the same return level curves.

In the block maxima context, a return level is associated with a (possibly non-integer) number $m > 1$ representing a duration as a multiple of the block duration. The $m$-block return level can be denoted as $x_{\texttt{GEV}}(m;\, \boldsymbol{\theta}^\star)$ or simply $x_{\texttt{GEV}}(m)$ and is the quantile of the GEV distribution associated with the probability $1 - 1/m$. It expresses as

$$x_{\texttt{GEV}}(m) = \mu^\star + \sigma^\star \phi(m^\star,\, \xi^\star), \qquad m^\star := \Big[-\log(1 - 1/m)\Big]^{-1}.$$

We have $m^\star \approx m$ for large $m$, so

$$x_{\texttt{POT}}(T) \approx x_{\texttt{GEV}}(T/w), \qquad \text{for large } T.$$

### Inference on return levels

Using the partial observations $M_b$ for a given block duration $w$, we have seen that two methods can be used to obtain the return level curve and the associated confidence intervals.

- **[A]** Fit a GEV distribution by ML. We get an estimated parameter vector $\widehat{\boldsymbol{\theta}}^\star$ along with a covariance matrix for the estimate $\text{Var}(\widehat{\boldsymbol{\theta}}^\star)$. We can compute the return levels $x_{\texttt{GEV}}(m;\, \widehat{\boldsymbol{\theta}}^\star)$ as well as confidence intervals for these quantities by using the delta method.
- **[B]** Fit an aggregated POT model, thus fixing the threshold $\mu$ regarded as known. We then get a ML estimates $\widehat{\boldsymbol{\theta}}_{(-\mu)}$ for the vector of the three parameters $\lambda$, $\sigma$ and $\xi$. We also get a covariance matrix $\text{Var}(\widehat{\boldsymbol{\theta}}_{(-\mu)})$. We can compute return levels $x_{\texttt{POT}}(T;\, \widehat{\boldsymbol{\theta}})$ and confidence intervals for these.

We have seen that by choosing the value $\widetilde{\mu} = \min M_b - \epsilon$ with $\epsilon > 0$ small, we can use all the available observations $M_b$ in POT and that the POT ML estimation allows the determination of the ML estimate for the GEV parameter $\widehat{\boldsymbol{\theta}}^\star$ as well as its asymptotic variance matrix. The return levels provided by the two methods are similar. Yet a question remains: do we obtain the same inference results for return levels and the same confidence intervals? The answer is *nearly yes*, provided that all parameters are used in the delta method. Small differences will be observed for small return levels, because the interpretation of the return levels differs in the two settings.

**Remark.** In the POT setting, the uncertainty on $\lambda$ has usually very little impact on the confidence intervals corresponding to large return periods, and the delta method may actually use the $2 \times 2$ covariance matrix for the excesses parameters estimates $\widehat{\sigma}$ and $\widehat{\xi}$. This seems due to the fact that $\lambda$ is small in practice, and it is no longer true when $\lambda$ is about 20 or larger.

■

Suppose we proceeded to **[A]**. Then we can choose any rate $\lambda > 0$ and use the Jacobian of the GEV to POT transform given in A.4 page 78 to obtain the variance of the POT parameters. The special choice $\lambda = 1/w$ is appealing because the POT parameters $\sigma$ and $\xi$ identical to their

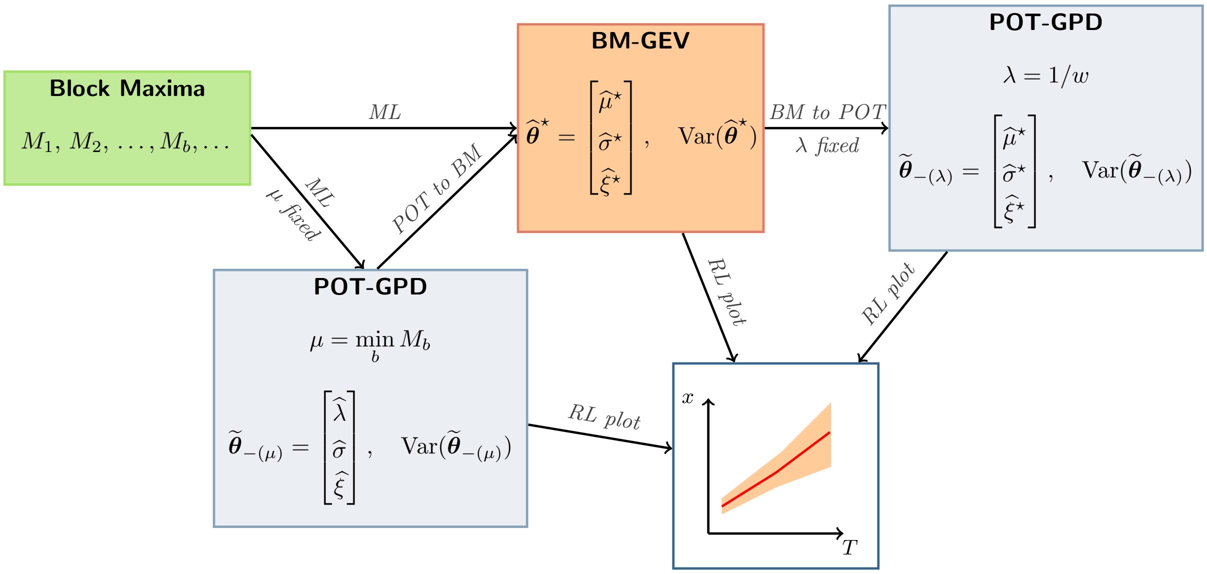



Figure 3.4: Fitting an aggregated POT with $\mu$ fixed (hence $\widehat{\lambda}$ random) leads after translation to the same GEV parameters and the same approximated variance as a BM fit would give. We have $\widehat{\lambda} > 1$ with probability 1 and $\mathbb{E}[\widehat{\lambda}] \approx \log B + \gamma$. The GEV parameters can be used to build a RL plot. Nearly the same RL plot (including confidence limits) will be obtained by the different methods. We can also fix a rate $\lambda > 0$ to build a POT model from the GEV parameters, as illustrated here with the special case $\lambda = 1/w$.

GEV equivalent, and the POT parameter vector is simply obtained by completing the GEV fit with $\lambda^\star := 1/w$

$$\widetilde{\boldsymbol{\theta}} := \left[\frac{1}{w}, \widehat{\mu}^\star, \widehat{\sigma}^\star, \widehat{\xi}^\star\right]^\top.$$

Yet we need use the Jacobian to obtain the variance of the vector of POT parameters.

### 3.2.2 Renext functions

The `Renouv` function in **Renext** can perform ML estimation from heterogeneous data including marked process data, block maxima data. Some utility functions can be used for parameter translations:

- `Ren2gev` and `gev2Ren` perform translations in the GPD/GEV case.
- `Ren2gumbel` and `gumbel2Ren` perform translations in the GPD/GEV case when $\xi = 0$.
- `gpd2lomax` and `gpd2maxlo` perform translations of GPD parameters $[\sigma, \xi]^\top$ with location $\mu = 0$ to Lomax $[\alpha, \beta]^\top$ (if $\xi > 0$) or maxlo $[\alpha, \beta]^\top$ parameters (if $\xi < 0$).

The relations between these functions is illustrated on Figure 3.5.

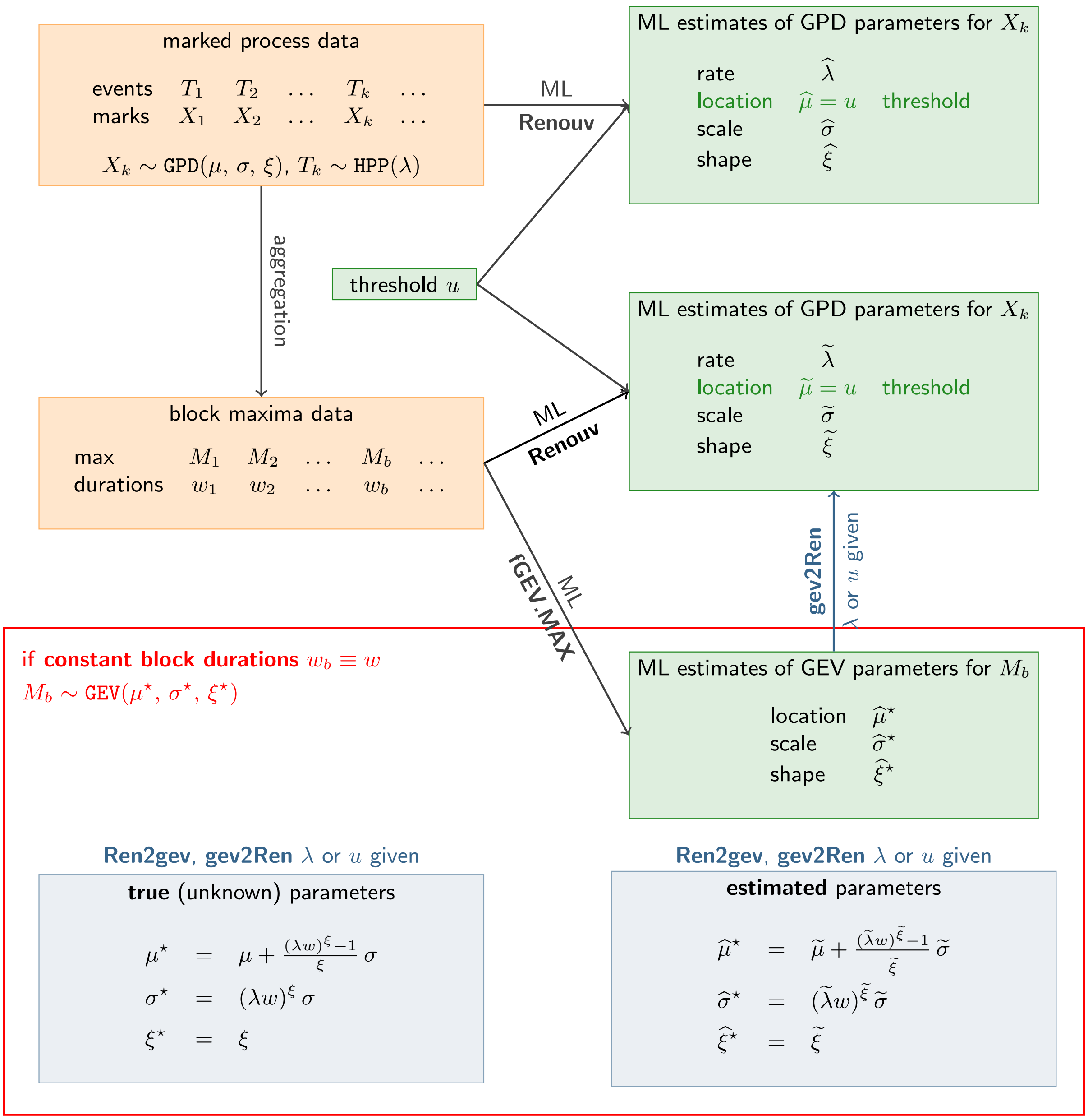



Figure 3.5: Temporal aggregation of the marked process and **Renext** functions. The ML estimation from block maxima in the constant duration case can be done by using the function `fGEV.MAX` from **Renext** as well as by using functions from other packages: **evd**, **ismev**, ...

## 3.3 Initial values for exponential marks/excesses

### 3.3.1 Problem

We consider the situation where the marks $X_k$ are exponentially distributed with unknown scale parameter $\sigma$ and where the marked process is partially observed within one of the three following schemes, each involving $B$ blocks.

1. **Block maxima.** For each block $b$, we know the block maximum $M_b$,
2. $r$ **largest.** For each block $b$, we know the $r_b$ largest order statistics $Z_{b,1} \geqslant Z_{b,2} \geqslant \cdots \geqslant Z_{b,r_b}$. Thus $Z_{b,1} = M_b$.
3. **OTS** data. For each block $b$, we know the $r_b$ observations $X_{b,i}$ that exceeded $u_b$, where $u_b \geqslant u$ is a threshold for the block.

We assume that the location parameter $\mu$ of the exponential distribution is fixed as the POT threshold, hence is known. We want to find cheap initial values, say $\widetilde{\lambda}$ and $\widetilde{\sigma}$ for the unknown parameters $\lambda$ and $\sigma$.

Note that the first index $b$ in the notation $Z_{b,i}$ for the order statistic is the block index, and not the (random) number of observations in the block as commonly used. An important point is that *we do not assume the blocks to have the same durations.*

### 3.3.2 Block maxima

Recall that $M_b \sim \texttt{Gumbel}(\mu_b^\star,\, \sigma_b^\star)$ where the location and shape parameters for block $b$ are given by

$$\mu_b^\star = \mu + \sigma\, \log(\lambda w_b), \qquad \sigma_b^\star = \sigma.$$

The expectation and the variance of $M_b$ are given by

$$\mathbb{E}(M_b) = \mu_b^\star + \gamma\, \sigma_b^\star = \mu + \sigma\, \log(\lambda w_b) + \gamma\, \sigma, \qquad \mathrm{Var}(M_b) = \frac{\pi^2}{6}\, \sigma^2$$

where $\gamma \approx 0.5772$ is Euler-Mascheroni constant[2]. The relation of the maxima to the log-durations takes the form of a linear regression where the slope coefficient $\beta$ and the standard deviation of the error are proportional

$$M_b = \alpha + \beta\, x_b + \varepsilon_b, \qquad \mathrm{Var}(\varepsilon_b) = \kappa^2\, \beta^2 \tag{3.15}$$

with $x_b := \log(w_b)$ and $\kappa^2 := \pi^2/6$. The two regression coefficients $\alpha$ and $\beta$ relate to the parameters according to

$$\alpha := \mu + \log(\lambda)\, \sigma + \gamma\, \sigma, \qquad \beta := \sigma. \tag{3.16}$$

Initial estimates $\widehat{\lambda}_{\texttt{reg}}$ and $\widehat{\sigma}_{\texttt{reg}}$ can be found by using estimates $\widehat{\alpha}_{\texttt{reg}}$ and $\widehat{\beta}_{\texttt{reg}}$ for the two parameters $\alpha$ and $\beta$ of this regression

$$\widehat{\sigma}_{\texttt{reg}} = \widehat{\beta}_{\texttt{reg}}, \qquad \widehat{\lambda}_{\texttt{reg}} = \exp\left\{[\widehat{\alpha}_{\texttt{reg}} - \mu]/\widehat{\sigma}_{\texttt{reg}} - \gamma\right\}. \tag{3.17}$$

[2]See e.g. Wikipedia's page for Gumbel distribution or Johnson, Kotz, and Balakrishnan (1995).

Assuming the regression error term $\varepsilon_b$ in (3.15) to be gaussian, the ML estimators would be easily found by writing the corresponding log-likelihood together with its two first order derivatives which must vanish. This leads to the two equations

$$\alpha = \bar{M} - \beta\,\bar{x}, \qquad \kappa^2\,\beta^2 + \mathrm{Cov}(x,\, M)\,\beta - \mathrm{Var}(M) = 0$$

where the number $B$ of blocks is used as denominator for the sample covariance and variance. The second equation for $\beta$ has two real roots with opposite signs, the largest of which gives the maximum likelihood estimate $\widehat{\beta}_{\texttt{reg}}$ for $\beta$. Then the ML estimate $\widehat{\alpha}_{\texttt{reg}}$ of $\alpha$ is given by the first equation.

**Remark.** When the block durations $w_b$ are constant, the initial values coincide with the moment estimates used in several R packages. Contrary to the ordinary regression context, the estimator $\widehat{\beta}_{\texttt{reg}}$ then still exists because the parameter $\beta$ can be identified through the error variance.

Empirical findings from simulations are that the resulting estimates of $\lambda$ and $\sigma$ are negatively correlated, and that $\sigma$ is moderately underestimated while $\lambda$ is overestimated.

### 3.3.3 $r$ largest

We will use a result concerning the *spacings* $Z_{b,i} - Z_{b,i+1}$ for $i = 1, 2, \ldots, r_b - 1$ and known as the Rényi's representation of sample from the exponential, see A.7 We have an equality in distribution

$$Z_{b,i} - Z_{b,i+1} \underset{\texttt{d}}{=} \frac{1}{i}\,E_i$$

where the $r_b - 1$ r.vs $E_i$ are i.i.d and follow the exponential distribution with mean $\sigma$. So we get an estimator of $\sigma$ as a weighted mean of the weighted spacings

$$\widehat{\sigma}_{\texttt{spac}} = \frac{1}{n_{\texttt{spac}}} \sum_{b=1}^{B} \sum_{i=1}^{r_b - 1} i \times \left[Z_{b,i} - Z_{b,i+1}\right], \qquad n_{\texttt{spac}} := \sum_{b=1}^{B} (r_b - 1) \tag{3.18}$$

where an empty sum for $r_b = 1$ is taken as 0. The variance of this estimator is easily computed as $\mathrm{Var}(\widehat{\sigma}_{\texttt{spac}}) = \sigma^2/n_{\texttt{spac}}$ and can be evaluated by replacing the unknown $\sigma$ by its estimate $\widehat{\sigma}_{\texttt{spac}}$.

We as well know the marginal distribution of the order statistic $Z_{b,r_b}$ and it is known that this r.v. is independent from the $r_b - 1$ spacings $Z_{b,i} - Z_{b,i+1}$ for $i = 1,\, 2,\, \ldots,\, r_b - 1$. We have

$$\mathbb{E}(Z_{b,r_b}) = \mathbb{E}(Z_{b,1}) - \sigma \times \sum_{i=1}^{r_b-1} \frac{1}{i}, \qquad \mathrm{Var}(Z_{b,r_b}) = \mathrm{Var}(Z_{b,1}) - \sigma^2 \times \sum_{i=1}^{r_b-1} \frac{1}{i^2}$$

As before with the block maxima, we may write a regression equation

$$Z_{b,r_b} = \alpha + \beta\,x_b + \varepsilon_b, \qquad \mathrm{Var}(\varepsilon_b) = \kappa_b^2\,\beta^2$$

with[3]

$$x_b := \log(w_b) - \sum_{i=1}^{r_b-1} 1/i, \qquad \kappa_b^2 := \pi^2/6 - \sum_{i=1}^{r_b-1} 1/i^2.$$

[3]Recall that $\sum_{i=1}^{\infty} 1/i^2 = \pi^2/6$.

Now regression estimates $\widehat{\alpha}_{\texttt{reg}}$ and $\widehat{\beta}_{\texttt{reg}}$ are found by replacing the maxima $M_b$ by $y_b := Z_{b,r_b}$ and by using the weights $v_b := 1/\kappa_b^2$ in the mean, variance and covariance. The estimate $\widehat{\beta}_{\texttt{reg}}$ is the positive root of the equation in $\beta$

$$\left\{\sum_b v_b\right\} \beta^2 + \left\{\sum_b v_b\,[y_b - \bar{y}][x_b - \bar{x}]\right\} \beta - \left\{\sum_b v_b\,[y_b - \bar{y}]^2\right\} = 0$$

where the means $\bar{x}$, $\bar{y}$ are now weighted with the weights $v_b$. The regression estimates of the coefficients $\lambda$ and $\sigma$ are still related to those of $\alpha$ and $\beta$ through (3.17) above.

We can use a weighted mean to combine the two estimators of $\sigma = \beta$

$$\widehat{\sigma} = \frac{n_{\texttt{spac}}}{n_{\texttt{spac}} + B/2}\,\widehat{\sigma}_{\texttt{spac}} + \frac{B/2}{n_{\texttt{spac}} + B/2}\,\widehat{\sigma}_{\texttt{reg}}$$

and then compute $\widehat{\lambda}$ using (3.17) with $\widehat{\sigma}_{\texttt{reg}}$ replaced by the previous combined estimate. The rationale for the cheap estimation $\widehat{\sigma}$ is that the regression uses $B$ independent observations for two estimated parameters while the estimate $\widehat{\sigma}_{\texttt{spac}}$ uses $n_{\texttt{spac}}$ observation for one parameter. Note that the two estimates for $\sigma$ are independent under our hypotheses since the regression response $y_b$ is independent from the spacings.

### 3.3.4 OTS

In this framework, $r_b$ is random and can be viewed as the realisation of a r.v. denoted by $R_b$. Since the marks $X_k$ and the events $T_k$ are independent, the observed events in the block form a peeled Poisson Process, which is a HPP with rate

$$\lambda_b := \lambda\, S_X(u_b) = \lambda\, \exp\left\{-\frac{1}{\sigma}\,[u_b - u]\right\}.$$

Hence, $R_b$ is Poisson with expectation $\mathbb{E}(R_b) = \lambda_b w_b$ and

$$\log \mathbb{E}(R_b) = \log w_b + \log \lambda_b = \log w_b + \log \lambda - \frac{1}{\sigma}\,[u_b - u].$$

Let us consider a Poisson regression model[4] with response $R_b$, covariate $u_b - u$ and offset $\log w_b$. The canonical log-link and the offset provide the relation

$$\log \mathbb{E}(R_b) = \underbrace{\log w_b}_{\text{offset}} + \alpha + \beta\,[u_b - u]$$

where $\alpha$ and $\beta$ are the linear parameters. Fitting this Poisson regression, we can find estimates for $\lambda$ and $\sigma$ as

$$\widehat{\lambda}_{\texttt{reg}} = \exp(\widehat{\alpha}_{\texttt{reg}}), \qquad \widehat{\sigma}_{\texttt{reg}} = -1/\widehat{\beta}_{\texttt{reg}}.$$

Note that while $\widehat{\lambda}_{\texttt{reg}}$ is always positive, the value $\widehat{\sigma}_{\texttt{reg}}$ may be negative. Moreover, this estimate of $\sigma$ makes no use of the available observations $X_{b,i}$ and thus must be of poor quality.

A simple estimate of $\sigma$ is obtained as the mean of the excesses $X_{b,i} - u_b$ where the threshold $u_b$ depends on the block $b$ in which the historical observation $X_{b,i}$ falls

$$\widehat{\sigma} = \frac{1}{\sum_{b=1}^{B} r_b} \sum_{b=1}^{B} \sum_{i=1}^{r_b} (X_{b,i} - u_b).$$

[4] See e.g. Wikipedia's page for Poisson regression.

# Chapter 4

# Plotting positions

## 4.1 Goals

### 4.1.1 Observations and censoring

**Renext** is mainly devoted to the analysis of data arising from a marked process. The usual *complete observations* consist in couples $[T_i, X_i]$ formed by an event time $T_i$ and the related mark $X_i$. The temporal aggregation of this marked process leads to what we may call *partial observations* such as maxima or $r$ largest observations on time intervals or blocks. Provided that the temporal aggregation is made on a regular basis (usually yearly) and that the marks or excesses follow a GEV distribution, the partial observations can equivalently be analysed using the classical block maxima (BM) or $r$ largest. In the BM context, the data simply consists in an ordinary sample, i.e. a vector of $n$ i.i.d random variables, and many graphical diagnostics are then available under the name of *diagnostic plots*.

**Renext** allows the use of *heterogeneous data*: one can use complete observations, partial observations as well as combination of these for different time intervals or periods. While the Maximum Likelihood context is well suited for the use of heterogeneous data in the estimation of the parameters, the derivation of diagnostic plots is more difficult. Return level plots, probability or quantile plots require a nonparametric estimation of the return level or survival function at the empirical points to provide suitable *plotting positions*. An adaptation of the plotting positions for BM context has been proposed by Hirsch and Stedinger (1987) to deal with the case where some *historical information* is available. These modified plotting positions are related to the general context of *censored observations* (Millard and Neerchal 2001) and seem quite popular in various fields of application.

We begin by recalling the derivation of Hirsch and Stedinger's plotting positions for the context of censored observations in an ordinary sample, as used in the BM context. Then we show that a simple adaptation can be derived for our POT/marked process context with heterogeneous data.

The plotting positions described here have been implemented in **Renext** (`>= 2.1-6`) in the function `SandT` used by several functions dedicated to graphics.

| plot | abscissa | ordinate |
|---|---|---|
| PP plot | $\widehat{F}(x)$ | $F(x;\, \widehat{\boldsymbol{\theta}})$ |
| QQ plot | $q(\widehat{F}(x);\, \widehat{\boldsymbol{\theta}})$ | $x$ |
| RL plot | $\widehat{H}(x)$ | $x$ |

Table 4.1: Main plot types used. The axes of the PP and QQ plots are sometimes used in the reverse order, as in Waller and Turnbull (1992).

### 4.1.2 Types of plots

In classical contexts of Extreme Value modelling, three diagnostic plots are commonly used to access the fit of a parametric distribution to data.

- Quantile-quantile plot (QQ-plot),
- Probability-probability plot (PP-plot),
- Return level plot (RL-plot).

Waller and Turnbull (1992) describe the first two types of plots – amongst others. These two plots compare two estimates of the distribution function at some chose levels $x$: a non-parametric one $\widehat{F}(x)$ and a parametric one $F(x;\, \widehat{\boldsymbol{\theta}})$, usually obtained by Maximum Likelihood, cf. table 4.1. The RL-plot displays the level (or quantile) $x$ against the *cumulative hazard* $H(x) = -\log S(x)$ where $S(x) = 1 - F(x)$ is the survival function. When the data is in good accordance with the chosen parametric distribution, both QQ and PP plots should display points which are nearly located on a straight line. A desirable property is that the positions are *unbiased*; however the unbiasedness property is not invariant under nonlinear transformations; for instance an unbiased estimate of $F(X_{(i)})$ will lead to an estimate of $\log[1 - F(X_{(i)})]$ which is no longer unbiased, but is only *asymptotically* such.

## 4.2 Plotting positions for ordinary sample

### 4.2.1 Outlook

In this section, we assume that $n$ levels $X_i$ are observed, corresponding to order statistics

$$X_{(1)} < X_{(2)} < \cdots < X_{(n)}.$$

This situation arise in the BM context: all blocks are then assumed to have the same duration $w^\star > 0$, usually one year. Several transformations of the data are related to the construction of the plots and lead to known expectations – see table 4.2.

We begin by discussing plotting positions in this context when no censoring occurs, then we turn to the problem of censored observations. The BM context is not used as such in **Renext**, but through an adaptation for the POT context as described in chap. 3 p. 32.

| variable | support | distribution | expectation |
|---|---|---|---|
| $F(X)$ | $[0,\,1]$ | uniform | $\mathbb{E}\left[F(X_{(i)})\right] = i/(n+1)$ |
| $S(X)$ | $[0,\,1]$ | uniform | $\mathbb{E}\left[S(X_{(i)})\right] = 1 - i/(n+1)$ |
| $H(X)$ | $[0,\,\infty)$ | exponential | $\mathbb{E}\left[H(X_{(i)})\right] = \sum_{k=1}^{i} 1/(n+1-k)$ |
| $1/S(X)$ | $[1,\,\infty)$ | Pareto, shape 1 | $\mathbb{E}\left[1/S(X_{(i)})\right] = \infty$ |

Table 4.2: Some transformations of a r.v. $X$ with distribution $F(x)$, and expectation of the order statistics for the transformed variables in a sample of size $n$. The Pareto distribution with shape 1 has density $t \mapsto 1/t^2$.

### 4.2.2 No censoring

#### PP-plot

The PP plot shows $n$ points with abscissas $F(x;\,\widehat{\boldsymbol{\theta}})$ for $x = X_{(i)}$. The ordinate is a non-parametric estimation for the distribution $F(x)$ at $x = X_{(i)}$ given by the formula proposed by Cunnane (1978)

$$\widehat{F}(X_{(i)}) = \frac{i-a}{n-2a+1} \qquad (1 \leqslant i \leqslant n) \tag{4.1}$$

where $a$ is a technical parameter with $0 \leqslant a < 1$. For instance $a = 0$ corresponds to so-called *Weibull* positions and provides the classical unbiased estimation. The R function `ppoints` from the **stats** package provides the vector of estimates for a given $n$; thus we will refer to (4.1) as the *p-points formula*. Another popular choice for the value of $a$ in (4.1) is $a = 0.5$, leading to the so-called *Hazen* plotting positions; this is the defaut value of the R function `pppoints` for $n \geqslant 10$.

#### RL-plot

Recall that the return period for a level $x$ is given by

$$T(x) = w^\star \times S(x)^{-1}. \tag{4.2}$$

Still assuming that $n$ maxima $X_i$ are observed; it will be convenient to use the notations $Z_i = X_{(n+1-i)}$ for $1 \leqslant i \leqslant n$, thus $Z_1 > Z_2 > \cdots > Z_n$. The corresponding estimation of the survival is

$$\widehat{S}(Z_i) = \frac{i-a}{n-2a+1} \qquad (1 \leqslant i \leqslant n) \tag{4.3}$$

leading to the following estimation for the return period

$$\widehat{T}(Z_i) = w^\star \times \frac{n+1-2a}{i-a}. \tag{4.4}$$

A number of remarks can be formulated concerning this formula based on $p$-points.

- The p-points formula (4.1) leads to the estimation (4.3) for the survival which has exactly the same form. It can be said that the p-points formula is *symmetric* (see the help `?ppoints`).

- For the smallest observation $Z_n$, the return period is $[1 + (1-a)/(n-a)]\, w^\star$, and thus is $\approx w^\star$ for large $n$.

- The return period for the largest observed level $Z_1$ is $\approx nw^\star/(1-a)$ and is therefore larger than the observation duration $nw^\star$, which yet may seem a 'natural' estimate. With the popular choice $a = 0.5$ the largest observation has a return period of $2nw^\star$, twice the observation duration.

- For $a = 0$, $\widehat{T}(Z_2)$ is half of $\widehat{T}(Z_1)$, even if the two largest levels $Z_1$ and $Z_2$ are close to each other. With $0 < a < 1$, the ratio of the two first return periods is even larger.

- When $n = 1$, the return period is always $2w^\star$, whatever be $a$.

Nelson's formula estimates the cumulative hazard $H(x) = -\log S(x)$ at the order statistics directly, without transforming an estimate of the distribution or of the survival (Nelson 2000). It writes as

$$\widehat{H}(X_{(i)}) = \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{n+1-i} = \sum_{k=1}^{i} \frac{1}{n+1-k}. \tag{4.5}$$

We may speak of the $n$ numbers $H_i := \widehat{H}(X_{(i)})$ as of the *H-points* since they relate to the cumulative hazard. The translation in the reverse order is

$$\widehat{H}(Z_i) = \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{i} = \sum_{k=i}^{n} \frac{1}{k}. \tag{4.6}$$

Nelson's formula can be shown to be unbiased for a sample of the exponential distribution. Indeed, this is a direct consequence of Rényi's representation since the r.vs $H(X_i)$ are i.i.d. and follow the standard exponential distribution see Appendix A.7 p. 86.

- In practice, the formula leads for $n \geqslant 30$ to positions which are very close to those arising from the $p$-points with $a = 0.5$, except for the largest observation.

- When $n$ is large, the largest cumulative hazard is $\widehat{H}(Z_1) \approx \gamma + \log n$ where $\gamma \approx 0.577$ is the Euler-Mascheroni constant. Consequently, by taking the exponential, the largest return period is $\exp(\gamma) nw^\star \approx 1.78\, nw^\star$.

- For $n = 1$ the unique observation has $\widehat{H} = 1$ and receives a return period of $\exp(\widehat{H})\, w^\star \approx 2.72\, w^\star$.

To illustrate the difference between different plotting positions, Figure 4.1 shows RL-plots obtained by averaging over a large number $m$ of simulated samples. More precisely, for different sizes $n$ and different shapes $\xi$, $m$ simulated samples of the GPD with shape $\xi$ and unit scale $\sigma = 1$ were drawn. For each sample, the maximum-likelihood estimate of the parameter vector $[\sigma,\, \xi]^\top$ was obtained and used to build the estimated RL curve. On each panel the average estimated RL curve is shown with the true curve and with the average plotting positions for three choices: the p-points formula with the default $a = 0.5$ (Hazen) and $a = 0$ (Weibull) and the H-points (Nelson). A log-scale is used for the $x$ axis. It transpires that the Weibull plotting positions lead to a biased empirical estimation of the RL curve. The difference between the Hazen and Nelson positions is quite small and is only visible for the largest observation. These two choices both provide on average a good estimation of the RL curve.

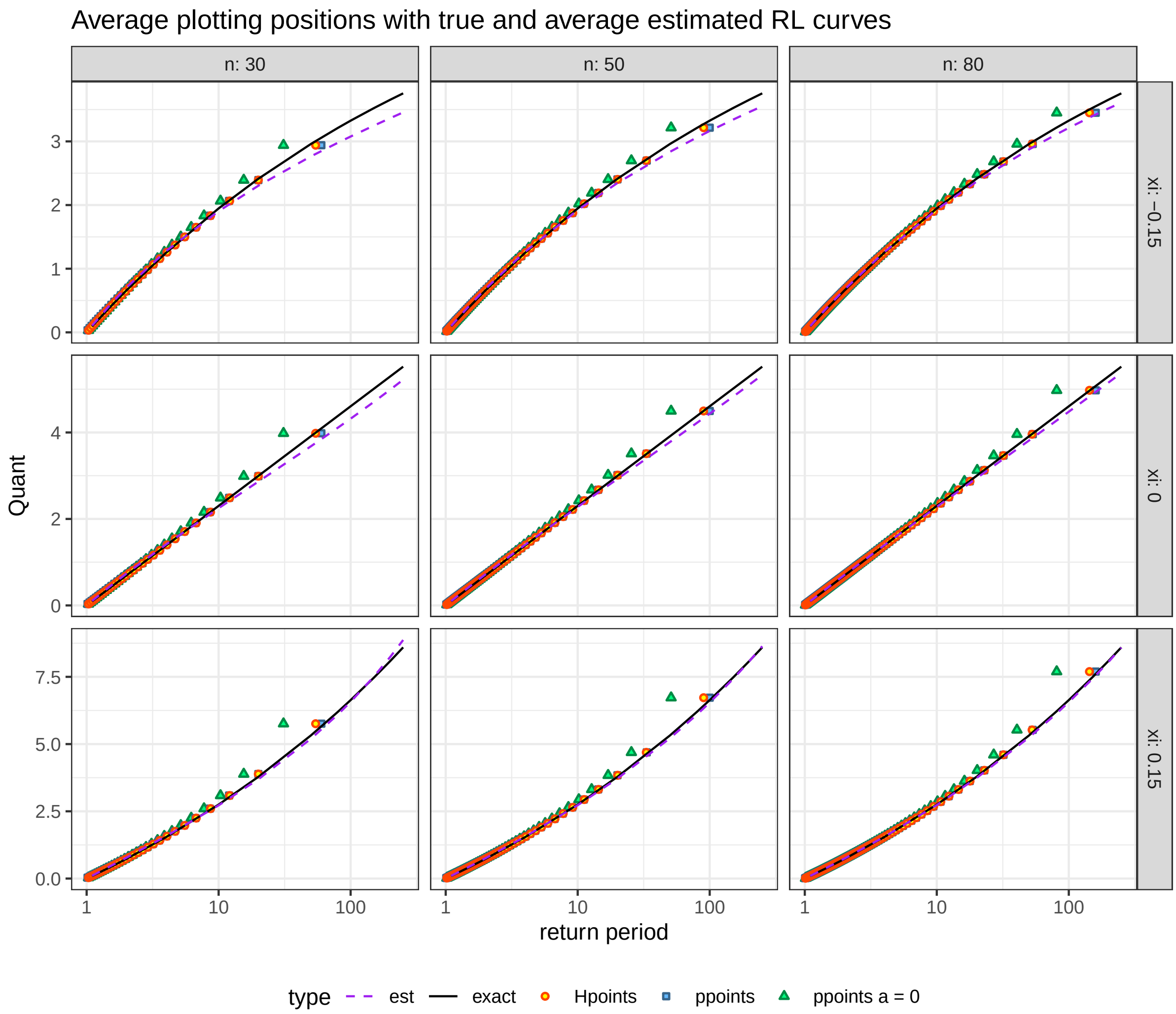



Figure 4.1: Average RL-plots for $m = 5000$ simulated samples of the GPD with size $n$ and shape $\xi$. On each panel the two lines show the ML estimate (dashed) and the true or "exact" RL curve (solid). These are nearly identical for $\xi = 0.15$. The points show the average plotting positions for three choices: H-points or “Nelson” (orange circles), p-points with the default $a = 0.5$ or “Hazen” (blue squares) and p-points with $a = 0$ or “Weibull” (green triangles). The Nelson and Hazen points are nearly identical, with a visible yet small difference only for the largest return period.

### 4.2.3 Type I censoring

#### Notations

We now describe the method of Hirsch and Stedinger (1987) for the estimation of the survival in a context of censored observations. The estimation is computed at the uncensored observations, and it can be used to get the return periods at these levels.

We assume here that the whole set of years $\{1, 2, \ldots, n\}$ is partitioned in $J$ "periods" for $j = 1, 2, \ldots, J$. By "period" we simply mean here a subset of the whole set of $n$ years, and there is no necessity that this subset should be formed of consecutive years, although this is simpler in graphical representations as time plots. For the period $j$, we have: a censoring threshold $u_j$, and a total number of observations $n_j$ among which $r_j$ are uncensored and $n_j - r_j$ are censored. For a censored observation, we only know that the level is greater than the threshold $u_j$.

Without loss of generality, we may assume the thresholds $u_j$ to be in strictly increasing order

$$u_1 < u_2 < \cdots < u_J.$$

Actually, although several periods may share the same threshold, then consider them as a single period with duration and number of observations obtained by summation. Moreover, we can assume that all the observations are greater than $u_1$, i.e. that $X_i > u_1$ for all $i$. It will be convenient to set $u_{J+1} := \infty$. Thus we have $J$ "slices" of levels, the slice number $j$ containing levels $x \in (u_j,\, u_{j+1}]$. We can think of a contingency table with rows matching slices and columns matching periods as in table 4.3 later. The problem is that we do not know the numbers below the minor diagonal, and thus we have to estimate or impute those numbers.

#### Estimation of the survival

We begin by estimating $S(u_j)$, noting that $S(u_1) = 1$ and $S(u_{J+1}) = 0$. We have for $1 \leqslant j \leqslant J$

$$\begin{aligned} S(u_j) &= \Pr\{X > u_j\} \\ &= \Pr\{X > u_{j+1}\} + \Pr\{u_j < X \leqslant u_{j+1}\} \\ &= \Pr\{X > u_{j+1}\} + \Pr\left[u_j < X \leqslant u_{j+1} \,|\, X \leqslant u_{j+1}\right] \Pr\{X \leqslant u_{j+1}\} \\ &= S(u_{j+1}) + \Pr\left[u_j < X \leqslant u_{j+1} \,|\, X \leqslant u_{j+1}\right] \left[1 - S(u_{j+1})\right]. \end{aligned}$$

The conditional probability can be estimated as

$$\widehat{\Pr}\left[u_j < X \leqslant u_{j+1} \,|\, X \leqslant u_{j+1}\right] = \frac{A_j}{A_j + B_j},$$

where

- $A_j$ is the number of uncensored observations in slice $j$, i.e. with level $X_i$ such that $u_j < X_i \leqslant u_{j+1}$,
- $B_j$ is the total *known* number of observations $X_i$ (censored or uncensored) such that $X_i \leqslant u_j$, i.e. falling in slices 1 to $j-1$.

The numbers $A_j$ and $B_j$ are computed on the periods for which the observations $u_j < X_i \leqslant u_{j+1}$ are known, i.e for periods $k = 1, 2, \ldots, j$, see section 4.2.4 below for an example. We have a recurrence relation

$$\widehat{S}(u_j) = \widehat{S}(u_{j+1}) + \widehat{\Pr}\left[u_j < X \leqslant u_{j+1} \,|\, X \leqslant u_{j+1}\right] \left[1 - S(u_{j+1})\right] \tag{4.7}$$

$j = J,\ J-1,\ \dots,\ 1$. It is easily checked by induction that $\widehat{S}(u_j) \leqslant 1$ and consequently that $\widehat{S}(u_j) \geqslant \widehat{S}(u_{j+1})$ for all $j \geqslant 1$.

Now the observations $X_i$ falling in $(u_j,\ u_{j+1}]$ can be seen as a sample of size $A_j$ from the conditional distribution with survival given by

$$\Pr\left[X > x \,|\, u_j < X \leqslant u_{j+1}\right] = \frac{S(x) - S(u_{j+1})}{S(u_j) - S(u_{j+1})}$$

For each of the $A_j$ observations $X_i$ falling in $(u_j,\ u_{j+1}]$ we may determine its (decreasing) rank $s_i$ in the restricted sample, with $1 \leqslant s_i \leqslant A_j$ and use

$$\frac{\widehat{S}(X_i) - \widehat{S}(u_{j+1})}{\widehat{S}(u_j) - \widehat{S}(u_{j+1})} = \frac{s_i - a}{A_j - 2a + 1} \tag{4.8}$$

from which we get our estimate $\widehat{S}(x)$ for $x = X_i$

$$\widehat{S}(X_i) = \widehat{S}(u_{j+1}) + \left[\widehat{S}(u_j) - \widehat{S}(u_{j+1})\right] \times \frac{s_i - a}{A_j - 2a + 1}. \tag{4.9}$$

### Remarks

- It can happen that no uncensored observation falls in the highest slice $(u_J,\ u_{J+1}]$, in which case $\widehat{S}(u_J) = 0$. We then have the information that no observation greater than the fixed level $u_J$ occurred during $n_J$ years. Although potentially very meaningful[1], such information will not be seen on the return level plot. For PP and QQ plots, the information will have an impact because it changes the ML estimate of the parameter $\boldsymbol{\theta}$.

- The parameter $a$ has no impact on the estimation of the survival at the thresholds $u_j$, and only impacts the inter-threshold interpolation.

- Consider the case when the number $A_j$ is large. Then for the largest of the $A_j$ observations, i.e. for $s_i = 1$, we get $\widehat{S}(X_i) \approx \widehat{S}(u_{j+1})$ and for the smallest one i.e. for $s_i = A_j$ we get $\widehat{S}(X_i) \approx \widehat{S}(u_j)$.

- When $A_j = 1$ the choice of $a$ has no impact, since the unique observation $X_i$ in slice $j$ then has an estimated survival

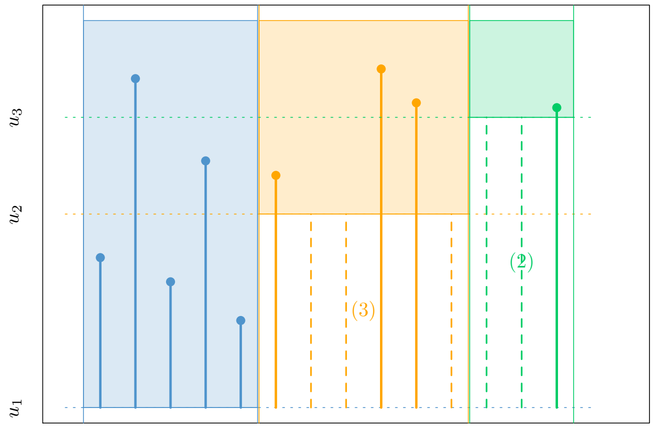
  $$\widehat{S}(X_i) = \left[\widehat{S}(u_j) + \widehat{S}(u_{j+1})\right]/2.$$

  For the highest slice containing observations, i.e. with $\widehat{S}(u_{j+1}) = 0$, the estimated survival at $X_i$ is half of $\widehat{S}(u_j)$.

### Estimation of the return period

Once the survival has been estimated at the observations $X_i$, the corresponding return periods are estimated as $\widehat{T}(X_i) = w/\widehat{S}(X_i)$, and $\log[\widehat{T}(X_i)]$ can be used as abscissa in the return level

[1] If $n_J$ is large.



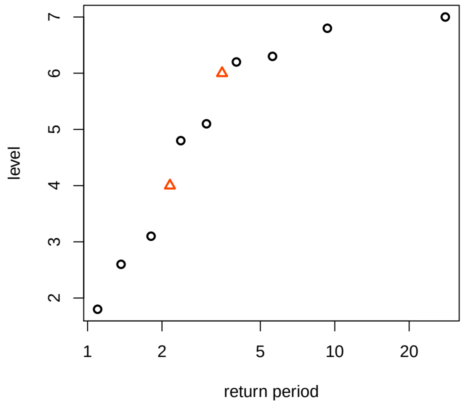




Figure 4.2: Left: example with three periods. Right: return level plot for the example (log scale for periods and $a = 0$ is used in the estimation); triangles show $\widehat{T}(u_j)$ for $j = 2, 3$ while circles show the estimations $\widehat{T}(X_i)$ for the 9 recorded levels $X_i$.

plot. It is easily checked that the return period is increasing with the level. Note that the return period of the threshold $u_J$ can be $\infty$ if no uncensored observation greater than $u_J$ is known.

**Remark.** It can be interesting in some cases to display on the return level plot the estimated return periods at the thresholds as is done at the right of Figure 4.2. The corresponding points should be displayed using special symbols differing from those used for interpolated points.

### 4.2.4 Worked example

We consider the example depicted on Figure 4.2. We have $n = 14$ blocks, say years, corresponding to periods of length (in years) $n_1 = 5$, $n_2 = 6$, and $n_3 = 3$. The numbers of uncensored observations in the three periods are $r_1 = 5$, $r_2 = 3$ and $r_3 = 1$. Note that for the period $k = 3$ we do not know how many observations $X_i$ are in the slice $u_j < M \leqslant u_{j+1}$ for $j = 1, 2$, and we only know the sum of these two numbers.

The computation of $A_j$ and $B_j$ is illustrated in table 4.3 where rows are associated to slices $j = 1, 2, 3$ and the 3 columns relate to periods $k = 1, 2, 3$. The greyed cells show combinations of slice and period for which the number of observations is generally unknown, but the column total is known since it equals the number $n_k$ of years in period $k$. The numbers $A_j$ is simply the row total over cells with known counts, i.e. for periods $k = 1$ to $j$. The number $B_j$ is the total *for the same columns* $k = 1$ to $j$, but for slices with number $< j$.

Using the numbers $A_j$ and $B_j$, the recursion is

$$\widehat{S}(u_3) = \widehat{S}(u_4) + \frac{A_3}{A_3 + B_3}\left[1 - \widehat{S}(u_4)\right] = 0 + \frac{4}{4+10}[1-0] = \frac{4}{14} \approx 0.29$$
$$\widehat{S}(u_2) = \widehat{S}(u_3) + \frac{A_2}{A_2 + B_2}\left[1 - \widehat{S}(u_3)\right] = \frac{4}{14} + \frac{2}{2+6}\left[1 - \frac{4}{14}\right] = \frac{13}{28} \approx 0.46$$
$$\widehat{S}(u_1) = \widehat{S}(u_2) + \frac{A_1}{A_1 + B_1}\left[1 - \widehat{S}(u_2)\right] = \frac{13}{28} + \frac{3}{3+0}\left[1 - \frac{13}{28}\right] = 1.$$

The return periods are shown on the right panel of Figure 4.2.

| | period 1 | period 2 | period 3 | |
|---|---|---|---|---|
| slice 3 | 1 | 2 | 1 | $A_3 = 4,\ B_3 = 10$ |
| slice 2 | 1 | 1 | (2) | $A_2 = 2,\ B_2 = 6$ |
| slice 1 | 3 | (3) | | $A_1 = 3,\ B_1 = 0$ |
| $n$ | 5 | 6 | 3 | 14 |

Table 4.3: Numbers $A_j$ for example. The greyed cells correspond to counts that are generally unknown. Numbers within parentheses relate to censored observations and correspond to the sum of counts in the greyed cells (same column). Both $A_j$ and $B_j$ are computed using periods $k = 1$ to $j$.

## 4.3 Plotting positions for POT

### 4.3.1 Outlook

In this section, we turn to the marked process context as used in **Renext**, and assume that $n$ marks or *levels* $X_i$ are observed on a period with duration $w$, corresponding to order statistics $Z_1 > Z_2 > \cdots > Z_n$. We first describe the derivation of plotting positions in this context when no censoring occurs, then we turn to the problem of censored observations.

### 4.3.2 No censoring

#### RL-plot

Recall that in the POT context the return period for a mark level $x$ is given by

$$T(x) = \lambda^{-1} \times S(x)^{-1}, \tag{4.10}$$

so we may consider estimating the rate and the survival to obtain the return period. However, given a *fixed* level $v$, the return period $T(v)$ is estimated simply by $\widehat{T}(v) = w/N(v)$ where $N(v)$ is the number of $Z_i > v$.

The rate $\lambda$ is estimated by $\widehat{\lambda} = n/w$. The survival $S(x)$ at $x = Z_i$ being estimated by the classical formula (4.3), we get

$$\widehat{T}(Z_i) = \frac{w}{n} \times \frac{n+1-2a}{i-a}, \tag{4.11}$$

which even for $a = 0$ is different from the result that we would get regarding $Z_i$ as a fixed level. Note that when $n$ is large the return period for the largest observed level $Z_1$ is $\approx w/(1-a)$, while the one for the smallest $Z_n$ is $\approx w/n$. The choice of $a = 0.5$ may seem appealing since it leads to the return period $w/\widetilde{N}$ where $\widetilde{N} := i - 0.5$ operates a kind of continuity correction between $N(Z_i - \epsilon) = i$ (for small $\epsilon > 0$) and $N(Z_i) = i - 1$.

In the POT context, remind that $n$ is random and can be zero in which case no estimation can be done.

#### QQ and PP-plots

The QQ and PP plots can be derived as in the ordinary sample context.

### 4.3.3 Type I censoring

#### Problem

The major difference with the ordinary sample context is that the number of observations for a censored period is no longer known, even though the period duration is known.

As we did in section 4.2.3, we consider $J$ "periods" $j = 1, 2, \ldots, J$ with threshold $u_j$, duration $w_j$ and number of observations $r_j$. By "period" we mean a finite set of time intervals with known total duration; such a period may not be a single time interval. As before, we may assume that the corresponding block thresholds $u_j$ are in increasing order

$$u_1 < u_2 < \cdots < u_J$$

and we set for convenience $u_{J+1} := \infty$. The lowest threshold $u_1$ is the ordinary POT threshold as used in the determination of the excesses.

If the numbers $n_j$ of observations were known for each period $j$, then one could proceed exactly as in the ordinary sample framework. First, we would then estimate the unknown rate $\lambda$ as $\widehat{\lambda} := (\sum n_j)/(\sum w_j)$. Then we could estimate the survivals $S(u_j)$ for $j = J, J-1, \ldots, 1$ as we did in the ordinary sample context.

However, it is very unlikely in practice that the number $n_j$ could be known. A natural approach is then to estimate the *return periods* $T(u_j)$ rather than the *survivals* $S(u_j)$ at the thresholds $u_j$ $(1 \leqslant j \leqslant J)$, because of the confusion between the rate and the survival. Without precautions in the estimation, we might get a sequence that is not increasing for increasing levels.

#### Example

Consider the Figure 4.3 involving $J = 2$ periods with durations $w_1 = 5$ and $w_2 = 6$. In the first period, we have only uncensored events and levels, with $r_1 = n_1 = 2$. In a second period, we have $r_2 = 4$ uncensored events and an unknown number $n_2 - r_2$ of censored events. Using the two periods, we find $\widehat{T}(u_2) = 11/5 = 2.2$, since 5 events with $X_i > u_2$ occurred during 11 years. Now, using only the first period, we would estimate the return period of $u_1$ as

$$\widehat{T}(u_1) = 5/2 = 2.5 > \widehat{T}(u_2)$$

which makes our estimates impracticable. A natural way to avoid this problem is to estimate the unknown number $n_2 - r_2$ using the rate of events with $u_1 < X_i \leqslant u_2$. Since 1 such event(s) occurred during the $n_1 = 5$ years, the rate estimate is $1/5$, and we expect to have $6 \times 1/5 = 1.2$ censored events on the second period and thus $\widehat{n}_2 = 4 + 1.2 = 5.2$. The return period at $u_1$ now becomes be $\widehat{T}(u_1) = 11/(2 + 5.2) = 1.52$ years.

#### Proposed algorithm

The computation of the $\widehat{T}(u_j)$ for $1 \leqslant j \leqslant J$ is generalised as in algorithm 1. The main idea is to estimate the unknown number of events with mark in slice $j$, i.e. such that $X_i \in (u_j,\, u_{j+1}]$ by using an estimate of the event rate $\lambda_j$ for the considered row only

$$\lambda_j := \text{rate of } X_i \text{ falling in } (u_j,\, u_{j+1}].$$

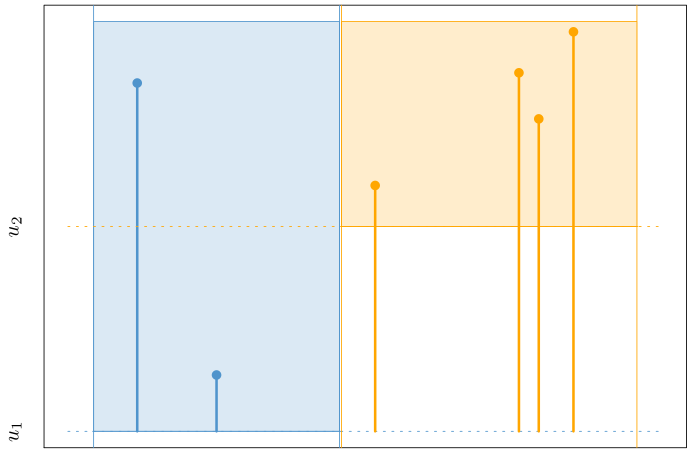



Figure 4.3: Hypothetical POT example. Without precautions, the estimated return periods could fail to be be increasing for increasing levels, e.g. $\widehat{T}(u_1) = 2.5$ and $\widehat{T}(u_2) = 2.2$

The whole rate $\lambda$ is the sum of the $\lambda_j$ for $j = 1, 2, \ldots, J$ and the same relation holds for the estimates

$$\widehat{\lambda} = \widehat{\lambda}_1 + \widehat{\lambda}_2 + \cdots + \widehat{\lambda}_J. \tag{4.12}$$

An estimated (or imputed) number $N(u_j)$ of events with marks $X_i > u_j$ can be obtained by estimating the number of events with marks in slice $j$ as

$$\widehat{N}(u_j) - \widehat{N}(u_{j+1}) = A_j + \widehat{\lambda}_j \, [W_J - W_j]$$

where $W_j := \sum_{i=1}^{j} w_i$ is the total duration of the periods with known number of observations in slice $j$. The two terms at the right hand side are numbers of events with marks in slice $j$: the first is the *known* number for the periods 1 to $j$, and the second is the *estimated* number for the periods $j + 1$ to $J$.

It is clear from algorithm 1 that the $\widehat{N}_j$ are in decreasing order so that the estimated return periods $\widehat{T}(u_j)$ are increasing as wanted.

Once the $J$ return periods $\widehat{T}(u_j)$ have been computed, we can determine the return period for each of the $A_j$ observations $X_i$ falling in an interval $(u_j, u_{j+1}]$. We first determine the rank $s_i$ of $X_i$ in the decreasing order, and then use an interpolation

$$\widehat{T}^{-1}(X_i) = \widehat{T}^{-1}(u_{j+1}) + \left[\widehat{T}^{-1}(u_j) - \widehat{T}^{-1}(u_{j+1})\right] \times \frac{s_i - a}{A_j - 2a + 1}. \tag{4.13}$$

The natural convention being now $\widehat{T}(u_{J+1}) := \infty$.

**Remark.** When $A_j$ is large, for the largest observation in the slice $(u_j, u_{j+1}]$ i.e. for $s_i = 1$ we get $\widehat{T}(X_i) \approx \widehat{T}(u_{j+1})$ and for the smallest $X_i$ in this slice, i.e. for $s_i = A_j$, we get $\widehat{T}(X_i) \approx \widehat{T}(u_j)$. ■

The rationale for formula (4.13) is better understood from the estimate of the survival function, which may also be needed e.g. to build a probability plot (PP-plot). Once the estimated inverse return periods and rates have been computed, we can first estimate the survival $S(u_j)$ at the threshold $u_j$ by

$$\widehat{S}(u_j) := \frac{1}{\widehat{\lambda}\,\widehat{T}(u_j)}. \tag{4.14}$$

The rate used here is the *total rate* from (4.12), namely the rate of the events with marks above the POT threshold. Then, we may use a survival interpolation formula as for BM, and obviously

$$\widehat{S}(X_i) := \frac{1}{\widehat{\lambda}\,\widehat{T}(X_i)}. \tag{4.15}$$

Note that the empirical survivals are decreasing as expected since the return periods are increasing. These estimated values of the survival can be used to build a probability plot.

Quite clearly, $\widehat{T}^{-1}(u_j)$ is an unbiased estimator of $T^{-1}(u_j)$ and $\widehat{\lambda}$ is an unbiased estimator of $\lambda$. Furthermore, these estimators are consistent, so $\widehat{S}(u_j)$ should not show a severe bias and the same can be said about the interpolated values $\widehat{T}^{-1}(X_i)$ or $\widehat{S}(X_i)$.

**Algorithm 1** Potting positions (for Return Level plot) .

*# Number of (uncensored) data in slices and periods*
Find the numbers $r_{j,k}$ of observations in slice $j$ and period $k$ for $1 \leqslant k \leqslant j \leqslant J$
*# Number and duration for uncensored data in periods*
Compute $A_j := \sum_{k=1}^{j} r_{j,k}$ and $W_j := \sum_{k=1}^{j} w_k$ for $1 \leqslant j \leqslant J$

*# $\widehat{N}_j$ is the (estimated) number of observations $> u_j$*
*# $\widehat{\lambda}_j$ is the (estimated) rate of observations $\in (u_j,\, u_{j+1}]$, $\widehat{\lambda}$ is the (estimated) whole rate.*
Set $\widehat{N}_{J+1} := 0$
Set $\widehat{\lambda} := 0$

**for** $j = J,\ J-1,\ \ldots,\ 1$ **do**
  $\widehat{\lambda}_j := A_j / W_j$
  $\widehat{\lambda} \leftarrow \widehat{\lambda} + \widehat{\lambda}_j$
  $\widehat{N}_j := \widehat{N}_{j+1} + A_j + \widehat{\lambda}_j\,[W_J - W_j]$
  $\widehat{T}(u_j) \leftarrow W_J / \widehat{N}_j$
**end for**

**for** $j = J,\ J-1,\ \ldots\ 1$ **do**
  **for** each of the $A_j$ observations $X_i \in (u_j,\, u_{j+1}]$ **do**
    Find the (decreasing) order $s_i$
    Compute the estimated return period $\widehat{T}(X_i)$ using formula (4.13)
    Compute the estimated survival $\widehat{S}(X_i)$ using formula (4.15)
  **end for**
**end for**

### Variant ("H" variant)

In the derivation above, a $p$-points formula was used for interpolation/extrapolation of the survival $S(x)$ at the observations. An alternative would be to interpolate/extrapolate the cumulative hazard $H(x)$ between the estimated values $H(u_j)$ at the thresholds. We only describe the extrapolation: let $j^\star$ be the index of the highest threshold with positive survival $S(u_{j^\star}) > 0$; normally $j^\star$ is equal to $J$. For an observation $X_i$ such that $X_i > u_{j^\star}$, we get

$$\widehat{H}(X_i) = \widehat{H}(u_{j^\star}) + \sum_{k=s_i}^{A_{j^\star}} \frac{1}{k} \qquad (X_i > u_{j^\star})$$

where $s_i$ is the rank of the observation $X_i$ in the reverse order.

**Remark.** With only one observation $A_j = 1$, we get $\widehat{H}(X_i) = \widehat{H}(u_{j^\star}) + 1$ hence $\widehat{S}(X_i) = e^{-1}\widehat{S}(u_{j^\star})$ and $\widehat{T}(X_i) = e\,\widehat{T}(u_{j^\star}) \approx 2.72\,\widehat{T}(u_{j^\star})$.

An interpolation formula could be derived and used for observations $\leqslant u_{j^\star}$, but it is likely to show only small differences with the interpolation described above, so it has not been implemented yet.

**Remark.** The interpolation formula requires computing the expectation of the order statistics for a sample of a doubly truncated exponential distribution. This can be done by using a recurrence relation.

### 4.3.4 Type II censoring

Keeping in mind the analysis of the likelihood of Chapter 1, it seems natural that for each MAX block (type II) with $r_b$ observed largest levels, the plotting positions should be identical to those for a OTS block (type I) with the same observations and with its threshold $u_b$ set just below the smallest observation $Z_{b,r_b}$. Doing so, an important question concerns the case where the data contains only $B$ MAX data periods with the same duration $w^\star$: how do our plotting positions compare then to the classical positions used for BM? We now focus on this specific case.

Since $r_b = 1$ for each block $b = 1, 2, \ldots, B$, the corresponding threshold is $u_b := Z_{b,1} - \epsilon_b$ where $\epsilon_b > 0$ is small. Following our notations of the previous section, we have $J := B$, the total duration being $w = Jw^\star$. We also have $W_j = jw^\star$ and $A_j = 1$ for $j = 1, \ldots, J$. In Algorithm 1 we find $\widehat{\lambda}_j = 1/(jw^\star)$, and the recursion for $\widehat{N}_j = \widehat{N}(u_j)$ writes as

$$N_j = N_{j+1} + A_j + \widehat{\lambda}_j\,[W_J - W_j] = N_{j+1} + 1 + \frac{1}{jw^\star}\,[Jw^\star - jw^\star] = N_{j+1} + J/j$$

for $j = J$ to 1, starting from $\widehat{N}_{J+1} = 0$. So

$$\widehat{N}_j = J \sum_{k=j}^{J} \frac{1}{k} \approx J \sum_{k=j}^{J} \int_k^{k+1} \frac{\mathrm{d}x}{x} = -J\,\log\left[\frac{j}{J+1}\right]. \tag{4.16}$$

The estimated return period $\widehat{T}(u_j)$ is given by $\widehat{T}(u_j) = Jw^\star/\widehat{N}_j$, with the special case $\widehat{T}(u_J) = Jw^\star$.

**Remark.** The estimate of the total rate $\lambda$ is such that $w^\star\widehat{\lambda} = \sum_{j=1}^{J} 1/j \approx \log J + \gamma$, consistently with theorem 3.3 of Chapter 3.

When $j < J$, by interpolating the inverse return period at the only observation $Z_{j,1}$ in the slice $(u_j,\, u_{j+1}]$ as in (4.13) we get

$$\widehat{T}^{-1}(Z_{j,1}) \approx \left[\widehat{T}^{-1}(u_j) + \widehat{T}^{-1}(u_{j+1})\right]/2.$$

Note that the choice of the parameter $a$ as used in the interpolation of the inverse return periods is immaterial here. For the largest observation $Z_{J,1}$, we use the same formula with $\widehat{T}^{-1}(u_{J+1}) = 0$, unless the "H" variant described above is used. In the later case we get

$$\widehat{S}(Z_{J,1}) \approx \frac{e^{-1}}{J}, \qquad \widehat{T}(Z_{J,1}) \approx 2.72\, Jw^{\star}.$$

Assume that the block duration is $w^{\star} = 1$ and remind that $u_j$ is by construction very close to the $j$-th order statistic $X_{(j)}$, so that estimating the return period or the value of the distribution function at $u_j$ or at $X_{(j)}$ are nearly the same thing. An interesting feature is that provided that $J \geqslant 30$ the following approximation holds for for $j$ between 1 and $J$

$$\log \widehat{T}(u_j) \approx -\log\left\{-\log\left(\frac{j}{J+1}\right)\right\} = q_{\texttt{Gum}}\{\widehat{F}(u_j)\}, \tag{4.17}$$

where $q_{\texttt{Gum}}(p)$ denotes the quantile function of the standard Gumbel distribution. This result is easily derived from (4.16); for $j \geq 3$ this approximation turns out to be quite good. In the aggregated marked process framework, we would use a *log* $x$-scale for a return level plot, and hence plot the $u_j$ against $\log \widehat{T}(u_j)$. However, in the classical block maxima framework we would use a *Gumbel* $x$-scale and hence plot $u_j$ against $q_{\texttt{Gum}}\{\widehat{F}(u_j)\}$. The approximation (4.17) means that the two resulting plots will be nearly the same: they differ essentially by the tickmarks displayed on the $x$-axis, see Figure 4.4 for an illustration.

So, in the case where the data are block maxima only, the plotting positions that we derived for the general aggregated marked process framework are different from the usual plotting positions usually retained for block maxima. However, the first plotting positions are designed to be used in a return level plot using a *log scale* rather than the *Gumbel scale* that should normally be retained for block maxima. The two resulting plots are then nearly identical.

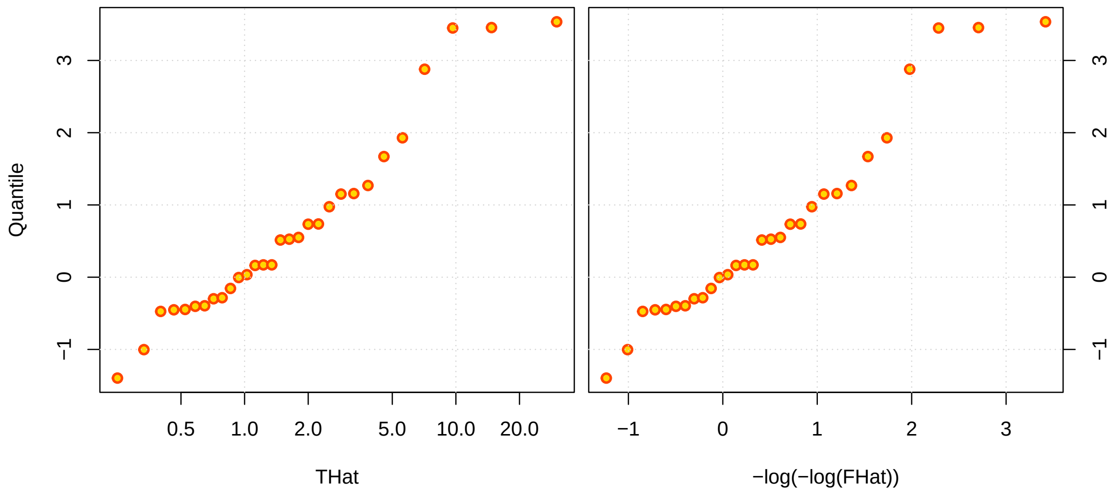



Figure 4.4: Simulated sample of length 30 from the Gumbel distribution. On left panel, the order statistics are plotted against the return periods $\widehat{T}(u_j)$ on log scale. On the right panel, the order statistics are plotted against $-\log\{-\log \widehat{F}(u_j)\}$ thus emulating a plot with Gumbel scale. The two plots are nearly identical; small differences can hardly be seen on the three leftmost points.

# Chapter 5

# Tests

## 5.1 GPD tests in POT

### 5.1.1 Tests of interest

An important concern with POT models is testing the exponentiality of the excesses against a GPD alternative, with a special interest on the Lomax (heavy tail) alternative. We assume that the excesses $Y_i$ are from $\texttt{GPD}(0,\, \sigma,\, \xi)$ with fixed location $\mu = 0$ and hence with two unknown parameters: the scale $\sigma > 0$ and the shape $\xi$. For $\xi = 0$ the distribution is exponential, and the parametric tests of interest are

$$\mathrm{H}_0 :\ \xi = 0 \text{ (exponential)} \quad \text{against} \qquad \mathrm{H}_1 : \begin{cases} \xi \neq 0 & \text{(GPD)}, \\ \xi > 0 & \text{(Lomax)}, \\ \xi < 0 & \text{(maxlo)}. \end{cases}$$

We will refer to the case where the set of observations boils down to $n$ i.i.d. excesses $Y_i$ as the *ordinary sample* context. We further want to test these hypotheses in the *heterogeneous data* context where MAX and/or OTS data are available. The Likelihood Ratio (LR) theory can be applied then, with some restrictions or adaptations.

### 5.1.2 Likelihood-ratio

The LR theory provides a fairly general framework for the derivation of tests, in connection with the asymptotic normality of log(LR). A good modern introduction can be found in the book by Davison (Davison 2003, chap. 4).

The distribution of interest depends on the parameter $\boldsymbol{\theta} := [\sigma,\, \xi]^\top$, which evolves in a domain $\Theta_F$ that depends on the distribution $F$ of the alternative hypothesis. For instance

$$\Theta_{\texttt{GPD}} = \{\boldsymbol{\theta} : \sigma > 0,\, \xi > -1\}, \qquad \Theta_{\texttt{Lomax}} = \{\boldsymbol{\theta} : \sigma > 0,\, \xi > 0\}.$$

The null hypothesis corresponds to the scalar restriction $\xi = 0$, while $H_1$ corresponds to $\boldsymbol{\theta} \in \Theta_F$ depending on the chosen distribution $F$. The LR statistic is

$$\mathrm{LR} = \frac{L(\widehat{\boldsymbol{\theta}}_0)}{L(\widehat{\boldsymbol{\theta}}_1)} = \frac{\text{maximal likelihood under } H_0}{\text{maximal likelihood under } H_1},$$

with values $\mathrm{LR} \leqslant 1$. It is often convenient to use the test statistic $W := -2\log \mathrm{LR}$, which takes values $W \geqslant 0$ and is the difference of the *deviances* $D := -2\log L$. A large value for $W$ tells that $H_0$ should be rejected. Under some general conditions it can be proved that $W$ has asymptotic distribution $\chi^2(r)$ where $r$ is the number of scalar restrictions imposed by the null hypothesis, so $r = 1$ in our case.

A first problem concerns the Lomax and the maxlo alternatives: under $H_0$, the parameter vector lies *on the boundary of the domain*. This leads to a distribution of *mixed type* (continuous-discrete) for the LR and the test statistic $W$. Indeed, $W$ can take any positive value but it can also be exactly equal to 0 with a positive probability. For instance, for a Lomax alternative, if it turns out that the ML estimate $\widehat{\xi}$ is negative, then the likelihood ratio is exactly equal to 1, meaning $W = 0$. This occurs with a positive probability

$$\Pr\{W_{\texttt{Lomax}} = 0\} = \Pr\{\widehat{\xi} < 0\}$$

which can be proved to tend to 1/2 for large $n$, see later. Note that the test statistics corresponding to the three hypotheses $H_1$ above are related through

$$W_{\texttt{GPD}} = \max\{W_{\texttt{Lomax}},\, W_{\texttt{maxlo}}\} = \begin{cases} W_{\texttt{Lomax}} & \text{if } \widehat{\xi} > 0, \\ W_{\texttt{maxlo}} & \text{if } \widehat{\xi} < 0. \end{cases}$$

Given two of the three distributions, the third follows.

A second – more difficult – problem is that the convergence to the asymptotic distribution is very slow, so a poor approximation of the $p$-value is to be feared even for $n \approx 60$ (say).

## 5.2 Lomax alternative: ordinary sample

Two classical tests will be described for the ordinary sample context: *Jackson's test* and *Wilk's exponentiality test* (or WE1 test). The later could as well be named "CV$^2$ test" since the test statistic is nothing but the squared coefficient of variation. As reported later, both of these tests have a good power for the Lomax alternative. Moreover, investigating the distributions of the test statistics will lead us to a quite general workable approximation formula.

In this section $Y_1, Y_2, \ldots, Y_n$ will denote a sample of the exponential distribution $\texttt{Exp}(\nu)$ with rate $\nu > 0$ and $Z_1 > Z_2 > \cdots > Z_n$ will be the order statistics.

### 5.2.1 Jackson's test

We know form Rényi's representation that $\mathbb{E}(Z_i) = c_i\,\mathbb{E}(Y)$ with

$$c_i := \frac{1}{i} + \frac{1}{i+1} + \cdots + \frac{1}{n}.$$

The coefficients $c_i$ relate to the H-points used by Nelson's plotting positions (see p. 48) since $c_i = H_{n+1-i}$. The Jackson's statistic is defined as

$$J_n := \frac{\sum_{i=1}^n c_i Z_i}{\sum_{j=1}^n Y_j}. \tag{5.1}$$

As noted by Beirlant, de Weit, and Goegebeur (2006), $J_n$ is a kind of correlation coefficient between the order statistics $Z_i$ and their expectations $c_i$ under the null hypothesis.

The distribution of $J_n$ for large $n$ is normal, cf. Beirlant *et al.* (2006) and given by the following

**Theorem 1.** *The distribution of $J_n$ is asymptotically normal with for large $n$*

$$\sqrt{n}\,(J_n - 2) \sim \texttt{Norm}(0,\,1).$$

*Hence we can reject exponentiality at level $\alpha$ when $\sqrt{n}\,(J_n - 2)$ is larger than the quantile $q_Z(1-\alpha)$ of the standard normal* $\texttt{Norm}(0,\,1)$.

Unfortunately, the convergence to the normal is very slow, and the normal approximation remains poor for, say, $n \leqslant 100$.

The statistic $J_n$ can be expressed using the normalised spacings $V_i^\star := i \times V_i$ where $V_i$ is the usual spacing[1] $V_i := Z_i - Z_{i-1}$. Indeed, it is easily shown that

$$J_n = \frac{\sum_{i=1}^n d_i V_i^\star}{\sum_{j=1}^n V_j^\star}$$

where $d_j := 1 + c_{j+1}$ and $c_{n+1} := 0$, see Beirlant *et al.* (2006). This formula is very useful to generate random values from the distribution of $J_n$ because the $n$ random variables $V_i^\star$ are independent and have the same exponential distribution. So we can simulate $J_n$ from the $n$ r.vs $V_j^\star$ without having to sort $n$ values as the original formula would require.

### 5.2.2 CV2 (or Wilk's) exponentiality test

The squared coefficient of variation used here is

$$\mathrm{CV}_n^2 := \frac{\sum_{i=1}^n \left[Y_i - \bar{Y}\right]^2}{n\,\bar{Y}^2} = \frac{\sum_i Y_i^2}{n\,\bar{Y}^2} - 1.$$

Note that the variance uses the denominator $n$ rather than the usual $n-1$. A well-known fact is that the distribution of $\mathrm{CV}_n^2$ is related to that of the *Greenwood's statistic* involving the spacings $U_{(i)} - U_{(i-1)}$ of a sample $U_i$ of size $n-1$ from the uniform distribution on $(0,\,1)$. More precisely, setting then $U_{(0)} := 0$ and $U_{(n)} := 1$ the Greenwood's statistic is defined as the sum of squares

$$G_n := \sum_{i=1}^n \left[U_{(i)} - U_{(i-1)}\right]^2,$$

and we have $\mathrm{CV}_n^2 = nG_n - 1$ in distribution. The statistic $G_n$ is often used to test the uniformity of the $U_i$.

We know that $\widehat{\xi} > 0$ is equivalent to $\texttt{CV}_n^2 > 1$, so

$$\Pr\{\widehat{\xi} > 0\} = \Pr\{\texttt{CV}_n^2 > 1\}.$$

The determination of this probability, say $p_n^\star$, is useful in several contexts. The `pGreenwood1` function of **Renext** provides an approximation.

[1] By convention $Z_0 := 0$.

**Theorem 2.** *The distribution of* $\mathrm{CV}_n^2$ *is the distribution of the* Greenwood's statistic. *For large* $n$*, the following normal approximation holds*

$$\frac{\sqrt{n}}{2}\left(\mathrm{CV}_n^2 - 1\right) \sim \mathtt{Norm}(0,\,1).$$

*Hence we can reject exponentiality at level* $\alpha$ *when* $\sqrt{n}\left(\mathrm{CV}_n^2 - 1\right)/2$ *is larger than the quantile* $q_Z(1-\alpha)$ *of the standard normal* $\mathtt{Norm}(0,\,1)$.

The distribution of the Greenwood's statistic is known to tend to the normal *very slowly.* Available approximations based on saddle point or Edgeworth expansions (Ghosh and Jammalamadaka 1998; Does, Helmers, and Klaassen 1988) were found to have only a low precision, and not to provide an exact second decimal of the tail distribution for $n \leqslant 100$. Royen (2010) provides some quantiles with a high precision (8 digits) obtained by using orthogonal polynomials. However the provided table does not cover all values needed in practice, and the approximation formula is far from easy to implement.

### 5.2.3 Likelihood-ratio

As mentioned before the distribution of the log-LR statistic $W_{\mathtt{Lomax}}$ can not be asymptotically $\chi^2(1)$ under $H_0$. Yet the asymptotic distribution is known and still relates to $\chi^2(1)$.

**Theorem 3.** *For large* $n$*, the distribution of* $W := -2\log \mathrm{LR}$ *for the Lomax alternative is that of the product* $BC$ *of two independent r.vs where* $B$ *has a Bernoulli distribution* $\mathtt{Ber}(p)$ *with* $p = 1/2$ *and* $C$ *has a* $\chi^2(1)$ *distribution with one degree of freedom*

$$W_{\mathtt{Lomax}} \xrightarrow{d} BC, \quad B \sim \mathtt{Ber}(1/2),\; C \sim \chi^2(1).$$

*This asymptotic distribution is a mixture* $1/2\,\delta_0 + 1/2\,\chi^2(1)$ *of a two distributions: a Dirac and a chi-square. For large* $n$*, we can reject exponentiality at level* $\alpha$ *when* $W$ *is larger than the quantile* $q_C(1-2\alpha)$ *of the chi-square distribution* $\chi^2(1)$.

See Kozubowski, Panorska, Qeadan, Gershunov, and Rominger (2009). The test implication is easy to check: under $H_0$, we have for large $n$ with $q := q_C(1-2\alpha)$

$$\begin{aligned}\Pr\{W > q\} &= \Pr\{BC > q\} = \Pr[C > q \,|\, B = 1]\ \Pr\{B = 1\}\\ &= \Pr\{C > q\}\ \Pr\{B = 1\} = 2\alpha \times \frac{1}{2} = \alpha.\end{aligned}$$

The good power of Jackson's and Wilk's test as found by simulations (Kozubowski *et al.* 2009) is enlightened by the fact that the test statistics $\mathrm{CV}_n^2$ and $J_n$ both are close to increasing functions of the LR, see Figure 5.1.

## 5.3 Workable approximation

### 5.3.1 Context

We describe here an approximation method that can work for the distributions of the test statistics: $J_n$, $\mathrm{CV}_n^2$, $W_{\mathtt{GPD}}$, $W_{\mathtt{Lomax}}$ and more. These distributions depend on $n$.

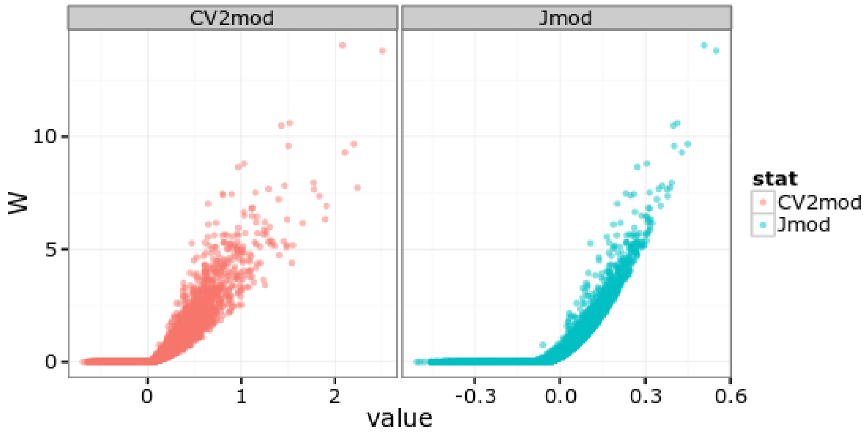



Figure 5.1: The statistic $W_{\mathtt{Lomax}}$ for $H_1 : \xi > 0$ against $\mathtt{CV}^2 - 1$ (left), and against $J_n - 2$ (right) for 15000 independent samples of the exponential distribution with $n = 40$. Note that $W_{\mathtt{Lomax}} > 0$ is equivalent to $\mathtt{CV}^2 - 1 >$ (left) but *not* to $J_n - 2 > 0$.

| Name | Statistic $S$ | normalisation | limit | $q_W(p, \infty)$ |
|---|---|---|---|---|
| Greenwood | $\mathtt{CV}_n^2$ | $W := \sqrt{n}\, \left[\mathrm{CV}_n^2 - 1\right]/2$ | $\mathtt{Norm}(0, 1)$ | $q_Z(p)$ |
| Jackson | $J_n$ | $W := \sqrt{n}\,[J_n - 2]$ | $\mathtt{Norm}(0, 1)$ | $q_Z(p)$ |
| GPD deviance | $W_{\mathtt{GPD}}$ | none | $\chi^2(1)$ | $q_{\chi^2(1)}(p)$ |
| Lomax deviance | $W_{\mathtt{Lomax}}$ | none | $1/2\,\delta_0 + 1/2\,\chi^2(1)$ | $q_{\chi^2(1)}(2p-1)$ for $p > 1/2$ |

Table 5.1: Normalising constants and asymptotic distributions.

Our interest is on a statistic $W$ tending to a non-degenerate known distribution such as $\mathtt{Norm}(0, 1)$, $\chi^2(1)$ or the mixed distribution used in theorem 3. This is the case for $W_{\mathtt{GPD}}$ and $W_{\mathtt{Lomax}}$. For $J_n$ and $\mathrm{CV}_n^2$, we have to use a normalisation of the original statistic $S$ according to $W := [S - \gamma]/\delta_n$ for some suitable known $\gamma$ and $\delta_n > 0$, as given by theorems 1 and 2 see table 5.1. We then consider

$$q_W(p, n) := \text{quantile function of } W.$$

Since rejection occurs when the statistic is large in all cases, the rejection at the level $\alpha$ occurs when $W > q_W(1 - \alpha, n)$.

### 5.3.2 The approximation formulas

We want to approximate the quantile function for a typical range of probability $0.001 \leqslant p \leqslant 0.999$. Let $q_Z(p)$ be the quantile of the standard normal distribution, and let us denote simply by $q$ its value for a given probability $p$. The corresponding range for $q$ is approximately $q = -2.33$ to $q = 2.33$. Depending on the sample size $n$, two distinct approximations will be used, both relying on B-splines and using the same basis functions. Let $m$ be the spline order (so $m = 4$ for a cubic spline), and consider a set of $r + m$ knots

$$\tau_1 < \tau_2 < \cdots < \tau_{r+m}$$

leading to a basis of $r$ B-splines $\phi_1(q),\, \phi_2(q),\, \dots,\quad \phi_r(q)$.

1. For $n_0 \leqslant n \leqslant n_1$ we use a regression spline in $q = q_Z(p)$

$$q_W(p,\, n) \approx \sum_{i=1}^{r} \alpha_{n,i}\, \phi_i(q) \tag{5.2}$$

   where the coefficients depend on the sample size.

2. For $n > n_1$, we regress the quantile on the tensor product of a spline in $q$ and of a polynomial in the variable $1/\sqrt{n}$. More precisely, the approximation writes as

$$q_W(p,\, n) \approx \sum_{k=0}^{3} b_k(p)\, n^{-k/2} \tag{5.3}$$

   where every coefficient $b_k(p)$ is expanded on the spline basis

$$b_k(p) = \beta_{k,1}\, \phi_1(q) + \beta_{k,2}\, \phi_2(q) + \cdots + \beta_{k,r}\, \phi_r(q) \tag{5.4}$$

For the first approximation, the coefficients can be stored in a matrix $\mathbf{A} = [\alpha_{n,i}]_{n,i}$ with $n_1 - n_0 + 1$ rows and $r$ columns. For the second approximation, we can use a matrix $\mathbf{B} = [\beta_{k,i}]_{n,i}$ with 4 rows and $r$ columns.

The motivation for using two formulas is that the second formula is not flexible enough for the small values of $n$ because for a fixed $p$ the quantile $q_W(p,\, n)$ has a strong variation and a non-monotone behaviour observed for small $n$, see Figure 5.2. On the opposite, the true quantile seems to have a smooth monotone behaviour for larger $n$ and hence to be easily interpolated there. It would be both costly and redundant to use a vector or $r$ coefficients for each value of $n$, with the additional burden of simulating samples for every value of $n$.

Quadratic splines (with order $m = 3$) seem smooth enough for our purpose, and a value of $r$ about 5 seems good. The number of knots used for the spline is about 8. Note that for a fixed probability $p$ the quantile formulas for $q_W(p,\, n)$ do not ensure a smooth behaviour for $n = n_1$, although this does not seem to be a concern for the fitted curves except perhaps for the highest probability, see Figure 5.2.

### 5.3.3 Getting the coefficients

The coefficients $\alpha_{n,i}$ and $\beta_{k,i}$ in the previous formula can be estimated by using a large number of simulated values $W^{[j]}$ for different values of the sample size $n$ and of the probability $p$. The values of $n$ and $p$ were chosen on an irregular grid with $n$ ranging from 2 to 500 and $p$ ranging from $p = 0.001$ tp 0.999, see Figure 5.2.

Ideally, a quantile regression could be used then. For the sake of simplicity, a method based on the empirical quantiles $\widetilde{q}_W(n,\, p)$ was used, and for each value of $p$ a spline smoothing was first used to denoise $\widetilde{q}_W(n,\, p)$. More precisely, a smoothing spline was fitted for $\widetilde{q}_W(n,\, p)$ against $\log n$; the degree of freedom of the spline was chosen in order to avoid the over-smoothing for the small values of $n$ that would have been observed with a standard determination of the smoothing level[2]. Then the smoothed quantiles are used as the response $q_W(n,\, p)$ in formulas (5.2) or (5.3)

[2] By cross-validation for instance.

and a Ordinary Least Squares fit is used to provide estimated values of the parameters $\alpha_{n,i}$ and $\beta_{k,i}$. The raw quantiles were computed using $N = 10^5$ samples for each sample size $n$; they are shipped as technical data[3] with **Renext** under the name `quantRaw`.

For the Greenwood's statistic the approximated quantiles where compared to those computed by Royen (2010) with a 8-digits precision using orthogonal polynomials. The approximation is quite good and show an accuracy better than two-digits for the probabilities that can be compared[4], cf. Figure 5.3.

### Limitation and possible improvements

The described approximation method is sufficient for its use in **Renext** but is clearly not intended to provide a high precision computation of the distributions as it could be needed in other contexts. Some affordable improvements could be obtained by ensuring a smooth connection between the two formula or by a using a 2D smoothing method preserving the monotonicity in $p$.

## 5.4 Test for GEV and Block Maxima

The problem of testing the 'Gumbelity' of the maxima $M_i$ in the Block Maxima context may appear at first glance as similar to testing the exponentiality of POT excesses. Assume that the $n$ r.vs $M_i$ are independently drawn from $\texttt{GEV}(\mu,\,\sigma,\xi)$, then the tests concern

$$\mathrm{H}_0:\ \xi = 0\ \text{(Gumbel)} \quad \text{against} \qquad \mathrm{H}_1: \begin{cases} \xi \neq 0 & \text{(GEV)},\\ \xi > 0 & \text{(Fréchet)},\\ \xi < 0 & \text{(reversed Weibull)}. \end{cases}$$

The LR framework can be applied as before. For the Fréchet and reversed Weibull alternatives, the distribution of the test statistic $W = -2\log \texttt{LR}$ will again be of mixed type. Then asymptotically (for large $n$) $W$ has the distribution of the product $BC$ of independent r.vs with $B \sim \texttt{Ber}(1/2)$ and $C \sim \chi^2(1)$.

It must be remarked that the ML estimation of a Gumbel distribution is more difficult than that of the exponential; thus investigating the distribution of the test statistics requires much heavier computations. Nevertheless, the method of approximation of the quantiles described above was worked out for the GEV context, see Figure 5.4. The determination of the raw quantiles from $N = 10^5$ Gumbel samples of size $n$ needed several hours of computation.

Fortunately enough, the convergence of the distributions is faster than for the GPD context, and it appears that the asymptotic distributions can be used as soon as $n \geqslant 30$.

**Remark.** This statement could be related to the broader rule that block maxima should be compared to a marked process (MP) with rate between 4 to 5, and not to a MP with unit rate. Roughly speaking: a BM observation should tell as much as 4 to 5 POT observations do. Quantiles curves shown on Figure 5.4 are nearly as flat for $n \geqslant n^\star$ as are the two bottom plots of Figure 5.2 for $n \geqslant 4.5n^\star$. ■

[3] In the `sysdata.rda` file of the `R` directory.

[4] The maximal absolute difference was found is $< 0.004$.

## 5.5 Heterogeneous data

### 5.5.1 Problem

With heterogeneous data including MAX or OTS blocks, the LR theory can still be considered. The distribution of the LR statistic no longer depends one a sole 'design' feature $n$. It depends then of extra features such as the number of blocks for each type as well as the duration and the number of observations of these blocks. This makes it difficult to formulate an asymptotic convergence theorem, and even more to prove such a thing. There is little hope of deriving an efficient approximation of the distribution of the LR.

The only simple solution is to perform Monte-Carlo simulations of the LR statistic under the null hypothesis. This simulation is conditional on what could be named the *design* – as in a regression context – namely, the non-random characteristics of the data[5]. Yet this is very computationally intensive, since every simulated value requires the maximisation of two likelihood functions. The computational burden could be reduced by a strategy of stratified sampling as described in the next section. Inasmuch as the asymptotic distribution was found to work fine for Block Maxima even with a limited number (say 30) of these, the straightforward asymptotic approximation was retained in **Renext** for the moment.

### 5.5.2 Stratified sampling ⋆

A possible reduction of the computational burden can result from using a modified Jackson's statistic as an instrument (or proxy) for the test statistic $W$. Admitting that we can find a statistic, say $J$, which has a high dependence on $W$ and which is easy to simulate from, we can use a stratified sampling to evaluate the $p$-value $\Pr\{W > W_{\texttt{obs}}\}$ for a given value $W_{\texttt{obs}}$ with a limited number of evaluations of $W$. By using a partition of the support of $J$ in $K$ intervals $\mathcal{J}_k$, we can first draw a large number of proxy values $J^{[m]}$ and estimate the probabilities of the strata $\pi_k := \Pr\{J \in \mathcal{J}_k\}$. Then using

$$\Pr\{W > W_{\texttt{obs}}\} = \sum_k \underbrace{\Pr[W > W_{\texttt{obs}} \mid J \in \mathcal{J}_k]}_{:=r_k} \underbrace{\Pr\{J \in \mathcal{J}_k\}}_{=\pi_k}.$$

we can sequentially proceed to evaluations of $W$ using successive iterations in which couples $[J, W]$ are drawn with $W$ evaluated only when needed. The number of evaluations of $W$ is allocated for $J$ in strate $k$ as $\propto \widehat{\pi}_j \sqrt{\widehat{r}_k(1-\widehat{r}_k)}$ where the estimation of $r_k$ is made from the previous iterations.

A modified Jackson's statistic seems a good candidate for a proxy for two reasons. Firstly the Jackson's statistic is a very good proxy of LR in the ordinary sample case, see Figure 5.1. Secondly, as remarked above in section 5.2.1, Jackson's statistic is related to the plotting positions. Based on the methods for censored data, good heuristic plotting positions were found for the considered heterogeneous data context, see chap. 4 p. 45. These plotting positions can be used to provide weights $c_i$ in the numerator of a modified Jackson's statistic (5.1).

[5] Number and durations of the blocks, ...

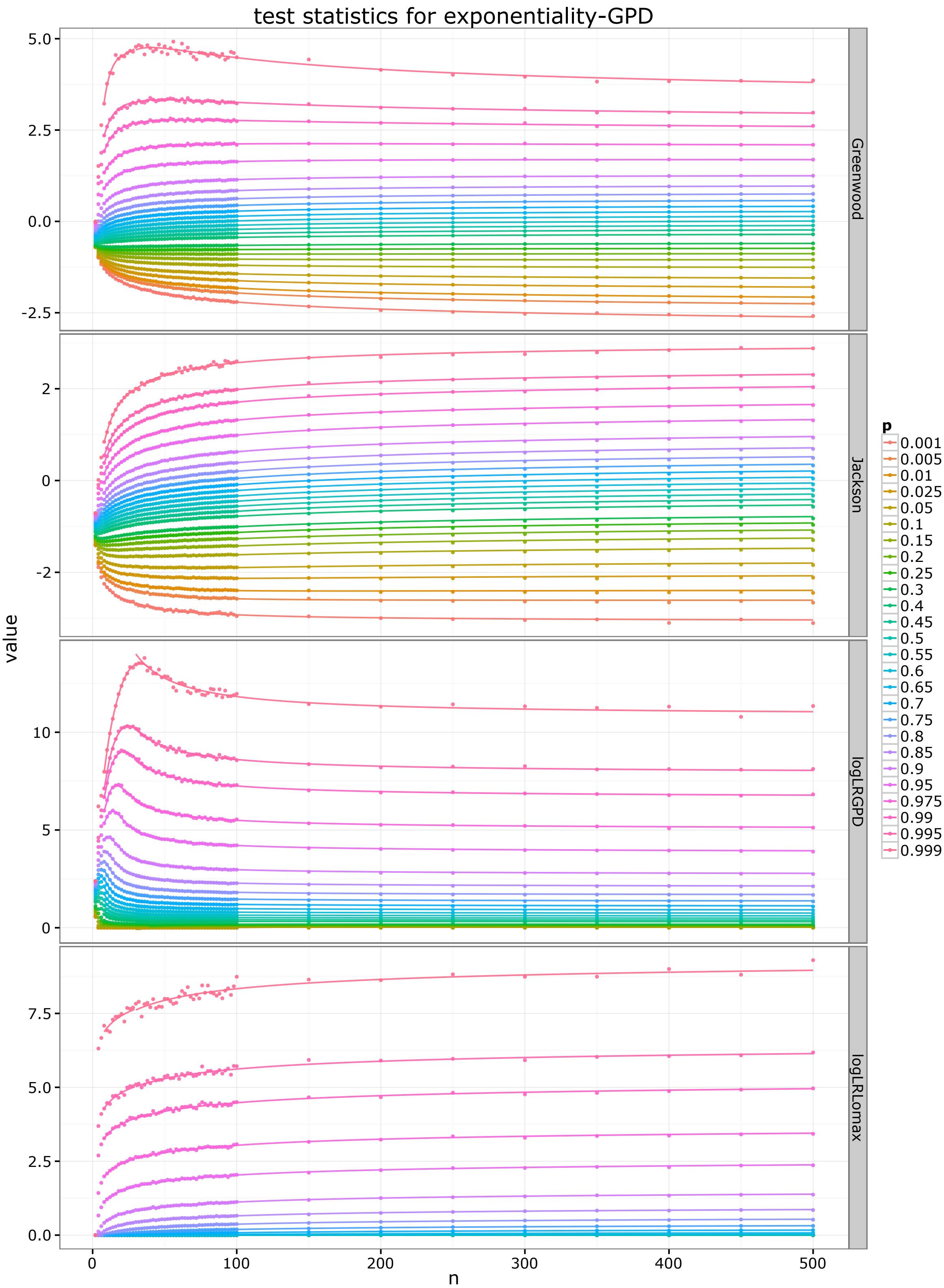



Figure 5.2: Raw quantiles (circles) and their approximation (solid lines) for a selected set of probabilities $p$. The statistics have been normalised as in table 5.1. It is known that each quantile curve tends to a limit for large $n$, but the convergence is very slow since non-negligible variations are seen for $n > 100$. In fact, the two approximations (small/large $n$) are shown as two curves. For the `logLRLomax` plot at the bottom, the exact quantiles $q_W(p, n)$ are zero for $p < 0.5$.

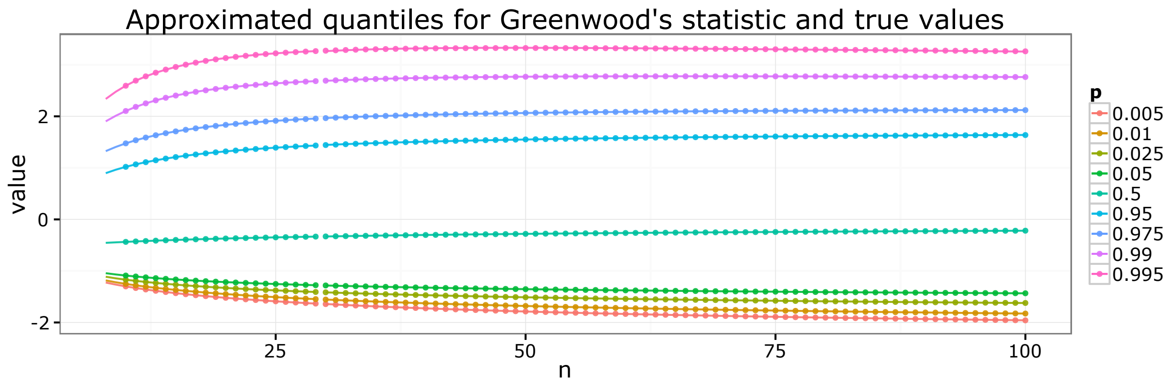



Figure 5.3: The approximation (solid lines) compared to the true values (points) as computed by Royen for a selected set of probabilities $p$.

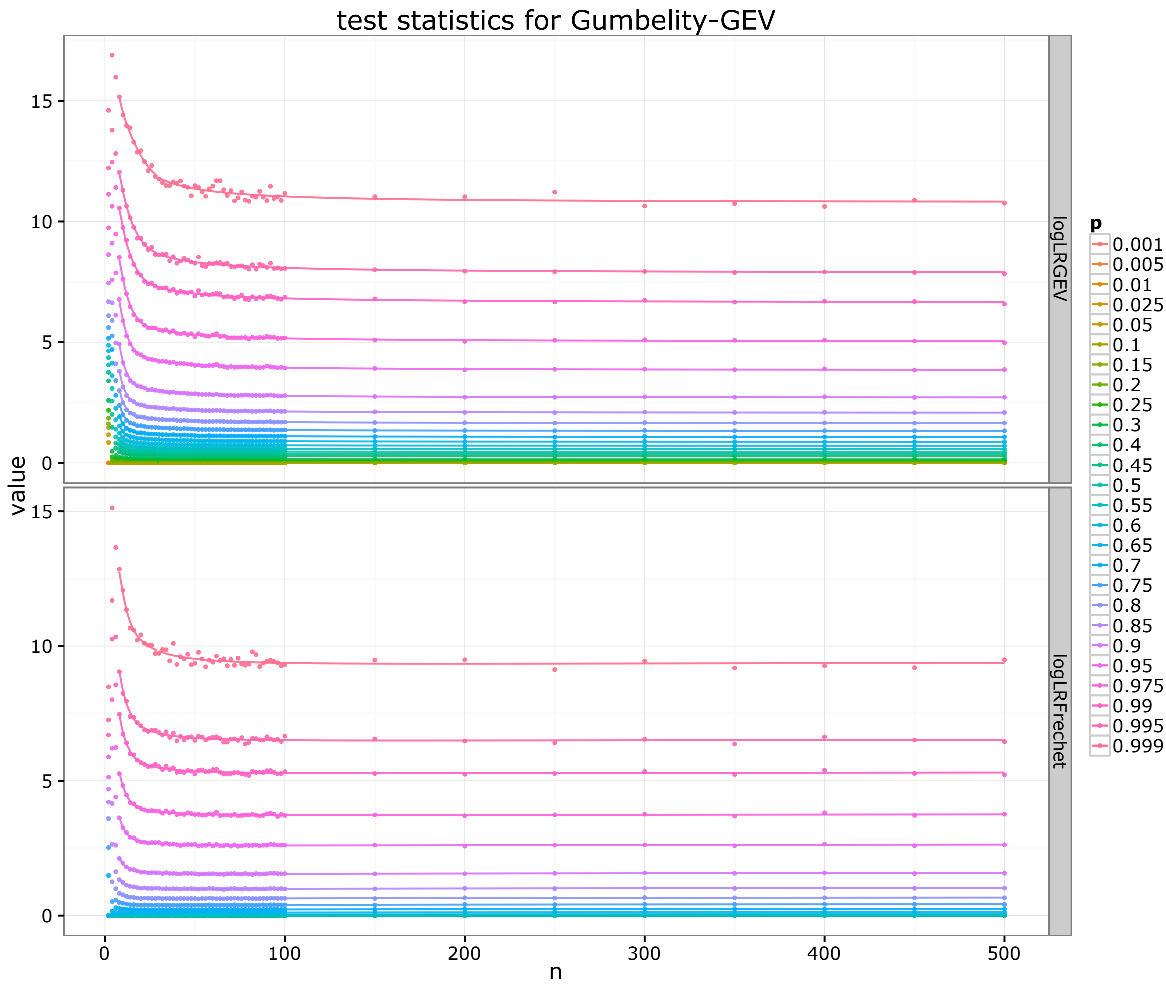



Figure 5.4: Raw quantiles (circles) and the approximation (solid lines) for a selected set of probabilities $p$. The asymptotic quantile provides a good approximation for $n > 50$ and seems practicable as soon as $n \geqslant 30$. For the `logLRFrechet` plot at the bottom, the exact quantiles $q_W(p, n)$ are zero for $p < 0.5$.

# Bibliography

Beirlant J, de Weit T, Goegebeur Y (2006). "A Goodness-of-fit Statistic for Pareto-Type Behaviour." *Journal of Computational and Applied Mathematics*, **186**(1), 99–116.

Coles S (2001). *An Introduction to Statistical Modelling of Extreme Values*. Springer.

Cunnane C (1978). "Unbiased Plotting Positions – a Review." *Journal of Hydrology*, **37**(3), 205–222.

Davison A (2003). *Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Davison A, Smith R (1990). "Models for Exceedances over High Thresholds." *J.R. Statist. Soc. B*, **52**(3), 393–442.

del Castillo J, Daoudi J (2009). "Estimation of the Generalized Pareto Distribution." *Statistics & Probability Letters*, **79**(5), 684–688. ISSN 0167-7152.

Deville Y, IRSN (2019). *Renext: Renewal Method for Extreme Values Extrapolation*. R package version 3.1-2.

Does R, Helmers R, Klaassen C (1988). "Approximating the Distribution of the Greenwood's Statistic." *Statistica Neerlandica*, **42**(3), 153–161.

Embrechts P, Klüppelberg C, Mikosch T (1996). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag.

Ghosh K, Jammalamadaka S (1998). "Theory and Applications Censored Failure Data." *Journal of Statistical Planning and Inference*, **69**(2), 245–261.

Giles D, Feng H, Godwin R (2013). "On the Bias of the Maximum Likelihood Estimator for the Two-Parameter Lomax Distribution." *Communications in Statistics - Theory and Methods*, **42**(11), 1934–1950.

Grimshaw S (1993). "Computing Maximum Likelihood Estimates for the Generalized Pareto Distribution." *Technometrics*, **35**(2), 185–191.

Hirsch R, Stedinger J (1987). "Plotting Positions for Historical Floods and their Precision." *Water Ressources Research*, **23**(4), 715–727.

Johnson N, Kotz S, Balakrishnan N (1994). *Continuous Univariate Distributions*, volume 1. 2nd edition. Wiley.

Johnson N, Kotz S, Balakrishnan N (1995). *Continuous Univariate Distributions*, volume 2. 2nd edition. Wiley.

Kozubowski T, Panorska AK, Qeadan F, Gershunov A, Rominger D (2009). "Testing Exponentiality Versus Pareto Distribution via Likelihood Ratio." *Communications in Statistics - Simulation and Computation*, **38**(1), 118–139.

Millard S, Neerchal N (2001). *Environmental Statistics with S-Plus*. CRC Press.

Miquel J (1981). *Guide d'Estimation des probabilités des débits de crue*. EDF, Direction des Études et Recherches.

Nelson W (2000). "Theory and Applications Censored Failure Data." *Technometrics*, **42**(1), 12–25.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org`.

Royen T (2010). "The Distribution of the Square Sum of Dirichlet Random Variables and a Table with Quantiles of the Greenwood's Statistic." *Far East Journal of Theoretical Statistics*, **33**, 57–80.

Waller L, Turnbull B (1992). "Probability Plotting With Censored Data." *The American Statistician*, **46**(1), 5–12.

# Appendix A

# Miscellaneous

## A.1 Relation between Extreme Value distributions

### A.1.1 GPD and GEV

The survival of the GP Distribution $\texttt{GPD}(\mu,\, \sigma,\, \xi)$ is given by

$$S_{\texttt{GPD}}(x) = \left[1 + \xi\,\frac{x-\mu}{\sigma}\right]_{+}^{-1/\xi}, \qquad x \geqslant \mu.$$

where $\mu$, $\sigma > 0$ and $\xi$ are the *location*, *scale* and *shape* parameters.

The distribution function of the GEV distribution $\texttt{GEV}(\mu,\, \sigma,\, \xi)$ is given by

$$F_{\texttt{GEV}}(x) = \exp\left\{-\left[1 + \xi\,\frac{x-\mu}{\sigma}\right]_{+}^{-1/\xi}\right\}.$$

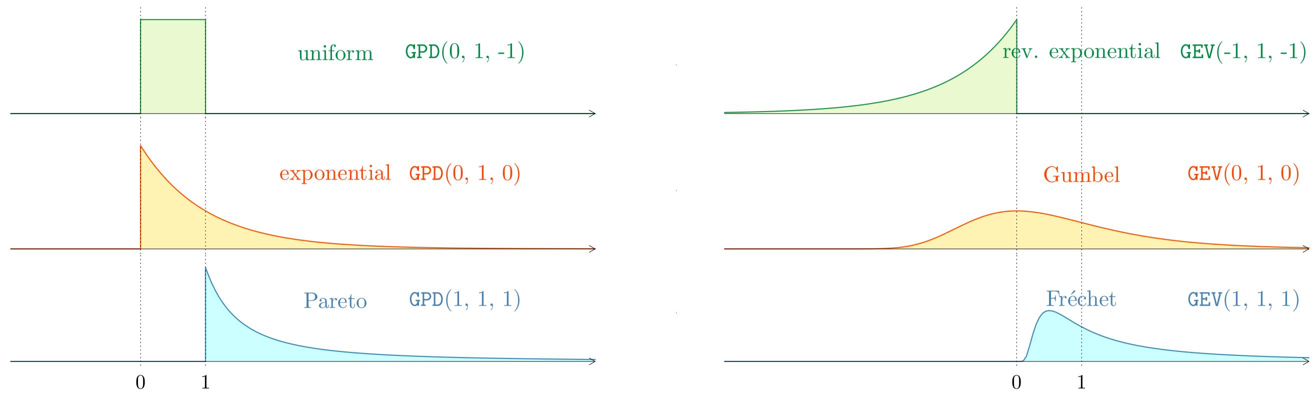
Special cases of these two families are given in table A.1. The support of the GPD and GEV distributions are illustrated on figure 3.2 page 35.

### A.1.2 Transformations

A number of transformations are commonly used to standardise a random variable having an extreme value distribution. Some of them are shown in table A.2.

| Name | | Support | Survival | Distribution |
|---|---|---|---|---|
| Uniform | $\texttt{GPD}(0,\, 1,\, -1)$ | $[0,\, 1]$ | $S(x) = 1 - x$ | |
| Exponential | $\texttt{GPD}(0,\, 1,\, 0)$ | $[0,\, +\infty)$ | $S(x) = e^{-x}$ | |
| Pareto | $\texttt{GPD}(1,\, 1,\, 1)$ | $[1,\, +\infty)$ | $S(x) = 1/x$ | |
| reversed exponential | $\texttt{GEV}(-1,\, 1,\, -1)$ | $(-\infty,\, 0]$ | | $F(x) = e^{x}$ |
| Gumbel | $\texttt{GEV}(0,\, 1,\, 0)$ | $(-\infty,\, +\infty)$ | | $F(x) = \exp\{-e^{-x}\}$ |
| Fréchet | $\texttt{GEV}(1,\, 1,\, 1)$ | $[0,\, +\infty)$ | | $F(x) = \exp\{-1/x\}$ |

Table A.1: Special GP and GEV distributions. They are sometimes called *standard* (exponential, Pareto, Fréchet). The expressions given for the survival or the distribution value are only valid for $x$ in the support. The densities are shown in table A.2.




| | | GP | | | GEV | | |
|---|---|---|---|---|---|---|---|
| | | `Unif` | `Exp` | `Pareto` | `revExp` | `Gumbel` | `Fréchet` |
| GP | `Unif` GPD(0, 1, −1) | $F := X$ | $Y := -\log X$ | $Y := 1/X$ | $Y := \log X$ | $Y := -\log(-\log X)$ | $Y := -1/\log X$ |
| | `Exp` GPD(0, 1, 0) | $S := e^{-X}$ | $Y := X$ | $Y := e^X$ | $Y := -X$ | $Y := -\log X$ | $Y := 1/X$ |
| | `Pareto` GPD(1, 1, 1) | $S := 1/X$ | $Y := \log X$ | $Y := X$ | $Y := -\log X$ | $Y := -\log(\log X)$ | $Y := 1/\log X$ |
| GEV | `revExp` GEV(1, 1, −1) | $F := e^X$ | $Y := -X$ | $Y := e^{-X}$ | $Y := X$ | $Y := -\log(-X)$ | $Y := -1/X$ |
| | `Gumbel` GEV(0, 1, 0) | $F = \exp(-e^{-x})$ | $Y := e^{-X}$ | $Y := \exp(e^{-X})$ | $Y := -e^{-X}$ | $Y := X$ | $Y := e^X$ |
| | `Fréchet` GEV(1, 1, 1) | $F := \exp(-1/X)$ | $Y := 1/X$ | $Y := \exp(1/X)$ | $Y := -1/X$ | $Y := \log X$ | $Y := X$ |

Table A.2: Transformations from a r.v. $X$ with special GP or GEV distribution to a r.v. $Y$, $F$ or $S$ with special GP or GEV distribution. The transformed r.vs $S$ and $F$ correspond to using the transformations $S_X$ and $F_X$.

## A.2 Derivatives of functions related to some distributions

### A.2.1 Exponential

For the exponential case with rate $\nu > 0$ with survival $S(x) = e^{-\nu x}$ for $x > 0$

$$H = \nu x$$

$\partial_\nu H = x$ and $\partial^2_{\nu,\nu} H = 0$. Then $\log f(x) = \log \nu - H$

$$\partial_\nu \log f = \frac{1}{\nu} - x \qquad \partial^2_{\nu,\nu} \log f = -\frac{1}{\nu^2}$$

The derivatives of $S$ are

$$\partial_\nu S = -x\, S \qquad \partial^2_{\nu,\nu} S = x^2\, S$$

### A.2.2 Weibull

For the Weibull case, it will be convenient to use the following quantities

$$H = (x/\beta)^\alpha \qquad U = \log(x/\beta)$$

Using abbreviated notations such as $\partial_\alpha = \partial/\partial\alpha$, we have

$$\frac{\partial H}{\partial \boldsymbol{\theta}} = \begin{array}{|c|c}\cline{1-1} UH & \alpha \\ \cline{1-1} -\dfrac{\alpha}{\beta}H & \beta \\ \cline{1-1}\end{array} \qquad \frac{\partial U}{\partial \boldsymbol{\theta}} = \begin{array}{|c|c}\cline{1-1} 0 & \alpha \\ \cline{1-1} -\dfrac{1}{\beta} & \beta \\ \cline{1-1}\end{array} \qquad \frac{\partial^2 H}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \begin{array}{|c|c|c}\multicolumn{1}{c}{\alpha} & \multicolumn{1}{c}{\beta} & \\ \cline{1-2} U^2 H & -\dfrac{1}{\beta} H\,[1+\alpha U] & \alpha \\ \cline{1-2} & \dfrac{\alpha(\alpha+1)}{\beta^2}\, H & \beta \\ \cline{1-2}\end{array}$$

The log-likelihood is given by

$$\log f(x) = \log \alpha - \log \beta + (\alpha - 1)\, U - H$$

hence

$$\frac{\partial \log f}{\partial \boldsymbol{\theta}} = \begin{array}{|c|c}\cline{1-1} \dfrac{1}{\alpha} + U - UH & \alpha \\ \cline{1-1} -\dfrac{\alpha}{\beta} + \dfrac{\alpha}{\beta} H & \beta \\ \cline{1-1}\end{array} \qquad \frac{\partial^2 \log f}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \begin{array}{|c|c|c}\multicolumn{1}{c}{\alpha} & \multicolumn{1}{c}{\beta} & \\ \cline{1-2} -\dfrac{1}{\alpha^2} - U^2 H & \dfrac{1}{\beta}\,[-1 + H + \alpha\, UH] & \alpha \\ \cline{1-2} & \dfrac{\alpha}{\beta^2}\,[1 - H - \alpha H] & \beta \\ \cline{1-2}\end{array}$$

The derivatives of $F$ can be computed using those of $H$ and the relations (1.5) above.

### A.2.3 GPD (two-parameter)

#### Non-zero shape

It will be convenient to use the two following quantities when $\xi \neq 0$

$$A = \log\left[1 + \xi\,\frac{x}{\sigma}\right] \qquad B = \frac{x}{\sigma}\,\left[1 + \xi\,\frac{x}{\sigma}\right]^{-1}$$

The log-density and the cumulative hazard are

$$\log f(x) = -\log\sigma - \frac{\xi+1}{\xi}\,A \qquad H(x) = \frac{1}{\xi}\,A$$

The quantity $B$ is such that

$$\frac{\sigma}{x}\,B^2 = B - \xi B^2$$

which can be verified by simple calculation. The derivatives of $A$, $B$ and $\log f$ are

$$\frac{\partial A}{\partial\boldsymbol{\theta}} = \begin{array}{|c|c} \hline -\dfrac{\xi}{\sigma}\,B & \sigma \\ \hline B & \xi \\ \hline \end{array} \qquad \frac{\partial B}{\partial\boldsymbol{\theta}} = \begin{array}{|c|c} \hline -\dfrac{1}{x}\,B^2 = -\dfrac{1}{\sigma}\,\left[B - \xi B^2\right] & \sigma \\ \hline -B^2 & \xi \\ \hline \end{array} \qquad \frac{\partial \log f}{\partial\boldsymbol{\theta}} = \begin{array}{|c|c} \hline -\dfrac{1}{\sigma}\,\left[1 - (\xi+1)B\right] & \sigma \\ \hline \dfrac{1}{\xi^2}\,A - \dfrac{\xi+1}{\xi}\,B & \xi \\ \hline \end{array}$$

$$\frac{\partial^2 \log f}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \begin{array}{|c|c|c} \multicolumn{1}{c}{\sigma} & \multicolumn{1}{c}{\xi} & \\ \hline \dfrac{1}{\sigma^2}\,\left[1 - 2(\xi+1)B + \xi(\xi+1)\,B^2\right] & \dfrac{1}{\sigma}\,\left[B - (\xi+1)B^2\right] & \sigma \\ \hline & \dfrac{1}{\xi^3}\,\left[-2A + 2\xi B + \xi^2(\xi+1)B^2\right] & \xi \\ \hline \end{array}$$

Using the expression $H = A/\xi$

$$\frac{\partial H}{\partial\boldsymbol{\theta}} = \begin{array}{|c|c} \hline -\dfrac{1}{\sigma}\,B & \sigma \\ \hline -\dfrac{1}{\xi^2}\,A + \dfrac{1}{\xi}\,B & \xi \\ \hline \end{array} \qquad \frac{\partial^2 H}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \begin{array}{|c|c|c} \multicolumn{1}{c}{\sigma} & \multicolumn{1}{c}{\xi} & \\ \hline \dfrac{1}{\sigma^2}\,\left[2B - \xi B^2\right] & \dfrac{1}{\sigma}\,B^2 & \sigma \\ \hline & \dfrac{1}{\xi^3}\,\left[2A - 2\xi B - \xi^2 B^2\right] & \xi \\ \hline \end{array}$$

The derivatives of $F$ can be computed using those of $H$ and the relations (1.5) above.

#### Zero shape

The formulas given for the non-zero shape case can no longer be used when $\xi = 0$. However all the derivatives of interest are continuous for $\xi = 0$ hence must be replaced by their limit for $\xi \to 0$. In practice, the problem occurs when the observed information matrix is needed and when the sample coefficient of variation is close to 1, see section 2.6.

Assume that the $n$ observations $X_i > 0$ have a unit (sample) coefficient of variation, i.e. that the (sample) moments $M_k$ are such that $M_2 = 2M_1$ and let us consider the mean value of the derivatives at $X_i$ with $\sigma$ replaced by its ML estimate $M_1$. Let $X_i^\star := X_i/M_1$ and

$$A_i := \log(1 + \xi\,X_i^\star) = X_i^\star\xi - \frac{X_i^\star}{2}\xi^2 + \frac{X_i^{\star 3}}{3}\,\xi^3 + o(\xi^3).$$

Let $M_k^\star := M_k/M_1^k$ be the moment of the standardised observations $X_i^\star$, so $M_1^\star = 1$ and $M_2^\star = 2$. The mean value of the $n$ second order derivatives at $X_i$ will require the mean values $\bar{A}$, $\bar{B}$ and $\bar{B^2}$ of the $n$ values $A_i$, $B_i$ and $B_i^2$ corresponding to the $X_i^\star$. Now

$$\bar{A} = \xi - \xi^2 + \frac{M_3^\star}{3}\xi^3 + o(\xi^3).$$

Similarly some simple algebra gives $\bar{B} = 1 - 2\xi + M_3^\star \xi^2 + o(\xi^2)$ and $\bar{B^2} = 2 - 2M_3^\star \xi + o(\xi)$, then

$$\frac{1}{\xi^3}\left[-2\bar{A} + 2\xi\bar{B} + \xi^2(\xi+1)\bar{B^2}\right] = -\frac{2}{3}\left[M_3^\star - 3\right] + o(1).$$

This gives the mean value of the $n$ quantities $\partial^2 \log f(X_i)/\partial\xi^2$. Similarily

$$\frac{1}{n}\sum_i \frac{\partial^2 H(X_i)}{\partial \xi^2} = \frac{2}{3}M_3^\star.$$

### A.2.4 Lomax

The density and survival are given for $x > 0$ by

$$f(x) = \frac{\alpha}{\beta}\left[1 + \frac{x}{\beta}\right]^{-\alpha-1} \qquad S(x) = \left[1 + \frac{x}{\beta}\right]^{-\alpha}$$

Then with $z := 1/[1 + x/\beta]$

$$\log f(x) = \log\alpha - \log\beta + (\alpha+1)\log z$$

and

$$\frac{\partial \log f}{\partial \boldsymbol{\theta}} = \begin{array}{|c|c} \hline \dfrac{1}{\alpha} + \log z & \alpha \\ \hline \dfrac{1}{\beta}\left[\alpha - (\alpha+1)z\right] & \beta \\ \hline \end{array} \qquad \frac{\partial^2 \log f}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \begin{array}{|c|c|c} \multicolumn{1}{c}{\alpha} & \multicolumn{1}{c}{\beta} & \\ \hline -\dfrac{1}{\alpha^2} & \dfrac{1}{\beta}\left[1 - z\right] & \alpha \\ \hline & \dfrac{1}{\beta^2}\left[1 - (\alpha+1)(1-z^2)\right] & \beta \\ \hline \end{array}$$

The derivatives of the cumulative hazard $H$ are

$$\frac{\partial H}{\partial \boldsymbol{\theta}} = \begin{array}{|c|c} \hline -\log z & \alpha \\ \hline -\dfrac{\alpha}{\beta}\left[1 - z\right] & \beta \\ \hline \end{array} \qquad \frac{\partial^2 H}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \begin{array}{|c|c|c} \multicolumn{1}{c}{\alpha} & \multicolumn{1}{c}{\beta} & \\ \hline 0 & -\dfrac{1}{\beta}\left[1 - z\right] & \alpha \\ \hline & \dfrac{\alpha}{\beta^2}\left[1 - z^2\right] & \beta \\ \hline \end{array}$$

### A.2.5 maxlo

The density and survival are given for $x > 0$ by

$$f(x) = \frac{\alpha}{\beta}\left[1 - \frac{x}{\beta}\right]^{\alpha-1} \qquad S(x) = \left[1 - \frac{x}{\beta}\right]^{\alpha}$$

Then with $z := 1/[1 - x/\beta]$

$$\log f(x) = \log\alpha - \log\beta - (\alpha - 1)\log z$$

and

$$\frac{\partial \log f}{\partial \boldsymbol{\theta}} = \begin{array}{|c|c} \hline \dfrac{1}{\alpha} - \log z & \alpha \\ \hline \dfrac{1}{\beta}\,[-\alpha + (\alpha - 1)z] & \beta \\ \hline \end{array} \qquad \frac{\partial^2 \log f}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} = \begin{array}{|c|c|c} \multicolumn{1}{c}{\alpha} & \multicolumn{1}{c}{\beta} & \\ \hline -\dfrac{1}{\alpha^2} & -\dfrac{1}{\beta}\,[1 - z] & \alpha \\ \hline & \dfrac{1}{\beta^2}\,\left[1 + (\alpha - 1)(1 - z^2)\right] & \beta \\ \hline \end{array}$$

The derivatives of the cumulative hazard $H$ are

$$\frac{\partial H}{\partial \boldsymbol{\theta}} = \begin{array}{|c|c} \hline \log z & \alpha \\ \hline \dfrac{\alpha}{\beta}\,[1 - z] & \beta \\ \hline \end{array} \qquad \frac{\partial^2 H}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} = \begin{array}{|c|c|c} \multicolumn{1}{c}{\alpha} & \multicolumn{1}{c}{\beta} & \\ \hline 0 & \dfrac{1}{\beta}\,[1 - z] & \alpha \\ \hline & -\dfrac{\alpha}{\beta^2}\,[1 - z^2] & \beta \\ \hline \end{array}$$

### A.2.6 Mixture of exponentials

The density is given by

$$f(x) = \alpha_1\lambda_1\, e^{-\lambda_1 x} + (1 - \alpha_1)\lambda_2\, e^{-\lambda_2 x} \qquad S(x) = \alpha_1\, e^{-\lambda_1 x} + (1 - \alpha_1)\, e^{-\lambda_2 x}$$

Then

$$\frac{\partial f}{\partial \boldsymbol{\theta}} = \begin{array}{|c|c} \hline \left[\lambda_1\, e^{-\lambda_1 x} - \lambda_2\, e^{-\lambda_2 x}\right] & \alpha_1 \\ \hline \alpha_1\,[1 - \lambda_1 x]\, e^{-\lambda_1 x} & \lambda_1 \\ \hline [1 - \alpha_2][1 - \lambda_2 x]\, e^{-\lambda_2 x} & \lambda_2 \\ \hline \end{array} \qquad \frac{\partial^2 f}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} = \begin{array}{|c|c|c|c} \multicolumn{1}{c}{\alpha_1} & \multicolumn{1}{c}{\lambda_1} & \multicolumn{1}{c}{\lambda_2} & \\ \hline 0 & [1 - \lambda_1 x]\, e^{-\lambda_1 x} & -[1 - \lambda_2 x]\, e^{-\lambda_2 x} & \alpha_1 \\ \hline & -\alpha_1\, x\,[2 - \lambda_1 x]\, e^{-\lambda_1 x} & 0 & \lambda_1 \\ \hline & & -[1 - \alpha_1]\, x\,[2 - \lambda_2 x]\, e^{-\lambda_2 x} & \lambda_2 \\ \hline \end{array}$$

For the survival, we have

$$\frac{\partial S}{\partial \boldsymbol{\theta}} = \begin{array}{|c|c} \hline \left[e^{-\lambda_1 x} - e^{-\lambda_2 x}\right] & \alpha_1 \\ \hline -\alpha_1\, x\, e^{-\lambda_1 x} & \lambda_1 \\ \hline -[1 - \alpha_1]\, x\, e^{-\lambda_2 x} & \lambda_2 \\ \hline \end{array} \qquad \frac{\partial^2 S}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} = \begin{array}{|c|c|c|c} \multicolumn{1}{c}{\alpha_1} & \multicolumn{1}{c}{\lambda_1} & \multicolumn{1}{c}{\lambda_2} & \\ \hline 0 & -x\, e^{-\lambda_1 x} & x\, e^{-\lambda_2 x} & \alpha_1 \\ \hline & \alpha_1\, x^2\, e^{-\lambda_1 x} & 0 & \lambda_1 \\ \hline & & [1 - \alpha_1]\, x^2\, e^{-\lambda_2 x} & \lambda_2 \\ \hline \end{array}$$

## A.3 Jacobian of the "POT to GEV" transform

### A.3.1 General case

The Jacobian matrix of the transformation

$$\boldsymbol{\theta} := [\lambda,\, \mu,\, \sigma,\, \xi]^\top \mapsto \boldsymbol{\theta}^\star := [\mu^\star,\, \sigma^\star,\, \xi^\star]^\top$$

can be used to compute an approximate covariance matrix for the GEV parameter vector. Using

$$C := [\lambda w]^\xi, \quad B := \frac{[\lambda w]^\xi - 1}{\xi},$$

we get the followinf formula for $\xi \neq 0$

$$\frac{\partial \boldsymbol{\theta}^\star}{\partial \boldsymbol{\theta}} = \begin{array}{|c||c|c|c|c|} \hline & \lambda & \mu & \sigma & \xi \\ \hline \mu^\star & \frac{1}{\lambda}\, C\sigma & 1 & B & \frac{\sigma}{\xi}\left\{C \log(\lambda w) - B\right\} \\ \hline \sigma^\star & \frac{\xi}{\lambda}\, C\, \sigma & 0 & C & C\, \sigma \log(\lambda w) \\ \hline \xi^\star & 0 & 0 & 0 & 1 \\ \hline \end{array} \tag{A.1}$$

For $\xi = 0$ we have $C = 1$ and the quantity $B$ must be replaced by $\log(\lambda w)$ which is the limit for $\xi \to 0$. Moreover the derivative of the GEV location $\mu^\star$ w.r.t the POT shape $\xi$ must be replaced by its limit for $\xi \to 0$

$$\frac{\partial \mu^\star}{\partial \xi}(\lambda,\, \sigma,\, 0) = \frac{\sigma}{2} \log^2(\lambda w),$$

which can be proved by using a Taylor expansion of $[\lambda w]^\xi = \exp\{\xi \log[\lambda w]\}$ at the second order near $\xi = 0$.

### A.3.2 Gumbel-Exponental case $\xi = 0$

When $\xi = 0$ the maxima follow a Gumbel distribution

$$\frac{\partial \boldsymbol{\theta}^\star}{\partial \boldsymbol{\theta}} = \begin{array}{|c||c|c|c|} \hline & \lambda & \mu & \sigma \\ \hline \mu^\star & \frac{\sigma}{\lambda} & 1 & \log(\lambda w) \\ \hline \sigma^\star & 0 & 0 & 1 \\ \hline \end{array} \tag{A.2}$$

## A.4 The "GEV to POT" transform and its Jacobian

### A.4.1 Location $\mu$ known

#### General GEV-GPD case

Using simple algebra, we find from the first relation in (3.4)

$$(\lambda w)^{-\xi^\star} = 1 + \xi^\star \, \frac{\mu - \mu^\star}{\sigma^\star}.$$

Since $\lambda w > 0$, the right hand side must be $> 0$. This leads to two constraints when $\xi \neq 0$

$$\begin{cases} \mu < \mu^\star - \sigma^\star/\xi^\star & \text{if } \xi^\star < 0, \\ \mu > \mu^\star - \sigma^\star/\xi^\star & \text{if } \xi^\star > 0. \end{cases} \tag{A.3}$$

The parameters $\lambda$, $\sigma$ and $\xi$ of the GPD are

$$\left\{ \begin{array}{lcl} \lambda & = & \frac{1}{w} C^{-1/\xi^\star} \\ \sigma & = & C\,\sigma^\star \\ \xi & = & \xi^\star \end{array} \right. \qquad \text{with} \qquad \left\{ \begin{array}{lcl} z & := & [\mu - \mu^\star]/\sigma^\star \\ C & := & [1 + \xi^\star z]_+ \,. \end{array} \right. \tag{A.4}$$

The gradient of $C$ is given by

$$\partial_{\mu^\star} C = -\xi^\star/\sigma^\star \quad \partial_{\sigma^\star} C = -\xi^\star z/\sigma^\star \quad \partial_{\xi^\star} C = z. \tag{A.5}$$

Using $D := C^{-1/\xi^\star - 1}$, the Jacobian matrix is given by

| | $\mu^\star$ | $\sigma^\star$ | $\xi^\star$ |
|---|---|---|---|
| $\lambda$ | $\frac{\lambda}{\sigma^\star C}$ | $z\,\frac{\lambda}{\sigma^\star C}$ | $\lambda\left[\log C - z\,\frac{\xi^\star}{C}\right]/\xi^{\star 2}$ |
| $\sigma$ | $-\xi^\star$ | 1 | $z\,\sigma^\star$ |
| $\xi$ | 0 | 0 | 1 |

(A.6)

where $\lambda$ at the right hand side is computed using (A.4) above.

For $\xi^\star = 0$ we have $C = 1$, and we must use $\lambda = e^{-z}/w$. The derivative of the rate $\lambda$ w.r.t the shape $\xi^\star$ must be replaced by

$$\frac{\partial \lambda}{\partial \xi^\star} = \lambda\, z^2/2,$$

which can be proved to be the limit of the formula above for $\xi^\star \to 0$.

**Gumbel-Exponential case $\xi = 0$**

In the exponential case $\xi^\star = 0$, the parameter $\lambda$ and $\sigma$ of the renewal model are found by

$$\left\{ \begin{array}{lcl} \lambda & = & \frac{1}{w} \exp\left\{-\frac{\mu-\mu^\star}{\sigma^\star}\right\} \\ \sigma & = & \sigma^\star \end{array} \right. \tag{A.7}$$

Using $E := \exp\{-[\mu - \mu^\star]/\sigma^\star\}$, the Jacobian matrix is

| | $\mu^\star$ | $\sigma^\star$ |
|---|---|---|
| $\lambda$ | $\frac{1}{w\sigma^\star}\,E$ | $\frac{\mu-\mu^\star}{w\sigma^{\star 2}}\,E$ |
| $\sigma$ | 0 | 1 |

(A.8)

**GEV-maxlo case $\xi < 0$**

**GEV-Lomax case $\xi > 0$**

The parameters $\lambda$, $\sigma$ and $\xi$ of the renewal model with Lomax excesses are found by using the general case $\xi \neq 0$ and then transforming the GPD vector parameters $[\sigma,\, \xi]^\top$

$$\left\{\begin{array}{lcl} \lambda & = & \frac{1}{w}\, C^{-1/\xi^\star} \\ \beta & = & C\,\sigma^\star \\ \alpha & = & 1/\xi^\star \end{array}\right. \qquad \text{with} \qquad C := \left[1 + \xi^\star\, \frac{\mu - \mu^\star}{\sigma^\star}\right]_+ . \tag{A.9}$$

### A.4.2 Rate $\lambda$ known

**General case $\xi \neq 0$**

Note that if $\mu \geqslant \mu^\star$, the right hand side is $S_{\texttt{GPD}}(\mu;\, \mu^\star,\, \sigma^\star,\, \xi^\star)$.

$$\left\{\begin{array}{lcl} \mu & = & \mu^\star - \frac{1-[\lambda w]^{-\xi^\star}}{\xi^\star}\, \sigma^\star \\ \sigma & = & [\lambda w]^{-\xi^\star}\, \sigma^\star \\ \xi & = & \xi^\star \end{array}\right. \tag{A.10}$$

The Jacobian matrix is

| | $\mu^\star$ | $\sigma^\star$ | $\xi^\star$ |
|---|---|---|---|
| $\mu$ | 1 | $-\frac{1-[\lambda w]^{-\xi^\star}}{\xi^\star}$ | $\left\{\frac{1-[\lambda w]^{-\xi^\star}}{\xi^\star} - \log(\lambda w)\,[\lambda w]^{-\xi^\star}\right\}\,\frac{\sigma^\star}{\xi^\star}$ |
| $\sigma$ | 0 | $[\lambda w]^{-\xi^\star}$ | $-\log(\lambda w)\,[\lambda w]^{-\xi^\star}\,\sigma^\star$ |
| $\xi$ | 0 | 0 | 1 |

(A.11)

The adaptations required for $\xi = 0$ are the same as in section A.3.1 p. 78.

**Gumbel-exponential case $\xi = 0$**

In the exponential case $\xi^\star = 0$

$$\left\{\begin{array}{lcl} \mu & = & \mu^\star - \log(\lambda w)\,\sigma^\star \\ \sigma & = & \sigma^\star \end{array}\right. \tag{A.12}$$

The Jacobian matrix is

| | $\mu^\star$ | $\sigma^\star$ |
|---|---|---|
| $\mu$ | 1 | $-\log(\lambda w)$ |
| $\sigma$ | 0 | 1 |

(A.13)

**GEV-maxlo case $\xi < 0$**

The parameters $\mu$, $\sigma$ and $\xi$ of the renewal model with maxlo excesses are found by using the general case $\xi \neq 0$ and then transforming the GPD vector of parameters $[\sigma,\, \xi]^\top$

$$\begin{cases} \mu &= \mu^\star - \frac{1-[\lambda w]^{-\xi^\star}}{\xi^\star}\,\sigma^\star \\ \beta &= -\frac{[\lambda w]^{-\xi^\star}}{\xi^\star}\,\sigma^\star \\ \alpha &= -1/\xi^\star \end{cases} \tag{A.14}$$

**GEV-Lomax case $\xi > 0$**

The parameters $\mu$, $\sigma$ and $\xi$ of the renewal model with Lomax excesses are found by using the general case $\xi \neq 0$ and then transforming the GPD vector of parameters $[\sigma,\, \xi]^\top$

$$\begin{cases} \mu &= \mu^\star - \frac{1-[\lambda w]^{-\xi^\star}}{\xi^\star}\,\sigma^\star \\ \beta &= \frac{[\lambda w]^{-\xi^\star}}{\xi^\star}\,\sigma^\star \\ \alpha &= 1/\xi^\star \end{cases} \tag{A.15}$$

## A.5 Bounds for $\widehat{\beta}$ in the Lomax case

### A.5.1 Goal

This rather technical section refers to the section 2.4 devoted to the Lomax distribution. The goal is to explain how one can compute an interval $(\beta_{\mathtt{L}},\, \beta_{\mathtt{U}})$ granted to contain the maximum likelihood estimate $\widehat{\beta}$ of the scale parameter $\beta$.

We use the notation $M_k$ for the empirical moment of order $k$, and it will be convenient to use $M_1$ in place of $\bar{X}$. We assume that the empirical coefficient of variation is greater than unity, i.e. $\mathrm{CV} > 1$, or equivalently that $M_2 > 2M_1^2$.

### A.5.2 Upper bound

We want to find a value $\beta_{\mathtt{U}}$ such that $\ell_{\mathtt{c}}'(\beta) < 0$ for all $\beta \geqslant \beta_{\mathtt{U}}$. Using (2.6), it will be enough to have

$$-[1+A]A' - A\beta^{-1} < 0$$

for every $\beta > \beta_{\mathtt{U}}$. Observe that

$$A'(\beta) = -\beta^{-1} + \beta^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+X_i/\beta}.$$

The following inequality is classical

$$1 - u + u^2 - u^3 \leqslant \frac{1}{1+u} \qquad (0 \leqslant u \leqslant 1).$$

Assume that $\beta > \max\{X_i\}$; using the previous inequality for $u := X_i/\beta$ and averaging over $i$, we get

$$1 - M_1\,\beta^{-1} + M_2\,\beta^{-2} - M_3\,\beta^{-3} \leqslant \frac{1}{n}\sum_{i=1}^{n}\frac{1}{1 + X_i/\beta}$$

and thus

$$-\,A'(\beta) \leqslant M_1\,\beta^{-2} - M_2\,\beta^{-3} + M_3\,\beta^{-4}. \tag{A.16}$$

Now the following inequalities are easily stated

$$u - \frac{1}{2}u^2 \leqslant \log(1+u) \leqslant u \qquad (0 \leqslant u \leqslant 1).$$

Let us still assume that $\beta > \max\{X_i\}$. Using the last inequalities for $u := X_i/\beta$ and averaging over $i$, we get

$$M_1\beta^{-1} - \frac{M_2}{2}\,\beta^{-2} \leqslant A(\beta) \leqslant M_1\,\beta^{-1}. \tag{A.17}$$

Since $A' \leqslant 0$, using (A.16) and the second inequality in (A.17) we get

$$-\,[1 + A]A' \leqslant \left[1 + M_1\,\beta^{-1}\right]\left[M_1\,\beta^{-2} - M_2\,\beta^{-3} + M_3\,\beta^{-4}\right]. \tag{A.18}$$

Using the first inequality in (A.17)

$$-\,A\,\beta^{-1} \leqslant -M_1\,\beta^{-2} + \frac{M_2}{2}\,\beta^{-3}. \tag{A.19}$$

Now summing (A.18) and (A.19) and rearranging the right hand side, we get

$$-[1 + A]\,A' - A\,\beta^{-1} \leqslant [M_1^2 - M_2/2]\,\beta^{-3} + [M_3 - M_1 M_2]\,\beta^{-4} + M_1 M_3\,\beta^{-5}.$$

Multiplying by $\beta^5$, we see that the right hand side will be negative as soon as

$$[M_1^2 - M_2/2]\,\beta^2 + [M_3 - M_1 M_2]\,\beta + M_1 M_3 < 0. \tag{A.20}$$

Since $M_2 - 2M_1^2 > 0$, the condition is true for $\beta$ large enough: (A.17) holds as soon as $\beta$ is greater than the largest root of the quadratic polynomial at the left hand side. So we may take $\beta_{\mathtt{U}}$ defined as

$$\beta_{\mathtt{U}} := \max\left\{\,\max\{X_i\},\ \beta_2\right\}$$

where $\beta_2$ is the largest root of the polynomial at the left hand side of (A.20).

### A.5.3 Lower bound

Although a *very small* value $\beta_{\mathtt{L}} > 0$ can theoretically be used as lower bound for $\beta$, this choice can be impracticable since the derivative of the concentrated likelihood $\ell'_{\mathtt{c}}(\beta)$ may be impossible to evaluate for small values. A better choice for $\beta_{\mathtt{L}}$ is now derived.

Mirroring the previous derivation, it will be enough to find $\beta_{\mathtt{L}}$ such that

$$\ell'_{\mathtt{c}} = -[1 + A]A' - A\beta^{-1} > 0$$

for every $\beta$ with $0 < \beta \leqslant \beta_{\mathrm{L}}$. Note that

$$A(\beta) = \frac{1}{n} \sum_i \log(\beta + X_i) - \log \beta.$$

Assuming that $0 < \beta < \min\{X_i\}$, we have for all $i$

$$\log(X_i) \leqslant \log(\beta + X_i) \leqslant \log(2X_i)$$

hence, denoting by $\bar{L}$ the mean of log observations $L_i := \log X_i$

$$\bar{L} - \log \beta \leqslant A(\beta) \leqslant \bar{L} + \log 2 - \log \beta. \tag{A.21}$$

The minus derivative $-A'(\beta)$ can be written as

$$-A'(\beta) = \beta^{-1} \frac{1}{n} \sum_i \frac{1}{1 + \beta/X_i}.$$

Using the inequality $1/(1+u) \geqslant 1 - u$ for $0 < u < 1$ with $u := \beta/X_i$ and averaging over $i$, we find

$$-A'(\beta) \geqslant \beta^{-1} [1 - M_{-1}\beta] = \beta^{-1} - M_{-1} \tag{A.22}$$

where $M_{-1}$ is the mean of the inverse observations $X_i^{-1}$. Since we assumed that $\beta < \min_i X_i$, and hence that $\beta^{-1} > X_i^{-1}$ for all $i$, we have $\beta^{-1} - M_{-1} \geqslant 0$. Now using the left inequality in (A.21) and (A.22)

$$-[1 + A]\, A' \geqslant \left[\bar{L} + 1 - \log \beta\right] \left[\beta^{-1} - M_{-1}\right],$$

and with the right inequality in (A.21)

$$-A\beta^{-1} \geqslant \left[-\bar{L} - \log 2 + \log \beta\right] \beta^{-1}.$$

Summing the last two inequalities and rearranging, we get

$$-[1 + A]\, A' - A\beta^{-1} \geqslant -\left[\bar{L} + 1\right] M_{-1} + M_{-1} \log \beta + [1 - \log 2]\, \beta^{-1}$$

or

$$-[1 + A]\, A' - A\beta^{-1} \geqslant \beta^{-1} \left\{-\left[\bar{L} + 1\right] M_{-1}\beta + M_{-1}\beta \log \beta + [1 - \log 2]\right\}.$$

Note that the first two terms of the sum between the curly brackets both tend to 0 for small $\beta$. To ensure that the right hand side is positive, it will be enough that the two following conditions hold

$$\left[\bar{L} + 1\right] M_{-1}\beta \leqslant \kappa, \qquad -M_{-1}\beta \log \beta \leqslant \kappa.$$

where $\kappa := [1 - \log 2]/2 \approx 0.153$. Note that $\bar{L} \leqslant \log(\bar{X})$ by Jensen inequality, and that $M_{-1}^{-1} \geqslant \min\{X_i\}$ (quite obviously). Temporarily assuming that $\bar{X} = 1$, hence that $\bar{L} \leqslant 0$, the first inequality can be successively replaced by the two stronger yet simpler conditions: $M_{-1}\beta \leqslant \kappa$, then: $\beta \leqslant \kappa \min\{X_i\}$. Further assuming that $\beta \leqslant e^{-1} \approx 0.368$, we have $-\beta \log \beta \leqslant \beta$ and the second sufficient condition can be replaced by the stronger $\beta \leqslant \kappa \min\{X_i\}$. Thus we can take

$$\beta_{\mathrm{L}} := 0.15 \times \min\{X_i\}.$$

The condition $\beta \leqslant e^{-1}$ is necessarily fulfilled when $\bar{X} = 1$, since then $\min\{X_i\} \leqslant 1$.

At a first glance, the previous derivation is only valid provided that the data are normalised to have unit mean, i.e. provided that the $X_i$ are replaced by $X_i^\star := X_i/\bar{X}$. However, the inequality $0.15 \times \min\{X_i\} \leqslant \widehat{\beta}$ holds as well for unnormalised data. Indeed, the ML estimate of the scale $\beta^\star$ for the normalised data is simply $\widehat{\beta}^\star = \widehat{\beta}/\bar{X}$, and the lower bound on the ML estimate remains after de-normalisation.

## A.6 Bounds for $\widehat{\beta}$ in the Maxlo case

### A.6.1 Goal

This section refers to the section 2.5 devoted to the Maxlo distribution. As it was done for the Lomax distribution, the goal is to explain how one can compute an interval $(\beta_{\mathtt{L}},\, \beta_{\mathtt{U}})$ granted to contain the maximum likelihood estimate $\widehat{\beta}$ of the scale parameter $\beta$.

We use the notation $M_k$ for the empirical moment of order $k$, and we use $M_1$ in place of $\bar{X}$. We assume that the empirical coefficient of variation is smaller than unity, i.e. CV $< 1$, or equivalently that $M_2 < 2M_1^2$.

### A.6.2 Upper bound

We want to find a value $\beta^\star$ such that the function $\ell_{\mathtt{c}}(\beta)$ is decreasing for $\beta > \beta^\star$. In other words, we want that $\ell_{\mathtt{c}}'(\beta) < 0$ for every $\beta > \beta^\star$ where the concentrated log-likelihood $\ell_{\mathtt{c}}(\beta)$ and its derivative were given in (2.10). We want

$$-[1-A]\,A' - A\beta^{-1} < 0$$

for every $\beta > \beta^\star$ where $A(\beta)$ was defined in (2.9). We will assume that $\beta > \max\{X_i\}$.

Observe that

$$A'(\beta) = \beta^{-1} - \beta^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1 - X_i/\beta}$$

We will use the following inequality

$$\frac{1}{1-u} \leqslant 1 + u + u^2 + 2u^3 \qquad \text{for } 0 \leqslant u < 1/2,$$

which is easily proved. Using this with $u := X_i/\beta$ and averaging over $i$, we get

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1 - X_i/\beta} \leqslant 1 + M_1\,\beta^{-1} + M_2\,\beta^{-2} + 2M_3\,\beta^{-3},$$

hence

$$-\,A'(\beta) \leqslant M_1\,\beta^{-2} + M_2\,\beta^{-3} + 2M_3\,\beta^{-4}. \tag{A.23}$$

Now for $0 \leqslant u < 1$, we have

$$u \leqslant -\log(1-u), \qquad u + \frac{1}{2}\,u^2 + \frac{1}{3}\,u^3 \leqslant -\log(1-u).$$

Using these inequalities for $u := X_i/\beta$ and averaging over $i$, we find

$$M_1\,\beta^{-1} \leqslant A, \qquad M_1\,\beta^{-1} + \frac{M_2}{2}\,\beta^{-2} + \frac{M_3}{3}\,\beta^{-3} \leqslant A. \tag{A.24}$$

Remind that $-A' \geqslant 0$. Using (A.23) and the first inequality in (A.24)

$$-\,[1-A]\,A' \leqslant \left[1 - M_1\beta^{-1}\right]\left[M_1\,\beta^{-2} + M_2\,\beta^{-3} + 2M_3\,\beta^{-4}\right]. \tag{A.25}$$

Using the second inequality in (A.24)

$$-A\beta^{-1} \leqslant -M_1\,\beta^{-2} - \frac{M_2}{2}\,\beta^{-3} - \frac{M_3}{3}\,\beta^{-4}. \tag{A.26}$$

Now summing (A.25) and (A.26) and rearranging the right hand side, we get

$$-[1-A]\,A' - A\,\beta^{-1} \leqslant \left[M_2/2 - M_1^2\right]\,\beta^{-3} + \left[5M_3/3 - M_1M_2\right]\,\beta^{-4} - 2M_1M_3\,\beta^{-5}.$$

Multiplying by $6\beta^5$, we see that the right hand side will be negative as soon as

$$3\left[M_2 - 2M_1^2\right]\beta^2 + 2\left[5M_3 - 3M_1M_2\right]\beta - 12M_1M_3 < 0 \tag{A.27}$$

Since $M_2 - 2M_1^2 < 0$, the condition is true for $\beta$ large enough: (A.24) holds as soon as $\beta$ is greater than the largest root of the quadratic polynomial at the left hand side. Recall that we required above that $X_i/\beta < 1/2$ for all $i$, so we may take $\beta^\star$ defined as

$$\beta^\star := \max\left\{2 \times \max\{X_i\},\, \beta_2\right\}$$

where $\beta_2$ is the largest root of the polynomial at the left hand side of (A.27).

### A.6.3 Lower bound

We now assume that $\alpha_{\mathrm{L}} > 1$ is given, and we prove that the concentrated log-likelihood $\ell_{\mathtt{c}}^\star(\beta)$ with the constraint $\alpha \geqslant \alpha_{\mathrm{L}}$ is increasing on the interval $(X_{(n)},\, \beta_{\mathrm{L}}]$ where $\beta_{\mathrm{L}}$ is given in (2.14).

We use the expression in (2.13) giving the log-likelihood for $X_{(n)} < \beta < \beta^\star$, that is

$$\ell_{\mathtt{c}}^\star(\beta) = n\log\alpha_{\mathrm{L}} - n\log\beta + (\alpha_{\mathrm{L}} - 1)\sum_i \log(1 - X_i/\beta).$$

Writing $\log(1 - X_i/\beta) = \log(\beta - X_i) - \log\beta$, we get after rearranging

$$\ell_{\mathtt{c}}^\star(\beta) = n\log\alpha_{\mathrm{L}} - n\,\alpha_{\mathrm{L}}\log\beta + (\alpha_{\mathrm{L}} - 1)\sum_i \log(\beta - X_i).$$

The derivative w.r.t. $\beta$ is

$$\ell_{\mathtt{c}}^{\star\prime}(\beta) = \frac{-n\,\alpha_{\mathrm{L}}}{\beta} + (\alpha_{\mathrm{L}} - 1)\sum_i \frac{1}{\beta - X_i}.$$

Ignoring all terms but one in the sum at the right hand side,

$$\ell_{\mathtt{c}}^{\star\prime}(\beta) \geqslant \frac{-n\,\alpha_{\mathrm{L}}}{\beta} + \frac{\alpha_{\mathrm{L}} - 1}{\beta - X_{(n)}} = \frac{n\alpha_{\mathrm{L}}X_{(n)} - (n\alpha_{\mathrm{L}} - \alpha_{\mathrm{L}} + 1)\beta}{\beta(\beta - X_{(n)})}.$$

The derivative is thus positive as soon as

$$\beta < \frac{n\alpha_{\mathrm{L}}}{n\alpha_{\mathrm{L}} - \alpha_{\mathrm{L}} + 1}\,X_{(n)} = \frac{1}{1 - (\alpha_{\mathrm{L}} - 1)/(n\alpha_{\mathrm{L}})}\,X_{(n)}$$

as claimed.

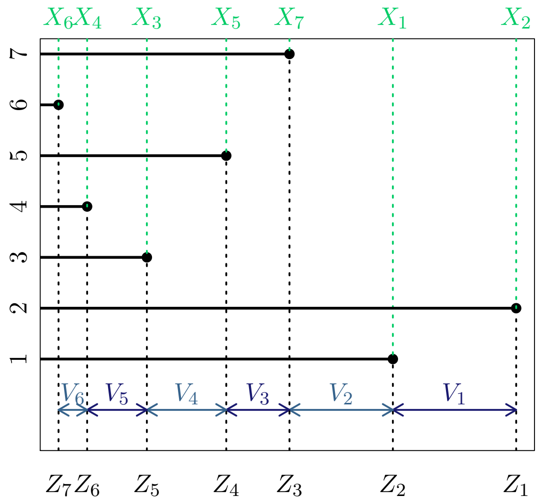



Figure A.1: Sample $X_i$, order statistics $Z_i$ and spacings $V_i$.

## A.7 Spacings of the exponential distribution

The following result is known as a the *Rényi's representation* of a sample from the exponential. It plays a major role in many problems involving the exponential distribution. See e.g. Embrechts *et al.* (1996, chap. 4) for more details.

Let $X_1, X_2, \ldots, X_n$ be a sample of size $n$ of the exponential distribution with rate $\nu > 0$. Consider the order statistics in the decreasing order

$$Z_n < Z_{n-1} < \cdots < Z_2 < Z_1$$

and the spacings $V_i = Z_i - Z_{i+1}$ for $i = 1, 2, \ldots, n$. We define an extra spacing $V_n$ using $Z_{n+1} := 0$ leading to $V_n := Z_n$. The distribution of $Z_n = V_n$ is easily found. Indeed, for $z \geqslant 0$ we have

$$\Pr\{Z_n > z\} = \Pr\{\text{all } X_i > z\} = \prod_{i=1}^{n} \Pr\{X_i > z\} = \prod_{i=1}^{n} \exp(-\nu z) = \exp(-n\nu z)$$

so $Z_n$ has an exponential distribution with rate $n\nu$. Note that the exponential distribution is *min-stable*[1].

**Theorem A.1.** *Let $X_i$ be a sample of the exponential distribution* $\texttt{Exp}(\nu)$ *with rate $\nu > 0$. Then the $n$ spacings $V_i$ are independent and*

$$V_i \underset{\text{dist}}{=} \frac{1}{i} E_i \qquad (1 \leqslant i \leqslant n)$$

*where $E_i$ is a sample of size $n$ from* $\texttt{Exp}(\nu)$. *In other words, the $n$ normalised spacings $V_i^{\star} := i \times V_i$ form a sample of* $\texttt{Exp}(\nu)$.

[1] As also is the Weibull.

Rather than a formal proof, consider an intuitive derivation based on the memorylessness property. Consider $n$ independent items with random lifetime having the same exponential distribution. Then starting from an initial time $t = 0$, $Z_n$ is the time at which the first failure occurs, $Z_{n-1}$ is the time of the second failure, and so on. We saw that $Z_n$ follows an exponential distribution with rate $n\nu$. When the first failure occurs at time $Z_n$, there remains $n-1$ items alive, which can be considered as new ones due to the memorylessness property. We are thus in the same situation as at time $t = 0$ but with $n-1$ items, the lifetime of which still having exponential distribution with rate $\nu$. The time $Z_{n-1} - Z_n$ to the next failure is therefore exponentially distributed with rate $(n-1)\nu$ and is independent from $Z_n$. Similarly $Z_{n-2} - Z_{n-1}$ is found to be independent for $Z_n$ and $Z_{n-1}$ and to follow an exponential distribution with rate $(n-2)\nu$, and so on.

Assume that $\nu = 1$. Since $Z_1 = V_1 + V_2 + \cdots + V_n$, we get

$$\mathbb{E}(Z_1) = \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n}$$

and for large $n$, we have $\mathbb{E}(Z_1) \approx \log n + \gamma$ where $\gamma \approx 0.577$ is the Euler-Mascheroni constant. Similarily

$$\text{Var}(Z_1) = \frac{1}{1^2} + \frac{1}{2^2} + \cdots + \frac{1}{n^2}$$

so $\text{Var}(Z_1) \approx \pi^2/6$ lor large $n$.

## A.8 Proof of theorem 3.3

To simplify the proof we will use two lemmas.

**Lemma 1.** *Assume that the rv.s $M_1$, $M_2$, ..., $M_B$ are i.i.d. with distribution* $\texttt{GEV}(\mu^\star, \sigma^\star, \xi^\star)$ *with $\xi^\star \neq 0$, and that $w > 0$ is fixed. Let $\widetilde{\mu}_B := \min_b M_b$.*

*Let the three parameters $\widetilde{\lambda}_B > 0$, $\widetilde{\sigma}_B > 0$ and $\widetilde{\xi}_B$ be the POT parameters $\lambda$, $\sigma$ and $\xi$ as (uniquely) determined from the GEV parameters $\mu^\star$, $\sigma^\star$ and $\xi^\star$ by using the "BM to POT transform", i.e. by solving the equations (3.4) where $\mu$ is replaced by $\widetilde{\mu}_B$. So, the three "parameters" $\widetilde{\mu}_B$, $\widetilde{\lambda}_B$, and $\widetilde{\sigma}_B$ are random variables functions of $\widetilde{\mu}_B$, while $\widetilde{\xi}_B = \xi^\star$.*

1. *The distribution of $\widetilde{\lambda}_B w$ is that of the maximum of $B$ i.i.d. random variables having a standard exponential distribution, hence*

$$w\,\mathbb{E}[\widetilde{\lambda}_B] \underset{B\to\infty}{\sim} \log B + \gamma, \qquad w^2\,\text{Var}[\widetilde{\lambda}_B] \underset{B\to\infty}{\sim} \frac{\pi^2}{6},$$

*and the random variable $\widetilde{\lambda}_B w - \log B$ converges in distribution to a standard Gumbel.*

2. *For large $B$ we have*

$$\frac{\log[\widetilde{\lambda}_B w]}{\log B} \xrightarrow{\text{Pr.}} 0, \qquad \frac{\widetilde{\lambda}_B w}{\log B} \xrightarrow{\text{Pr.}} 1. \tag{A.28}$$

*Proof.* 1. By construction, $\widetilde{\mu}_B$ is in the interior of the support $\mathcal{I}_{\texttt{GEV}}(\boldsymbol{\theta}^\star)$ with probability one, so the determination of the POT parameter vector $\boldsymbol{\theta}_{(-\mu)}$ is possible. As in the proof of result 3.2 we have

$$\widetilde{\lambda}_B w = -\log F_{\text{GEV}}(\widetilde{\mu}_B;\, \boldsymbol{\theta}^\star) = -\log F_M(\widetilde{\mu}_B)$$

where $F_M$ is the GEV distribution function of the $M_b$. But since $\widetilde{\mu}_B$ is the minimum of the $B$ random variables $M_b$, we see that $\widetilde{\lambda}_B w$ is the maximum of the $B$ rv.s $E_b := -\log F_M(M_b)$ which are independent and have the same standard exponential distribution. The expectation and variance of $\widetilde{\lambda}_B w$ are given by Rényi's representation A.7.

2. We can prove the two convergence results *in distribution* since the limit is constant (non-random) in both case. Consider the first statement in (A.28). For simplicity let $U_B := -\log[\widetilde{\lambda}_B w]$ and $V_B := U_B/\log B$. Recall that if $E$ has a standard exponential distribution, $-\log E$ follows the standard Gumbel distribution. Hence we know from the first part of this lemma that $U_B$ has the same distribution as the minimum of $B$ i.i.d. random variables with standard Gumbel distribution. From this, we easily find the log-survival of $V_B$

$$\log S_{V_B}(v) = B\, \log\left[1 - \exp(-B^{-v})\right], \qquad (-\infty < v < \infty).$$

We want to prove that $S_{V_B}(v)$ tends to 1 for $v < 0$ and to 0 for $v > 0$. Equivalently, we want

$$\log S_{V_B}(v) \to 0 \quad (v < 0), \qquad \text{and} \qquad \log S_{V_B}(v) \to -\infty \quad (v > 0).$$

These two assertions are easily verified using the closed form above.

The proof for the second statement is similar: the distribution function of $Z_B := \widehat{\lambda}_B w/\log B$ is found to be given by $F_{Z_B}(z) = [1 - B^{-z}]^B$ for $z > 0$, and it easily checked that this tend to 0 for $z < 1$ and to 1 for $z > 1$.

□

In the following lemma and the subsequent proofs, we will use $Z_B \xrightarrow{\text{Pr.}} \infty$ only for a sequence of positive r.vs $Z_B$. This is equivalent to $1/Z_B \xrightarrow{\text{Pr.}} 0$.

**Lemma 2.** *Assume that the rv.s $M_1$, $M_2$, ..., $M_B$ are i.i.d. with distribution $\texttt{GEV}(\mu^\star,\, \sigma^\star,\, \xi^\star)$ with $\xi^\star \neq 0$, $\xi^\star > -1/2$, and that $w > 0$ is fixed. Let $\widetilde{\mu}_B := \min_b M_b$, and let the three POT parameters $\widetilde{\lambda}_B > 0$, $\widetilde{\sigma}_B > 0$ and $\widetilde{\xi}_B$ be (uniquely) determined by the "BM to POT transform", as in lemma 1 above. Let $\widehat{\mu}^\star_B$, $\widehat{\sigma}^\star_B$ and $\widehat{\xi}^\star_B$ be the ML estimates of the GEV parameters, assumed to correspond to a unique maximum.*

1. *Let $I_B := [\widetilde{\lambda}_B w]^{[\widehat{\xi}^\star_B - \xi^\star]}$. We have $I_B \xrightarrow{\text{Pr.}} 1$ for large $B$, and moreover there exists $\beta > 0$ such that $B^\beta\, |I_B - 1| \xrightarrow{\text{Pr.}} 0$*

2. *For any $\alpha > 0$ fixed, we have $B^\alpha\, [\widetilde{\lambda}_B w]^{-\widehat{\xi}^\star_B} \xrightarrow{\text{Pr.}} \infty$ for large $B$.*

*Proof.* 1. We have

$$B^\beta \log I_B = \frac{\log B}{B^{\rho-\beta}} \times \left\{B^\rho\, [\widehat{\xi}^\star_B - \xi^\star]\right\} \times \left\{\frac{\log[\widetilde{\lambda}_B w]}{\log B}\right\}$$

with $\rho > 0$ arbitrary. We know that $\sqrt{B}\,[\widehat{\xi}^\star_B - \xi^\star]$ converges to a non-degenerate normal distribution because the regularity conditions hold when $\xi^\star > -1/2$. Hence if $\rho$ is chosen with $0 < \rho < 1/2$ the first expression between curly brackets $\{\,\}$, namely $B^\rho\,[\widehat{\xi}^\star_B - \xi^\star]$ tends to 0 in probability. The second expression between curly brackets is known to tend to 0 in probability by lemma 1 above. So if $0 < \beta < \rho < 1/2$, we have $B^\beta \log I_B \overset{\text{Pr.}}{\to} 0$, which proves that $I_B \overset{\text{Pr.}}{\to} 1$ and moreover since $I_B - 1 \approx \log I_B$ for large $B$, assertion 1. follows.

2. Similarly, with $J_B := B^\alpha\,[\widetilde{\lambda}_B w]^{-\widehat{\xi}^\star_B}$

$$J_B = B^\alpha \times [\widetilde{\lambda}_B w]^{-\xi^\star} \times \underbrace{[\widetilde{\lambda}_B w]^{-[\widehat{\xi}^\star_B - \xi^\star]}}_{=I_B \to 1}$$

where the convergence under the braces is in probability and was proved before. Hence it will be enough to prove that $J^\dagger_B := B^\alpha\,[\widetilde{\lambda}_B w]^{-\xi^\star}$ tends to $\infty$ in probability. But

$$\log J^\dagger_B = \alpha \log B \left[1 - \frac{\xi^\star}{\alpha}\,\frac{\log[\widetilde{\lambda}_B w]}{\log B}\right],$$

and since $\log[\widetilde{\lambda}_B w]/\log B$ tends to 0 in probability by lemma 1, the result 2. follows. $\square$

We can now proceed to the proof of result 3.3. We know that (3.11) holds with $\mu$ replaced by the fixed value $\widetilde{\mu}_B$. As in the proof of result 3.2, this implies that $\widehat{\lambda}_B w = -\log F_{\texttt{GEV}}(\widetilde{\mu}_B;\, \widehat{\boldsymbol{\theta}}^\star_B)$. Thus

$$\left[\widehat{\lambda}_B w\right]^{-\widehat{\xi}^\star_B} = 1 + \widehat{\xi}^\star_B\,\frac{\widetilde{\mu}_B - \widehat{\mu}^\star_B}{\widehat{\sigma}^\star_B}, \qquad \left[\widetilde{\lambda}_B w\right]^{-\xi^\star} = 1 + \xi^\star\,\frac{\widetilde{\mu}_B - \mu^\star}{\sigma^\star}.$$

By eliminating $\widetilde{\mu}_B$ between these two equations, we get

$$\sigma^\star\,\frac{[\widetilde{\lambda}_B w]^{-\xi^\star} - 1}{\xi^\star} - \widehat{\sigma}^\star_B\,\frac{[\widehat{\lambda}_B w]^{-\widehat{\xi}^\star_B} - 1}{\widehat{\xi}^\star_B} = -\mu^\star + \widehat{\mu}^\star_B \tag{A.29}$$

For the sake of simplicity we will now only consider the case $\xi^\star \neq 0$. By multiplying both sides of (A.29) by $\sqrt{B}$, we know that the right-hand side converges for $B \to \infty$ to a non-degenerate distribution: this is indeed a consequence of the asymptotic behaviour of the ML estimation for the GEV distribution. This implies that for any $\alpha$ with $0 < \alpha < 1/2$

$$B^\alpha \left\{\frac{\sigma^\star}{\xi^\star}\,[\widetilde{\lambda}_B w]^{-\xi^\star} - \frac{\widehat{\sigma}^\star_B}{\widehat{\xi}^\star_B}\,[\widehat{\lambda}_B w]^{-\widehat{\xi}^\star_B}\right\} \overset{\text{Pr.}}{\to} 0$$

i.e. after rearranging

$$\underbrace{B^\alpha\,\left[\widetilde{\lambda}_B w\right]^{-\widehat{\xi}^\star_B}}_{=J_B \to \infty} \times \frac{\widehat{\sigma}^\star_B}{\widehat{\xi}^\star_B} \times \left\{\frac{\sigma^\star \widehat{\xi}^\star_B}{\widehat{\sigma}^\star_B \xi^\star}\,[\widetilde{\lambda}_B w]^{[\widehat{\xi}^\star_B - \xi^\star]} - \left[\widehat{\lambda}_B/\widetilde{\lambda}_B\right]^{-\widehat{\xi}^\star_B}\right\} \xrightarrow{\text{Pr.}} 0. \tag{A.30}$$

where the indicated limit (in probability) is the second affirmation of lemma 2. We thus know that the expression between the curly brackets must tend to 0 in probability. Using the first affirmation of lemma 2 and the fact that $B^{\beta'}|\{\widehat{\sigma}_B^\star \xi^\star\}/\{\sigma^\star \widehat{\xi}_B^\star\} - 1| \xrightarrow{\text{Pr.}} 0$ for some $\beta' > 0$ we see that $[\widehat{\lambda}_B/\widetilde{\lambda}_B]^{-\widehat{\xi}_B^\star} \xrightarrow{\text{Pr.}} 1$ implying that $\widehat{\lambda}_B/\widetilde{\lambda}_B \xrightarrow{\text{Pr.}} 1$.

We finish by justifying the fact that $\widehat{\lambda}_B w - \log B$ tends for large $B$ to the standard Gumbel distribution, as does $\widetilde{\lambda}_B w - \log B$. We have

$$\widehat{\lambda}_B w - \log B = \frac{\widetilde{\lambda}_B w}{\log B} \times \log B \times \left[\widehat{\lambda}_B/\widetilde{\lambda}_B - 1\right] + \left\{\widetilde{\lambda}_B w - \log B\right\}.$$

By proceeding as in the proofs above, it is not difficult to show that $\log B \times [\widehat{\lambda}_B/\widetilde{\lambda}_B - 1] \xrightarrow{\text{Pr.}} 0$. Since $\widetilde{\lambda}_B w/\log B \xrightarrow{\text{Pr.}} 1$ by the second part of lemma 1), we see that the difference between the two variables $\widehat{\lambda}_B w - \log B$ and $\widetilde{\lambda}_B w - \log B$ tends to 0. But $\widetilde{\lambda}_B w - \log B$ tends in distribution to a standard Gumbel, and so does $\widehat{\lambda}_B w - \log B$ by Slutsky's Theorem.

# Index