University of Padua
Department of Statistical Sciences

Master's Degree in
Statistical Sciences

Empirical Bayes inference for the Peaks Over a Threshold method

Supervisor Prof. Nicola Sartori
Department of Statistical Sciences

Co-Supervisor Prof. Simone Padoan
Department of Decision Sciences, Bocconi University

Graduand Pietro Scanzi
Serial number 2018901

Academic Year 2021/2022

Contents

Introduction

Extreme Value Theory (EVT) is a branch of statistical theory aimed at quantifying the stochastic behaviour of a process at unusually large (or small) levels. In particular, it aims to estimate the probability that events more extreme than those that have already been observed with the the available dataset take place. EVT can be implemented in scientific areas where accurate descriptions of rare phenomena are needed, since when these events occur, they have a strong impact on the society and therefore, extreme predictions may turn out to be very helpful. Examples of fields of application are oceanography, wind engineering, finance, telecommunications, biomedical data processing, thermodynamics of earthquakes, food science and so on.

Extreme data can be statistically analysed through two main approaches: the Block Maxima (BM) method which fits the Generalised Extreme Value (GEV) distribution to the sample maxima of different portions of the initial data and the Peaks Over a Threshold (POT) method, which relies on the Generalised Pareto (GP) distribution for modeling those observation which overcome a chosen threshold. Considering that the main goal of the EVT is the prediction of events that are expected to fall far beyond the observed data, the probabilistic models and statistical tools that it provides are asymptotically motivated. From a statistical point of view, such events are represented by the quantiles corresponding to an exceeding probability equal to or smaller than $1/n$ (where n is the sample size) of the unknown unconditional distribution that has generated the data. These are called extreme quantiles. Since to date, most of the real applications focused on extreme events concerned with the exceedances of a very large threshold, in this work we concentrate on the POT approach. Nevertheless, the GP distribution is an asymptotic distribution for the conditional distribution of those data conditioned on having exceeded a high threshold. Therefore, in order to achieve the prediction of extreme quantiles of their unconditional distribution, a two-step procedure is typically needed.

In alternative to the standard POT based on the GP limit, in this work we propose to rely on an adapted version of the asymptotic GEV distribution which also provides a suitable approximation for the tail of the unknown unconditional distribution of the data. On this basis, we perform inference through the so-called censored likelihood function, which depends on the GEV distribution and exploits the entire dataset. Note that, for making inference on the extreme quantile, the most important observations are still the large ones and so, for this reason, we still refer to this approach as a POT type method. Specifically, we name our approach as the Censored Peaks Over a Threshold (CPOT) approach.

This work develops a new Bayesian framework for the CPOT. In particular, we use an empirical Bayes prior for the parameters of the assumed GEV model, considering that these quantities are not fixed but vary with the sample size n. We implemented this new inferential method using software `R`. We sample from the posterior distribution of the parameters through an Adaptive Gaussian Random Walk Metropolis-Hastings (RWMH) algorithm, in order to obtain a prefixed optimal Overall Acceptance Probability (OAP). We test the performances of the Empirical Bayes CPOT comparing the frequentist coverage probabilities of credible intervals concerning extreme quantiles with their nominal level in a simulation study. Quantile, HPD and Gaussian approximation based intervals are taken into account.

The Thesis is organized as follows. Chapter 1 introduces the main ideas behind the univariate Extreme Value Theory and its asymptotic justification. The BM, POT and CPOT approaches are presented theoretically and from a likelihood based inferential point of view. Particular attention is given to the location and scale normalizing constants and to the concepts of extreme quantile and return level.

Chapter 2 is concerned with the empirical Bayes CPOT approach, which is described in detail and justified. We then introduce the Adaptive Gaussian RWMH algorithm that we used to draw samples from the posterior distribution. As an illustration, three complete analyses are presented in order to show the behaviour of the method on simulated extreme data pertaining to the three possible types of extreme distribution.

Chapter 3 describes the simulation study we made to test the performances of the empirical Bayes CPOT in terms of frequentist coverage probabilities of 95% quantile, HPD and normality-based credible intervals, for different sample sizes n and proportions of exceedances k, and for nine different extreme models. We report the simulation results

and discuss the performance of the new method, its problems and some possible future developments.

The Appendix contains the `R` code used in the thesis.

Chapter 1

Univariate Extreme Value Theory

1.1 Foundational concepts

In the following, we give a brief introduction to Extreme Value Theory, mainly based on Chapter 1 of Coles (2001).

Suppose that, as part of its design criteria for coastal defense, a sea-wall is required to protect against all sea-levels that it is likely to experience within its projected life span of 100 years. Local data on sea-levels might be available, but for a much shorter period of 10 years. Our purpose is to estimate what sea-levels might occur over the next 100 years given the 10-year history. Extreme value theory provides a framework that enables extrapolations of this type: it supplies standard models from asymptotic arguments that are useful in these peculiar contexts.

Suppose we have gathered the hourly sea-levels $x_1, \ldots, x_n$ over the past 10 years and we have organized these data into 10 blocks of m observations each. Taking directly into account the univariate random variables underlying the observed data, we denote by $X_{1,1}, X_{2,1}, \ldots, X_{m,1}$ the sea-levels related to the first year and, proceding in the same way, $X_{1,10}, X_{2,10}, \ldots, X_{m,10}$ indicate the sea-levels related to the tenth year. Then,

$$M_{m,1} = \max\{X_{1,1}, \ldots, X_{m,1}\}, \quad \ldots, \quad M_{m,10} = \max\{X_{1,10}, \ldots, X_{m,10}\}$$

is the sequence of yearly maximum sea-levels, considering a year as an "m-observation" period. If we knew the probability distribution of the X_is, we could derive the corresponding probability distribution of M_m. Unfortunately, this is not our case, since we put ourselves in the most general situation where the distribution of the X_is is left unspecified. However, under suitable assumptions, the approximate behaviour of M_m

for large values of m (asymptotic distribution) follows from detailed limit arguments by letting $m \to \infty$, leading to a family of models whose parameters can be estimated by the observed values $m_{m,1}, \ldots, m_{m,10}$. This reasoning is termed as the extreme value paradigm, since it comprises a principle for model extrapolation based on the implementation of mathematical limits as finite-level approximations.

1.2 Block Maxima (BM) approach

This section refers to Chapter 3 of Coles (2001) and Chapter 1 of De Haan & Ferreira (2006).

1.2.1 Model formulation

We consider a sequence of independent univariate random variables $X_1, \ldots, X_m$ having a common distribution F and their maximum

$$M_m = \max\{X_1, \ldots, X_m\}. \tag{1.1}$$

In applications, the X_is usually represent values of a process measured on a regular time-scale, perhaps hourly measurements of sea-level, or daily mean temperatures, so that M_m constitutes the maximum of the process over m time units of observations.

When F is known and $X_1, \ldots, X_m$ are i.i.d., the distribution of M_m can be derived directly for every value of m, in fact:

$$\begin{aligned} F^m(z) = Pr(M_m \leq z) &= Pr(X_1 \leq z, \ldots, X_m \leq z) \\ &= Pr(X_1 \leq z) \cdot \ldots \cdot Pr(X_m \leq z) \\ &= (F(z))^m . \end{aligned} \tag{1.2}$$

However, in most applications the distribution function F is unknown and so is the correct statistical behaviour of M_m. One could use standard statistical techniques to estimate F from the observed data, and then substitute this estimate into (1.2) to obtain the distribution function of M_m. Unfortunately, very small discrepancies in the estimate of F can lead to substantial discrepancies in F^m.

Alternatively, we can work under the assumption that F is unknown and look for approximate families of models for F^m, which can be estimated on the basis of extreme data only. We procede by searching the asymptotic distribution of M_m as $m \to \infty$. As a matter of fact, if we call z_+ the upper end-point of F, i.e. $z_+ := \sup\{z \in \mathbb{R} : F(z) < 1\}$,

which may be infinite, it can be demonstrated (page 3 of De Haan & Ferreira, 2006) that

$$M_m \xrightarrow{p} z_+ \qquad \text{as } m \to \infty, \tag{1.3}$$

where $\xrightarrow{p}$ stands for convergence in probability. So, F^m converges to 0 for $z < z_+$ and to 1 for $z = z_+$. In order to avoid a degenerate limit distribution, a linear renormalization of the variable M_m is necessary:

$$M_m^* = \frac{M_m - b_m}{a_m},$$

for sequences of scale $\{a_m > 0\}$ and location $\{b_m \in \mathbb{R}\}$ constants, both depending on the block size m. We therefore seek limit distributions for M_m^*, with appropriate choices of $\{a_m\}$ and $\{b_m\}$, rather than the one of M_m.

Theorem 1.1. (Fisher & Tippett, 1928; Gnedenko, 1943)
If there exist sequences of constants $\{a_m > 0\}$ *and* $\{b_m\}$ *such that*

$$Pr\left(\frac{M_m - b_m}{a_m} \leq z\right) = F^m(a_m z + b_m) \longrightarrow G_\gamma(z) \qquad \text{as } m \to \infty \tag{1.4}$$

for a non-degenerate distribution function G, then G is a member of the Generalised Extreme Value (GEV) family

$$G_\gamma(z) = \exp\left\{-(1+\gamma\, z)^{-\frac{1}{\gamma}}\right\}, \tag{1.5}$$

defined on $\{z : 1 + \gamma\, z > 0\}$.

In addition, expressions (1.4) and (1.5) can be furtherly developed into (page 4 of De Haan & Ferreira (2006)):

$$m \log F(a_m z + b_m) \longrightarrow \log G_\gamma(z) = -(1+\gamma\, z)^{-\frac{1}{\gamma}} \quad \text{as } m \to \infty \tag{1.6}$$

and

$$m\,\{1 - F(a_m z + b_m)\} \longrightarrow -\log G_\gamma(z) = (1+\gamma\, z)^{-\frac{1}{\gamma}} \quad \text{as } m \to \infty. \tag{1.7}$$

The GEV cumulative distribution function G_γ takes three different forms depending on the value assumed by the shape parameter γ, also called tail (or extreme-value) index (Fisher & Tippett, 1928), a location parameter $\mu \in \mathbb{R}$ and a scale parameter $\delta > 0$:

1. Fréchet type ($\gamma > 0$):

$$G_\gamma(z) = \begin{cases} 0 & \text{if } z \leq \mu - \frac{\delta}{\gamma}, \\ \exp\left\{-\left[1+\gamma\left(\frac{z-\mu}{\delta}\right)\right]^{-\frac{1}{\gamma}}\right\} & \text{otherwise;} \end{cases}$$

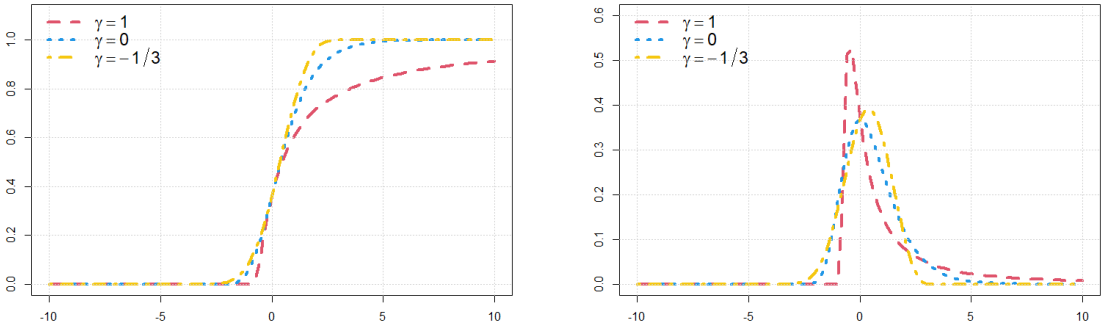
2. Gumbel (or double-exponential) type ($\gamma = 0$):

$$G(z) = \exp\left\{-\exp\left(-\frac{z-\mu}{\delta}\right)\right\} \qquad z \in \mathbb{R};$$

3. Reverse-Weibull type ($\gamma < 0$):

$$G_\gamma(z) = \begin{cases} \exp\left\{-\left[1+\gamma\left(\frac{z-\mu}{\delta}\right)\right]^{-\frac{1}{\gamma}}\right\} & \text{if } z \leq \mu - \frac{\delta}{\gamma}, \\ 1 & \text{otherwise.} \end{cases}$$

In other words, the GEV family includes three classes of distributions, termed as the extreme value distributions, widely known as the Fréchet, Gumbel and Reverse-Weibull families respectively. The three types of extreme value distribution are the only possible limits for the distribution of the rescaled maximum M_m^*, regardless of the distribution F of the population. We provide a graphical representation of the three types of GEV distributions in Figure 1.1. The `R` code for the plots is in Appendix A.



((A)) GEV c.d.f. ((B)) GEV p.d.f.

FIGURE 1.1: GEV cumulative density function ((a)) and probability density function ((b)) for $\mu = 0$, $\delta = 1$ and different values of γ.

The class of distribution functions F satisfying (1.4) is called max-domain of attraction of G_γ, and we write $F \in \mathcal{D}(G_\gamma)$. We now give necessary and sufficient conditions for a distribution function F to belong to the max-domain of attraction of G_γ.

Theorem 1.2. (Theorem 1.2.1 of De Haan & Ferreira, 2006)
The distribution function F is in the domain of attraction of the generalized extreme value distribution G_γ if and only if

1. *for $\gamma > 0$ (Fréchet case): $z_+ = \sup\{z \in \mathbb{R} : F(z) < 1\}$ is infinite and*

$$\lim_{t\to\infty} \frac{1 - F(tz)}{1 - F(t)} = z^{-\frac{1}{\gamma}} \tag{1.8}$$

   for all $z > 0$. This means that the function $1 - F$ is regularly varying at infinity with index $-1/\gamma$;

2. *for $\gamma = 0$ (Gumbel case): z_+ can be finite or infinite and*

$$\lim_{t\uparrow z_+} \frac{1 - F(t + zf(t))}{1 - F(t)} = e^{-z} \tag{1.9}$$

   for all real z, where f is a suitable positive function. If (1.9) holds for some f, then $\int_t^{z_+}(1 - F(s))ds < \infty$ for $t < z_+$ and (1.9) holds with

$$f(t) := \frac{\int_t^{z_+}(1 - F(s))ds}{1 - F(t)}\,; \tag{1.10}$$

3. *for $\gamma < 0$ (Reverse-Weibull case): z_+ is finite and*

$$\lim_{t\downarrow 0} \frac{1 - F(z_+ - tz)}{1 - F(z_+ - t)} = z^{-\frac{1}{\gamma}} \tag{1.11}$$

   for all $z > 0$.

The three types of limits that arise in Theorems 1.1 and 1.2 have distinct forms of behaviour, corresponding to different functions F of the X_is. This can be made precise by considering the behaviour of the limit distribution G_γ at z_+, its upper end-point. For the Reverse-Weibull class z_+ is finite, while for both the Fréchet and Gumbel classes z_+ is infinite. Still, the density of G_γ decays exponentially for the Gumbel distribution and polynomially for the Fréchet distribution, corresponding to relatively different rates of decay in the tail of F. It follows that in applications the three different families give quite different representations of extreme value behaviour. Actually, the extreme-value index γ describes the heaviness of the tail of the distribution G_γ, indeed distributions pertaining to the GEV class are called heavy-tailed if $\gamma > 0$, light-tailed if $\gamma = 0$ and short-tailed if $\gamma < 0$. Assuming the general model specified in formula (1.5), we can determine the most appropriate type of tail behaviour through inference on γ.

Uncertainty in the inferred value of γ measures the lack of certainty as to which of the original three types is most appropriate for a given dataset.

The Generalised Extreme Value (GEV) distribution is a probability distribution for modeling extreme values which generally depends on the parameter $\theta = (\gamma, \mu, \delta)$, where $\gamma \in \mathbb{R}$ is a shape parameter, $\mu \in \mathbb{R}$ is a location parameter and $\delta > 0$ is a scale parameter. Its probability density function is here:

$$G_\gamma(z; \gamma, \mu, \delta) = \left\{ -\left[1 + \gamma\left(\frac{z-\mu}{\delta}\right)\right]^{-\frac{1}{\gamma}} \right\}, \tag{1.12}$$

defined on $\left\{z : 1 + \gamma\left(\frac{z-\mu}{\delta}\right) > 0\right\}$. Equation (1.4) states that, if $F \in \mathcal{D}(G_\gamma)$, the renormalized block maximum M_m^* is asymptotically distributed as a $GEV(\gamma, 0, 1)$ random variable. If we interpret the limit in Theorem 1.1 as an approximation for large values of m, we can use the GEV family as a proper statistical model for the distribution of maxima of long sequences. The apparent difficulty arisen by the lack of knowledge of the normalizing constants in practice can be easily resolved. Indeed, if for large m

$$Pr(M_m^* \leq z) = Pr\left(\frac{M_m - b_m}{a_m} \leq z\right) \approx G_\gamma(z),$$

then, equivalently

$$Pr(M_m \leq z) = (F(z))^m \approx G_\gamma\left(\frac{z - b_m}{a_m}\right) = G_\gamma^*(z), \tag{1.13}$$

where G_γ^* is another member of the GEV family, depending on the parameter $\theta = (\gamma, \mu^*, \delta^*)$, which can be estimated by standard techniques.

This argument leads to the following approach for modeling extremes of a series of independent observations $x_1, \ldots, x_n$, the so called Block Maxima (BM) approach. Data are blocked into sequences of observations of length m, for some large value of m, generating a series of block maxima $m_{m,1}, \ldots, m_{m,q}$ ($n = m \times q$), say, to which the GEV distribution can be fitted. Often the blocks are chosen to correspond to a time period of length one year. Estimates of the extreme quantiles of the block (annual) maximum distribution are obtained by inverting Equation (1.12):

$$z_p = \begin{cases} \mu - \frac{\delta}{\gamma}\left[1 - \{-\log(1-p)\}^{-\gamma}\right] & \text{if } \gamma \neq 0, \\ \mu - \delta \log\{-\log(1-p)\} & \text{if } \gamma = 0, \end{cases} \tag{1.14}$$

where $G_\gamma(z_p) = 1 - p$. z_p is commonly named return level associated with the return

period $1/p = T$, since the level z_p is expected to be exceeded on average once every T units of time (years) or, equivalently, $Pr(M_n > z_p) = p = 1/T$.

We now provide two additional ingredients that are helpful in the proof of Theorem 1.1, which is the core of extreme value theory.

Definition 1.3. (Max-stability)
A distribution G is said to be max-stable if, for every $n = 2, 3, \ldots$, there are constants $\alpha_n > 0$ and β_n such that

$$G^n(\alpha_n z + \beta_n) = G(z).$$

Theorem 1.4. (Theorem 3.2 of Coles, 2001)
A distribution is max-stable if and only if it is a generalised extreme value (GEV) distribution.

It is possible to demonstrate that the limit distribution G in (1.4) is max-stable and thus, thanks to Theorem 1.4 we conclude that G must belong to the GEV family.

1.2.2 Likelihood based inference

Let $X_1, \ldots, X_n$ be univariate independent random variables from unknown F and $Z_1, \ldots, Z_q$ the derived block maxima of size m. Independence between the Z_is is guaranteed and, thanks to Theorem 1.1, we can assume for the block maxima a GEV distribution. When $\gamma \neq 0$ (Fréchet and Reverse-Weibull cases) the log-likelihood for the GEV parameters is

$$l(\gamma, \mu, \delta; z) = \begin{cases} -\infty \quad \text{if } \exists i \in \{1, \ldots, q\} : 1 + \gamma\left(\frac{z_i - \mu}{\delta}\right) \leq 0, \\ -q\log(\delta) - \left(1 + \frac{1}{\gamma}\right)\sum_{i=1}^{q} \log\left[1 + \gamma\left(\frac{z_i - \mu}{\delta}\right)\right] - \sum_{i=1}^{q}\left[1 + \gamma\left(\frac{z_i - \mu}{\delta}\right)\right]^{-\frac{1}{\gamma}} \text{ otherwise.} \end{cases} \tag{1.15}$$

When $\gamma = 0$ (Gumbel case) the log-likelihood for the GEV parameters is

$$l(\mu, \delta; z) = -q\log(\delta) - \sum_{i=1}^{q}\left(\frac{z_i - \mu}{\delta}\right) - \sum_{i=1}^{q}\exp\left\{-\left(\frac{z_i - \mu}{\delta}\right)\right\}. \tag{1.16}$$

Unfortunately, with the GEV model the regularity conditions that are required for the usual asymptotic properties associated with the maximum likelihood estimator to be valid do not hold, since here the support depends on the parameter. Specifically, the end-points of the GEV distribution are functions of the parameter values: $\mu - \delta/\gamma$ is an upper end-point when $\gamma < 0$ and a lower end-point when $\gamma > 0$. Smith (1985) studied this problem in detail and obtained the following results:

- when $\gamma > -0.5$, maximum likelihood estimators are regular, in the sense of having the usual asymptotic properties;
- when $-1 < \gamma \leq -0.5$, maximum likelihood estimators are generally obtainable, but do not have the standard asymptotic properties;
- when $\gamma < -1$, maximum likelihood estimators are unlikely to be obtainable.

The case $\gamma \leq -0.5$ corresponds to distributions with a very short bounded upper tail, a situation which is rarely encountered in applications on extreme value modeling.

Maximization of the log-likelihoods (1.15) and (1.16) with respect to the parameter vector $\theta = (\gamma, \mu, \delta)$ leads to the maximum likelihood estimate (MLE) with respect to the entire GEV family. There is no analytical solution, but for any given dataset the maximization is straightforward using standard numerical optimization algorithms. Some care is needed to ensure that such algorithms do not move to parameter combinations violating the condition in (1.15) of non-null likelihood. Furthermore, (1.15) presents numerical difficulties when evaluated in a neighbourhood of $\gamma = 0$, but this problem is easily solved using (1.16) in place of (1.15).

Dombry & Ferreira (2019) studied the existence and asymptotic normality of the maximum likelihood estimator $\hat{\theta} = (\hat{\gamma}, \hat{\mu}, \hat{\delta})$ under the following conditions:

1. First order condition:

$$F \in \mathcal{D}\left(G_{\gamma^0}\right) \qquad \text{with} \qquad \gamma^0 > -\frac{1}{2},$$

   where γ^0 is the true value of the tail index γ.

2. We set $V = \left(-\frac{1}{\log F}\right)^{\leftarrow}$, the left-continuous inverse of the function $-\frac{1}{\log F}$, such that $\left(-\frac{1}{\log F(x)}\right)^{\leftarrow} = \inf\left\{y : -\frac{1}{\log F(y)} \geq x\right\}$. Second order condition: for some positive function a and some positive or negative function A with $\lim_{t\to\infty} A(t) = 0$,

$$\lim_{t\to\infty} \frac{\frac{V(tx)-V(t)}{a(t)} - \frac{x^{\gamma^0}-1}{\gamma^0}}{A(t)} = \int_1^x s^{\gamma^0-1} \int_1^s u^{\rho-1} du ds = H_{\gamma^0,\rho}(x), \quad x > 0, \qquad (1.17)$$

   with $\gamma^0 > -\frac{1}{2}$. Note that necessarily $\rho \geq 0$ and $|A|$ is regularly varying with index ρ.

3. Asymptotic growth for the number q of of blocks and block size m:

$$q = q_n \to \infty, \quad m = m_n \to \infty \text{ and } \sqrt{q}A(m) \to \lambda \in \mathbb{R}, \text{ as } n \to \infty. \qquad (1.18)$$

In the following, Q_{γ^0} denotes the quantile function of the extreme value distribution G_{γ^0}, i.e.

$$Q_{\gamma^0}(p) = \frac{(-\log p)^{-\gamma^0} - 1}{\gamma^0}, \quad p \in (0,1), \tag{1.19}$$

and $i(\theta^0)$ stands for the Fisher information matrix computed in the true value of the parameter θ. We introduce the asymptotic bias b, i.e.

$$b = b(\gamma^0, \rho) = \int_0^1 \frac{\partial^2 l}{\partial x \partial \theta} \left(\theta^0, Q_{\gamma^0}(s)\right) H_{\gamma^0,\rho}\left(\frac{1}{-\log s}\right) ds. \tag{1.20}$$

Theorem 1.5. (Existence and asymptotic normality of the MLE)
Assume conditions (1.17) and (1.18).

- *There exists a sequence of estimators* $\hat{\theta}_n = \left(\hat{\gamma}_n, \hat{\mu}_n, \hat{\delta}_n\right)$, $n \geq 1$, *such that*

$$\lim_{n \to +\infty} Pr\left(\hat{\theta}_n \text{ is a MLE}\right) = 1 \tag{1.21}$$

  and

$$\sqrt{q}\left(\hat{\gamma}_n - \gamma^0, \frac{\hat{\mu}_n - b_m}{a_m}, \frac{\hat{\delta}_n}{a_m} - 1\right) \xrightarrow{d} \mathcal{N}_3\left(\lambda i(\theta^0)^{-1} b, i(\theta^0)^{-1}\right). \tag{1.22}$$

- *The MLE* $\hat{\theta} = \left(\hat{\gamma}, \hat{\mu}, \hat{\delta}\right)$ *is unique.*

Equation (1.18) requires that both the number of blocks q and the block size m go to infinity with a relative rate measured by the second order scaling function A and a parameter λ. When $\lambda = 0$, the bias term disappears in (1.22); this corresponds to the situation where m grows to infinity very quickly with respect to q so that the block size is large enough and the GEV approximation (1.4) is very good. From (1.22) we can directly derive the asymptotic distribution for the MLE $\hat{\theta} = \left(\hat{\gamma}, \hat{\mu}, \hat{\delta}\right)$, where a_m, b_m and $i(\theta^0)$ can be substituted in the formula by suitable estimates. Confidence intervals and other forms of inference follow immediately from the approximate normality of the estimator.

The maximum likelihood estimate of the $1/p = T$ return level z_p can be obtained by substituting the maximum likelihood estimates of the GEV parameters into (1.14):

$$\hat{z}_p = \begin{cases} \hat{\mu} - \frac{\hat{\delta}}{\hat{\gamma}}\left[1 - \{-\log(1-p)\}^{-\hat{\gamma}}\right] & \text{for } \hat{\gamma} \neq 0, \\ \hat{\mu} - \hat{\delta}\log\{-\log(1-p)\} & \text{if } \hat{\gamma} \to 0. \end{cases} \tag{1.23}$$

Furthermore, deriving the variance-covariance matrix $Var(\hat{\theta})$ of the MLE from (1.22), by the delta method,

$$Var(\hat{z}_p) \approx \Delta z_p^T \, Var(\hat{\theta}) \, \Delta z_p, \tag{1.24}$$

where $\Delta z_p^T = \left[\frac{\partial z_p}{\partial \gamma}, \frac{\partial z_p}{\partial \mu}, \frac{\partial z_p}{\partial \delta}\right]$ evaluated at $(\hat{\gamma}, \hat{\mu}, \hat{\delta})$. If $\hat{\gamma} < 0$ it is also possible to make inferences on the upper end-point of the distribution, which is effectively the "infinite observation return period", corresponding to z_p with $p = 0$. Its maximum likelihood estimate is

$$\hat{z}_0 = \hat{\mu} - \frac{\hat{\delta}}{\hat{\gamma}},$$

and the variance around this estimate can still be found with the delta method.

1.3 Classical Peaks Over a Threshold (POT) approach

This section refers to Chapter 4 of Coles (2001) and Chapters 3 and 4 of De Haan & Ferreira (2006).

1.3.1 Model formulation

As anticipated in Sections 1.1 and 1.2, the BM approach to extreme values is incredibly wasteful, since it really models only the maxima of the blocked sequences and leaves out the other data. Therefore, a new procedure of analysis of extreme values avoiding the blocking process is needed.

Let $X_1, \ldots, X_n$ be a sequence of univariate independent and identically distributed random variables with marginal distribution function F. We can classify as extreme events those of the X_is that exceed some high threshold u. In this way, extreme data are nothing but peaks over a threshold (POT). Following this rationale, a description of the stochastic behaviour of extreme events is given by this conditional probability:

$$Pr\,(X > u + y | X > u) = \frac{1 - F(u + y)}{1 - F(u)}, \qquad y > 0. \tag{1.25}$$

If the parent distribution F was known, the distribution of threshold exceedances in (1.25) would also be known. In practice we prefer avoiding restrictive (and often wrong) assumptions, thus we look for asymptotic approximations which are broadly applicable for high values of the threshold.

Theorem 1.6. (Pickands' Theorem: theorem 4.1 of Coles, 2001)
Let $X_1, \ldots, X_m$ be a sequence of independent random variables with common distribution F, and let

$$M_m = \max\{X_1, \ldots, X_m\}.$$

Denote an arbitrary term in the X_i sequence by X, and suppose that F satisfies Theorem 1.1, so that for large m,

$$Pr\left(M_m \leq z\right) \approx G_\gamma(z),$$

where

$$G_\gamma(z) = \exp\left\{-\left(1+\gamma\, z\right)^{-\frac{1}{\gamma}}\right\}$$

for some γ. Then, for large enough u, the distribution function of $(X-u)$, conditional on $X > u$, is approximately

$$H(y) = 1 - \left(1 + \frac{\gamma\, y}{\tilde{\delta}}\right)^{-\frac{1}{\gamma}} \tag{1.26}$$

defined in $\left\{y : y > 0 \text{ and } \left(1+\gamma\,\frac{y}{\tilde{\delta}}\right) > 0\right\}$, where

$$\tilde{\delta} = \gamma\, u. \tag{1.27}$$

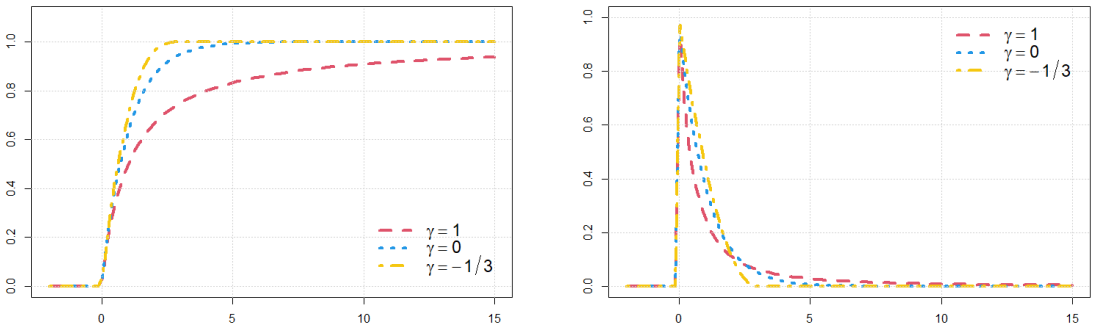
The family of distributions defined by (1.26) is called generalised Pareto family (GP) and by Theorem 1.6 it turns out to be the asymptotic distribution of threshold exceedances as the threshold $u \to \infty$. If $\gamma < 0$ the generalised Pareto distribution has an upper bound of $u - \tilde{\delta}/\gamma$, while it is unbounded when $\gamma \geq 0$. When $\gamma = 0$, taking the limit for $\gamma \to 0$ in (1.26), we get

$$H(y) = 1 - \exp\left(-\frac{y}{\tilde{\delta}}\right), \qquad y > 0, \tag{1.28}$$

corresponding to an exponential distribution with scale parameter $1/\tilde{\delta}$.

Theorem 1.6 implies that, if block maxima have approximating distribution G, then threshold excesses have a corresponding approximate distribution within the generalized Pareto family. Moreover, the parameters of the generalized Pareto distribution of threshold excesses are uniquely determined by those of the associated GEV distribution of block maxima. In particular, the parameter γ in (1.26) is equal to that of the corresponding GEV distribution. Choosing a different, but still large, block size m would affect the values of the GEV parameters, but not those of the corresponding generalized Pareto distribution of threshold excesses: γ is invariant to block size, and so is the

calculation of $\tilde{\delta}$ in (1.27). We provide in Figure 1.2 some plots of the generalized Pareto distribution. The `R` code used to create the plots is in Appendix B. From now on, for simplicity of notation, we substitute the expression of the scale parameter $\tilde{\delta}$ with just δ.



((A)) GP c.d.f. ((B)) GP p.d.f.

FIGURE 1.2: GP cumulative density function ((a)) and probability density function ((b)) for $\tilde{\delta} = 1$ and different values of γ.

We now clarify the concept of return level in the Peaks Over a Threshold (POT) context. Suppose that a generalized Pareto distribution with parameters γ and δ is a suitable model for exceedances of a threshold u by a random variable X. That is, putting together equations (1.25), (1.26) and (1.28), for $x > u$, we have

$$Pr\left(X > x | X > u\right) = \frac{Pr\left(X > x\right)}{Pr\left(X > u\right)} = \begin{cases} \left[1 + \gamma\left(\frac{x-u}{\delta}\right)\right]^{-\frac{1}{\gamma}} & \text{if } \gamma \neq 0, \\ \exp\left(-\frac{x-u}{\delta}\right) & \text{if } \gamma = 0. \end{cases}$$

It follows that, recalling that $Pr\left(X > u\right) = \zeta_u$,

$$Pr\left(X > x\right) = \begin{cases} \zeta_u \left[1 + \gamma\left(\frac{x-u}{\delta}\right)\right]^{-\frac{1}{\gamma}} & \text{if } \gamma \neq 0, \\ \zeta_u \exp\left(-\frac{x-u}{\delta}\right) & \text{if } \gamma = 0. \end{cases} \tag{1.29}$$

Hence, the r-observation return level is, by construction, the level x_r that is exceeded on average once every r observations, and can be found by solving the equation $Pr\left(X > x\right) = 1/r$. Provided that r is sufficiently large to ensure that $x_r > u$, the return level takes the expression

$$x_r = \begin{cases} u + \frac{\delta}{\gamma}\left[(r\zeta_u)^{\gamma} - 1\right] & \text{if } \gamma \neq 0, \\ u + \delta \log\left(r\zeta_u\right) & \text{if } \gamma = 0. \end{cases} \tag{1.30}$$

On an annual scale, the N-year return level is the level expected to be exceeded once every N years. If there are n_y observations per year, this corresponds to the r-observation return level, where $r = N \times n_y$. Therefore, the N-year return level is defined by:

$$z_N = \begin{cases} u + \frac{\delta}{\gamma}\left[(Nn_y\zeta_u)^\gamma - 1\right] & \text{if } \gamma \neq 0, \\ u + \delta \log\left(Nn_y\zeta_u\right) & \text{if } \gamma = 0. \end{cases}$$

1.3.2 Likelihood based inference

We collect the raw univariate data $x_1, \ldots, x_n$, consisting of a sequence of independent measurements from the random variables $X_1, \ldots, X_n$ whose common distribution F is in the max-domain of attraction of the GEV distribution. Following the classical POT framework, extreme events are identified by defining a high threshold u, for which the exceedances are $\{x_i : x_i > u\}$. These exceedances are labeled by $x_{(1)}, \ldots, x_{(k)}$ and threshold excesses are defined by $y_j = x_{(j)} - u$, for $j = 1, \ldots, k$. By Theorem 1.6, the y_j may be regarded as independent realizations of a random variable whose distribution can be approximated by a member of the generalized Pareto family. The standard practice is to adopt as low a threshold u as possible, subject to the limit model providing a reasonable approximation. Inference consists of fitting the generalized Pareto family to the observed threshold exceedances.

The parameters of the generalized Pareto distribution can be estimated by maximum likelihood. When $\gamma \neq 0$, the log-likelihood is obtained from (1.26) as:

$$l(\gamma, \delta; y) = \begin{cases} -\infty & \text{if } \exists\, i \in \{1, \ldots, k\} : \left(1 + \frac{\gamma\, y_i}{\delta}\right) \leq 0, \\ -k \log \delta - \left(1 + \frac{1}{\gamma}\right) \sum_{i=1}^{k} \log\left(1 + \frac{\gamma\, y_i}{\delta}\right) & \text{otherwise.} \end{cases} \tag{1.31}$$

If $\gamma = 0$, the log-likelihood is obtained from (1.28) as:

$$l(\delta; y) = -k \log(\delta) - \frac{1}{\delta} \sum_{i=1}^{k} y_i. \tag{1.32}$$

Analytical maximization of the log-likelihood is not possible, so numerical techniques are again required, taking care to avoid numerical instabilities when $\gamma \approx 0$ in (1.31), and ensuring that the algorithm does not fail due to evaluation outside of the allowable parameter space. Standard errors and confidence intervals for the generalized Pareto distribution are obtained in the usual way from standard likelihood theory.

Estimation of return levels requires the substitution of the maximum likelihood estimates of the parameters $\hat{\gamma}$ and $\hat{\delta}$ into (1.30). An estimate of ζ_u, the probability of

an individual observation exceeding the threshold u, is also needed. This has natural estimator of the sample proportion of points exceeding u, i.e.,

$$\hat{\zeta}_u = \frac{k}{n}. \tag{1.33}$$

Since it is reasonable to assume $K \sim Bi(n, \zeta_u)$, the estimate in (1.33) is also maximum likelihood estimate for ζ_u. Standard errors and confidence intervals for x_r can be derived using the delta method, including in the calculation the uncertainty in the estimate of ζ_u, which is, by standard properties of the binomial distribution, $Var(\hat{\zeta}_u) \approx \hat{\zeta}_u(1-\hat{\zeta}_u)/n$. Assuming independence between $\hat{\zeta}_u$ and the pair $(\hat{\gamma}, \hat{\delta})$ and denoting with V the variance-covariance matrix of $(\hat{\gamma}, \hat{\delta}, \hat{\zeta}_u)$, by delta method we compute

$$Var(\hat{x}_r) \approx \Delta x_r^T \; V \; \Delta x_r,$$

where $\Delta x_r^T = \left[\frac{\partial x_r}{\partial \gamma}, \frac{\partial x_r}{\partial \delta}, \frac{\partial x_r}{\partial \zeta_u}\right]$ evaluated at $(\hat{\gamma}, \hat{\delta}, \hat{\zeta}_u)$.

One serious competitor to the maximum likelihood estimator is the probability weighted moments (PWM) method of Hosking & Wallis (1997). The PMW method was earlier developed for the GEV distribution, where it was shown by simulation to be in general more efficient than the MLE for the central range of γ ($-0.2 < \gamma < 0.2$) and moderate values of n (up to 100). However, PWMs are much less flexible than MLEs as a general estimation method.

1.4 Censored Peaks Over a Threshold (CPOT) method

This section refers to Prescott & Walden (1983), Smith (1994) and Beranger et al. (2021).

Our purpose it to define an alternative Peaks Over a Threshold framework which exploits the asymptotic results of Block Maxima, defined in Theorem 1.1. We consider a set of independent and identically distributed extreme data $x_1, \ldots, x_n$ from $F \in \mathcal{D}(G_\gamma)$, we divide them into q blocks of size m and we set $y_m = a_m x + b_m$, which is a large value. From now on, we denote a_m with $a(m)$ and b_m with $b(m)$ to underline the crucial dependence of these constants on m. Then, from (1.4) we can state

$$Pr(\max\{X_1, \ldots, X_m\} \leq y_m) \approx \exp\left\{-\left[1 + \gamma\left(\frac{y_m - \mu}{\delta}\right)\right]^{-\frac{1}{\gamma}}\right\} \quad \text{for large } m, \tag{1.34}$$

where $\mu = \mu_m = b(m)$ and $\delta = \delta_m = a(m)$.

Here, having n overall sample units and k number of units exceeding a threshold u, we set $s = n/k$, which is the inverse of the proportion of threshold exceedances. This quantity s is important because it establishes a link between the BM and POT approaches to extreme values, i.e. it plays the role of the block size m. From a logical point of view, a rule for splitting units into homogeneous blocks of size s could be creating k subsets containing one u-threshold exceeding unit and $s-1$ non-exceeding units. We point out that, under this blocking scheme, the random variable $M_s = \max\{X_1, \dots, X_s\}$ selects the unique peak over the threshold of the set. For this reason, in this context the k block maxima $m_{s,1}, \dots, m_{s,k}$ are exactly the k threshold exceedances. From an analytical point of view, $s = n/k \to \infty$ as $n \to \infty$, just like $m \to \infty$ as $n \to \infty$, and consequently the asymptotic approximations of Theorem 1.1 still hold with s in place of m. Hence, rearranging equation (1.7) we obtain:

$$
\begin{aligned}
& s\left[1 - Pr\left(\frac{M_s - b(s)}{a(s)} \leq z\right)\right] \longrightarrow (1+\gamma z)^{-\frac{1}{\gamma}} \quad \text{as } s \to \infty \\
& \quad \Longrightarrow Pr\left(\frac{M_s - b(s)}{a(s)} \leq z\right) \approx 1 - \frac{1}{s}(1+\gamma z)^{-\frac{1}{\gamma}} \quad \text{for large } s.
\end{aligned}
$$

First of all, fixing $v_s = b(s) + a(s)z$ (which is a large value), $\tilde{\mu} = \tilde{\mu}_s = b(s)$ and $\tilde{\delta} = \tilde{\delta}_s = a(s)$, we compute:

$$
\begin{aligned}
Pr(M_s \leq v_s) &\approx 1 - \frac{1}{s}\left(1 + \gamma\,\frac{v_s - \tilde{\mu}}{\tilde{\delta}}\right)^{-\frac{1}{\gamma}} \\
&\approx \exp\left\{-\frac{1}{s}\left(1 + \gamma\,\frac{v_s - \tilde{\mu}}{\tilde{\delta}}\right)^{-\frac{1}{\gamma}}\right\} \\
&= \exp\left\{-\left[1 + \gamma\left(\frac{v_s - \tilde{\mu}}{\tilde{\delta}}\right)\right]^{-\frac{1}{\gamma}}\right\}^{\frac{1}{s}}
\end{aligned} \tag{1.35}
$$

where, between the first and the second step, we have used a Taylor expansion of the exponential function and, thanks to the result in (1.35), we find that $M_s \dot{\sim} GEV^{\frac{1}{s}}(\gamma, \tilde{\mu}, \tilde{\delta})$ for large s. We highlight that the parameters $\tilde{\mu}$ and $\tilde{\delta}$ are nothing but the aforementioned location and scale norming constants $b(s)$ and $a(s)$, unavoidable for the convergence results of Theorem 1.1 to the Generalised Extreme Value family. In the next section we will supply some operating methods for finding such normalizing constants.

Secondly, resuming the first line of the previous calculations we have

$$
\begin{aligned}
Pr(M_s \leq v_s) \approx 1 - \frac{1}{s}\left(1+\gamma\,\frac{v_s-\tilde{\mu}}{\tilde{\delta}}\right)^{-\frac{1}{\gamma}} &= 1 - \left(\frac{\tilde{\delta}+\gamma(v_s-\tilde{\mu})}{s^{-\gamma}\tilde{\delta}}\right)^{-\frac{1}{\gamma}} \\
&= 1 - \left(\frac{s^{-\gamma}\tilde{\delta} - s^{-\gamma}\tilde{\delta} + \tilde{\delta} + \gamma(v_s-\tilde{\mu})}{s^{-\gamma}\tilde{\delta}}\right)^{-\frac{1}{\gamma}} \\
&= 1 - \left(1 + \frac{\gamma\left(v_s - \tilde{\mu} + \frac{\tilde{\delta}(1-s^{-\gamma})}{\gamma}\right)}{s^{-\gamma}\tilde{\delta}}\right) \\
&= 1 - \left(1+\frac{\gamma(v_s-\bar{\mu})}{\bar{\delta}}\right)^{-\frac{1}{\gamma}} \\
&\approx \exp\left\{-\left[1+\gamma\left(\frac{v_s-\bar{\mu}}{\bar{\delta}}\right)\right]^{-\frac{1}{\gamma}}\right\}, \qquad (1.36)
\end{aligned}
$$

where, between the last two steps we have used again a Taylor expansion of the exponential function and (1.36) states that $M_s \dot{\sim} GEV(\gamma, \bar{\mu}, \bar{\delta})$ for large s with a new parameterization, which depends on the previous one in the following way

$$
\bar{\mu} = \begin{cases} \tilde{\mu} - \tilde{\delta}\,\frac{1-s^{-\gamma}}{\gamma} = \tilde{\mu} - \tilde{\delta}\,\frac{1-\left(\frac{n}{k}\right)^{-\gamma}}{\gamma} & \text{if } \gamma \neq 0, \\ \tilde{\mu} - \tilde{\delta}\log(s) = \tilde{\mu} - \tilde{\delta}\log\left(\frac{n}{k}\right) & \text{if } \gamma = 0; \end{cases} \qquad (1.37)
$$

$$
\bar{\delta} = \begin{cases} s^{-\gamma}\tilde{\delta} = \left(\frac{n}{k}\right)^{-\gamma}\tilde{\delta} & \text{if } \gamma \neq 0, \\ \tilde{\delta} & \text{if } \gamma = 0. \end{cases} \qquad (1.38)
$$

Having said this, we are ready to define the new Censored Peaks Over a Threshold (CPOT) inferential method. Suppose we have gathered a sample of data $x_1, \ldots, x_n$ and we want to use them to perform an extreme value analysis of the POT type. We then define a high threshold u, e.g. the 90%, 95% or 99% percentile of the observed sample, and, following the results in (1.35) and (1.36), we know we can properly model the peaks with (a transformation of) the GEV distribution. In particular, in order not to waste information, we regard at the units overcoming the threshold u as actually observed, while we consider left-censored the units that fall before u. At this point we can choose whether to exploit result (1.35) with the *tilde* parameterization or to follow result (1.36) with the *bar* parameterization. In the remainder of this section, for simplicity of notation, we will use the symbol G_γ to refer to the simple GEV distribution function free of scale and location parameters, defined in (1.5).

- In the *tilde* parameterization we define the Censored POT log-likelihood as:

$$l(\gamma, \tilde{\mu}, \tilde{\delta};\, x) = \sum_{i=1}^{n} \log \mathcal{L}(\gamma, \tilde{\mu}, \tilde{\delta};\, x_i), \tag{1.39}$$

where $\tilde{\mu}$ and $\tilde{\delta}$ are the usual scale and location norming constants and, for $i = 1, \ldots, n$

$$\mathcal{L}(\gamma, \tilde{\mu}, \tilde{\delta};\, x_i) = \begin{cases} G_\gamma\left(\frac{u-\tilde{\mu}}{\tilde{\delta}}\right)^{\frac{1}{s}} & \text{if } x_i \leq u, \\ \frac{d}{dx}\left[G_\gamma\left(\frac{x_i-\tilde{\mu}}{\tilde{\delta}}\right)^{\frac{1}{s}}\right] & \text{if } x_i > u, \end{cases} \tag{1.40}$$

where

$$\frac{d}{dx}\left[G_\gamma\left(\frac{x_i-\tilde{\mu}}{\tilde{\delta}}\right)^{\frac{1}{s}}\right] = \begin{cases} \frac{1}{\tilde{\delta}}\,\frac{1}{s}\,G_\gamma\left(\frac{x_i-\tilde{\mu}}{\tilde{\delta}}\right)^{\frac{1}{s}}\left[1+\gamma\left(\frac{x_i-\tilde{\mu}}{\tilde{\delta}}\right)\right]^{-\frac{1}{\gamma}-1} & \text{if } \gamma \neq 0, \\ \frac{1}{\tilde{\delta}}\,\frac{1}{s}\,G\left(\frac{x_i-\tilde{\mu}}{\tilde{\delta}}\right)^{\frac{1}{s}}\exp\left[-\left(\frac{x_i-\tilde{\mu}}{\tilde{\gamma}}\right)\right] & \text{if } \gamma = 0. \end{cases} \tag{1.41}$$

- In the *bar* parameterization we define the Censored POT log-likelihood as:

$$l(\gamma, \bar{\mu}, \bar{\delta};\, x) = \sum_{i=1}^{n} \log \mathcal{L}(\gamma, \bar{\mu}, \bar{\delta};\, x_i), \tag{1.42}$$

where $\bar{\mu}$ and $\bar{\delta}$ are determined by (1.37) and (1.38) and, for $i = 1, \ldots, n$

$$\mathcal{L}(\gamma, \bar{\mu}, \bar{\delta};\, x_i) = \begin{cases} G_\gamma\left(\frac{u-\bar{\mu}}{\bar{\delta}}\right) & \text{if } x_i \leq u, \\ \frac{d}{dx}\left[G_\gamma\left(\frac{x_i-\bar{\mu}}{\bar{\delta}}\right)\right] & \text{if } x_i > u, \end{cases} \tag{1.43}$$

where

$$\frac{d}{dx}\left[G_\gamma\left(\frac{x_i-\bar{\mu}}{\bar{\delta}}\right)\right] = \begin{cases} \frac{1}{\bar{\delta}}\left[1+\gamma\left(\frac{x_i-\bar{\mu}}{\bar{\delta}}\right)\right]^{-\frac{1}{\gamma}-1}\exp\left\{-\left[1+\gamma\left(\frac{x_i-\bar{\mu}}{\bar{\delta}}\right)\right]^{-\frac{1}{\gamma}}\right\} & \text{if } \gamma \neq 0, \\ \frac{1}{\bar{\delta}}\exp\left\{-\frac{x_i-\bar{\mu}}{\bar{\delta}} - \exp\left(-\frac{x_i-\bar{\mu}}{\bar{\delta}}\right)\right\} & \text{if } \gamma = 0. \end{cases} \tag{1.44}$$

We compute the maximum likelihood estimates of the parameters maximizing either (1.39) for the *tilde* parameterization or (1.42) for the *bar* parameterization. Anyway, the solution of the likelihood equation must be found numerically. Prescott & Walden (1983) suggest a suitable iterative Newton-Raphson type procedure (Ypma, 1995). It is possible for this procedure to produce a solution which is a local maximum or to fail to converge. Nevertheless, this occurs rarely and with sample configurations which are atypical of samples from extreme-value distributions. Other considerations about the obtained inference are the same as in Section 1.2.2.

We now make a comparison between the Censored Peaks Over a Threshold likelihood for parameters (γ, μ, δ) in the *tilde* and *bar* parameterizations. In order to compare them graphically, we have computed the profile likelihood for each of the three possible couples of parameters and we have then derived three contour plots. This procedure has been done using a sample of $n = 1000$ data generated from a distribution pertaining to the GEV family. In particular, in Figures 1.3 and 1.4 (at the end of this chapter) are shown the contour plots for the standardized unit-Fréchet ($\gamma = 1$) and unit-exponential ($\gamma = 0$) model. We can see that, in general, the likelihood for the *tilde* parameterization is flatter than the one of the *bar* parameterization, which is more curved. This can be interpreted into an approximate orthogonality between *tilde* parameters, while this does not hold for the *bar* parameters. This consideration will be taken into account later in the simulation studies of Chapter 3. The code of the used `R` functions for the CPOT likelihood and profile likelihoods is in Appendix C.

1.5 Norming constants

As introduced in Sections 1.2 and 1.4, both the Block Maxima and Censored Peaks Over a Threshold approaches are based on the normalization of the extremes by location and scale sequences, according to the asymptotic theory. Since Chapter 3 of the present thesis concerns with a simulation study for testing the performances of the proposed Empirical Bayes inferential procedure, then here we explicitly report the norming constants that are needed there. In this situation, it seems necessary to present some tools for computing these constants, which will be used in the following.

We start by considering a pure Block Maxima approach with block size m. We set the function $\mho(t) = F^{\leftarrow}\left(1 - \frac{1}{t}\right)$, with F denoting the distribution function of the random variable X and $F^{\leftarrow}$ left-continuous inverse of F, such that $F^{\leftarrow}(x) = \inf\{y : F(y) \geq x\}$. Then, a suitable choice for the norming sequences $a_m = a(m)$ and $b_m = b(m)$ is:

$$b(m) = \mho(m) \qquad \text{and} \qquad a(m) = m\, \mho'(m) \qquad \text{as } m \to \infty. \tag{1.45}$$

This choice is suggested by the Von Mises' sufficient conditions described by Theorem 1.1.8 of De Haan & Ferreira (2006). However, there are other choices for $a(m)$ and $b(m)$ that are asymptotically equivalent to the selection in (1.45). For the purposes of this work, another appropriate option for the norming constants is (De Haan & Ferreira,

2006):

$$b(m) = \mho(m) \quad \text{and} \quad a(m) = \begin{cases} \gamma\, \mho(m) & \text{if } \gamma > 0, \\ \mho(m) - m^{-1} \int_0^m \mho(x)\, dx & \text{if } \gamma = 0, \\ -\gamma\, (\mho(\infty) - \mho(m)) & \text{if } \gamma < 0, \end{cases} \quad \text{as } m \to \infty. \tag{1.46}$$

We recall that this theory works for the block size m that grows appropriately. Specifically, m depends on the sample size n so that $m = m(n)$, $m \to \infty$ as $n \to \infty$ and $m/n \to 0$ as $n \to \infty$. Hence, in practice, we choose the location and scale constants whose required calculations are made more straightforward by the problem under study.

In our project we consider a Peaks Over a Threshold (POT) approach, rather than the BM method. The POT setting assumes to work with k exceedances, whose number also depends on the sample size n, i.e. $k = k(n)$ and $k \to \infty$ as $n \to \infty$. In our setting it is like we see the block size m as the inverse of the proportion of the exceedances $s = n/k$, and so $s = n/k \to \infty$ as $n \to \infty$ and $s/n = 1/k \to 0$ as $n \to \infty$.

In conclusion, we can convert the choice (1.45) in:

$$b(s) = \mho(s) \qquad \text{and} \qquad a(s) = s\, \mho'(s) \qquad \text{as } s \to \infty, \tag{1.47}$$

and the choice in (1.46) in:

$$b(s) = \mho(s) \quad \text{and} \quad a(s) = \begin{cases} \gamma\, \mho(s) & \text{if } \gamma > 0, \\ \mho(s) - s^{-1} \int_0^s \mho(x)\, dx & \text{if } \gamma = 0, \\ -\gamma\, (\mho(\infty) - \mho(s)) & \text{if } \gamma < 0, \end{cases} \quad \text{as } s \to \infty. \tag{1.48}$$

We now provide detailed computations of the norming sequences for some parametric families of distributions that are in the domain of attraction of the GEV distribution in order to exploit them in the simulation study presented in Chapter 3. In particular, we pick the Fréchet, Pareto and Half-Cauchy models for the Fréchet class, the Gumbel, Exponential and Gamma models for the Gumbel class and finally the Power-Law, Reverse-Weibull and Beta models for the Reverse-Weibull class. We present these sequences in a listing format where 1. refers to (1.47) and 2. corresponds to (1.48). In the following, F is the cumulative distribution function, F' is the density function and q is the quantile function.

Fréchet constants

$$F(x;\alpha) = e^{-x^{-\alpha}},\ x>0,\ \alpha>0,\ \gamma = 1/\alpha \iff x = F^{-1}(u;\alpha) = (-\log(u))^{-\frac{1}{\alpha}};$$
$$\frac{d}{dm}\left(-\log\left(1-\frac{1}{m}\right)\right)^{-\frac{1}{\alpha}} = \frac{1}{\alpha m(m-1)}\left(\log(m)-\log(m-1)\right)^{-\frac{1}{\alpha}-1}.$$

1.

$$b\left(n/k\right) = \mho\left(\frac{n}{k}\right) = \left(\log\left(\frac{n}{k}\right) - \log\left(\frac{n}{k}-1\right)\right)^{-\frac{1}{\alpha}} = \left(\log(n)-\log(n-k)\right)^{-\frac{1}{\alpha}},$$

$$\begin{aligned} a(n/k) = \frac{n}{k}\cdot\mho'\left(\frac{n}{k}\right) &= \frac{n}{k}\frac{1}{\frac{n}{k}\left(\frac{n}{k}-1\right)}\left(\log\left(\frac{n}{k}\right)-\log\left(\frac{n}{k}-1\right)\right)^{-\frac{1}{\alpha}-1} \\ &= \frac{1}{\alpha\left(\frac{n}{k}-1\right)}\left(\log(n)-\log(n-k)\right)^{-\frac{1}{\alpha}-1}; \end{aligned}$$

2.

$$\begin{aligned} b(n/k) &= \left(\log(n)-\log(n-k)\right)^{-\frac{1}{\alpha}}, \\ a(n/k) &= \gamma\, b(n/k) = \frac{1}{\alpha}\left(\log(n)-\log(n-k)\right)^{-\frac{1}{\alpha}}. \end{aligned}$$

Pareto constants

$$F(x;\alpha) = 1-x^{-\alpha},\ x>1,\ \alpha>0,\ \gamma = 1/\alpha \implies 1-x^{-\alpha} = 1-\frac{1}{t} \iff x = t^{\frac{1}{\alpha}};$$
$$\mho'(t) = \frac{1}{\alpha}\,t^{\frac{1}{\alpha}-1}.$$

1.

$$\begin{aligned} b(n/k) &= \mho\left(\frac{n}{k}\right) = \left(\frac{n}{k}\right)^{\frac{1}{\alpha}}, \\ a(n/k) &= \frac{n}{k}\,\mho'\left(\frac{n}{k}\right) = \frac{n}{k}\,\frac{1}{\alpha}\,\left(\frac{n}{k}\right)^{\frac{1}{\alpha}-1} = \frac{1}{\alpha}\,\left(\frac{n}{k}\right)^{\frac{1}{\alpha}}; \end{aligned}$$

2.

$$\begin{aligned} b(n/k) &= \left(\frac{n}{k}\right)^{\frac{1}{\alpha}}, \\ a(n/k) &= \frac{1}{\alpha}\,\mho'\left(\frac{n}{k}\right) = \frac{1}{\alpha}\,\left(\frac{n}{k}\right)^{\frac{1}{\alpha}}. \end{aligned}$$

Half-Cauchy constants

$$
\begin{aligned}
F^{'}(x) &= \frac{2}{\pi(1+x^2)} - 1,\ x > 0;\ \nu = 1,\ \gamma = \frac{1}{\nu} = 1;\\
1 - F(x) &\approx \frac{2}{\pi\, x} = \frac{1}{t} \Longleftrightarrow x \approx \frac{2}{\pi}\, t \qquad \text{as } x \to \infty;\\
\mho^{'}(t) &\approx \frac{2}{\pi}.
\end{aligned}
$$

1.

$$
\begin{aligned}
b(n/k) &= \mho\left(\frac{n}{k}\right) = \frac{2}{\pi}\,\frac{n}{k},\\
a(n/k) &= \frac{n}{k}\mho^{'}\left(\frac{n}{k}\right) = \frac{2}{\pi}\,\frac{n}{k};
\end{aligned}
$$

2.

$$
\begin{aligned}
b(n/k) &= \frac{2}{\pi}\,\frac{n}{k},\\
a(n/k) &= \frac{2}{\pi}\,\frac{n}{k}.
\end{aligned}
$$

Gumbel constants

$$
\begin{aligned}
F(x) &= \exp\left(-\exp\left(-x\right)\right),\ x \in \mathbb{R},\ \gamma = 0;\\
e^{-e^{-x}} &= 1 - \frac{1}{t} \iff x = -\log\left(-\log\left(\frac{t-1}{t}\right)\right);\\
\mho^{'}(t) &= \frac{1}{t(t-1)(\log(t) - \log(t-1))}.
\end{aligned}
$$

1.

$$
\begin{aligned}
b(n/k) = \mho\left(\frac{n}{k}\right) &= -\log\left(\log\left(\frac{n}{k}\right) - \log\left(\frac{n-k}{k}\right)\right)\\
&= -\log\left(\log\left(n\right) - \log\left(n-k\right)\right),\\
a(n/k) = \frac{n}{k}\,\mho^{'}\left(\frac{n}{k}\right) &= \frac{n}{k}\cdot\frac{1}{\frac{n}{k}\left(\frac{n}{k}-1\right)\left(\log(n) - \log(n-k)\right)}\\
&= \frac{k}{(n-k)\left(\log(n) - \log(n-k)\right)};
\end{aligned}
$$

2.

$$
\begin{aligned}
b(n/k) &= -\log\left(\log\left(n\right) - \log\left(n-k\right)\right),\\
a(n/k) &= \mho\left(\frac{n}{k}\right) - \frac{k}{n}\int_1^{\frac{n}{k}} \log\left(\log(s) - \log(s-1)\right) ds.
\end{aligned}
$$

Exponential constants

$$F(x;\lambda) = 1 - e^{-\lambda x},\ x \geq 0,\ \lambda > 0,\ \gamma = 0;$$
$$1 - e^{-\lambda x} = 1 - \frac{1}{t} \Longleftrightarrow x = \frac{\log(t)}{\lambda};$$
$$\mho'(t) = \frac{1}{\lambda t}.$$

1.

$$b(n/k) = \mho\left(\frac{n}{k}\right) = \frac{\log(n) - \log(k)}{\lambda},$$
$$a(n/k) = \frac{n}{k}\mho'\left(\frac{n}{k}\right) = \frac{1}{\lambda};$$

2.

$$\begin{aligned} b(n/k) &= \frac{\log(n) - \log(k)}{\lambda}, \\ a(n/k) &= \mho\left(\frac{n}{k}\right) - \frac{k}{n}\int_0^{\frac{n}{k}} \mho(s)ds \\ &= \frac{\log(n) - \log(k)}{\lambda} - \frac{k}{n}\frac{1}{\lambda}(s\log(s) - s)\Big|_0^{\frac{n}{k}} \\ &= \frac{\log(n) - \log(k)}{\lambda} - \frac{k}{n}\frac{1}{\lambda}\frac{n}{k}\left(\log\left(\frac{n}{k} - 1\right)\right) = \frac{1}{\lambda}. \end{aligned}$$

Gamma constants

$$F'(x;\alpha,\beta) = \frac{x^{\alpha-1}e^{-\beta x}\beta^{\alpha}}{\Gamma(\alpha)},\ x > 0,\ \alpha > 0,\ \beta > 0,\ \gamma = 0;$$
$$\mho\left(\frac{n}{k}\right) = F^{-1}\left(1 - \frac{k}{n}\right) = q_{\alpha,\beta}\left(1 - \frac{k}{n}\right);$$
$$\mho'\left(1 - \frac{k}{n}\right) = q'_{\alpha,\beta}\left(1 - \frac{k}{n}\right)\left(\frac{k}{n}\right)^2 = \frac{1}{F'\left(q_{\alpha,\beta}\left(1 - \frac{k}{n}\right)\right)}\left(\frac{k}{n}\right)^2.$$

1.

$$b(n/k) = \mho\left(\frac{n}{k}\right) = q_{\alpha,\beta}\left(1 - \frac{k}{n}\right),$$
$$a(n/k) = \frac{n}{k}\mho'\left(\frac{n}{k}\right) = \frac{k}{nF'\left(q_{\alpha,\beta}\left(1 - \frac{k}{n}\right)\right)};$$

2.

$$b(n/k) = q_{\alpha,\beta}\left(1-\frac{k}{n}\right),$$
$$a(n/k) = q_{\alpha,\beta}\left(1-\frac{k}{n}\right) - \frac{k}{n}\int_0^{\frac{n}{k}} q_{\alpha,\beta}\left(1-\frac{1}{s}\right) ds.$$

Power-Law constants

$$\begin{aligned} F(x;c,x^*) =& 1-c(x^*-x)^\alpha,\ \alpha>0,\ x\le x^*,\ \gamma=-1/\alpha;\\ & 1-c(x^*-x)^\alpha = 1-\frac{1}{t} \Longleftrightarrow x = x^*-(ct)^{-\frac{1}{\alpha}};\\ & \mho^{'}(t) = \frac{c}{\alpha}(ct)^{-\frac{1}{\alpha}-1} \end{aligned}$$

1.

$$b(n/k) = \mho\left(\frac{n}{k}\right) = x^* - \left(c\,\frac{n}{k}\right)^{-\frac{1}{\alpha}},$$
$$a(n/k) = \frac{n}{k}\mho^{'}\left(\frac{n}{k}\right) = \frac{1}{\alpha}\left(c\,\frac{n}{k}\right)^{-\frac{1}{\alpha}};$$

2.

$$b(n/k) = x^* - \left(c\,\frac{n}{k}\right)^{-\frac{1}{\alpha}},$$
$$a(n/k) = -\gamma\left(\mho(\infty) - \mho\left(\frac{n}{k}\right)\right) = \frac{1}{\alpha}\left(c\,\frac{n}{k}\right)^{-\frac{1}{\alpha}}.$$

Reverse-Weibull constants

$$\begin{aligned} F(x;\alpha) =& \exp\left(-(-x)^\alpha\right),\ x<0,\ \alpha>0,\ \gamma=-\frac{1}{\alpha};\\ & \exp\left(-(-x)^\alpha\right) = 1-\frac{1}{t} \Longleftrightarrow x = -\left(-\log\left(1-\frac{1}{t}\right)\right)^{\frac{1}{\alpha}};\\ & \mho^{'}(t) = \frac{1}{\alpha t(1-t)}\left(-\log\left(1-\frac{1}{t}\right)\right)^{\frac{1}{\alpha}-1}. \end{aligned}$$

1.

$$b(n/k) = \mho\left(\frac{n}{k}\right) = -\left(-\log\left(1-\frac{k}{n}\right)\right)^{\frac{1}{\alpha}},$$
$$a(n/k) = \frac{n}{k}\mho^{'}\left(\frac{n}{k}\right) = \frac{k}{\alpha(n-k)}\left(-\log\left(1-\frac{k}{n}\right)\right)^{\frac{1}{\alpha}-1};$$

2.

$$
\begin{aligned}
b(n/k) &= -\left(-\log\left(1-\frac{k}{n}\right)\right)^{\frac{1}{\alpha}},\\
a(n/k) &= \mho\left(\frac{n}{k}\right) - \frac{n}{k}\int_0^{\frac{n}{k}} \mho(s)ds\\
&= -\left(-\log\left(1-\frac{k}{n}\right)\right)^{\frac{1}{\alpha}} - \frac{n}{k}\int_0^{\frac{n}{k}} -\left(-\log\left(1-\frac{k}{n}\right)\right)^{\frac{1}{\alpha}} d\,\frac{n}{k}.
\end{aligned}
$$

Beta constants

$$
\begin{aligned}
F^{'}(x;\alpha,\beta) &= \frac{x^{\alpha-1}(1-x)^{\alpha-1}}{B(\alpha,\beta)},\ 0<x<1,\ \alpha>0,\ \beta>0,\ \gamma=-\frac{1}{\beta};\\
\mho\left(\frac{n}{k}\right) &= F^{-1}\left(1-\frac{k}{n}\right) = q_{\alpha,\beta}\left(1-\frac{k}{n}\right);\\
\mho^{'}\left(1-\frac{k}{n}\right) &= q^{'}_{\alpha,\beta}\left(1-\frac{k}{n}\right)\left(\frac{k}{n}\right)^2 = \frac{1}{F'\left(q_{\alpha,\beta}\left(1-\frac{k}{n}\right)\right)}\left(\frac{k}{n}\right)^2.
\end{aligned}
$$

1.

$$
\begin{aligned}
b(n/k) &= \mho\left(\frac{n}{k}\right) = q_{\alpha,\beta}\left(1-\frac{k}{n}\right),\\
a(n/k) &= \frac{n}{k}\,\mho^{'}\left(\frac{n}{k}\right) = \frac{k}{nF'\left(q_{\alpha,\beta}\left(1-\frac{k}{n}\right)\right)};
\end{aligned}
$$

2.

$$
\begin{aligned}
b(n/k) &= q_{\alpha,\beta}\left(1-\frac{k}{n}\right),\\
a(n/k) &= q_{\alpha,\beta}\left(1-\frac{k}{n}\right) - \frac{k}{n}\int_0^1 q_{\alpha,\beta}\left(1-\frac{1}{s}\right)ds.
\end{aligned}
$$

1.6 Extreme Quantile and Return Level

One of the main goal of the Extreme Value Theory is to predict extreme events. Statistically speaking, one possible way to quantify extreme events is by computing Extreme Quantiles and Return Levels. These two values refer to the same concept, that is they are utmost quantiles and provide the expected extreme value related to a given return probability p, or equivalently, to a return period T. On the other hand, they differ for some more specific aspects.

Regarding the initial distribution of the data F and given a small exceeding probability p, the extreme quantile is the quantile of level $1-p$. Indeed, it is the value y_p

such that

$$1 - p = F(y_p) \iff y_p = F^{-1}(1 - p).$$

Since F is unknown and we know from Theorem 1.1 that

$$F(y) \approx (G_\gamma(y; a_m, b_m))^{\frac{1}{m}} \qquad \text{for } m \to \infty,$$

then we have to solve with respect to y_p the equation

$$1 - p \approx \exp\left\{-\frac{1}{m}\left(1 + \gamma\left(\frac{y_p - b_m}{a_m}\right)\right)^{-\frac{1}{\gamma}}\right\},$$

from which we obtain

$$y_p \approx b_m + a_m \frac{(-m\log(1-p))^{-\gamma} - 1}{\gamma} \approx b_m + a_m \frac{(mp)^{-\gamma} - 1}{\gamma},$$

by a Taylor expansion for $-p \to 0$. This approximation is applicable since p is by definition a small probability and, furthermore, by the asymptotic theory of De Haan & Ferreira (2006) $p = p_n$ depends on n in a way that $p \to 0$ as $n \to \infty$.

Now, to work with the Block Maxima approach with block size m is equivalent to work with the Peaks Over a Threshold method with number of exceedances k and sample size n. As previously stated, these two are linked by the relation $m = \frac{n}{k} = s$. Hence, in our context it is like we operate with $k = \frac{n}{m}$ exceedances, and with a suitable substitution we arrive to the final formula for the extreme quantile (Beranger et al., 2021):

$$y_p = b(n/k) + a(n/k)\frac{\left(\frac{np}{k}\right)^{-\gamma} - 1}{\gamma}. \tag{1.49}$$

Considering the distribution of the m-block maxima F^m and given a (relatively) high return period T, the return value is the level $1 - \frac{1}{T}$ quantile. In formulas, we have that

$$1 - \frac{1}{T} = F^m(y_T) \iff y_T = (F^m)^{-1}\left(1 - \frac{1}{T}\right).$$

Since also F^m is unknown, we know from the abovementioned theorem that for $m \to \infty$

$$F^m(y) \approx G_\gamma(y; a_m, b_m) = \exp\left\{-\left(1 + \gamma\left(\frac{y - b_m}{a_m}\right)\right)^{-\frac{1}{\gamma}}\right\},$$

with $1+\gamma\frac{y-b_m}{a_m}>0$. Therefore, we solve with respect to y_T the equation

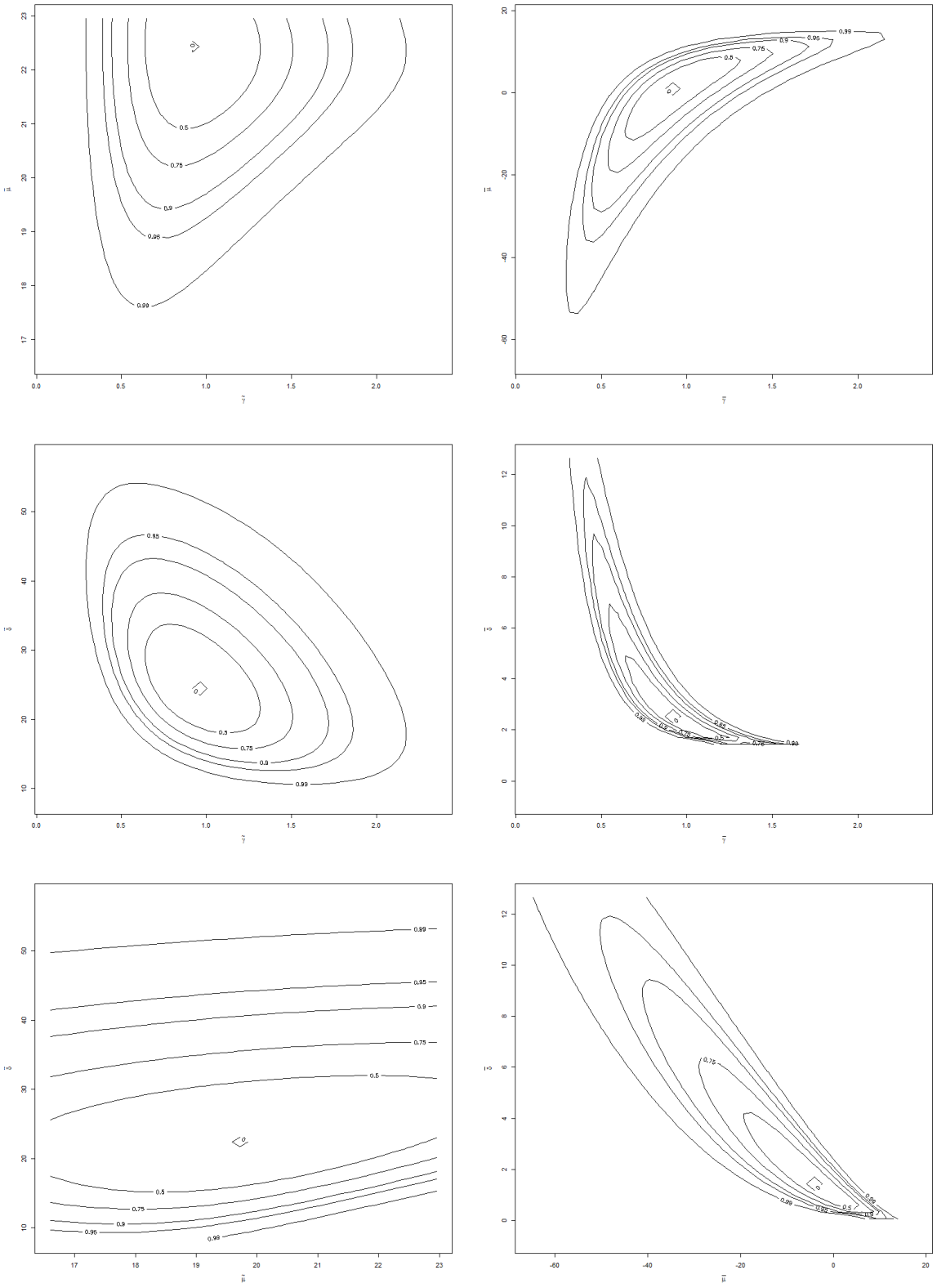
$$1-\frac{1}{T}=\exp\left\{-\left(1+\gamma\left(\frac{y_T-b_m}{a_m}\right)\right)^{-\frac{1}{\gamma}}\right\},$$

from which we obtain the return level y_T associated with the return period T:

$$y_T=b_m+a_m\frac{\left(-\log\left(1-\frac{1}{T}\right)\right)^{-\gamma}-1}{\gamma}.$$

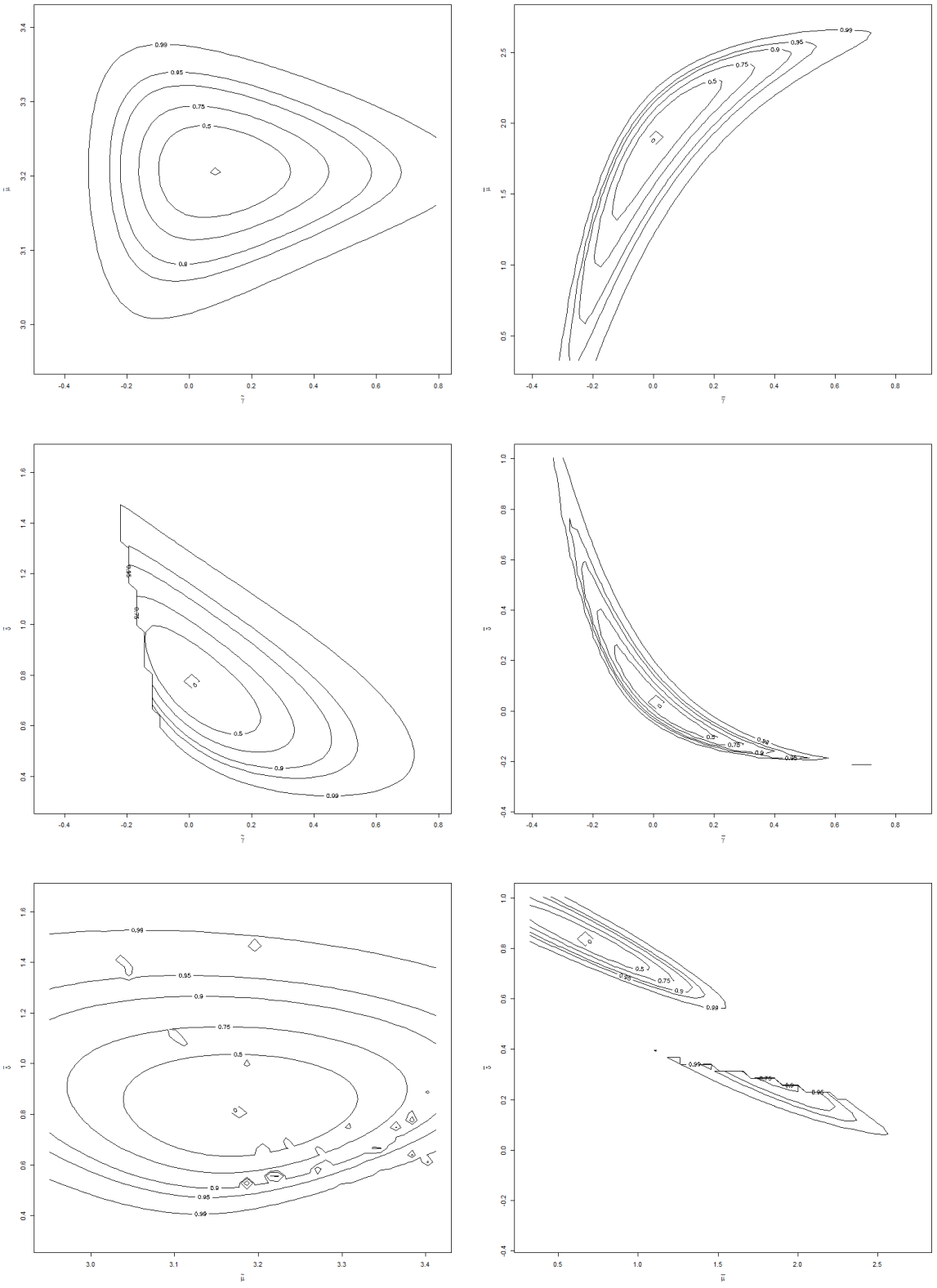
In POT terms, this quantile is

$$y_T=b(n/k)+a(n/k)\frac{\left(-\log\left(1-\frac{1}{T}\right)\right)^{-\gamma}-1}{\gamma}. \tag{1.50}$$



((A)) *tilde* parameterization ((B)) *bar* parameterization

FIGURE 1.3: Contour plots of the profile likelihood for (γ, μ), (γ, δ) and (μ, δ) in the *tilde* ((a)) and *bar* ((b)) parameterizations for a sample of size $n = 1000$ from the standardized unit-Fréchet distribution.



((A)) *tilde* parameterization ((B)) *bar* parameterization

FIGURE 1.4: Contour plots of the profile likelihood for (γ, μ), (γ, δ) and (μ, δ) in the *tilde* ((a)) and *bar* ((b)) parameterizations for a sample of size $n = 1000$ from the unit-exponential distribution.

Chapter 2

Empirical Bayes inferential method for Peaks Over a Threshold

2.1 The Bayesian framework

This section is inspired by the lectures about Bayesian Inference of the course Theory and Methods of Inference by Professors Salvan and Sartori. For a more thorough discussion on Bayesian inference we refer, for instance, to Davison (2003), Chapter 11.

2.1.1 Bayesian Inference

The observed data $x = (x_1, \ldots, x_n)$ are realization of a random vector X whose probability distribution is (partly) unknown. Data are used to reconstruct the distribution of X. We assume x as a realization of $X \sim p^0(x)$, $x \in \mathcal{X}$, where $p^0(x)$ represents the unknown probability density function (p.d.f.), with respect to a suitable measure, and where $\mathcal{X}$ is the sample space. The aim of statistical inference is to reconstruct $p^0(x)$ on the basis of both data and suitable assumptions and, on the ground of previous information, extract quantitative insights from the behaviour of the observed data, interpretate some aspects of the phenomenon under study and possibly make predictions for hypothesized future data from the same generating process. In other words, we want to make an inference.

We assume a parametric statistical model $\mathcal{F}$ for the sample of data at hand, such that

$$\mathcal{F} = \left\{ p_X(x|\theta),\, x \in \mathcal{X} \subset \mathbb{R}^n,\, \theta \in \Theta \subset \mathbb{R}^d \right\}.$$

Generally, we further assume for $\mathcal{F}$ to be correctly specified with identifiable parameter θ. In this respect, $p^0(x) = p(x|\theta^0)$ for some $\theta^0 \in \Theta$. In an extreme value context,

the GEV class is an asymptotic (not exact) model for block maxima or threshold exceedances, thus $\mathcal{F}$ is actually misspecified.

The Bayesian approach to statistical inference views the probability as an evaluation of uncertainty, both concerning observables, like x or future observations x^*, and unobservables, like θ. We therefore additionally assume that the parameter θ is a realization of a random variable with prior density $\pi(\theta)$ over Θ which summarizes prior knowledge about the parameter. In this regard, we hypothesize for continuity of notation the existence of a true parameter value $\theta^0 \in \Theta$, where such value is drawn from $\pi(\theta)$. By Bayes theorem, information about θ is updated after the observation of x according to

$$\pi(\theta|x) = \frac{\pi(\theta)p_X(x|\theta)}{\int_\Theta \pi(\theta)p_X(x|\theta)d\theta} = \frac{\pi(\theta)\mathcal{L}(\theta;x)}{\int_\Theta \pi(\theta)\mathcal{L}(\theta;x)d\theta}, \tag{2.1}$$

the posterior density of θ given x. In (2.1) $p_X(x|\theta)$ stands for the probability distribution of the data, conditioned to the parameter. Indeed, it has the same expression of the likelihood function $\mathcal{L}(\theta;x)$ for θ with data x, apart for possible multiplicative constants depending on x. Our objective is to perform computations to obtain summaries of $\pi(\theta|x)$, such as estimates and credible intervals. For instance, with a scalar θ, a posterior credible bound for θ, with posterior probability $1-\alpha$ is the $1-\alpha$ quantile of $\pi(\theta|x)$, $\theta^{1-\alpha}(x)$,

$$Pr\left(\theta \leq \theta^{1-\alpha}(x)|x\right) = \int_{-\infty}^{\theta^{1-\alpha}(x)} \pi(\theta|x)d\theta = 1-\alpha.$$

The choice of the prior distribution $\pi(\theta)$ for the parameter is called prior specification. This can be worked out in several ways, such as using information from previous studies or expert personalistic opinions. In the absence of prior information, non-informative or "objective" priors could be used. In any case, often mathematically convenient prior distributions are used, as with conjugate priors. Ideally, the prior distributions should not depend on the data. An exception is given by the empirical Bayes approach (Robbins, 1956), which will be discussed more in detail in Section 2.2, and that will be used in our extreme value setting.

The computation of $\pi(\theta|x)$ can be sorted out:

- analytically, with the use of conjugate priors, if available;
- through asymptotic approximations due to the Bernstein-Von Mises theorem (Reid, 1996; Van der Vaart, 2000), valid under general regularity likelihood conditions

and for every smooth $\pi(\theta)$. We consider the quantity

$$\Delta_n(\theta^0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l^* \left(\theta^0; x_i\right) i_1 \left(\theta^0\right)^{-1}, \tag{2.2}$$

where $l^* \left(\theta^0; x_i\right)$ and $i_1(\theta^0)$ are the score function and the Fisher information matrix of a single observation evaluated at the true parameter value θ^0. Hence, the total variation distance between the posterior distribution of $\sqrt{n}\left(\pi\left(\theta_n|x\right) - \theta^0\right)$ and the random distribution $\mathcal{N}\left(\Delta_n\left(\theta^0\right), i_1\left(\theta^0\right)^{-1}\right)$ converges to 0 as $n \to \infty$, therefore

$$\theta_n|x \dot{\sim} \mathcal{N}\left(\theta_0 + \frac{\Delta_n(\theta_0)}{\sqrt{n}}, i\left(\theta^0\right)^{-1}\right) \qquad \text{for large } n, \tag{2.3}$$

where $i\left(\theta^0\right)$ is the full sample Fisher information matrix. We point out that the bias term in (2.3) fades out with the speed of $\sqrt{n}$, i.e. it is an $O_p\left(n^{-\frac{1}{2}}\right)$. Operationally, θ^0 in Equation (2.3) can be substituted by a suitable estimate, like the posterior mode θ^* or the maximum likelihood estimate $\hat{\theta}$. Similarly, $i\left(\theta^0\right)^{-1}$ can be substituted by $j\left(\hat{\theta}\right)^{-1}$ or $j^*\left(\theta^*\right)^{-1}$, which is the inverse of the negative log-posterior hessian evaluated at the mode;

- via Monte Carlo methods like rejection sampling, importance sampling or Markov Chain Monte Carlo methods (Robert & Casella, 2010), which will be discussed in detail in Section 2.3.

A $1-\alpha$ credible region is the set $\mathcal{C} \subset \Theta$ such that $Pr(\theta \in \mathcal{C}|x) = 1-\alpha$. For the scalar case, we can compute credible intervals (θ_L, θ_U) in various ways, e.g.:

- Equi-tailed quantile intervals, i.e.:

$$(\theta_L, \theta_U) \qquad \text{with} \quad Pr(\theta < \theta_L|x) = \frac{\alpha}{2} \quad \text{and} \quad Pr(\theta < \theta_U|x) = 1 - \frac{\alpha}{2}; \tag{2.4}$$

- higher posterior density (HPD) credible intervals:

$$\begin{aligned}(\theta_L, \theta_U) \qquad & \text{with} \quad \pi(\theta|x) \geq \pi(\theta'|x), \quad \theta \in (\theta_L, \theta_U), \ \theta' \notin (\theta_L, \theta_U) \\ & \text{and} \quad Pr\left(\theta \in (\theta_L, \theta_U)|y\right) = 1-\alpha.\end{aligned} \tag{2.5}$$

- Intervals based on the asymptotic normality of the posterior distribution, hence referring to (2.3):

$$(\theta_L,\, \theta_U) = \left(\theta^* + \frac{\Delta_n\left(\theta^*\right)}{\sqrt{n}} - z_{1-\frac{\alpha}{2}}\sqrt{j^*\left(\theta^*\right)^{-1}},\ \theta^* + \frac{\Delta_n\left(\theta^*\right)}{\sqrt{n}} + z_{1-\frac{\alpha}{2}}\sqrt{j^*\left(\theta^*\right)^{-1}}\right) \quad \text{or} \tag{2.6}$$

$$(\theta_L,\, \theta_U) = \left(\hat{\theta} - z_{1-\frac{\alpha}{2}}\sqrt{j(\hat{\theta})^{-1}},\ \hat{\theta} + z_{1-\frac{\alpha}{2}}\sqrt{j(\hat{\theta})^{-1}}\right), \tag{2.7}$$

  where $z_{1-\frac{\alpha}{2}}$ is the $1-\frac{\alpha}{2}$ quantile from a standard normal distribution $Z \sim \mathcal{N}(0,\, 1)$.

2.1.2 Bayesian Censored Peaks Over a Threshold

Starting from a sample of observed extreme data $x_1, \ldots, x_n$, we assume that the underlying univariate random variables $X_1, \ldots, X_n$ are i.i.d. with common unknown distribution F, with F belonging to the max-domain of attraction of G_γ. Following the abovementioned asymptotic theory, we consider a Censored Peaks Over a Threshold (CPOT) misspecified model $\mathcal{F}$. Here, the function of the model is $p_X(x|\tilde{\theta})$ in (1.41) if we choose to opt for the $\tilde{\theta} = \left(\gamma,\, \tilde{\mu},\, \tilde{\delta}\right) \in (-1, +\infty) \times \mathbb{R} \times (0, +\infty)$ parameterization, or $p_X(x|\bar{\theta})$ in (1.44) in the other $\bar{\theta} = \left(\gamma,\, \bar{\mu},\, \bar{\delta}\right) \in (-1, +\infty) \times \mathbb{R} \times (0, +\infty)$ parameterization. Therefore, we are able to compute a *tilde* likelihood $\mathcal{L}(\gamma, \tilde{\mu}, \tilde{\delta}; x)$ in (1.40) or a *bar* likelihood $\mathcal{L}(\gamma, \bar{\mu}, \bar{\delta}; x)$ in (1.43).

We assume for the *tilde* parameters a data-dependent prior $\pi_{\tilde{\theta}}\left(\gamma, \tilde{\mu}, \tilde{\delta}\right)$, as in Padoan & Rizzelli (2022) for block maxima models. Justifications for the choice of this prior will be discussed in the next section. Let $\gamma, \tilde{\mu}$ and $\tilde{\delta}$ be prior independent, i.e. $\pi_{\tilde{\theta}}\left(\gamma, \tilde{\mu}, \tilde{\delta}\right) = \pi_{\tilde{\theta}}\left(\gamma\right) \times \pi_{\tilde{\theta}}\left(\tilde{\mu}\right) \times \pi_{\tilde{\theta}}\left(\tilde{\delta}\right)$ where:

- γ has density function

$$\begin{aligned}\pi_{\tilde{\theta}}\left(\gamma\right) &= \left(1 - T_1(-1)\right)^{-1} t_1(\gamma)\mathbb{I}\left(-1 < \gamma < \infty\right) \\ &= \left(\frac{1}{2} - \frac{1}{\pi}\arctan(-1)\right)^{-1} \frac{1}{\pi\left(1+\gamma^2\right)}\mathbb{I}(-1 < \gamma < +\infty),\end{aligned} \tag{2.8}$$

  where T_1 and t_1 stand for the Cauchy c.d.f. and p.d.f respectively, and $\mathbb{I}(\cdot)$ is the indicator function;

- $\tilde{\mu}$ is distributed according to a transformation of the Gaussian density with location $\hat{\hat{\mu}}$ and scale $\hat{\hat{\delta}}^2$, indeed

$$\pi_{\tilde{\theta}}(\tilde{\mu}) \propto \frac{1}{\hat{\hat{\delta}}^2} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left(\frac{\tilde{\mu} - \hat{\hat{\mu}}}{\hat{\hat{\delta}}} \right)^2 \right\}, \tag{2.9}$$

  where $\hat{\hat{\mu}} = \hat{b}(n/k)$ and $\hat{\hat{\delta}} = \hat{a}(n/k)$ are the maximum likelihood estimates of the norming costants of the assumed underlying GEV model;

- $\tilde{\delta}$ is distributed according to a transformation of the exponential density with scale parameter $\hat{\hat{\delta}}$, i.e.

$$\pi_{\tilde{\theta}}\left(\tilde{\delta}\right) \propto \frac{1}{\hat{\hat{\delta}}^2} \exp\left\{ -\frac{1}{\hat{\hat{\delta}}} \tilde{\delta} \right\} \mathbb{I}(0 < \tilde{\delta} < +\infty). \tag{2.10}$$

In order to derive the prior $\pi_{\bar{\theta}}\left(\gamma, \bar{\mu}, \bar{\delta}\right)$ in the *bar* parameterization, we have to employ the transformation formula between densities of random variables, starting from $\pi_{\tilde{\theta}}\left(\gamma, \tilde{\mu}, \tilde{\delta}\right)$ and exploiting the relations defined in equations (1.37) and (1.38). In particular, we consider the one-to-one transformation

$$g = (g_1, g_2, g_3)^T : (-1, +\infty) \times \mathbb{R} \times (0, +\infty) \longrightarrow (-1, +\infty) \times \mathbb{R} \times (0, +\infty)$$

between $(\gamma, \bar{\mu}, \bar{\delta})^T$ and $\left(\gamma, \tilde{\mu}, \tilde{\delta}\right)^T$ and its inverse $g^{-1} = \left(g_1^{-1}, g_2^{-1}, g_3^{-1}\right)^T$. These transformations are defined by:

$$\gamma = g_1\left(\gamma, \tilde{\mu}, \tilde{\delta}\right) = \gamma \quad \Longleftrightarrow \quad \gamma = g_1^{-1}\left(\gamma, \bar{\mu}, \bar{\delta}\right) = \gamma\,;$$

$$\bar{\mu} = g_2\left(\gamma, \tilde{\mu}, \tilde{\delta}\right) = \begin{cases} \tilde{\mu} - \tilde{\delta}\,\frac{1-s^{-\gamma}}{\gamma} & \text{if } \gamma \neq 0, \\ \tilde{\mu} - \tilde{\delta}\log(s) & \text{if } \gamma \to 0 \end{cases} \quad \Longleftrightarrow$$

$$\tilde{\mu} = g_2^{-1}\left(\gamma, \bar{\mu}, \bar{\delta}\right) = \begin{cases} \bar{\mu} + s^{\gamma}\bar{\delta}\,\frac{1-s^{-\gamma}}{\gamma} & \text{if } \gamma \neq 0, \\ \bar{\mu} + \bar{\delta}\log(s) & \text{if } \gamma \to 0; \end{cases}$$

$$\bar{\delta} = g_3\left(\gamma, \tilde{\mu}, \tilde{\delta}\right) = \begin{cases} s^{-\gamma}\tilde{\delta} & \text{if } \gamma \neq 0, \\ \tilde{\delta} & \text{if } \gamma \to 0 \end{cases} \quad \Longleftrightarrow \quad \tilde{\delta} = g_3^{-1}\left(\gamma, \bar{\mu}, \bar{\delta}\right) = \begin{cases} s^{\gamma}\bar{\delta} & \text{if } \gamma \neq 0, \\ \bar{\delta} & \text{if } \gamma \to 0, \end{cases}$$

where $s = \frac{n}{k}$ is the inverse of the proportion of the exceedances. Moreover, we compute the matrix of the first derivatives of g^{-1} with respect to $\left(\gamma, \bar{\mu}, \bar{\delta}\right)^T$ and its determinant:

$$\frac{\partial g^{-1}\left(\gamma, \bar{\mu}, \bar{\delta}\right)}{\partial \left(\gamma, \bar{\mu}, \bar{\delta}\right)^T} = \begin{cases} \begin{bmatrix} 1 & 0 & 0 \\ \frac{\bar{\delta}(s^\gamma \log(s) - s^\gamma + 1)}{\gamma} & 1 & s^\gamma \\ \bar{\delta} s^\gamma \log(s) & 0 & s^\gamma \end{bmatrix} & \text{if } \gamma \neq 0, \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \log(s) \\ 0 & 0 & 1 \end{bmatrix} & \text{if } \gamma \to 0; \end{cases}$$

$$\left| \frac{\partial g^{-1}\left(\gamma, \bar{\mu}, \bar{\delta}\right)}{\partial \left(\gamma, \bar{\mu}, \bar{\delta}\right)^T} \right| = \begin{cases} s^\gamma & \text{if } \gamma \neq 0, \\ 1 & \text{if } \gamma \to 0. \end{cases}$$

Then, the expression of $\pi_{\bar{\theta}}\left(\gamma, \bar{\mu}, \bar{\delta}\right)$ can be obtained through the transformation formula. In particular, if $\gamma \neq 0$, i.e. in the heavy and short-tailed cases:

$$\begin{aligned} \pi_{\bar{\theta}}\left(\gamma, \bar{\mu}, \bar{\delta}\right) &= \left| \frac{\partial g^{-1}\left(\gamma, \bar{\mu}, \bar{\delta}\right)}{\partial \left(\gamma, \bar{\mu}, \bar{\delta}\right)^T} \right| \pi_{\tilde{\theta}}\left(g^{-1}\left(\gamma, \bar{\mu}, \bar{\delta}\right)\right) \\ &= \left| \frac{\partial g^{-1}\left(\gamma, \bar{\mu}, \bar{\delta}\right)}{\partial \left(\gamma, \bar{\mu}, \bar{\delta}\right)^T} \right| \pi_{\tilde{\theta}}\left(g_1^{-1}\left(\gamma, \bar{\mu}, \bar{\delta}\right)\right) \times \pi_{\tilde{\theta}}\left(g_2^{-1}\left(\gamma, \bar{\mu}, \bar{\delta}\right)\right) \times \pi_{\tilde{\theta}}\left(g_3^{-1}\left(\gamma, \bar{\mu}, \bar{\delta}\right)\right) \\ &\propto s^\gamma \left(\frac{1}{2} - \frac{1}{\pi}\arctan(-1)\right)^{-1} \frac{1}{\pi\left(1+\gamma^2\right)} \mathbb{I}(-1 < \gamma < +\infty) \times \qquad (2.11) \\ &\quad \frac{1}{\hat{\bar{\delta}}^2} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left(\frac{\bar{\mu} + s^\gamma \bar{\delta} \frac{1-s^{-\gamma}}{\gamma} - \hat{\bar{\mu}}}{\hat{\bar{\delta}}} \right)^2 \right\} \times \qquad (2.12) \\ &\quad \frac{1}{\hat{\bar{\delta}}^2} \exp\left\{ -\frac{1}{\hat{\bar{\delta}}} s^\gamma \bar{\delta} \right\} \mathbb{I}(0 < \bar{\delta} < +\infty). \qquad (2.13) \end{aligned}$$

On the other hand, if $\gamma \to 0$, i.e. in the light-tailed case:

$$\begin{aligned}
\pi_{\bar{\theta}}\left(\gamma, \bar{\mu}, \bar{\delta}\right) &= \left|\frac{\partial g^{-1}\left(\gamma, \bar{\mu}, \bar{\delta}\right)}{\partial\left(\gamma, \bar{\mu}, \bar{\delta}\right)^{T}}\right| \pi_{\tilde{\theta}}\left(g^{-1}\left(\gamma, \bar{\mu}, \bar{\delta}\right)\right) \\
&= \left|\frac{\partial g^{-1}\left(\gamma, \bar{\mu}, \bar{\delta}\right)}{\partial\left(\gamma, \bar{\mu}, \bar{\delta}\right)^{T}}\right| \pi_{\tilde{\theta}}\left(g_1^{-1}\left(\gamma, \bar{\mu}, \bar{\delta}\right)\right) \times \pi_{\tilde{\theta}}\left(g_2^{-1}\left(\gamma, \bar{\mu}, \bar{\delta}\right)\right) \times \pi_{\tilde{\theta}}\left(g_3^{-1}\left(\gamma, \bar{\mu}, \bar{\delta}\right)\right) \\
&\propto \left(\frac{1}{2} - \frac{1}{\pi}\arctan(-1)\right)^{-1} \frac{1}{\pi\left(1+\gamma^2\right)} \mathbb{I}(-1 < \gamma < +\infty) \times \qquad (2.14) \\
&\quad \frac{1}{\hat{\tilde{\delta}}^2} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\bar{\mu} + \bar{\delta}\log(s) - \hat{\tilde{\mu}}}{\hat{\tilde{\delta}}}\right)^2\right\} \times \qquad (2.15) \\
&\quad \frac{1}{\hat{\tilde{\delta}}^2} \exp\left\{-\frac{1}{\hat{\tilde{\delta}}}\bar{\delta}\right\} \mathbb{I}(0 < \bar{\delta} < +\infty). \qquad (2.16)
\end{aligned}$$

We deduce that the *bar* parameters are no longer prior independent. This prior specification is in accordance with the likelihood shapes of Figures 1.3 and 1.4, and in general with the considerations about parameter orthogonality for the *tilde* and *bar* parameters made in the last paragraph of Section 1.4.

The posterior distributions of the parameters given the data x can be calculated in either parameterization:

$$\pi_{\tilde{\theta}}\left(\gamma, \tilde{\mu}, \tilde{\delta} | x\right) = \frac{\pi_{\tilde{\theta}}\left(\gamma, \tilde{\mu}, \tilde{\delta}\right) \mathcal{L}\left(\gamma, \tilde{\mu}, \tilde{\delta}; x\right)}{\int_{\tilde{\Theta}} \pi_{\tilde{\theta}}\left(\gamma, \tilde{\mu}, \tilde{\delta}\right) \mathcal{L}\left(\gamma, \tilde{\mu}, \tilde{\delta}; x\right) d\tilde{\theta}} \qquad \text{or} \qquad (2.17)$$

$$\pi_{\bar{\theta}}\left(\gamma, \bar{\mu}, \bar{\delta} | x\right) = \frac{\pi_{\bar{\theta}}\left(\gamma, \bar{\mu}, \bar{\delta}\right) \mathcal{L}\left(\gamma, \bar{\mu}, \bar{\delta}; x\right)}{\int_{\bar{\Theta}} \pi_{\bar{\theta}}\left(\gamma, \bar{\mu}, \bar{\delta}\right) \mathcal{L}\left(\gamma, \bar{\mu}, \bar{\delta}; x\right) d\bar{\theta}}. \qquad (2.18)$$

In the following, we will draw dependent samples from the posterior (2.17) or (2.18) via an Adaptive Multivariate Random-Walk Metropolis-Hastings algorithm, which will be better described in Section 2.3. Thus, we can isolate from the sampled joint posterior distribution the marginal posterior distributions of the single parameters and compute credible intervals, as in (2.4), (2.6) and (2.5). The `R` code for the prior and posterior distributions is in Appendix D.

2.2 Empirical Bayes approach

This section sets out some ideas about the Empirical Bayes (EB) approach and presents the EB framework of Padoan & Rizzelli (2022).

Suppose we are in a classical Bayesian inference framework, as described in Section 2.1.1, and the parameter of interest θ is priorly distributed according to $\pi(\theta;\lambda)$, which expresses a parametric model indexed by a hyperparameter λ. This prior needs to be completely specified by tuning λ with a reasonable value resulting from previous studies (if avalilable) or by adding another prior on this hyperparameter, which brings more complexity to the model and a higher computational cost. An alternative solution would be to use the Empirical Bayes (EB) approach.

Robbins (1956) initially proposed EB as a technique for estimating non-parametrically (some functionals of) the prior distribution $\pi(\theta)$ from the marginal distribution of the data. In the parametric context, the same approach can be helpful in tuning the hyperparameter λ with an estimate from the observed data $x_1, \ldots, x_n$, computed on the marginal distribution of X, i.e.

$$p_X(x;\lambda) = \int_\Theta p_X(x|\theta)\pi(\theta;\lambda)d\theta. \tag{2.19}$$

For instance, λ can be easily estimated from $p_X(x;\lambda)$ using maximum likelihood if λ is finite-dimensional. The estimate $\hat{\lambda}$ is then plugged-in into the posterior for θ, giving $\pi(\theta|x;\hat{\lambda}) \propto \pi(\theta;\hat{\lambda})\mathcal{L}(\theta;x)$, from which inference will be based. The Empirical Bayes approach is not accepted by many Bayesians since data are used twice and uncertainty about λ is not accounted for.

We have applied an EB reasoning in the Bayesian CPOT method (described in Section 2.1.2) for the task of choosing the hyperparameters for the prior distributions of $\tilde{\mu}$ and $\tilde{\delta}$. In particular, we can imagine that the *tilde* parameters were orginally distributed according to

$$\pi_{\tilde{\theta}}\left(\tilde{\mu};\, b(n/k), a(n/k)\right) \propto \frac{1}{a^2(n/k)} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\tilde{\mu} - b(n/k)}{a(n/k)}\right)^2\right\} \tag{2.20}$$

$$\text{and} \qquad \pi_{\tilde{\theta}}\left(\tilde{\delta};\, a(n/k)\right) \propto \frac{1}{a^2(n/k)} \exp\left\{-\frac{1}{a(n/k)}\,\tilde{\delta}\right\} \mathbb{I}(0 < \tilde{\delta} < +\infty), \tag{2.21}$$

since it seems logical to think at the location constant $b(n/k)$ as the mean of $\tilde{\mu}$ and at the scale constant $a(n/k)$ as the mean of $\tilde{\delta}$ and the standard deviation of $\tilde{\mu}$. In a pure frequentist framework, these constants can be consistently estimated from the data by maximization of the likelihood in (1.39). Recalling that $\hat{b}(n/k) = \hat{\tilde{\mu}}$ and $\hat{a}(n/k) = \hat{\tilde{\delta}}$ and plugging-in these estimates into (2.20) and (2.21), by EB we arrive at the priors in (2.9) and (2.10), respectively.

Further justifications on the use of EB in the extreme value setting can be found

in Section 3 of Padoan & Rizzelli (2022) for block maxima models. The main results of this theory imply that, under some mild conditions on the data generating process (satisfied by all extreme value distributions considered in this work) and on the prior distribution (satisfied by the aforementioned prior), the posterior distributions for $\tilde{\theta}$ in Equation (2.17) and for $\bar{\theta}$ in (2.18):

- provide consistent estimation of the unknown true parameter $\left(\gamma^0, \tilde{\mu}^0, \tilde{\delta}^0\right)$ (or $\left(\gamma^0, \bar{\mu}^0, \bar{\delta}^0\right)$). In addition, we have that the asymptotic models $GEV^{\frac{1}{s}}\left(\gamma^0,\, \tilde{\mu}^0,\, \tilde{\delta}^0\right)$ and $GEV\left(\gamma^0,\, \bar{\mu}^0,\, \bar{\delta}^0\right)$ give consistent estimation of the unknown density of the peaks over a threshold $m_{s,1}, \ldots, m_{s,k}$;
- are asymptotically normal as $s \to \infty$.

Furthermore, also the derived marginal posterior distribution for the return level y_p, i.e. $\pi\left(y_p|x\right)$ concentrates around the true return level value y_p^0. The degree of this concentration depends on the tail heaviness of F: the lighter is the tail of F, the narrower is the neighbourhood of y_p^0 on which $\pi\left(y_p|x\right)$ concentrates. Also, the marginal posterior distribution of the return level is asymptotically normal as $s \to \infty$.

2.3 Computational Aspects

In this section we will discuss some algorithmic and computational aspects of the development of our Empirical Bayes Censored Peaks Over a Threshold (CPOT) approach, fully described in Sections 1.4, 2.1.2 and 2.2. This tools have been applied to assess the frequentist performances of the derived inference (empirical coverage of 95% credible intervals) from our approach in the simulation study of Chapter 3.

Given a sample of extreme data $x_1, \ldots, x_n$, in order to obtain the posterior distribution of the *tilde* and *bar* parameters defined in Equations (2.17) and (2.18), the Markov Chain Monte Carlo (MCMC) procedure explained in Section 4 of Padoan & Rizzelli (2022) has been adapted in the present context. In particular, we sampled from the posterior distribution through a Random-Walk Metropolis-Hastings (RWMH) algorithm (Metropolis et al., 1953; Hastings, 1970), refined in a way that it adapts the scaling parameter κ and the covariance matrix Σ of the proposal distribution $\mathcal{N}_3\left(\theta^{(i)},\, \kappa^{(i)}\Sigma^{(i)}\right)$ at each iteration $i+1$ of the algorithm with the objective of reaching a fixed optimal Overall Acceptance Probability (OAP) η^*. Theoretical justifications behind this Adaptive MH algorithm can be found in Garthwaite et al. (2016) and Haario et al. (2001).

2.3.1 The Metropolis-Hastings (MH) algorithm

This section refers to Chapter 6 of Robert & Casella (2010).

MCMC methods exploit simulation from Markov chains to approximate the shape of an objective distribution, e.g. a posterior distribution whose integral in the denominator is too difficult to compute, either analytically or numerically.

A Markov chain $\{X^{(t)}\}$ is a sequence of dependent random variables $X^{(0)}, X^{(1)}, \ldots, X^{(t)}, \ldots$ over the sample space $\mathcal{X}$ such that the probability distribution of $X^{(t)}$ given the past variables depends only on $X^{(t-1)}$. This conditional probability distribution is called a transition kernel or a Markov kernel K; that is $X^{(t+1)}|X^{(0)}, X^{(1)}, \ldots, X^{(t)} \sim K(X^{(t)}, X^{(t+1)})$. For example, a simple random walk Markov chain satisfies $X^{(t+1)} = X^{(t)} + \epsilon_t$, where $\epsilon_t \sim \mathcal{N}(0,1)$, indipendently of $X^{(t)}$; therefore, the Markov kernel $K(X^{(t)}, X^{(t+1)})$ corresponds to a $\mathcal{N}(X^{(t)}, 1)$ density. A stationary probability distribution f exists for a Markov chain $\{X^{(t)}\}$ if $X^{(t)} \sim f$ implies $X^{(t+1)} \sim f$, i.e. if

$$\int_{\mathcal{X}} K(x,y)f(x)dx = f(y).$$

A Markov chain is recurrent if it returns to any arbitrary nonnegligible set an infinite number of times. Recurrent chains are ergodic, i.e. the stationary distribution f is also a limiting distribution in the sense that the limiting distribution of $X^{(t)}$ is f for almost any initial value $X^{(0)}$. From a simulation point of view, if a given kernel K produces an ergodic Markov chain with stationary distribution f, then generating a chain from K yields to a sequence of dependent observation with f as marginal distribution. In particular, it holds for this Markov chain the Law of Large Numbers for dependent variables (the Ergodic Theorem, proved by Birkhoff (1931)):

$$\frac{1}{T}\sum_{t=1}^{T} h\left(X^{(t)}\right) \xrightarrow{p} \mathbb{E}_f\left[h(X)\right], \tag{2.22}$$

where h is an integrable function.

The Metropolis-Hastings (MH) algorithm is a MCMC method that, for almost any target density f, is able to build a Markov kernel K with stationary distribution f and then generate from K a Markov chain $\{X^{(t)}\}$ with limiting distribution f. Averages taken from the observed chain fulfill the Ergodic Theorem. The algorithm do not actually simulates from f, but from the proposal conditional density $q(x|y)$, which is easier to simulate. The proposal q must cover the entire support of f and the ratio $f(x)/q(x|y)$ must be known up to a constant independent of y. Under these moderate requirements, q can be chosen arbitrarily. Hence, provided the length of the desired chain R and the

initial value $x^{(0)}$ of the chain, a general MH algorithm follows the structure in Algorithm 1:

Algorithm 1: Metropolis-Hastings

Set R and $x^{(0)}$;

for $i = 1$ **to** R **do**

draw proposal $X^* \sim q\left(x^*|x^{(i)}\right)$;

compute acceptance probability $\eta^{(i)} = \min\left(\frac{f(x^*)}{f(x^{(i)})}\frac{q(x^{(i)}|x^*)}{q(x^*|x^{(i)})}, 1\right)$;

take $X^{(i+1)} = \begin{cases} x^* & \text{with probability } \eta^{(i)}, \\ x^{(i)} & \text{with probability } 1-\eta^{(i)}. \end{cases}$

Algorithm 1 satisfies the so-called detailed balance condition $f(x^*)K(x^{(i)}|x^*) = f(x^{(i)})K(x^*|x^{(i)})$

$$\implies f(x^*)\min\left(\frac{f(x^{(i)})}{f(x^*)}\frac{q(x^*|x^{(i)})}{q(x^{(i)}|x^*)}, 1\right) = f(x^{(i)})\min\left(\frac{f(x^*)}{f(x^{(i)})}\frac{q(x^{(i)}|x^*)}{q(x^*|x^{(i)})}, 1\right),$$

from which we can deduce that f is the stationary distribution of the chain $\{X^{(t)}\}$. When q is symmetric. i.e. $q(x^*|x^{(i)}) = q(x^{(i)}|x^*)$, the acceptance probability simplifies into

$$\eta^{(i)} = \min\left(\frac{f(x^*)}{f(x^{(i)})}, 1\right). \tag{2.23}$$

One important property of the Metropolis-Hastings algorithm is that it depends on f only through the ratio $f(x^*)/f(x^{(i)})$ in the computation for $\eta^{(i)}$. This means that it is not necessary to completely specify f, indeed its norming constant can be neglected and this is the crucial issue in Bayesian inference. Via the MH algorithm one can simulate a sample of slightly positive dependent data from any given posterior whose just the numerator is known.

The Overall Acceptance Probability (OAP) of the algorithm is nothing but the average through the R iterations of the acceptance probabilities $\eta^{(i)}$. In general, the acceptance probability $\eta^{(i)}$ is high when the proposed value x^* is near the current last element of the chain $x^{(i)}$ and it is low when they are distant. The aim is to build a Markov chain respecting the following properties:

1. it has entirely explored the support of the target distribution f or, in other words, it has performed a good mixing. The marginal trace plots graphically check this condition;

2. it has reached stationarity and therefore presents f as marginal distribution. It is common use, after the run of the MH algorithm, to discard the burn-in, an initial piece of the obtained chain (whose length must be chosen) in order to consider only the values that have reached stationarity. An histogram of relative frequencies of the post-burn-in chain can underline its resemblance with f;

3. once stationary, it mantains a sufficiently low degree of autocorrelation between consecutives observations (at different lags). This element can be pointed out by an autocorrelation plot, which should show an exponential decay through increasing lags, typical behaviour in stationary series.

Section 2.4 will give plenty of such diagnostic plots.

In practice, the performance of the Metropolis-Hastings strongly depends on the structure of the proposal distribution q. In particular, the proposal can be chosen in order to take into account the value previously simulated to generate the following value. Random walks (RW) are proposals of this type, like the guassian RW where $X^* \sim \mathcal{N}\left(x^{(i)}, \sigma^2\right)$ and the uniform RW where $X^* \sim \mathcal{U}\left(x^{(i)} - \sigma, x^{(i)} + \sigma\right)$, in unidimensional settings. The additional parameter σ has to be tuned and it controls the spread of the proposed values around the latest element of the chain. These proposals induce a local stepwise exploration of the sample space $\mathcal{X}$, which is appropriate when a good initial value $x^{(0)}$ is known and for target distributions f with even complex shapes. In fact, the local moves proposed by q avoid that the chain remains stuck into single values for a long time, but favor more homogeneous explorations of the support of f. In addition, random walk proposals are often symmetric distributions, leading to the simplification in Equation (2.23).

The parameter σ regulates the closeness between successive steps of a RWMH algorithm. Low variance of q induces the next proposed value to fall in a narrow neighbourhood of the latest value of the chain and to be accepted with high probability. Resulting chains have high OAP, are highly autocorrelated and require a relatively big amount of time for completely exploring $\mathcal{X}$. On the other hand, fixing σ to a too high value induces the proposal to pick up values in the support of f which may be really far from the latest element of the chain, and hence, are very unlikely to be accepted. Resulting chains have low OAP, are highly autocorrelated and remain stuck into single values of $\mathcal{X}$ for several iterations, requiring more time for a good mixing, as well. The choice of the tuning parameter σ must reflect a trade-off between these two opposite situations, aimed at obtaining from a RWMH algorithm Markov chains with the right properties in

a reasonable amount of time. Theory supplies some optimal Overall Acceptance Probabilities that have to be looked for in tuning the proposal's dispersion parameter σ. Roberts & Rosenthal (2001) suggest a value of the OAP of 0.44 for univariate RWMH, instead Gelman et al. (1997) recommend 0.234 for the multivariate case.

Adaptive Random-Walk Metropolis-Hastings algorithms avoid the necessity of tuning σ, adapting its value through iterations in a way that induces a required OAP. For this reason, in the simulation study of Chapter 3 we have sampled from the *tilde* and *bar* posterior distributions with an Adaptive RWMH, which will be described in detail in the next section.

2.3.2 An Adaptive Metropolis-Hastings (AMH) algorithm

This section refers to Padoan & Rizzelli (2022), Garthwaite et al. (2016) and Haario et al. (2001).

An efficient way to draw samples from the posterior distributions for $\tilde{\theta}|x$ and $\bar{\theta}|x$ defined in (2.17) and (2.18) is by the means of an Adaptive Gaussian Random-Walk Metropolis-Hastings (RWMH) algorithm. Padoan & Rizzelli (2022) used the algorithmic scheme discussed in Garthwaite et al. (2016), which is a special case of the AMH class of algorithms introduced by Haario et al. (2001), for the Bayesian Block Maxima approach to extreme values. Here, we extend its use to the Censored POT approach. This Adaptive Gaussian Random-Walk Metropolis-Hastings algorithm is summarised below in Algorithm 2.

Algorithm 2: Adaptive Gaussian Random-Walk Metropolis-Hastings

Initialize: Set R, $\theta^{(0)}$, $\kappa^{(0)}$ and $\Sigma^{(0)}$;

for $i = 1$ **to** R **do**

- draw proposal $\theta^* \sim \mathcal{N}_d\left(\theta^{(i)}, \kappa^{(i)}\Sigma^{(i)}\right)$;
- compute acceptance probability $\eta^{(i)} = \min\left(\frac{\pi(\theta^*)\mathcal{L}(\theta^*;x)}{\pi(\theta^{(i)})\mathcal{L}(\theta^{(i)};x)}, 1\right)$;
- draw $U \sim \mathcal{U}(0,1)$;
- **if** $\eta^{(i)} > U$ **then**
  - set $\theta^{(i+1)} = \theta^*$;
- **else**
  - set $\theta^{(i+1)} = \theta^{(i)}$;
- update $\Sigma^{(i+1)}$ according to Equation (2.25);
- update $\kappa^{(i+1)}$ according to Algorithm 3.

A short summary of the algorithm is as follows. From now on, with the generic symbol θ we will relate to either $\tilde{\theta} = \left(\gamma, \tilde{\mu}, \tilde{\delta}\right)$ or $\bar{\theta} = \left(\gamma, \bar{\mu}, \bar{\delta}\right)$. The current state of the chain $\theta^{(i)}$ at iteration i is potentially updated by the proposal $\theta^* \sim \mathcal{N}_3\left(\theta^{(i)}, \kappa^{(i)}\Sigma^{(i)}\right)$. Given the symmetry of the proposal, the acceptance probability of the update $\theta^{(i+1)} = \theta^*$ reduces to

$$\eta^{(i)} = \min\left(\frac{\pi(\theta^*)\mathcal{L}(\theta^*;\, x)}{\pi(\theta^{(i)})\mathcal{L}(\theta^{(i)};\, x)},\, 1\right), \tag{2.24}$$

otherwise set $\theta^{(i+1)} = \theta^{(i)}$ with probability $1 - \eta^{(i)}$.

Following Haario et al. (2001), the proposal covariance matrix $\Sigma^{(i+1)}$ is specified in order to be proportional to the current estimate of the covariance matrix of the running chain (Craiu et al., 2009) plus a noise term which avoids singularity. The update of $\Sigma^{(i+1)}$ is

$$\Sigma^{(i+1)} = \begin{cases} \Sigma^{(0)} & \text{if } i \leq i_t, \\ \frac{1}{i-1}\sum_{s=1}^{i}\left(\theta^{(s)} - \bar{\theta}^{(i)}\right)\left(\theta^{(s)} - \bar{\theta}^{(i)}\right)^T + (1/i)I_d & \text{if } i > i_t, \end{cases} \tag{2.25}$$

where $d = 3$ is the dimension of the parameter θ, I_d is the d-dimensional identity matrix and $\bar{\theta}^{(i)} = \frac{\theta^{(1)}+\cdots+\theta^{(i)}}{i}$ is the average of the obtained chain until iteration i. The bound i_t reflects our trust in the initial covariance matrix $\Sigma^{(0)}$ and in our algorithm we have set it to 100. We recall that, for $i > i_t$, the updates for $\bar{\theta}^{(i)}$ and $\Sigma^{(i+1)}$ can be efficiently computed via recursion, i.e.

$$\bar{\theta}^{(i)} = \frac{(i-1)\bar{\theta}^{(i-1)} + \theta^{(i)}}{i}, \tag{2.26}$$

$$\Sigma^{(i+1)} = \frac{i-2}{i-1}\,\Sigma^{(i)} + \bar{\theta}^{(i-1)}\left(\bar{\theta}^{(i-1)}\right)^T - \frac{i}{i-1}\,\bar{\theta}^{(i)}\left(\bar{\theta}^{(i)}\right)^T + \frac{1}{i-1}\,\theta^{(i)}\left(\theta^{(i)}\right)^T + \frac{1}{d}\,I_d. \tag{2.27}$$

Since the update of $\Sigma^{(i+1)}$ involves information on the running chain, the AMH algorithm is non-Markovian, but Haario et al. (2001) established that it has the correct ergodic properties anyway.

When d is large, a substantial number of iterations of the Markov chain may be needed before the update $\Sigma^{(i+1)}$ stabilizes. Monitoring of the trace plots of the parameters should be carried out to ensure that the updates of $\Sigma^{(i+1)}$ have stabilized.

The scaling parameter $\kappa^{(i)} > 0$ affects the acceptance probability $\eta^{(i)}$ and its starting value $\kappa^{(1)}$ can be chosen arbitrarily, e.g. $\kappa^{(1)} = 1$. According to Garthwaite et al. (2016), $\kappa^{(i+1)}$ can be adaptively updated using a Robbins-Monro process. The Robbins-Monro process (Robbins & Monro, 1951) is a stochastic search algorithm that, essentially,

increases $\kappa^{(i+1)}$ if the previous MCMC proposed move was accepted and decreases it otherwise, in a way that the obtained sequence of $\kappa^{(i)} \to \kappa^*$, the value of the scaling parameter that yields the desired OAP η^*. The update of $\kappa^{(i+1)}$ is such that

$$\kappa^{(i+1)} = \exp\left\{\log\left(\kappa^{(i)}\right) + a\frac{\eta^{(i)} - \eta^*}{\max\{200,\, i/d\}}\right\}, \tag{2.28}$$

where, in our case, $\eta^* = 0.234$ (Gelman et al., 1997),

$$a = \left(1 - \frac{1}{d}\right)\sqrt{2\pi}\,\frac{\exp\left\{\frac{\zeta_0^2}{2}\right\}}{2\zeta_0} + \frac{1}{d\eta^*\left(1-\eta^*\right)}$$

is a steplength constant with $\zeta_0 = -\Phi^{-1}\left(\eta^*/2\right)$ and Φ represents the c.d.f. of a univariate standard normal random variable. The multiplier of a is called step size and reduces its magnitude as i increases.

The Robbins-Monro process can be monitored and a search for $\kappa^{(i+1)}$ restarted if the starting value k_s seems poor (Garthwaite, 1996). Otherwise $\kappa^{(i)} \to \kappa^*$ can take a long time to converge as the step size decreses with i, as shown in (2.28). On a restart, the most recent update $\kappa^{(i)}$ is taken as the starting value and the value of i is reset. We start (and restart) a search with $i = n_0$, where n_0 is a moderate size so as to avoid too rapid steplength a changes in the early stages of the search. We choose n_0 to be the integer closest to $5/\eta^*(1-\eta^*)$, which typically works well in practice. We also choose to restart the search if the update $\kappa^{(i+1)}$ changes by a factor of 3 from its value from when the search started (or last restarted). To ensure that the search do not continually restart, the algorithm records the number of restarts resulting from κ tripling and reducing in value by two-thirds. Should both these numbers reach 5, then the process is not restarted again. We also do not restart if more than 100 steps have been taken since the last restart, as taking 100 steps without restarting suggests a reasonable starting point has been used. These decision rules are arbitrary but typically work well in practice.

The sequence of operations needed for an update $\kappa^{(i+1)}$ with restart control has been summarised in Algorithm 3.

Algorithm 3: Adaptive κ update via Robbins-Monro process with restart control

Initialize: Set $\kappa_s = \kappa^{(1)}$, $iMax = 100$, $n.up = 0$, $n.low = 0$, $t.up = 0$ and $t.low = 0$;

set OAP $\eta^* = 0.234$;

set $n_0 = \lfloor 5/(\eta^*(1-\eta^*)) \rceil$;

set $\zeta_0 = -\Phi^{-1}(\eta^*/2)$;

set $a = \left(1 - \frac{1}{d}\right)\sqrt{2\pi}\,\frac{e^{\frac{\zeta_0^2}{2}}}{2\zeta_0} + \frac{1}{d\eta^*(1-\eta^*)}$;

if $i > n_0$ **then**

  update $\kappa^{(i+1)} = \exp\left(\log(\kappa^{(i)}) + a(\eta - \eta^*)/\max\{200, i/d\}\right)$;

  set $k = \kappa^{(i+1)}$;

  if $(i \leq iMax + n_0)$ & $(n.up < 5 \,|\, n.low < 5)$ **then**

    $t.up = \kappa > 3\kappa_s$;

    $t.low = \kappa < \kappa_s/3$;

    if $t.up \,|t.low$ **then**

      $n.big = n.big + t.big$;

      $n.low = n.low + t.low$;

      $i = n_0$;

      $\kappa_s = k$.

For the process to converge to the correct value of κ^*, it must not converge before the update of $\Sigma^{(i+1)}$ stabilizes. Similarly, many sampler iterations are typically needed to effectively explore the parameter space Θ when there are many parameters, and proposal acceptance probabilities $\eta^{(i)}$ may vary dramatically over this space. Hence, even after the update of $\Sigma^{(i+1)}$ is fairly stable, the update of $\kappa^{(i+1)}$ should converge slowly if it is to reflect the overall optimum for the parameter space. To achieve this, as it is possible to see in (2.28), the magnitude of the steps $\eta^{(i)} - \eta^*$ of the process is not reduced below some pre-fixed limit until the update of $\Sigma^{(i+1)}$ is reasonably stable, and after that the step size is reduced slowly.

2.4 Three complete analyses on simulated data

We now perform three complete Bayesian analyses aimed at applying our Empirical Bayes POT method to large samples of simulated extreme data from each one of the classes of distributions belonging to the GEV family. In particular, we consider the unit Fréchet model for the Fréchet type ($\gamma > 0$), the unit Exponential model for the Gumbel

type ($\gamma = 0$) and the Power-Law model for the Reverse-Weibull type ($\gamma < 0$). We only report our analyses in the *tilde* parameterization, as we will compare the performances between the two different parameterizations later in Chapter 3. Every computation has been carried out with the software `R` (R Core Team, 2022). The analyses are focused on:

- checking the validity of the asymptotic theory, especially results in (1.35) and (1.36) of Section 1.4 and considerations at the end of Section 2.2. To do this, we generally rely on histograms of the generated samples and of the obtained posterior chains;

- computing MCMC posterior chains using the Adaptive RWMH Algorithm 2, whose `R` code is in Appendix E. Specifically, we are interested in the posterior distribution for the parameters γ, $\tilde{\mu}$ and $\tilde{\delta}$, the level p extreme quantile $\tilde{y}_p$ (1.49) and the return level $\tilde{y}_T$ associated with the return period T (1.50). Trace plots, autocorrelation plots and histograms of the posterior chains are provided;

- examining the adaptive path of the algorithm. To underline this aspect, we render series plots of the scaling parameter k and of the Overall Acceptance Probability of the latter half of the running iterations;

- Making Bayesian inference and evaluating its performances. In this context, we use a function that performs frequentist or Bayesian inference assuming the Censored POT model, whose code is in Appendix F. In particular, we inspect whether the true (and known) parameter value is included in the corresponding HPD 95% credible interval (2.5);

- monitoring the convergence of our MCMC algorithm. We sample different chains from different starting points. The computational burden of this operation may be lightened through parallelization. Convergence plots and diagnostics are obtained via the functions from the `coda` package by Plummer et al. (2006). Special attention is paid to the Gelman & Rubin (1992) diagnostic statistics, e.g. the potential scale reduction factor and the effective sample size.

  The potential scale reduction factor idicates if the obtained parallel chains mix in the same entire support of the target posterior distribution or they have not converged yet. Absence of convergence may be due to the fact that different chains may get stuck into local maxima of the target distribution near their starting point. In this situation an increase in the chains' length R can be helpful in reaching

convergence. The potential scale reduction factor is calculated in this way:

$$PSRF = \sqrt{\frac{\frac{(R-1)W+B}{R}}{W}}, \tag{2.29}$$

where W stands for the variability within the chains and B for the variance between. If a good mixing has been performed, the fraction of between variance $B/R \to 0$, as $R \to \infty$, and therefore our desired $PSRF$ is as close to 1 as possible. When $PSRF > 1.1$ or $PSRF > 1.2$, we should increase R in order to improve convergence.

We recall that MCMC chains are not i.i.d., but positively correlated, therefore a sample of size R contains less information than an i.i.d. sample of the same size. The effective sample size is the i.i.d. sample numerosity that would represent the obtained chains in terms of informativity. It is defined by

$$EFS = R\,\frac{1-\hat{\rho}}{1+\hat{\rho}}, \tag{2.30}$$

where $\hat{\rho}$ is an estimate of the correlation between the states of the chain assuming an AR(1) process with parameter ρ. We wish for high EFS.

We set the sample size $n = 10000$ and the number of exceedances $k = 500$, hence the inverse of the proportion of excesses is $s = 20$. We set the level of the extreme quantile $p = 0.0001$ and the return period $T = 50$. We finally set the length of posterior chains $R = 60000$, the burn-in at 10000 units and we simulate 5 different chains for convergence diagnostics.

2.4.1 Fréchet model

We recall the Fréchet density function and distribution function:

$$f(x;\alpha,\mu,\sigma)=\frac{\alpha}{\sigma}\left(\frac{x-\mu}{\sigma}\right)^{-1-\alpha}e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha}},\qquad F(x;\alpha,\mu,\sigma)=e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha}},$$

where $x>\mu$.

We set the true parameter value $(\alpha,\mu,\sigma)=(1,0,1)$ and we can derive through calculations in Section 1.5 the true values of the norming constants $b(n/k)=19.50$ and $a(n/k)=20.00$. The true parameter value of the *tilde* asymptotic model for threshold exceedances $GEV^{\frac{1}{s}}\left(\gamma,\tilde{\mu},\tilde{\delta}\right)$ is $(1,19.50,20.00)$, while its value for the *bar* asymptotic model for the excesses $GEV\left(\gamma,\bar{\mu},\bar{\delta}\right)$ is $(1,0.49,1.00)$. The true value of the requested extreme quantile is $y_p=9999.5$, while the true desired return level is $y_T=989.67$.

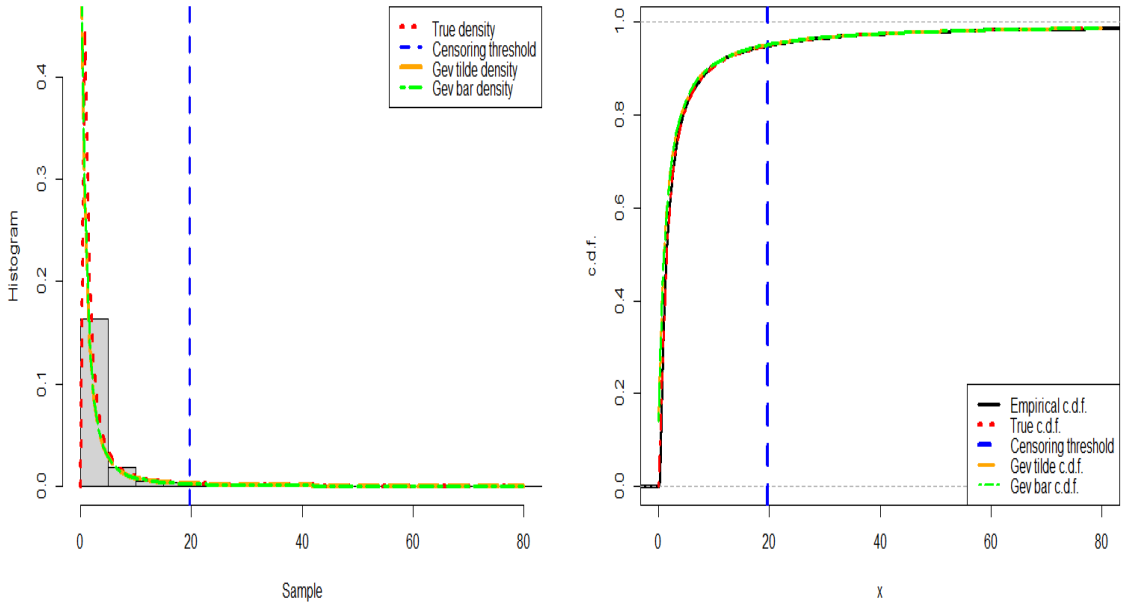
We simulate a random sample of size 10000 from the Fréchet distribution via the `R` package `evd` (Stephenson, 2002). We set the threshold u at the 95th percentile of the observed sample, so $u=19.71$.

Figure 2.1 shows the empirical distribution of the data at hand. In addition, the blue line is the censoring threshold u, the red curve relates to the true generating process $Fre(1,0,1)$ and the orange and green curves stand for the true asymptotic models for threshold exceedances $GEV^{\frac{1}{20}}(1,19.50,20.00)$ and $GEV(1,0.49,1.00)$. Here, we can see that the true model and the asymptotic models overlap in the entire support and become indistinguishable beyond the threshold. This fact supports the asymptotic theory.

From now on, we assume for the observed sample the *tilde* Censored POT model with likelihood in (1.39). We perform an initial frequentist analysis. We compute by maximum likelihood a point estimate of the parameters, the extreme quantile and the return level and we obtain $(0.98,20.24,21.02)$, 9576.81 and 987.05.

If we further assume the Empirical Bayes prior of Section 2.1.2, we can operate a full Bayesian analysis. We set the starting point for the adaptive MH algorithm at $(0.1,0.1,1)$, we compute posterior chains of length 60000 in 342.22 seconds and we only keep the 50000 post burn-in samples. The obtained Overall Acceptance Probability is 0.1986. From Figure 2.2 we see that the scaling parameter κ of the proposal collapses towards 0 in the very first iterations, but then increases roughly linearly. The running OAP seems to stabilize to a lower value than the desired $\eta^*=0.234$ through iterations of the algorithm. Here, adaptivity is not completely fulfilled.

The posterior mode is $(0.98,20.20,20.90)$, while the posterior mean is $(0.94,20.26,22.03)$. Figures 2.3 and 2.4 display trace plots, autocorrelation plots and histograms of the posterior distributions for $\gamma,\tilde{\mu},\tilde{\delta},\tilde{y}_{0.0001}$ and $\tilde{y}_{50}$. We generally denote a good



((A)) Fréchet p.d.f. ((B)) Fréchet c.d.f.

FIGURE 2.1: Simulated density and distribution function with their threshold exceedances asymptotic approximations GEV *tilde* and GEV *bar*.

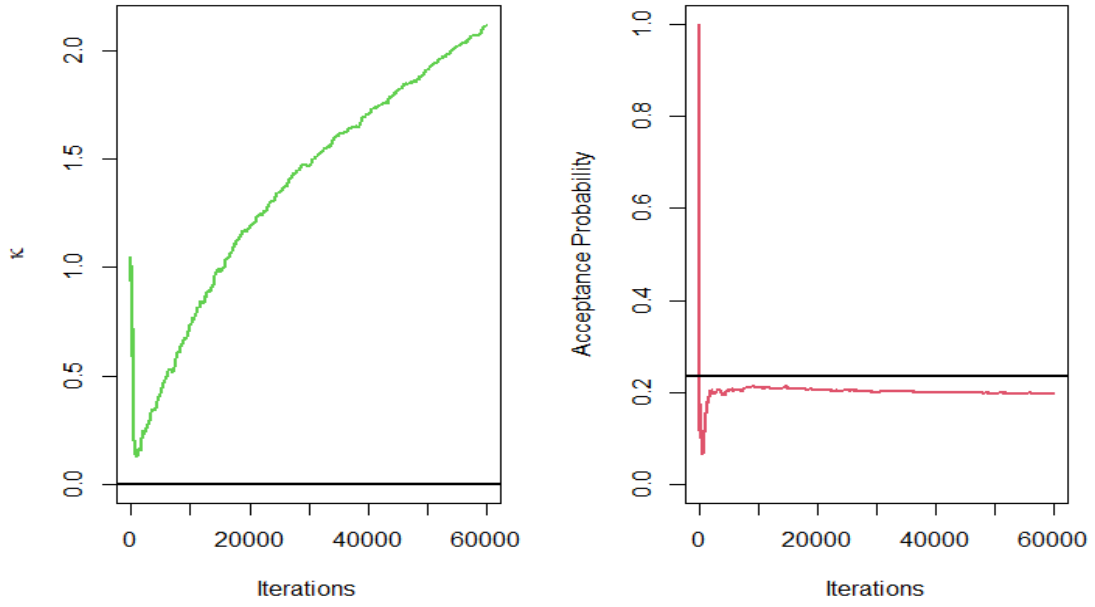


FIGURE 2.2: Fréchet model: scaling constant κ and Overall Acceptance Probability (OAP) through the iterations of the running adaptive MCMC algorithm.

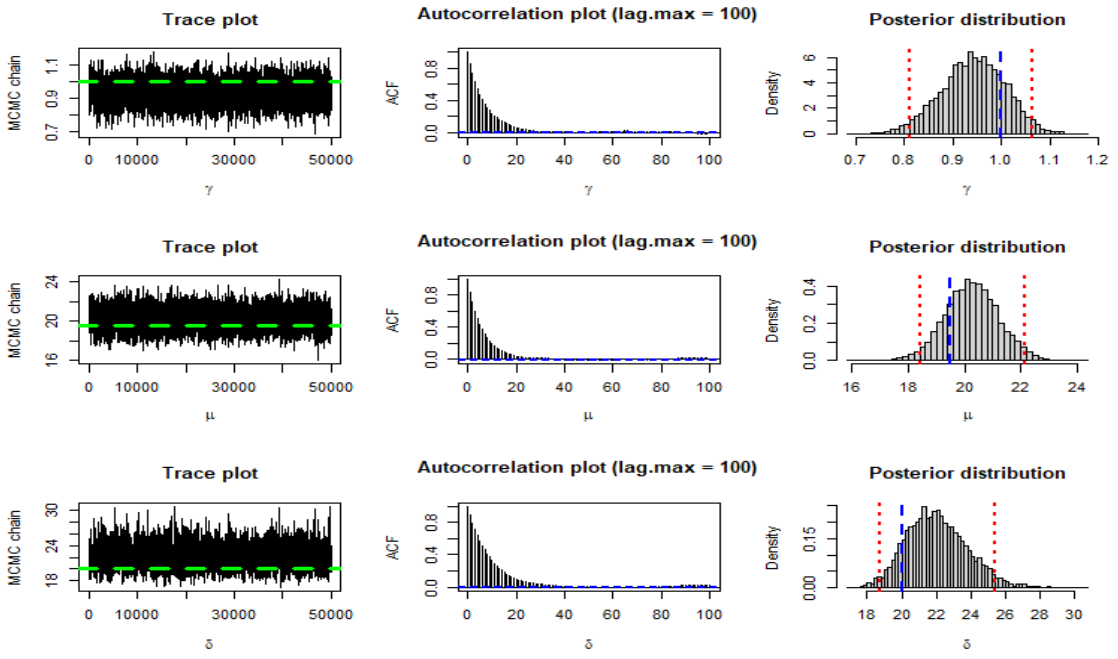
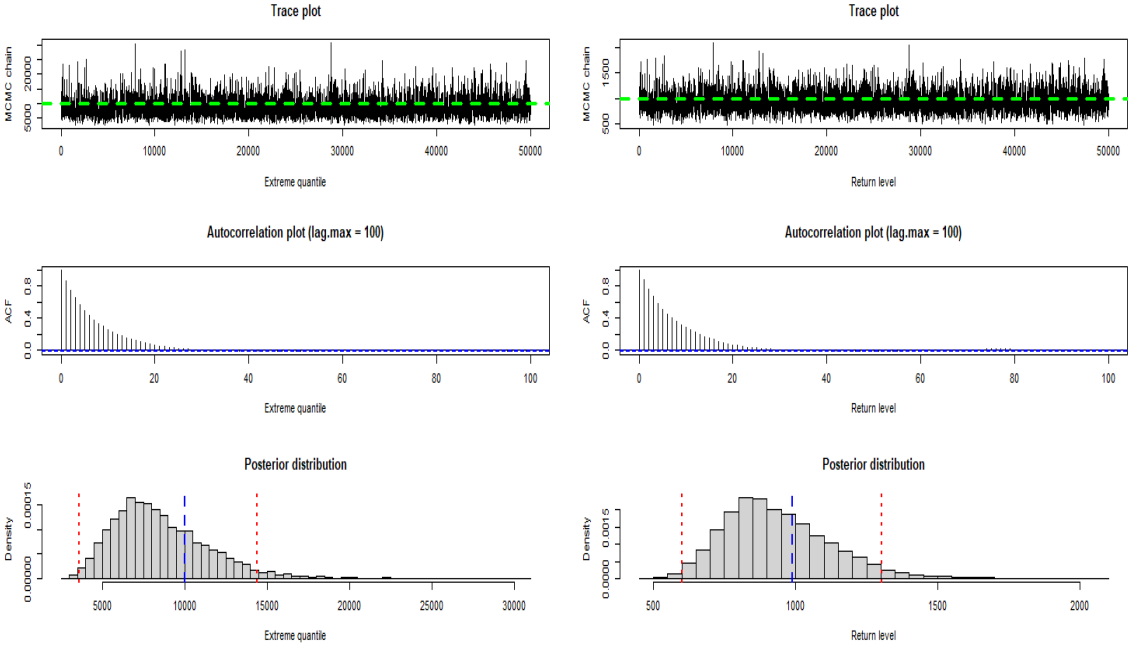
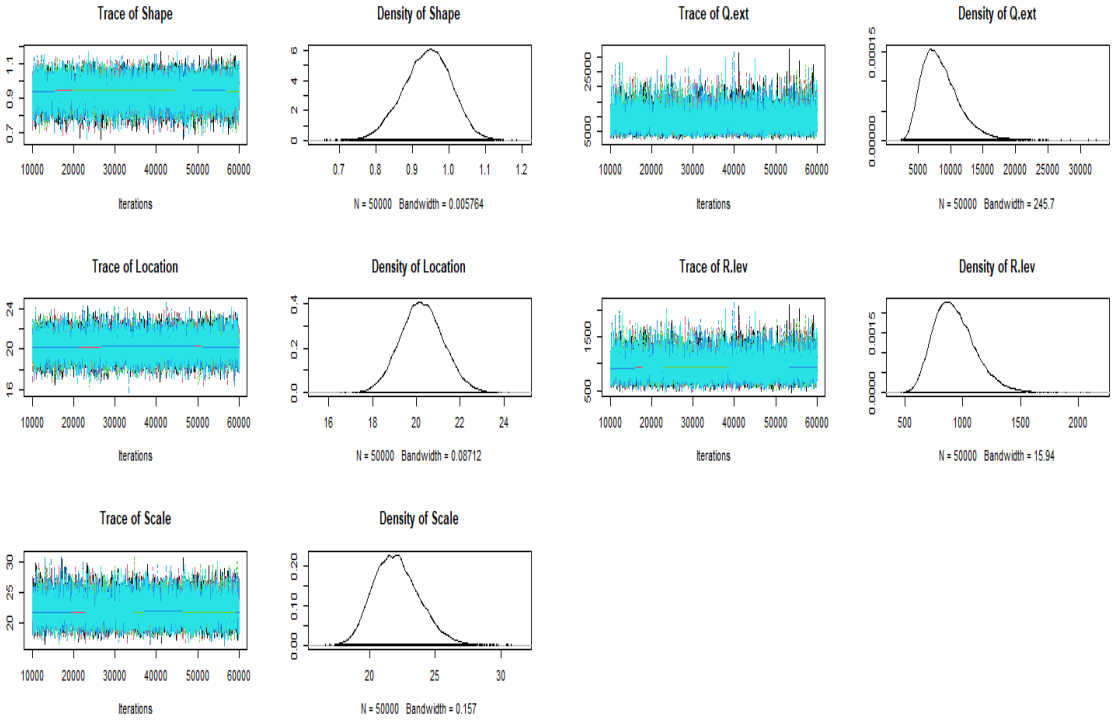


FIGURE 2.3: Fréchet model: trace plot, autocorrelation plot and histogram of the posterior chains for γ, $\tilde{\mu}$ and $\tilde{\delta}$. Dashed lines stand for the true parameter value, red dotted lines for the 95% HPD credible interval.



((A)) $\tilde{y}_p$ ((B)) $\tilde{y}_T$

FIGURE 2.4: Fréchet model: trace plot, autocorrelation plot and histogram of the posterior chains for the extreme quantile and the return level. Dashed lines stand for the true parameter value, red dotted lines for the 95% HPD credible interval.



((A)) $\gamma, \tilde{\mu}, \tilde{\delta}$ ((B)) $\tilde{y}_p, \tilde{y}_T$

FIGURE 2.5: Fréchet model: trace plots and histograms of the 5 parallel chains.

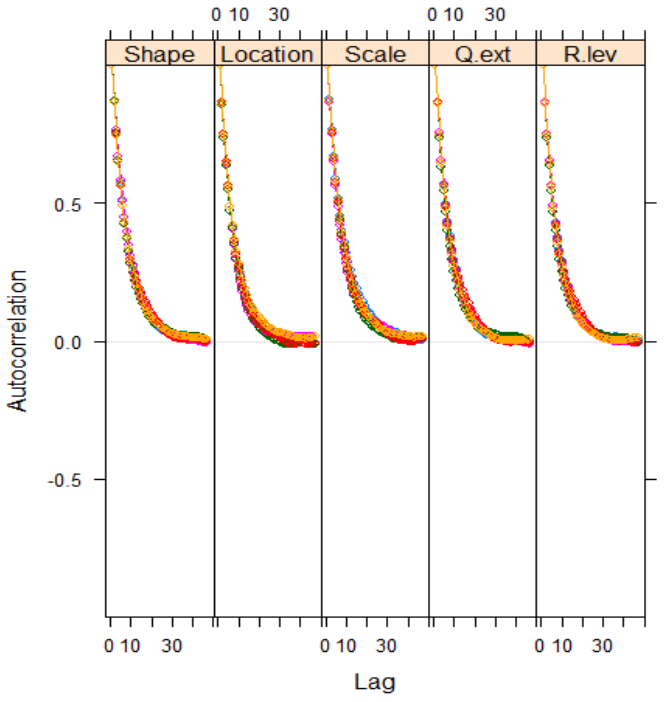


FIGURE 2.6: Fréchet model: autocorrelation plot of the 5 parallel chains.

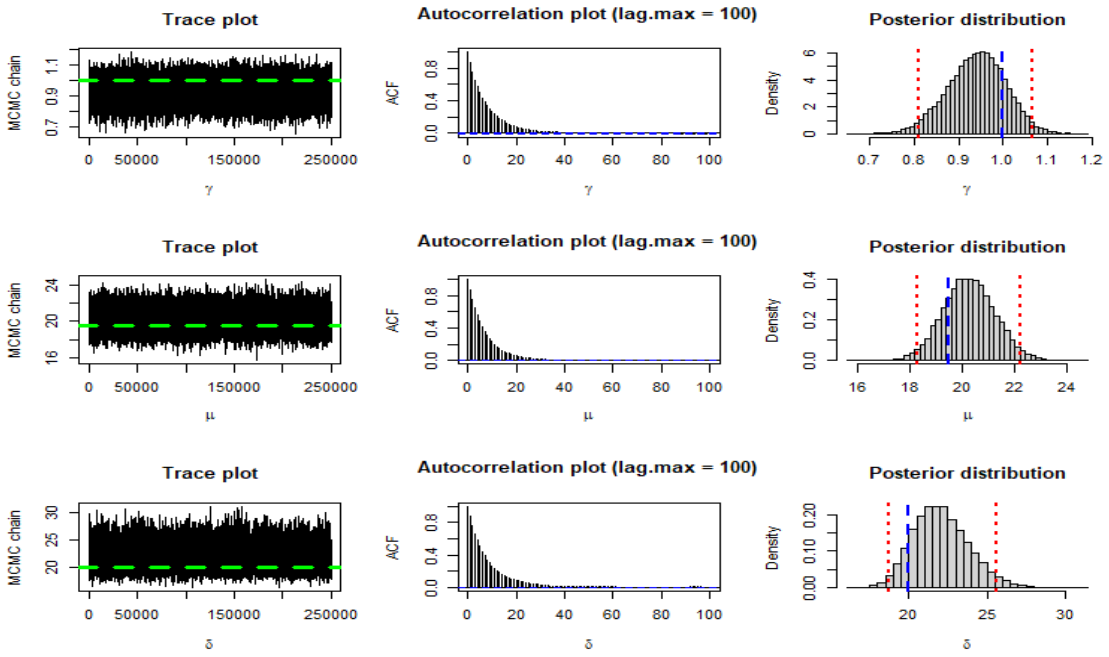
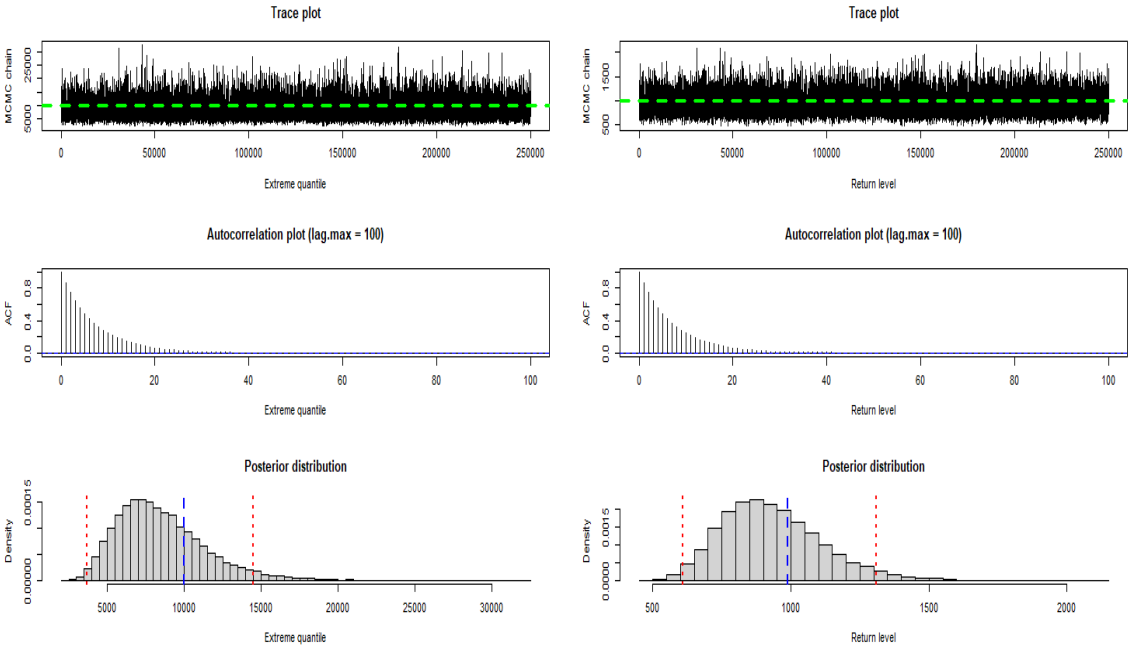


FIGURE 2.7: Fréchet model: trace plot, autocorrelation plot and histogram of the final posterior chains for γ, $\tilde{\mu}$ and $\tilde{\delta}$. Dashed lines stand for the true parameter value, red dotted lines for the 95% HPD credible interval.



((A)) $\tilde{y}_p$ ((B)) $\tilde{y}_T$

FIGURE 2.8: Fréchet model: trace plot, autocorrelation plot and histogram of the final posterior chains for the extreme quantile and the return level. Dashed lines stand for the true parameter value, red dotted lines for the 95% HPD credible interval.

mixing and appropriate correlation structure for each one of these quantities. We can observe by histograms that none of the posterior distributions is centered at its true parameter value. In addition, these empirical distributions are all skewed, expect for $\tilde{\mu}$. Nevertheless, 95% HPD credible intervals contain the true parameter value.

To assess the convergence of the algorithm we sample 5 different analogue chains from different starting points. These points are generated from the large sample approximation of the posterior distribution in (2.3) evaluated at the posterior mode θ^*. Figures 2.5 and 2.6 show a good mixing of the chains in the entire support of the posterior distribution and a similar stationary behaviour of autocorrelation. The potential scale reduction factor is 1 with upper confidence limit equal to 1 for every one of the 5 quantities under investigation. With this information, we are able to affirm that these chains seem to have reached stationarity and they seem to have properly explored the same support of the posterior distribution. The averaged effective sample size of these 5 parallel (post burn-in) chains of length 50000 is 17042.77 for γ, 18206.81 for $\tilde{\mu}$, 16708.51 for $\tilde{\delta}$, 17959.65 for the extreme quantile and 17981.84 for the return level. Althought the evident loss of information with respect to an i.i.d. sample, the dependent chains are still worthy for exctracting inference from them.

Therefore, we put together the post burn-in portions of these 5 chains to obtain posterior samples of size 250000. Figures 2.7 and 2.8 show trace plots, autocorrelation plots and histograms with overwritten the corresponding true parameter value and the 95% HPD credible interval. In addition to what has already been said we point out that these posterior samples present the same bias of the previous smaller samples, though their asymmetry has been faded in favor of a smoother shape. Every high posterior density 95% credible interval still includes its true parameter value.

2.4.2 Exponential model

We recall the exponential density function and distribution function:

$$f(x;\sigma) = \frac{1}{\sigma}\, e^{-\frac{1}{\sigma}\, x}, \qquad F(x;\sigma) = 1 - e^{-\frac{1}{\sigma}\, x},$$

where $x > 0$.

We set the true parameter value $\sigma = 1$ and we can derive through calculations in Section 1.5 the true values of the norming constants $b(n/k) = 3.00$ and $a(n/k) = 1$. The true parameter value of the *tilde* asymptotic model for threshold exceedances $GEV^{\frac{1}{s}}\left(\gamma, \tilde{\mu}, \tilde{\delta}\right)$ is $(0, 3.00, 1)$, while its value for the *bar* asymptotic model for the excesses $GEV\left(\gamma, \bar{\mu}, \bar{\delta}\right)$ is $(0, 0, 1)$. The true value of the requested extreme quantile is $y_p = 9.21$, while the true desired return level is $y_T = 6.90$.

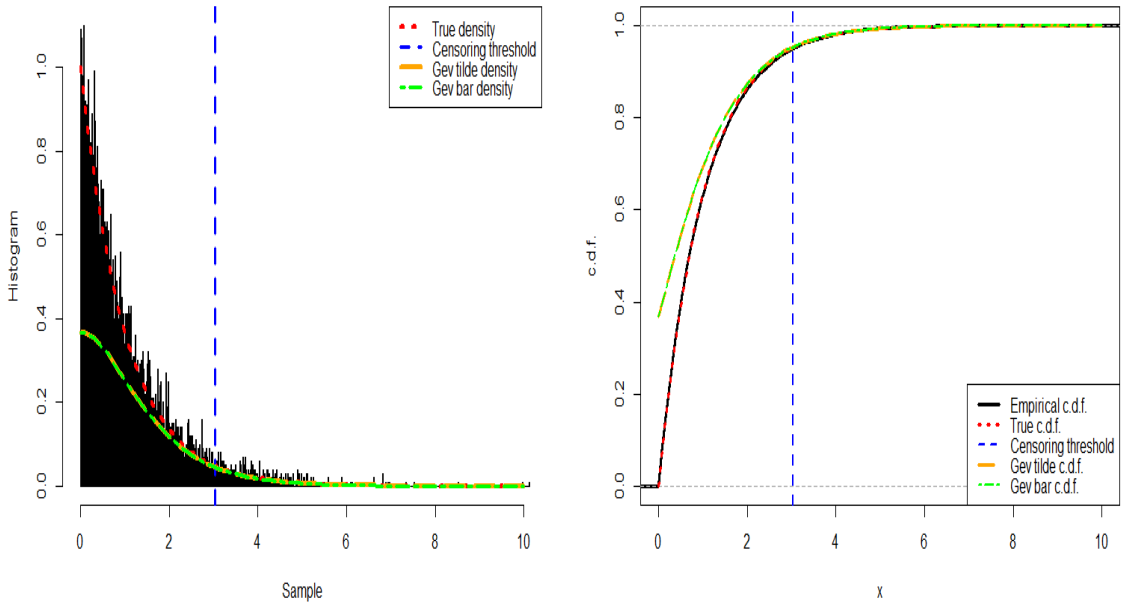
We simulate a random sample of size 10000 from the exponential distribution. We set the threshold u at the 95th percentile of the observed sample, so $u = 3.04$.

Figure 2.9 shows the empirical distribution of the data at hand. In addition, the blue line is the censoring threshold u, the red curve relates to the true generating process $Exp(1)$ and the orange and green curves stand for the true asymptotic models for threshold exceedances $GEV^{\frac{1}{20}}\,(0, 3.00, 1)$ and $GEV\,(0, 0, 1)$. Here, we can see that the true model and the asymptotic models start overlapping after the value of the threshold. This fact supports the asymptotic theory.

From now on, we assume for the observed sample the *tilde* Censored POT model with likelihood in (1.39). We perform an initial frequentist analysis. We compute by maximum likelihood a point estimate of the parameters, the extreme quantile and the return level and we obtain $(-0.0049, 3.06, 0.98)$, 9.04 and 6.84.

If we further assume the Empirical Bayes prior of Section 2.1.2, we can operate a full Bayesian analysis. We set the starting point for the adaptive MH algorithm at $(0.1, 1, 1)$, we compute posterior chains of length 60000 in 438.39 seconds and we only keep the 50000 post burn-in samples. The obtained Overall Acceptance Probability is 0.2455. From Figure 2.10 we see that the scaling parameter κ of the proposal collapses towards 0 in the very first iterations, but then increases with a decreasing rate. The OAP seems to converge from above to $\eta^* = 0.234$ through iterations of the algorithm. Here, we get a good adaptivity.

The posterior mode is $(-0.0030, 3.06, 0.98)$, while the posterior mean is $(0.0023, 3.06, 0.98)$. Figures 2.11 and 2.12 display trace plots, autocorrelation plots and histograms of the posterior distributions for $\gamma, \tilde{\mu}, \tilde{\delta}, \tilde{y}_{0.0001}$ and $\tilde{y}_{50}$. We generally denote a good mixing and appropriate correlation structure for each one of these quantities. We can



((A)) Exponential p.d.f. ((B)) Exponential c.d.f.

FIGURE 2.9: Simulated density and distribution function with their threshold exceedances asymptotic approximations GEV *tilde* and GEV *bar*.

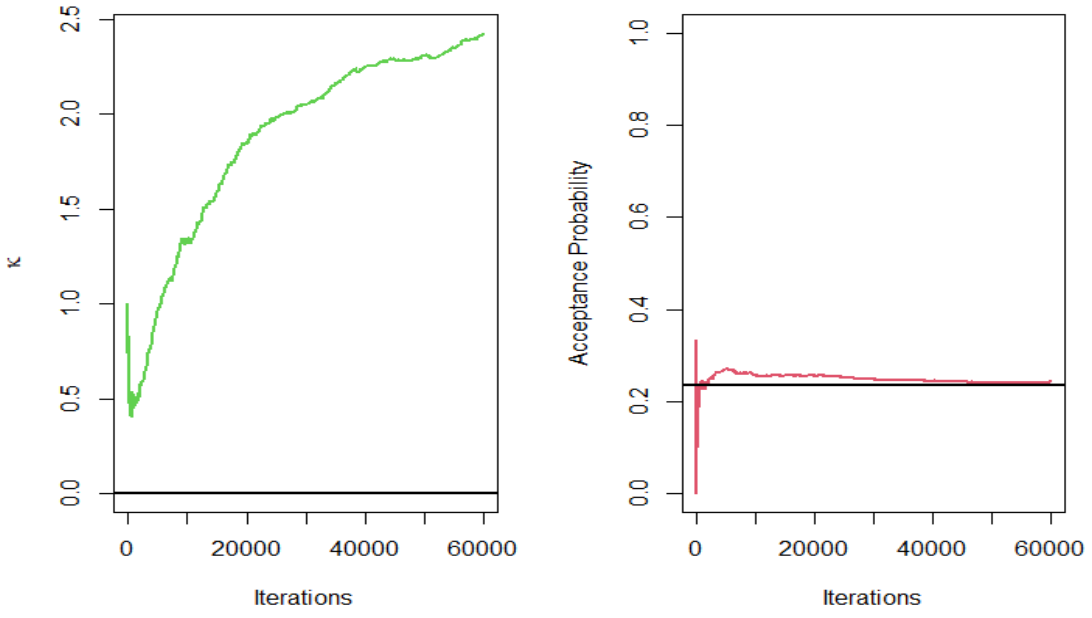


FIGURE 2.10: Exponential model: scaling constant κ and Overall Acceptance Probability (OAP) through the iterations of the running adaptive MCMC algorithm.

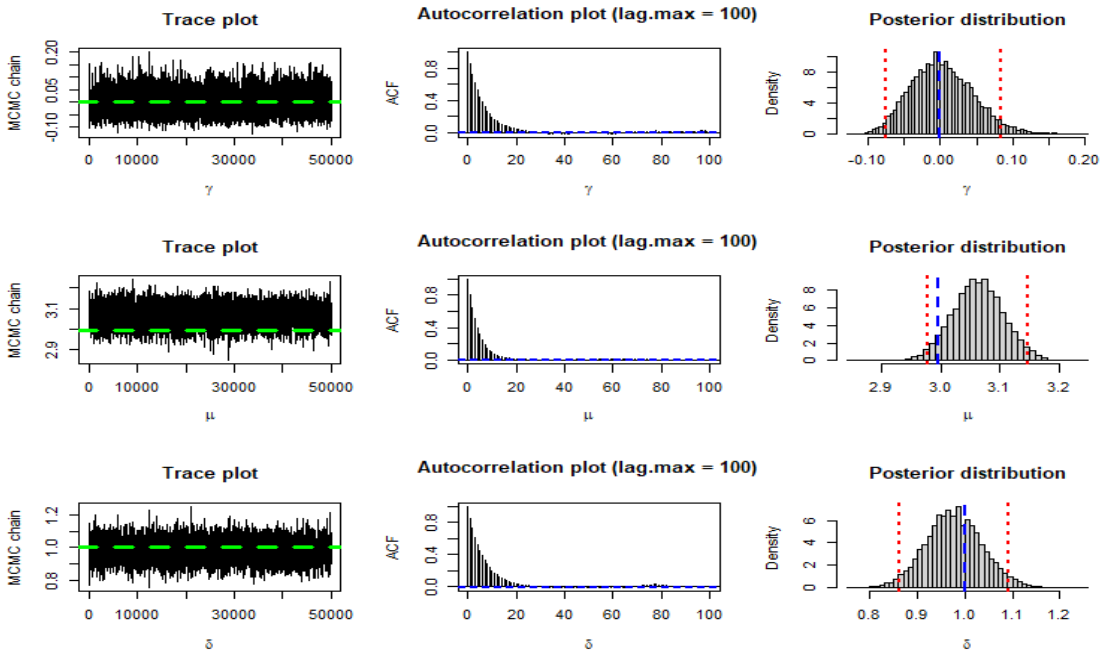
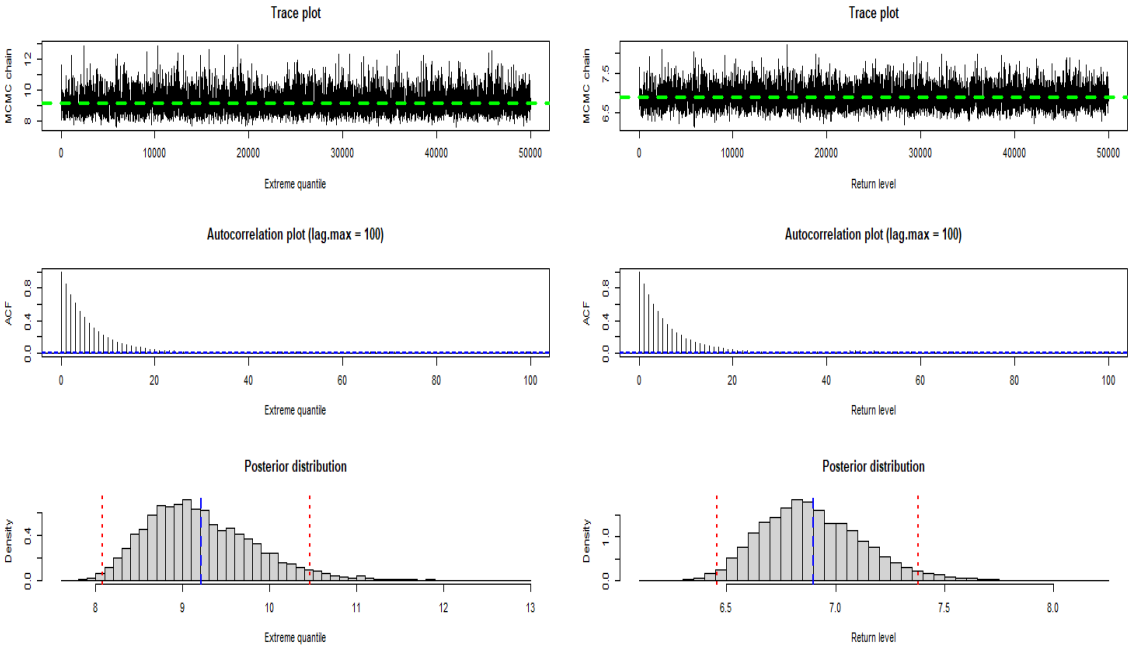
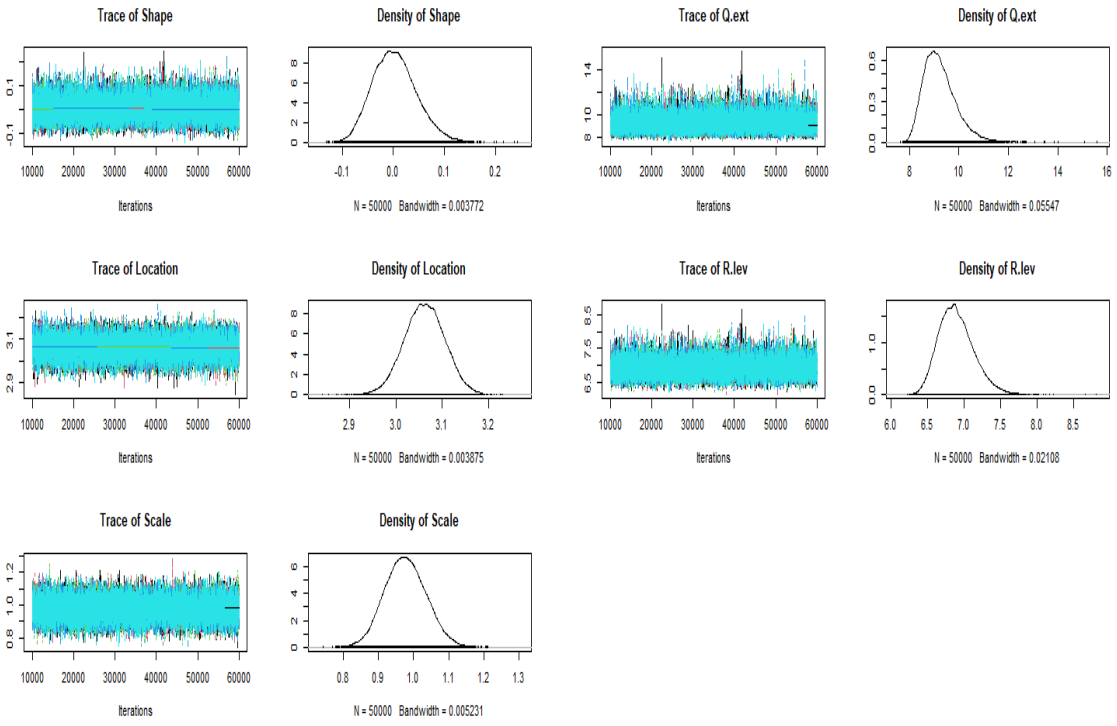


FIGURE 2.11: Exponential model: trace plot, autocorrelation plot and histogram of the posterior chains for γ, $\tilde{\mu}$ and $\tilde{\delta}$. Dashed lines stand for the true parameter value, red dotted lines for the 95% HPD credible interval.



((A)) $\tilde{y}_p$ ((B)) $\tilde{y}_T$

FIGURE 2.12: Exponential model: trace plot, autocorrelation plot and histogram of the posterior chains for the extreme quantile and the return level. Dashed lines stand for the true parameter value, red dotted lines for the 95% HPD credible interval.



((A)) $\gamma, \tilde{\mu}, \tilde{\delta}$ ((B)) $\tilde{y}_p, \tilde{y}_T$

FIGURE 2.13: Exponential model: trace plots and histograms of the 5 parallel chains.

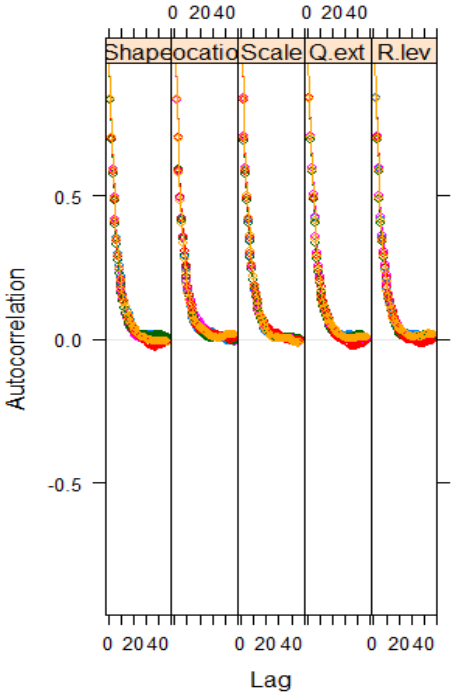


FIGURE 2.14: Exponential model: autocorrelation plot of the 5 parallel chains.

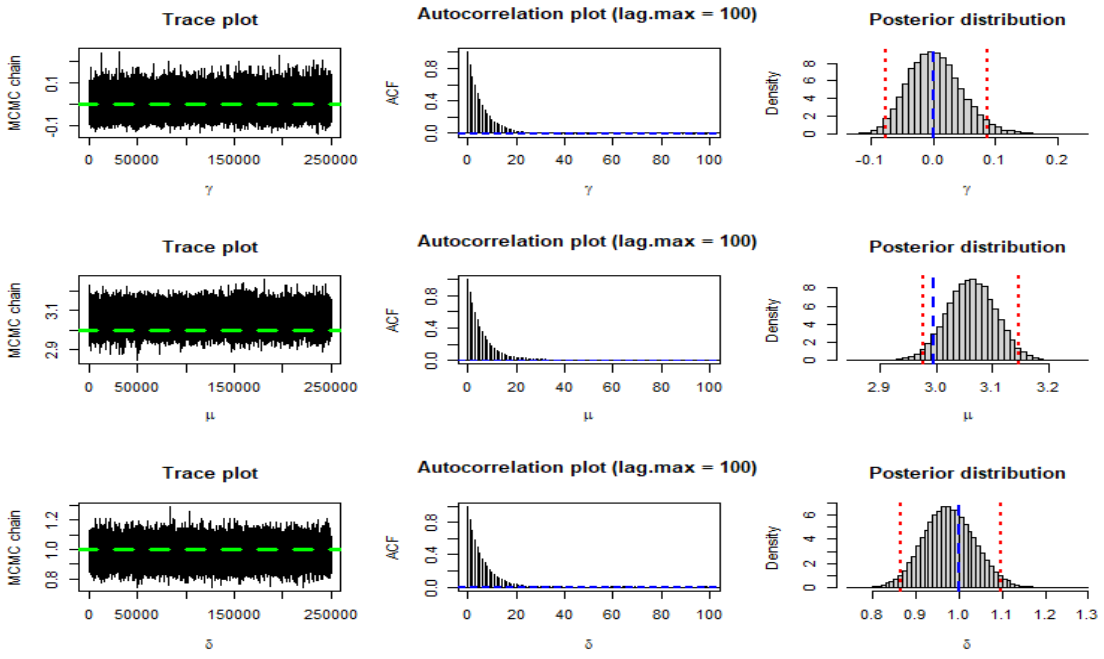
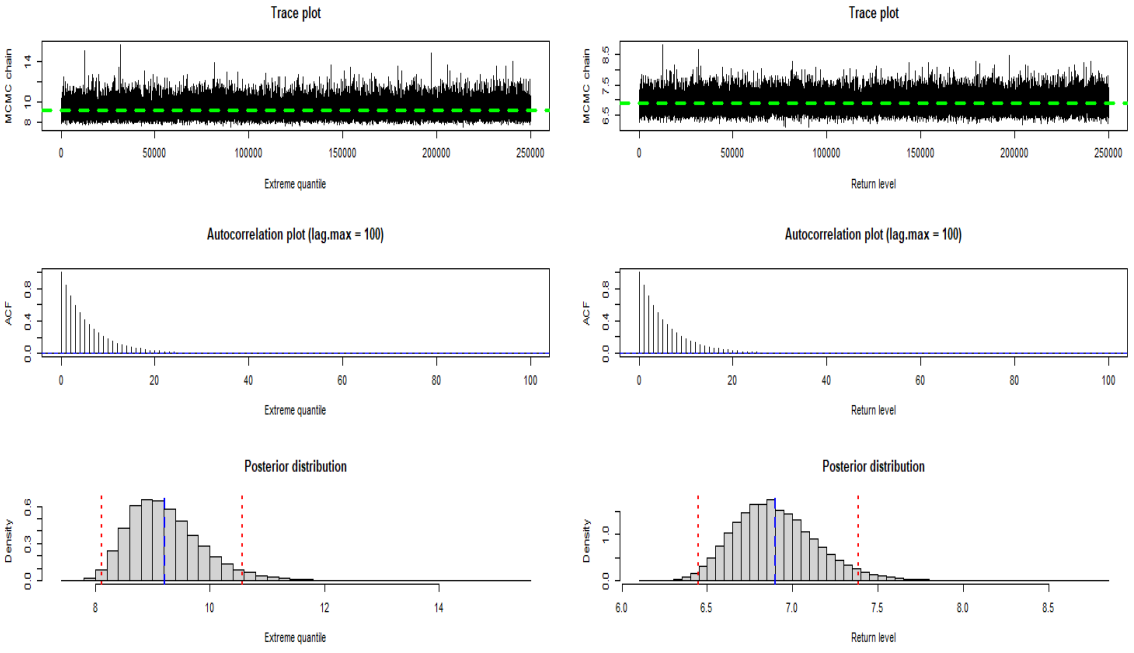


FIGURE 2.15: Exponential model: trace plot, autocorrelation plot and histogram of the final posterior chains for γ, $\tilde{\mu}$ and $\tilde{\delta}$. Dashed lines stand for the true parameter value, red dotted lines for the 95% HPD credible interval.



((A)) $\tilde{y}_p$ ((B)) $\tilde{y}_T$

FIGURE 2.16: Exponential model: trace plot, autocorrelation plot and histogram of the final posterior chains for the extreme quantile and the return level. Dashed lines stand for the true parameter value, red dotted lines for the 95% HPD credible interval.

observe by histograms that only the posterior for $\tilde{\mu}$ is particularly uncentered at its true parameter value. In addition, these empirical distributions seem to already behave normally at different levels. Furthermore, 95% HPD credible intervals contain the true parameter value.

To assess the convergence of the algorithm we sample 5 different analogue chains from different starting points. These points are generated from the large sample approximation of the posterior distribution in (2.3) evaluated at the posterior mode θ^*. Figures 2.13 and 2.14 show a good mixing of the chains in the entire support of the posterior distribution and a similar stationary behaviour of autocorrelation. The potential scale reduction factor is 1 with upper confidence limit equal to 1 for every one of the 5 quantities under investigation. With this information, we are able to affirm that these chains seem to have reached stationarity and they seem to have properly explored the same support of the posterior distribution. The averaged effective sample size of these 5 parallel (post burn-in) chains of length 50000 is 22267.20 for γ, 21939.58 for $\tilde{\mu}$, 21864.77 for $\tilde{\delta}$, 21878.25 for the extreme quantile and 21968.94 for the return level. Althought the evident loss of information with respect to an i.i.d. sample, the dependent chains are still worthy for extracting inference from them.

Therefore, we put together the post burn-in portions of these 5 chains to obtain posterior samples of size 250000. Figures 2.15 and 2.16 show trace plots, autocorrelation plots and histograms with overwritten the corresponding true parameter value and the 95% HPD credible interval. In addition to what has already been said we point out that the posterior sample for $\tilde{\mu}$ presents the same bias of the previous smaller sample and the smoothness of every posterior distribution is evidently improved. Every high posterior density 95% credible interval still includes its true parameter value.

2.4.3 Power-Law model

We recall the Power-Law density function, distribution function and quantile function:

$$f(x; x^*, \alpha, K) = K\,\alpha\,(x^* - x)^{\alpha-1}, \qquad F(x; x^*, \alpha, K) = 1 - K\,(x^* - x)^{\alpha},$$
$$q(p; x^*, \alpha, K) = x^* - \left(\frac{1-p}{K}\right)^{\frac{1}{\alpha}},$$

where $x < x^*$ and $K = 1/9$.

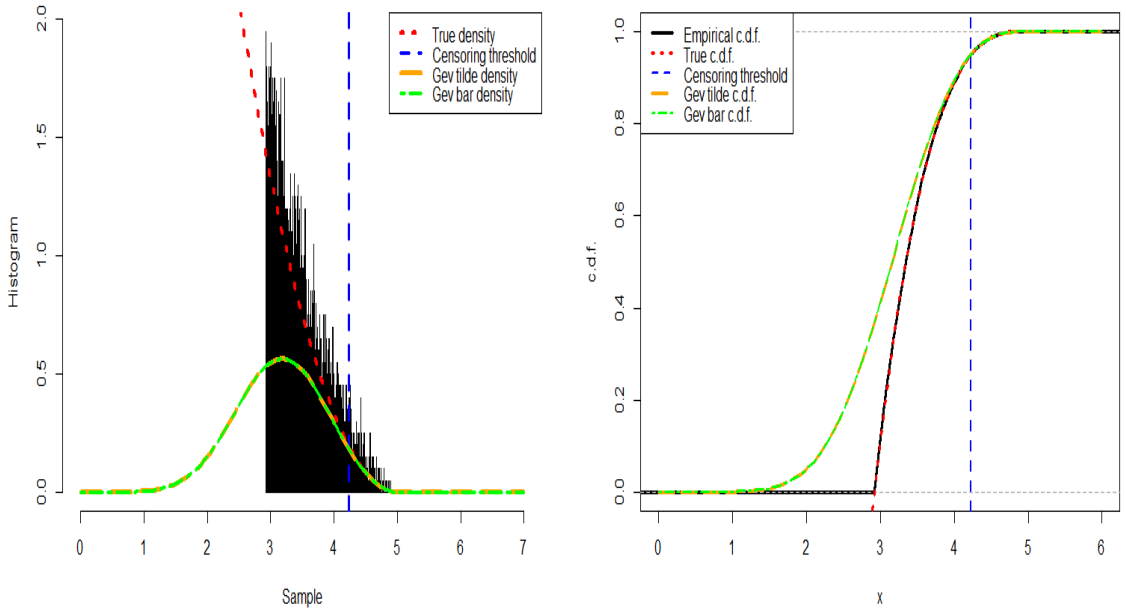
We set the true parameter value $(x^*, \alpha) = (5, 3)$ and we can derive through calculations in Section 1.5 the true values of the norming constants $b(n/k) = 4.23$ and $a(n/k) = 0.26$. The true parameter value of the *tilde* asymptotic model for threshold exceedances $GEV^{\frac{1}{s}}\left(\gamma, \tilde{\mu}, \tilde{\delta}\right)$ is $(-0.33, 4.23, 0.26)$, while its value for the *bar* asymptotic model for the excesses $GEV\left(\gamma, \bar{\mu}, \bar{\delta}\right)$ is $(-0.33, 2.92, 0.69)$. The true value of the requested extreme quantile is $y_p = 4.90$, while the true desired return level is $y_T = 4.79$.

We simulate a random sample of size 10000 from the Power-Law distribution through its quantile function by the inversion method (Section 2.1.2 of Robert & Casella (2010)). We set the threshold u at the 95th percentile of the observed sample, so $u = 4.23$.

Figure 2.17 shows the empirical distribution of the data at hand. In addition, the blue line is the censoring threshold u, the red curve relates to the true generating process $PL(5, 3, 1/9)$ and the orange and green curves stand for the true asymptotic models for threshold exceedances $GEV^{\frac{1}{20}}(-0.33, 4.23, 0.26)$ and $GEV(-0.33, 2.92, 0.69)$. Here, we can see that the true model and the asymptotic models start overlapping after the value of the threshold. This fact supports the asymptotic theory.

From now on, we assume for the observed sample the *tilde* Censored POT model with likelihood in (1.39). We perform an initial frequentist analysis. We compute by maximum likelihood a point estimate of the parameters, the extreme quantile and the return level and we obtain $(-0.34, 4.24, 0.26)$, 4.88 and 4.78.

If we further assume the Empirical Bayes prior of Section 2.1.2, we can operate a full Bayesian analysis. We set the starting point for the adaptive MH algorithm at $(0.1, 2, 1)$, we compute posterior chains of length 60000 in 411.56 seconds and we only keep the 50000 post burn-in samples. The obtained Overall Acceptance Probability is 0.2433. From Figure 2.18 we see that the scaling parameter κ of the proposal collapses towards 0 in the very first iterations, but then increases linearly. The OAP seems to converge from above to $\eta^* = 0.234$ through iterations of the algorithm. Here, we get a good adaptivity.



((A)) Power-Law p.d.f. ((B)) Powerl-Law c.d.f.

FIGURE 2.17: Simulated density and distribution function with their threshold exceedances asymptotic approximations GEV *tilde* and GEV *bar*.

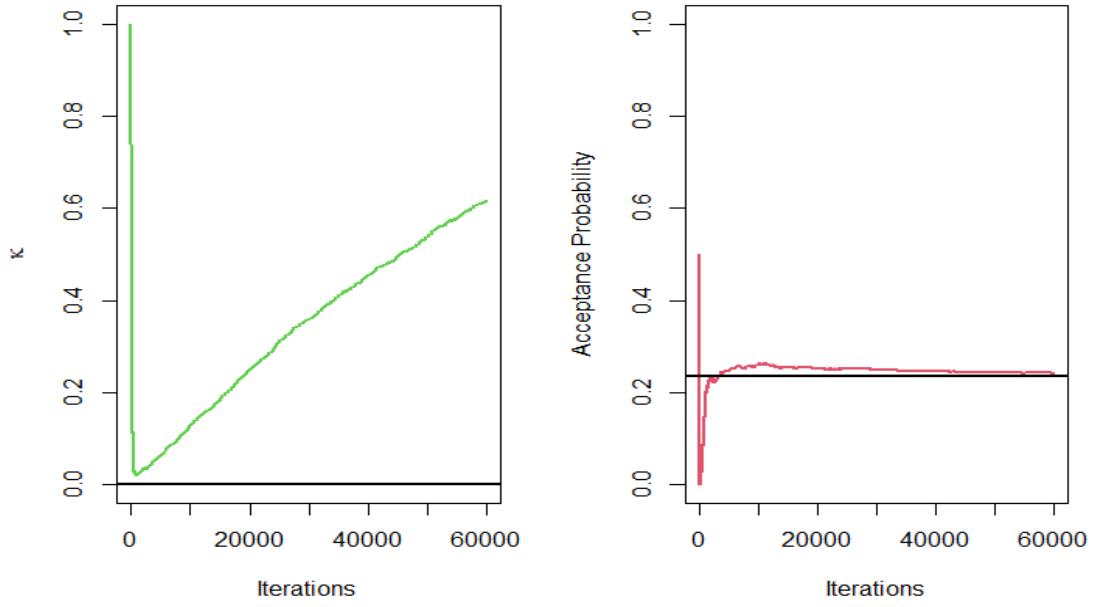


FIGURE 2.18: Power-Law model: scaling constant κ and Overall Acceptance Probability (OAP) through the iterations of the running adaptive MCMC algorithm.

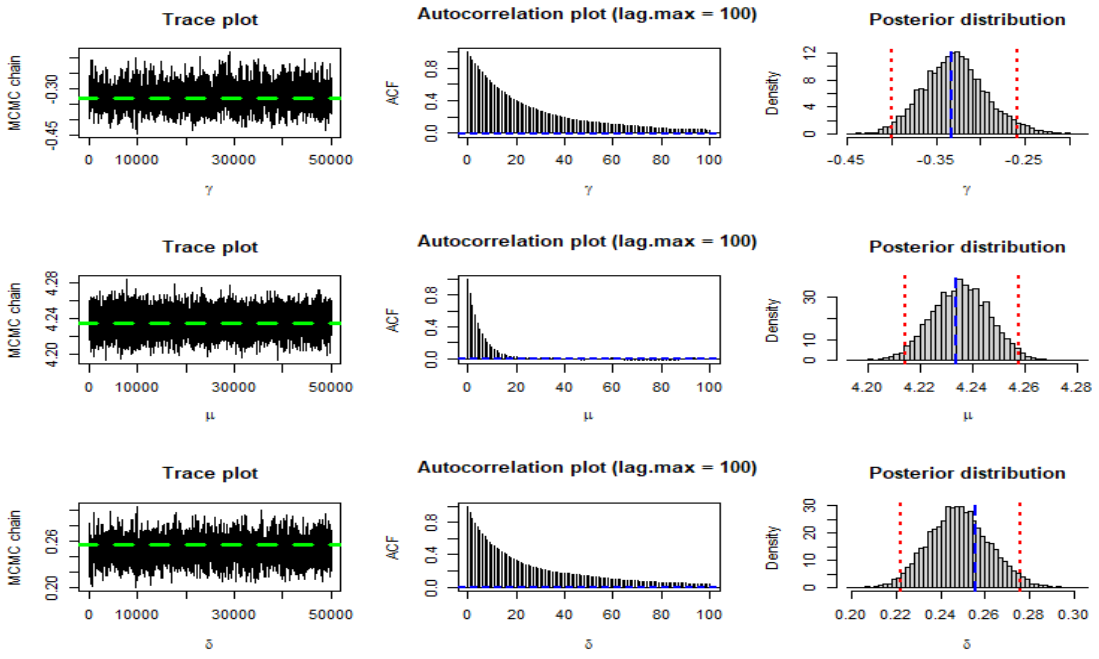
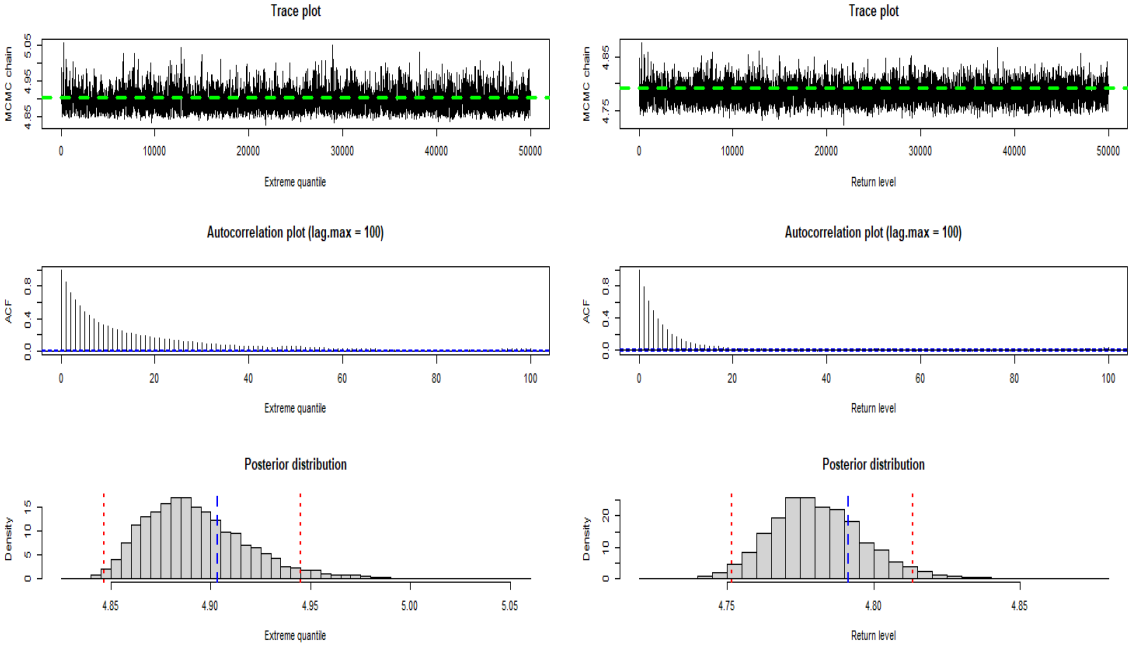
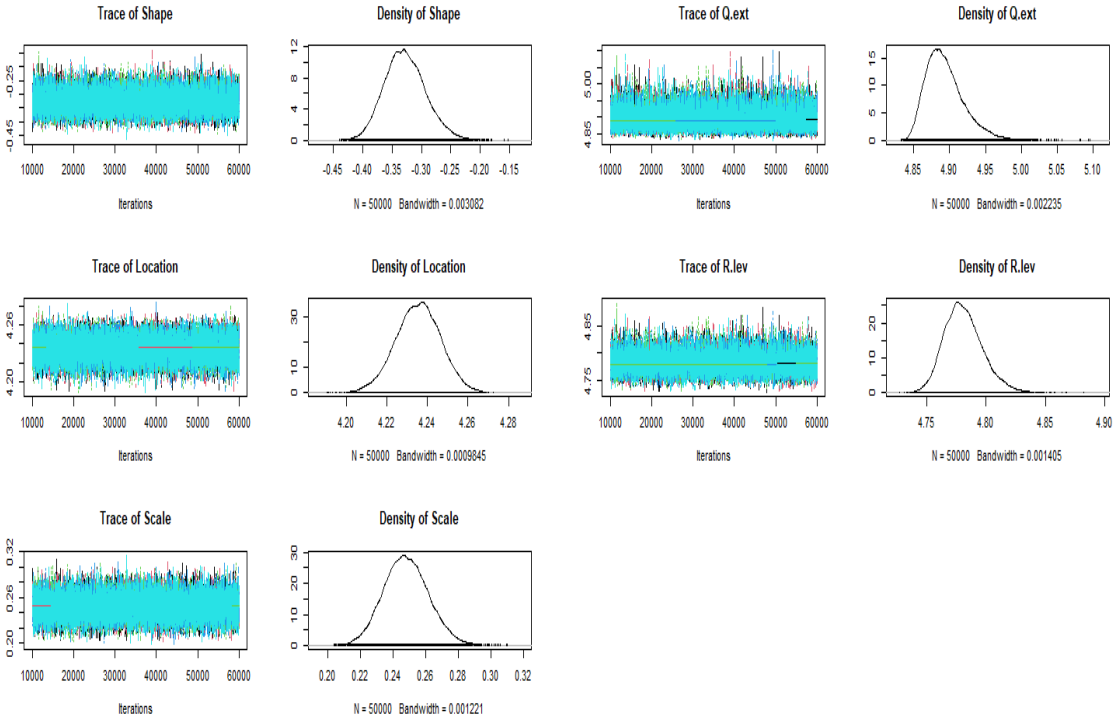


FIGURE 2.19: Power-Law model: trace plot, autocorrelation plot and histogram of the posterior chains for γ, $\tilde{\mu}$ and $\tilde{\delta}$. Dashed lines stand for the true parameter value, red dotted lines for the 95% HPD credible interval.



((A)) $\tilde{y}_p$ ((B)) $\tilde{y}_T$

FIGURE 2.20: Power-Law model: trace plot, autocorrelation plot and histogram of the posterior chains for the extreme quantile and the return level. Dashed lines stand for the true parameter value, red dotted lines for the 95% HPD credible interval.



((A)) $\gamma, \tilde{\mu}, \tilde{\delta}$ ((B)) $\tilde{y}_p, \tilde{y}_T$

FIGURE 2.21: Power-Law model: trace plots and histograms of the 5 parallel chains.

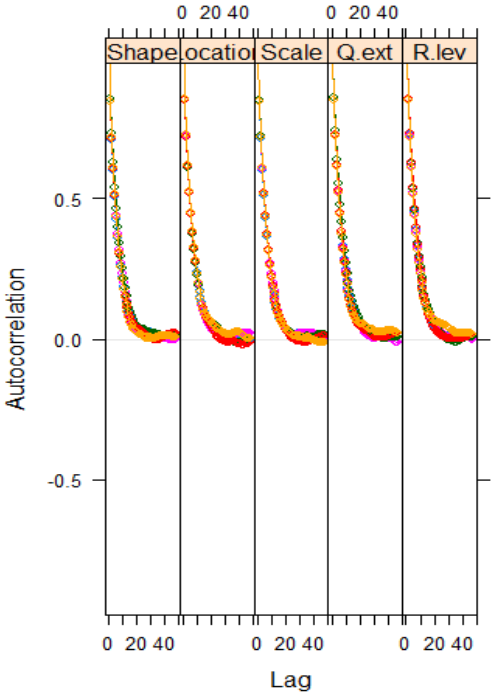


FIGURE 2.22: Power-Law model: autocorrelation plot of the 5 parallel chains.

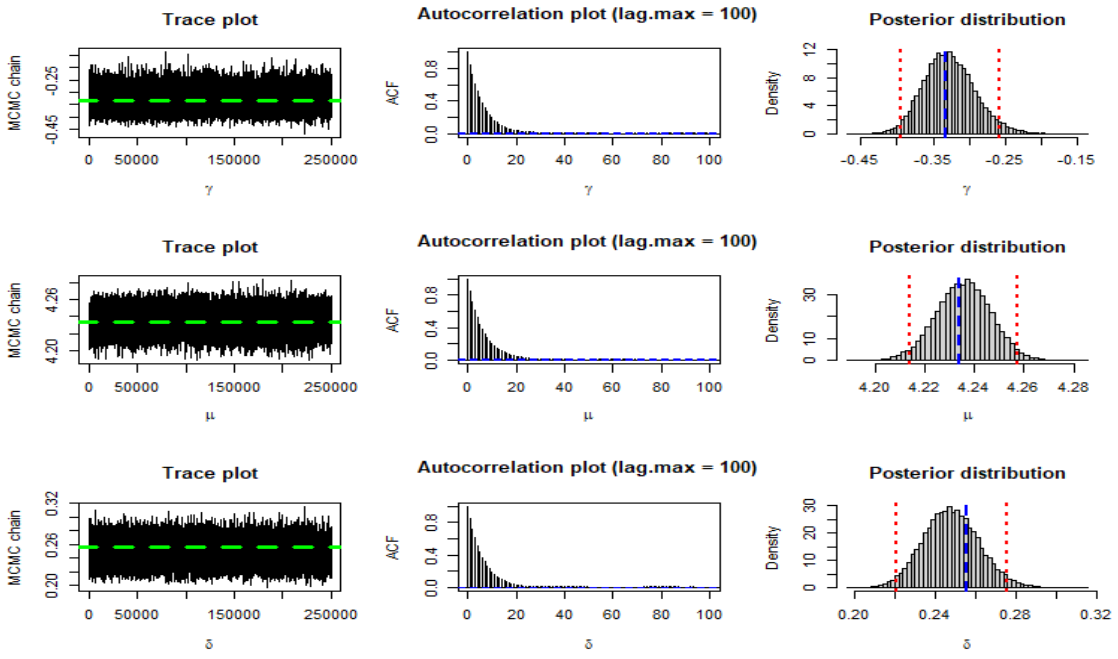
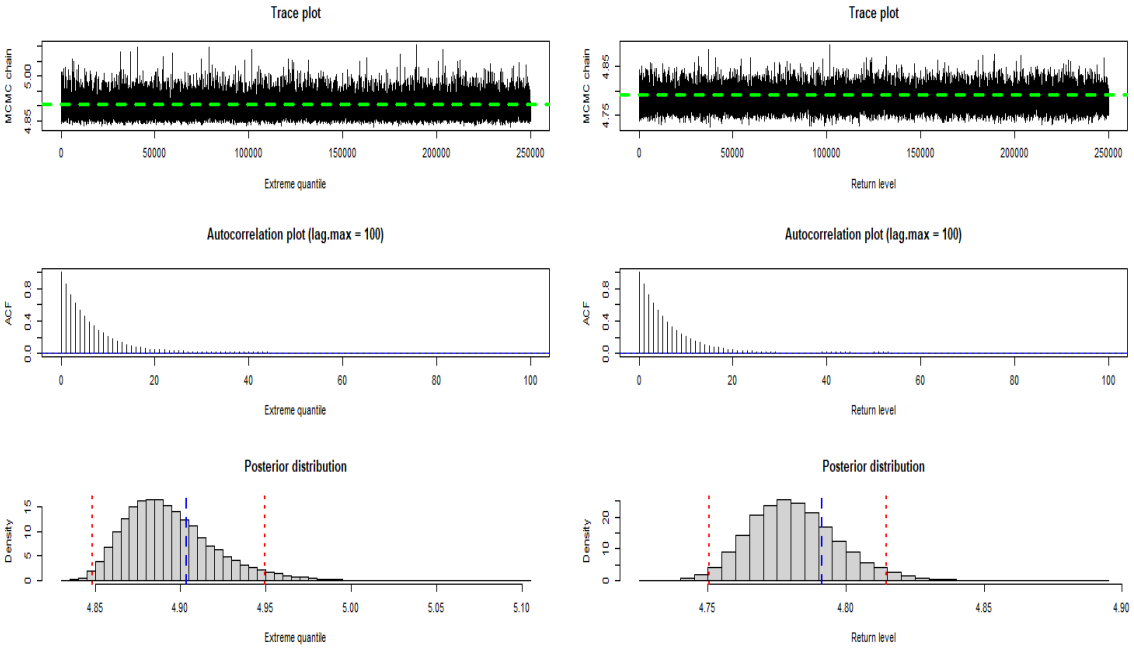


FIGURE 2.23: Power-Law model: trace plot, autocorrelation plot and histogram of the final posterior chains for γ, $\tilde{\mu}$ and $\tilde{\delta}$. Dashed lines stand for the true parameter value, red dotted lines for the 95% HPD credible interval.



((A)) $\tilde{y}_p$ ((B)) $\tilde{y}_T$

FIGURE 2.24: Power-Law model: trace plot, autocorrelation plot and histogram of the final posterior chains for the extreme quantile and the return level. Dashed lines stand for the true parameter value, red dotted lines for the 95% HPD credible interval.

The posterior mode is $(-0.33, 4.24, 0.25)$, while the posterior mean is $(-0.33, 4.24, 0.25)$. Figures 2.19 and 2.20 display trace plots, autocorrelation plots and histograms of the posterior distributions for $\gamma, \tilde{\mu}, \tilde{\delta}, \tilde{y}_{0.0001}$ and $\tilde{y}_{50}$. We generally denote a good mixing and appropriate correlation structure for each one of these quantities. The autocorrelation decay seems slower for γ, $\tilde{\delta}$ and $\tilde{y}_p$ with respect to $\tilde{\mu}$ and $\tilde{y}_T$. We can observe by histograms that every posterior distribution is nearly centered at its true parameter value. In addition, these empirical distributions seem to already behave normally at different levels. Furthermore, 95% HPD credible intervals contain the true parameter value.

To assess the convergence of the algorithm we sample 5 different analogue chains from different starting points. These points are generated from the large sample approximation of the posterior distribution in (2.3) evaluated at the posterior mode θ^*. Figures 2.21 and 2.22 show a good mixing of the chains in the entire support of the posterior distribution and a similar stationary behaviour of autocorrelation. The potential scale reduction factor is 1 with upper confidence limit equal to 1 for every one of the 5 quantities under investigation. With this information, we are able to affirm that these chains seem to have reached stationarity and they seem to have properly explored the same support of the posterior distribution. The averaged effective sample size of these 5 parallel (post burn-in) chains of length 50000 is 20601.15 for γ, 20246.50 for $\tilde{\mu}$, 20673.23 for $\tilde{\delta}$, 19805.41 for the extreme quantile and 19929.44 for the return level. Althought the evident loss of information with respect to an i.i.d. sample, the dependent chains are still worthy for extracting inference from them.

Therefore, we put together the post burn-in portions of these 5 chains to obtain posterior samples of size 250000. Figures 2.23 and 2.24 show trace plots, autocorrelation plots and histograms with overwritten the corresponding true parameter value and the 95% HPD credible interval. In addition to what has already been said we point out that the smoothness of every posterior distribution is evidently improved. Every high posterior density 95% credible interval still includes its true parameter value.

Chapter 3

Simulation Study

3.1 Simulation setting

We construct a simulation study to test the frequentist accuracy of credible intervals based of the Empirical Bayes Censored Peaks Over a Threshold (CPOT) method. We are interested in the marginal posterior distribution of the shape parameter γ, location parameter μ, scale parameter δ, the extreme quantile y_p (1.49) and the return level y_T (1.50) in order to evaluate the corresponding quantile (2.4), Gaussian approximation based (2.7) and high posterior density (HPD) (2.5) 95% credible intervals. To this aim, we draw $N = 1000$ independent random samples of increasing size n from 9 distributions pertaining to the max-domain of attraction of the GEV distribution and we compute coverage probabilities of credible intervals over the N iterations. We consider 3 heavy-tailed distributions, 3 light-tailed distributions and 3 short-tailed distributions, i.e.:

1. Fréchet with shape, location and scale parameter $(1, 0, 1)$, Pareto with shape, location and scale parameter $(1, 0, 1)$ and Half-Cauchy with scale parameter 1. Each of these distributions has tail index $\gamma = 1$;

2. Gumbel with location and scale parameter $(0, 1)$, Exponential with rate parameter 1 and Gamma with shape and scale parameter $(2, 2)$. Each of these distributions has tail index $\gamma = 0$;

3. Power-Law with upper bound, shape and K parameter $(5, 3, 1/9)$, Reverse-Weibull with shape parameter 3 and Beta with shape parameter $(1, 3)$. Each of these distributions has tail index $\gamma = -1/3$.

Computations of the posterior chains are performed through the Adaptive Gaussian RWMH algorithm 2, either in the *tilde* and *bar* parameterization. We use the posterior

mode or the MLE as starting point for the algorithm. We set the posterior sample size to $R = 10000$ and the burn-in to 10000. Finally, we set the return period to $T = 50$.

We study 4 different simulation scenarios, characterized by the sample size n, the fixed number of threshold excesses k and consequently by the inverse of the proportion of threshold exceedances s:

- small extreme sample: $n = 800$ and $k = 20$, then $s = 40$;
- medium extreme sample: $n = 1800$ and $k = 30$, then $s = 60$;
- large extreme sample: $n = 5450$ and $k = 50$, then $s = 109$;
- "big data"-type extreme sample: $n = 23400$ and $k = 100$, then $s = 234$.

We point out some implications due to these distinct simulation settings:

- the CPOT asymptotic theory works with $s \to \infty$, hence we expect improving performances from the small to the "big-data"-type sample configuration;
- the location and scale normalizing constants $b(n/k)$ and $a(n/k)$ increase as s increase, and so do the true parameter values $\tilde{\theta}^0$ and $\bar{\theta}^0$ and their maximum likelihood estimates from the data $\hat{\tilde{\theta}}$ and $\hat{\bar{\theta}}$. In other words, the Empirical Bayes hyperparameters of the prior distributions defined in (2.9) and (2.10) increase as s grows. This is necessary in an Extreme Values context in order to avoid infinite (and mathematically incorrect) priors;
- we set the probability p of the extreme quantile y_p to the fraction $1/n$, thus it assumes the decreasing values of 0.00125, 0.000555, 0.000183 and 0.0000427 over the 4 scenarios.

Every computation has been carried out with the software `R`. Because of computational problems, simulations for the Gumbel model in the *tilde* parameterization are $N = 500$.

3.2 Results

We present the coverage probabilities resulting from the simulation study in a cross-table format, in which rows represent the (transformation of the) parameter of interest and columns are the 4 simulation scenarios. Firstly, we report tables for quantile credible intervals, secondly the normality based intervals and then the HPD intervals. Corresponding tables for the *tilde* and *bar* parameterizations are reported side by side. Lastly,

we order the results' discussion according to the class of extreme value distribution of the generated data: the Fréchet family (Fréchet, Pareto and Half-Cauchy samples) is first, the Gumbel family (Gumbel, Exponential and Gamma samples) comes second and the Reverse-Weibull family (Power-Law, Reverse-Weibull and Beta samples) is at the very end.

3.2.1 Fréchet family

First of all, we note that the coverage probabilities between the *tilde* and *bar* parameterizations do not differ considerably, but the ones for the *tilde* parameterization seem to be a little higher. Generally, these empirical coverages are far from 95% for every considered (transformation of the) parameter and for every credible interval, except for μ. They range between 72% and 99%. Despite they are always higher than their nominal level, empirical coverages for the location parameter are not bad, though none of the three credible interval types seems to outstand the other competitors. We also highlight that intervals based on the normal approximation provide good results for the scale parameter δ. Globally, coverage probabilities tend to improve as n grows for γ, μ and δ, while the extreme quantile and the return level show a similar irregular behaviour with the return level that varies more with n. However, improvements are always moderate. Parameter δ is best covered by Gaussian intervals, then by HPD intervals and the worst are the quantile intervals. For other quantities, best performances are brought by the the quantile intervals, then by the Guassian and by the HPD intervals. Results for the Pareto model are better than those for the Half-Cauchy, that in turn are better than those for the Fréchet. We underline an overall unsatisfactory performance of the Empirical Bayes CPOT method.

The tail index γ reaches the best coverage of 85% for $n = 5450$ and $n = 23400$, in the Pareto model, with quantile credible intervals in the *tilde* parameterization, while its worst is of 72% for $n = 800$ and the *bar* parameterization, in the Fréchet model with normal and HPD intervals and in the Half-Cauchy model with normal intervals. The location parameter μ gets the best coverage of 96% for $n = 800$, in the Pareto model, with Gaussian and HPD credible intervals in the *tilde* parameterization, while its worst is of 99% for $n = 5450$ and the *tilde* parameterization, in the Half-Cauchy model with quantile intervals. The scale parameter δ has the best coverage of 95% for every n, in the Fréchet and Half-Cauchy model ($n = 5450$), with Gaussian credible intervals in the *tilde* parameterization (Fréchet model) and *bar* parameterization, while its worst is of 79% for $n = 1800$ and the *bar* parameterization, in the Fréchet model with quantile intervals. The extreme quantile y_p gets the best coverage of 90% for $n = 5450$ and

$n = 23400$, in the Pareto model ($n = 23400$) and Half-Cauchy model, with quantile credible intervals in the *tilde* parameterization, while its worst is of 80% for $n = 800$ and the *bar* parameterization, in the Pareto model with HPD intervals. The return level y_T gets the best coverage of 91% for $n = 23400$, in the Pareto model, with quantile credible intervals in the *tilde* parameterization, while its worst is of 77% for $n = 800$ and the *bar* parameterization, in the Pareto and Half-Cauchy models with HPD intervals.

TABLE 3.1: Fréchet model: coverage probabilities of quantile 95% credible intervals.

n	800	1800	5450	23400
γ	0.82	0.80	0.84	0.82
$\tilde{\mu}$	0.98	0.98	0.98	0.98
$\tilde{\delta}$	0.82	0.83	0.83	0.86
$\tilde{y}_p$	0.89	0.88	0.89	0.88
$\tilde{y}_T$	0.87	0.87	0.89	0.88

n	800	1800	5450	23400
γ	0.78	0.80	0.81	0.83
$\bar{\mu}$	0.97	0.98	0.98	0.98
$\bar{\delta}$	0.82	0.79	0.83	0.85
$\bar{y}_p$	0.87	0.88	0.89	0.88
$\bar{y}_T$	0.86	0.87	0.89	0.88

TABLE 3.2: Fréchet model: coverage probabilities of Guassian approximation based 95% credible intervals.

n	800	1800	5450	23400
γ	0.77	0.75	0.81	0.81
$\tilde{\mu}$	0.97	0.98	0.98	0.97
$\tilde{\delta}$	0.97	0.96	0.96	0.95
$\tilde{y}_p$	0.87	0.85	0.85	0.83
$\tilde{y}_T$	0.87	0.85	0.85	0.84

n	800	1800	5450	23400
γ	0.72	0.76	0.78	0.82
$\bar{\mu}$	0.97	0.98	0.98	0.98
$\bar{\delta}$	0.95	0.95	0.95	0.93
$\bar{y}_p$	0.82	0.84	0.83	0.82
$\bar{y}_T$	0.80	0.83	0.84	0.83

TABLE 3.3: Fréchet model: coverage probabilities of HPD 95% credible intervals.

n	800	1800	5450	23400
γ	0.77	0.76	0.81	0.81
$\tilde{\mu}$	0.97	0.98	0.98	0.98
$\tilde{\delta}$	0.90	0.90	0.90	0.91
$\tilde{y}_p$	0.82	0.81	0.83	0.81
$\tilde{y}_T$	0.79	0.80	0.83	0.84

n	800	1800	5450	23400
γ	0.72	0.75	0.78	0.80
$\bar{\mu}$	0.97	0.98	0.98	0.98
$\bar{\delta}$	0.88	0.88	0.88	0.89
$\bar{y}_p$	0.81	0.83	0.83	0.82
$\bar{y}_T$	0.78	0.82	0.83	0.83

TABLE 3.4: Pareto model: coverage probabilities of quantile 95% credible intervals.

n	800	1800	5450	23400
γ	0.82	0.80	0.85	0.85
$\tilde{\mu}$	0.97	0.98	0.98	0.98
$\tilde{\delta}$	0.81	0.84	0.82	0.86
$\tilde{y}_p$	0.89	0.88	0.90	0.90
$\tilde{y}_T$	0.87	0.87	0.90	0.91

n	800	1800	5450	23400
γ	0.77	0.79	0.80	0.78
$\bar{\mu}$	0.98	0.98	0.98	0.98
$\bar{\delta}$	0.82	0.83	0.83	0.87
$\bar{y}_p$	0.86	0.87	0.86	0.85
$\bar{y}_T$	0.84	0.87	0.86	0.86

TABLE 3.5: Pareto model: coverage probabilities of Guassian approximation based 95% credible intervals.

n	800	1800	5450	23400
γ	0.76	0.76	0.82	0.84
$\tilde{\mu}$	0.96	0.97	0.98	0.98
$\tilde{\delta}$	0.97	0.97	0.96	0.94
$\tilde{y}_p$	0.86	0.86	0.86	0.85
$\tilde{y}_T$	0.88	0.86	0.86	0.86

n	800	1800	5450	23400
γ	0.73	0.74	0.78	0.77
$\bar{\mu}$	0.97	0.98	0.98	0.98
$\bar{\delta}$	0.96	0.94	0.94	0.94
$\bar{y}_p$	0.82	0.83	0.81	0.77
$\bar{y}_T$	0.79	0.81	0.81	0.79

TABLE 3.6: Pareto model: coverage probabilities of HPD 95% credible intervals.

n	800	1800	5450	23400
γ	0.77	0.76	0.82	0.84
$\tilde{\mu}$	0.96	0.98	0.98	0.97
$\tilde{\delta}$	0.90	0.90	0.90	0.90
$\tilde{y}_p$	0.82	0.83	0.84	0.84
$\tilde{y}_T$	0.79	0.81	0.84	0.86

n	800	1800	5450	23400
γ	0.73	0.74	0.77	0.76
$\bar{\mu}$	0.97	0.98	0.98	0.98
$\bar{\delta}$	0.89	0.88	0.89	0.90
$\bar{y}_p$	0.80	0.82	0.81	0.77
$\bar{y}_T$	0.77	0.79	0.81	0.80

TABLE 3.7: Half-Cauchy model: coverage probabilities of quantile 95% credible intervals.

n	800	1800	5450	23400
γ	0.79	0.83	0.84	0.84
$\tilde{\mu}$	0.97	0.98	0.99	0.98
$\tilde{\delta}$	0.81	0.83	0.85	0.85
$\tilde{y}_p$	0.87	0.88	0.90	0.89
$\tilde{y}_T$	0.86	0.87	0.90	0.90

n	800	1800	5450	23400
γ	0.78	0.81	0.82	0.81
$\bar{\mu}$	0.97	0.98	0.98	0.97
$\bar{\delta}$	0.82	0.82	0.85	0.84
$\bar{y}_p$	0.87	0.87	0.88	0.87
$\bar{y}_T$	0.85	0.86	0.88	0.88

TABLE 3.8: Half-Cauchy model: coverage probabilities of Guassian approximation based 95% credible intervals.

n	800	1800	5450	23400
γ	0.74	0.78	0.82	0.83
$\tilde{\mu}$	0.97	0.98	0.98	0.98
$\tilde{\delta}$	0.96	0.96	0.96	0.96
$\tilde{y}_p$	0.86	0.87	0.87	0.84
$\tilde{y}_T$	0.87	0.87	0.87	0.85

n	800	1800	5450	23400
γ	0.72	0.76	0.79	0.80
$\bar{\mu}$	0.97	0.98	0.98	0.97
$\bar{\delta}$	0.96	0.96	0.95	0.93
$\bar{y}_p$	0.83	0.82	0.82	0.81
$\bar{y}_T$	0.81	0.81	0.82	0.83

3.2.2 Gumbel family

First of all, we note that in this case the coverage probabilities between the *tilde* and *bar* parameterizations do not differ considerably. Generally, the empirical coverages reach 95% at least once for every considered (transformation of the) parameter and for every credible interval, for every generating process and for every sample size n. They range between 87% and 99%. Globally, coverage probabilities tend to improve as n grows for γ, μ and δ, the return level and the extreme quantile, in the *tilde* and *bar*

TABLE 3.9: Half-Cauchy model: coverage probabilities of HPD 95% credible intervals.

n	800	1800	5450	23400
γ	0.75	0.79	0.81	0.82
$\tilde{\mu}$	0.97	0.98	0.98	0.98
$\tilde{\delta}$	0.89	0.90	0.91	0.91
$\tilde{y}_p$	0.82	0.83	0.84	0.82
$\tilde{y}_T$	0.79	0.81	0.84	0.85

n	800	1800	5450	23400
γ	0.73	0.75	0.77	0.78
$\bar{\mu}$	0.97	0.98	0.98	0.97
$\bar{\delta}$	0.89	0.89	0.91	0.88
$\bar{y}_p$	0.82	0.81	0.82	0.81
$\bar{y}_T$	0.77	0.80	0.82	0.83

parameterization, for every type of credible interval and for every starting light-tailed distribution. Results for the Exponential model are better than those for the Gamma, that in turn are better than those for the Gumbel. We underline an overall accurate performance of the Empirical Bayes CPOT method.

The tail index γ reaches the worst coverage probability of 88% with $n = 800$, in the *bar* parameterization, in the Gumbel model with quantile intervals. The location parameter μ gets the worst coverage of 98%, at least once for every sample size n, in the *tilde* and *bar* parameterizations, in the Gumbel model with quantile, normality based or HPD intervals. The scale parameter δ has the worst coverage of 87% for $n = 800$ and in the *bar* parameterization, in the Gumbel model with quantile intervals. The extreme quantile y_p gets the worst coverage of 90% for every possible n except 23400, with the *tilde* parameterization, in the Gumbel model and with every type of considered credible interval. The return level y_T gets the worst coverage of 89% for $n = 800$ or $n = 1800$ and both parameterizations, in the Gumbel, Exponential and Gamma models with Gaussian or HPD intervals.

TABLE 3.10: Gumbel model: coverage probabilities of quantile 95% credible intervals.

n	800	1800	5450	23400
γ	0.96	0.94	0.96	0.96
$\tilde{\mu}$	0.98	0.97	0.98	0.98
$\tilde{\delta}$	0.89	0.91	0.91	0.94
$\tilde{y}_p$	0.90	0.92	0.92	0.93
$\tilde{y}_T$	0.92	0.91	0.92	0.93

n	800	1800	5450	23400
γ	0.92	0.94	0.94	0.95
$\bar{\mu}$	0.98	0.98	0.97	0.96
$\bar{\delta}$	0.87	0.90	0.93	0.95
$\bar{y}_p$	0.92	0.91	0.94	0.94
$\bar{y}_T$	0.92	0.92	0.94	0.95

3.2.3 Reverse-Weibull family

First of all, we note that the coverage probabilities between the *tilde* and *bar* parameterizations do not differ considerably, except for parameters γ and δ in the Reverse-Weibull

TABLE 3.11: Gumbel model: coverage probabilities of Guassian approximation based 95% credible intervals.

n	800	1800	5450	23400
γ	0.94	0.93	0.95	0.94
$\tilde{\mu}$	0.97	0.97	0.98	0.97
$\tilde{\delta}$	0.99	0.97	0.96	0.96
$\tilde{y}_p$	0.91	0.91	0.90	0.93
$\tilde{y}_T$	0.89	0.90	0.90	0.93

n	800	1800	5450	23400
γ	0.89	0.92	0.92	0.94
$\bar{\mu}$	0.98	0.98	0.97	0.96
$\bar{\delta}$	0.96	0.96	0.96	0.95
$\bar{y}_p$	0.93	0.92	0.91	0.94
$\bar{y}_T$	0.91	0.91	0.90	0.93

TABLE 3.12: Gumbel model: coverage probabilities of HPD 95% credible intervals.

n	800	1800	5450	23400
γ	0.93	0.94	0.95	0.95
$\tilde{\mu}$	0.98	0.97	0.98	0.97
$\tilde{\delta}$	0.94	0.94	0.92	0.95
$\tilde{y}_p$	0.91	0.90	0.90	0.93
$\tilde{y}_T$	0.89	0.90	0.90	0.92

n	800	1800	5450	23400
γ	0.88	0.92	0.92	0.94
$\bar{\mu}$	0.98	0.98	0.96	0.96
$\bar{\delta}$	0.91	0.94	0.94	0.94
$\bar{y}_p$	0.91	0.91	0.91	0.94
$\bar{y}_T$	0.89	0.90	0.91	0.93

TABLE 3.13: Exponential model: coverage probabilities of quantile 95% credible intervals.

n	800	1800	5450	23400
γ	0.95	0.96	0.96	0.97
$\tilde{\mu}$	0.97	0.97	0.96	0.96
$\tilde{\delta}$	0.91	0.93	0.95	0.96
$\tilde{y}_p$	0.91	0.92	0.95	0.95
$\tilde{y}_T$	0.91	0.93	0.95	0.94

n	800	1800	5450	23400
γ	0.95	0.95	0.95	0.96
$\bar{\mu}$	0.96	0.97	0.97	0.95
$\bar{\delta}$	0.92	0.92	0.93	0.94
$\bar{y}_p$	0.91	0.93	0.95	0.94
$\bar{y}_T$	0.93	0.93	0.95	0.95

TABLE 3.14: Exponential model: coverage probabilities of Guassian approximation based 95% credible intervals.

n	800	1800	5450	23400
γ	0.92	0.94	0.95	0.95
$\tilde{\mu}$	0.97	0.97	0.96	0.96
$\tilde{\delta}$	0.98	0.98	0.98	0.97
$\tilde{y}_p$	0.92	0.93	0.94	0.95
$\tilde{y}_T$	0.91	0.93	0.94	0.94

n	800	1800	5450	23400
γ	0.93	0.92	0.94	0.94
$\bar{\mu}$	0.96	0.97	0.97	0.95
$\bar{\delta}$	0.96	0.96	0.94	0.95
$\bar{y}_p$	0.94	0.93	0.95	0.95
$\bar{y}_T$	0.93	0.93	0.95	0.94

model for $n = 23400$, where the *bar* parameterization shows an evident decrease. Generally, these empirical coverages reach 95% at least once for every considered (transformation of the) parameter and for every credible interval, for every generating process and for every sample size n. They range between 78% and 99%. Globally, coverage probabilities tend to improve as n grows for γ, μ and δ, the return level and the extreme quantile, in the *tilde* and *bar* parameterization, for every type of credible interval and for every starting short-tailed distribution. The Beta model has empirical coverages that generally show a different behaviour than the more similar Power-Law and

TABLE 3.15: Exponential model: coverage probabilities of HPD 95% credible intervals.

n	800	1800	5450	23400
γ	0.92	0.95	0.96	0.96
$\tilde{\mu}$	0.97	0.97	0.96	0.95
$\tilde{\delta}$	0.95	0.96	0.97	0.96
$\tilde{y}_p$	0.91	0.92	0.94	0.94
$\tilde{y}_T$	0.89	0.92	0.94	0.94

n	800	1800	5450	23400
γ	0.93	0.92	0.94	0.94
$\bar{\mu}$	0.96	0.97	0.97	0.95
$\bar{\delta}$	0.94	0.94	0.93	0.94
$\bar{y}_p$	0.92	0.93	0.94	0.94
$\bar{y}_T$	0.92	0.93	0.94	0.94

TABLE 3.16: Gamma model: coverage probabilities of quantile 95% credible intervals.

n	800	1800	5450	23400
γ	0.95	0.95	0.96	0.95
$\tilde{\mu}$	0.97	0.97	0.96	0.96
$\tilde{\delta}$	0.91	0.93	0.94	0.96
$\tilde{y}_p$	0.91	0.92	0.94	0.95
$\tilde{y}_T$	0.92	0.92	0.94	0.94

n	800	1800	5450	23400
γ	0.96	0.95	0.95	0.95
$\bar{\mu}$	0.98	0.96	0.97	0.96
$\bar{\delta}$	0.93	0.92	0.95	0.94
$\bar{y}_p$	0.92	0.92	0.92	0.94
$\bar{y}_T$	0.94	0.91	0.93	0.93

TABLE 3.17: Gamma model: coverage probabilities of Guassian approximation based 95% credible intervals.

n	800	1800	5450	23400
γ	0.93	0.95	0.94	0.94
$\tilde{\mu}$	0.97	0.97	0.96	0.96
$\tilde{\delta}$	0.99	0.98	0.98	0.97
$\tilde{y}_p$	0.93	0.92	0.94	0.93
$\tilde{y}_T$	0.91	0.90	0.91	0.90

n	800	1800	5450	23400
γ	0.94	0.92	0.93	0.93
$\bar{\mu}$	0.97	0.96	0.97	0.95
$\bar{\delta}$	0.97	0.96	0.96	0.95
$\bar{y}_p$	0.94	0.92	0.92	0.93
$\bar{y}_T$	0.91	0.90	0.91	0.92

TABLE 3.18: Gamma model: coverage probabilities of HPD 95% credible intervals.

n	800	1800	5450	23400
γ	0.93	0.95	0.95	0.94
$\tilde{\mu}$	0.97	0.97	0.96	0.96
$\tilde{\delta}$	0.96	0.96	0.96	0.96
$\tilde{y}_p$	0.91	0.91	0.94	0.93
$\tilde{y}_T$	0.90	0.89	0.92	0.91

n	800	1800	5450	23400
γ	0.93	0.91	0.93	0.93
$\bar{\mu}$	0.98	0.96	0.97	0.96
$\bar{\delta}$	0.95	0.94	0.96	0.94
$\bar{y}_p$	0.92	0.91	0.92	0.93
$\bar{y}_T$	0.91	0.91	0.90	0.91

Reverse-Weibull models. Results for the Power-Law model are better than those for the Beta, that in turn are better than those for the Reverse-Weibull. We underline an overall accurate performance of the Empirical Bayes CPOT method.

The tail index γ reaches the worst coverage probability of 78% for $n = 23400$ and the *bar* parameterization, in the Reverse-Weibull model with Gaussian intervals. The location parameter μ gets the worst coverages of 97% or 93% for every n except 5450, with the *tilde* and *bar* parameterizations, in every model and with every credible interval. The scale parameter δ has the worst coverage of 84% for $n = 23400$ and the *bar*

parameterization, in the Reverse-Weibull model with normality based intervals. The extreme quantile y_p gets the worst coverage of 89% for $n = 23400$, with the *tilde* parameterization, in the Beta model with quantile intervals. The return level y_T gets the worst coverage of 90% for $n = 23400$ and both parameterizations, in the Reverse-Weibull and Beta models with quantile intervals.

TABLE 3.19: Power-Law model: coverage probabilities of quantile 95% credible intervals.

n	800	1800	5450	23400
γ	0.99	0.98	0.95	0.95
$\tilde{\mu}$	0.97	0.96	0.95	0.95
$\tilde{\delta}$	0.97	0.96	0.95	0.94
$\tilde{y}_p$	0.95	0.94	0.95	0.93
$\tilde{y}_T$	0.97	0.94	0.95	0.93

n	800	1800	5450	23400
γ	0.96	0.94	0.93	0.92
$\bar{\mu}$	0.96	0.95	0.95	0.94
$\bar{\delta}$	0.93	0.94	0.93	0.92
$\bar{y}_p$	0.94	0.93	0.94	0.94
$\bar{y}_T$	0.94	0.94	0.94	0.95

TABLE 3.20: Power-Law model: coverage probabilities of Guassian approximation based 95% credible intervals.

n	800	1800	5450	23400
γ	0.97	0.96	0.95	0.94
$\tilde{\mu}$	0.97	0.96	0.95	0.96
$\tilde{\delta}$	0.99	0.98	0.96	0.95
$\tilde{y}_p$	0.96	0.97	0.96	0.96
$\tilde{y}_T$	0.97	0.98	0.96	0.95

n	800	1800	5450	23400
γ	0.93	0.91	0.92	0.93
$\bar{\mu}$	0.96	0.95	0.95	0.93
$\bar{\delta}$	0.96	0.95	0.94	0.93
$\bar{y}_p$	0.96	0.96	0.96	0.96
$\bar{y}_T$	0.97	0.96	0.96	0.96

TABLE 3.21: Power-Law model: coverage probabilities of HPD 95% credible intervals.

n	800	1800	5450	23400
γ	0.96	0.95	0.94	0.93
$\tilde{\mu}$	0.97	0.96	0.96	0.95
$\tilde{\delta}$	0.98	0.97	0.95	0.94
$\tilde{y}_p$	0.96	0.95	0.96	0.94
$\tilde{y}_T$	0.97	0.96	0.95	0.94

n	800	1800	5450	23400
γ	0.93	0.91	0.90	0.91
$\bar{\mu}$	0.96	0.95	0.96	0.94
$\bar{\delta}$	0.93	0.94	0.92	0.92
$\bar{y}_p$	0.96	0.95	0.95	0.96
$\bar{y}_T$	0.96	0.96	0.95	0.95

TABLE 3.22: Reverse-Weibull model: coverage probabilities of quantile 95% credible intervals.

n	800	1800	5450	23400
γ	0.99	0.98	0.95	0.91
$\tilde{\mu}$	0.96	0.97	0.95	0.94
$\tilde{\delta}$	0.97	0.95	0.94	0.92
$\tilde{y}_p$	0.92	0.93	0.92	0.91
$\tilde{y}_T$	0.94	0.94	0.93	0.92

n	800	1800	5450	23400
γ	0.95	0.94	0.93	0.80
$\bar{\mu}$	0.96	0.96	0.96	0.96
$\bar{\delta}$	0.95	0.94	0.93	0.84
$\bar{y}_p$	0.93	0.94	0.94	0.90
$\bar{y}_T$	0.94	0.94	0.94	0.90

TABLE 3.23: Reverse-Weibull model: coverage probabilities of Guassian approximation based 95% credible intervals.

n	800	1800	5450	23400
γ	0.97	0.95	0.94	0.90
$\tilde{\mu}$	0.96	0.97	0.96	0.94
$\tilde{\delta}$	0.99	0.98	0.96	0.92
$\tilde{y}_p$	0.96	0.96	0.95	0.93
$\tilde{y}_T$	0.96	0.96	0.95	0.94

n	800	1800	5450	23400
γ	0.92	0.92	0.92	0.81
$\bar{\mu}$	0.97	0.96	0.96	0.95
$\bar{\delta}$	0.97	0.96	0.94	0.86
$\bar{y}_p$	0.95	0.96	0.94	0.91
$\bar{y}_T$	0.95	0.95	0.94	0.92

TABLE 3.24: Reverse-Weibull model: coverage probabilities of HPD 95% credible intervals.

n	800	1800	5450	23400
γ	0.96	0.94	0.93	0.90
$\tilde{\mu}$	0.97	0.97	0.95	0.94
$\tilde{\delta}$	0.98	0.96	0.95	0.92
$\tilde{y}_p$	0.94	0.94	0.93	0.92
$\tilde{y}_T$	0.94	0.94	0.93	0.93

n	800	1800	5450	23400
γ	0.92	0.92	0.90	0.78
$\bar{\mu}$	0.96	0.96	0.96	0.95
$\bar{\delta}$	0.96	0.95	0.93	0.84
$\bar{y}_p$	0.95	0.95	0.94	0.90
$\bar{y}_T$	0.95	0.94	0.94	0.91

TABLE 3.25: Beta model: coverage probabilities of quantile 95% credible intervals.

n	800	1800	5450	23400
γ	0.98	0.97	0.96	0.90
$\tilde{\mu}$	0.96	0.96	0.96	0.93
$\tilde{\delta}$	0.97	0.96	0.96	0.91
$\tilde{y}_p$	0.93	0.95	0.94	0.89
$\tilde{y}_T$	0.94	0.95	0.94	0.90

n	800	1800	5450	23400
γ	0.97	0.96	0.95	0.93
$\bar{\mu}$	0.94	0.94	0.95	0.94
$\bar{\delta}$	0.95	0.95	0.93	0.92
$\bar{y}_p$	0.94	0.96	0.96	0.95
$\bar{y}_T$	0.94	0.96	0.96	0.95

TABLE 3.26: Beta model: coverage probabilities of Guassian approximation based 95% credible intervals.

n	800	1800	5450	23400
γ	0.96	0.95	0.95	0.88
$\tilde{\mu}$	0.96	0.96	0.96	0.94
$\tilde{\delta}$	0.99	0.98	0.96	0.91
$\tilde{y}_p$	0.97	0.97	0.96	0.91
$\tilde{y}_T$	0.98	0.97	0.96	0.92

n	800	1800	5450	23400
γ	0.95	0.95	0.94	0.94
$\bar{\mu}$	0.94	0.95	0.95	0.94
$\bar{\delta}$	0.96	0.96	0.93	0.93
$\bar{y}_p$	0.96	0.97	0.97	0.96
$\bar{y}_T$	0.97	0.97	0.97	0.96

TABLE 3.27: Beta model: coverage probabilities of HPD 95% credible intervals.

n	800	1800	5450	23400
γ	0.96	0.94	0.93	0.88
$\tilde{\mu}$	0.96	0.97	0.96	0.93
$\tilde{\delta}$	0.98	0.97	0.95	0.91
$\tilde{y}_p$	0.96	0.95	0.95	0.90
$\tilde{y}_T$	0.96	0.95	0.95	0.91

n	800	1800	5450	23400
γ	0.94	0.94	0.92	0.92
$\bar{\mu}$	0.94	0.94	0.95	0.93
$\bar{\delta}$	0.95	0.95	0.92	0.92
$\bar{y}_p$	0.95	0.97	0.97	0.96
$\bar{y}_T$	0.95	0.97	0.96	0.96

Conclusions

In this work, we have developed a Bayesian version of the POT method for analysing extreme data through the censored GEV likelihood. Figures 2.1, 2.9 and 2.17 highlight that the $GEV^{\frac{1}{s}}\left(\tilde{\theta}\right)$ and $GEV\left(\bar{\theta}\right)$ are appropriate asymptotic models for the unconditional distribution of threshold exceedances. The Empirical Bayes procedure has been employed to tune from the observed data the prior distribution for $\tilde{\theta}$ or $\bar{\theta}$ with suitable hyperparameters, which increase as n grows. We have sampled efficiently from the posterior distribution of the parameters via an AMH algorithm. Afterwards, we have tested the performance of the new method checking the frequentist coverage probabilities of credible intervals.

The Empirical Bayes CPOT method shows overall good performances for the Gumbel and Reverse-Weibull families, while in the Fréchet class the results are not yet as satisfactory as we would like. We also recall that even the adaptivity of the used RWMH algorithm is not adequate in the Fréchet model. This discrepancy could be better inspected through a graphical study of the posterior distributions in the three families. The main reason behind these problematic simulation results is that so far, the number of exceedances k to consider, given a sample of n observations, has been set quite arbitrarily. On the contrary, with intent to reach a good performance, it would be necessary to determine a precise relationship between k and n using a criterion that takes into account the bias term arising from the use of the asymptotic model, as it has been done in Padoan & Rizzelli (2022).

Furthermore, we could compare the Empirical Bayes CPOT's inferential performances with the ones of the Bayesian Classical POT approach in order to see if the censored additional data carry an actual benefit or a drawback. We could also question if in this case a Bayesian analysis is really appropriate through a comparison with coverage probabilities of frequentist confidence intervals computed from the likelihood of the CPOT model. This comparisons could be made either on simulated data or real extreme datasets. Finally, computational difficulties of the Adaptive RWMH algorithm for the Gumbel model could be deeper examined.

Appendix

Appendix A

```
#GEV distribution
#Probability density function

dgev1 <- function(x, par, log = FALSE){
        gam <- par[1]; mu <- par[2]; delta <- par[3]
        pdf <- 0
        if(delta <= 0){
                if(log == FALSE) return(pdf)
                if(log == TRUE) return(log(pdf))
        }
        else{
                z <- (x - mu)/delta
                #Define 0 neighbourhood
                eps <- .Machine$double.eps^.3
                #the smallest positive floating-point
                #number x such that 1 + x != 1
                if(abs(gam) <= eps){
                        pdf <- (1/delta)*
                                exp(-(z + exp(-z)))
                }
                else{
                if((1 + gam*z) > 0){
                        gev.k <- (1 + gam*z)^(-1/gam)
                        pdf <- (1/delta)*
                                (gev.k^(gam + 1))*
                                exp(-gev.k)
                        }
```

```
                }
                if(log == FALSE) return(pdf)
                if(log == TRUE) return(log(pdf))
        }
}
dgev.V <- Vectorize(dgev1, "x")

#Cumulative density function

pgev1 <- function(x, par, log = FALSE){
        gam <- par[1]; mu <- par[2]; delta <- par[3]
        cdf <- 0
        if(delta <= 0){
                if(log == FALSE) return(cdf)
                if(log == TRUE) return(log(cdf))
        }
        else{
                z <- (x - mu)/delta
                #Define 0 neighbourhood
                eps <- .Machine$double.eps^.3
                #the smallest positive floating-point
                #number x such that 1 + x != 1
                if(abs(gam) <= eps){
                        cdf <- exp(-exp(-z))
                }
                else{
                        gev.k <- (1 + gam*z)^(-1/gam)
                        cdf <- exp(-gev.k)
                        if(gam > 0 & ((1 + gam*z) <= 0))
                                cdf <- 0
                        if(gam < 0 & ((1 + gam*z) <= 0))
                                cdf <- 1
                }
                if(log == FALSE) return(cdf)
                if(log == TRUE) return(log(cdf))
        }
}
pgev.V <- Vectorize(pgev1, "x")
```

```
#GEV distribution in the tilde parameterization
#Probability density function

dgev1.tilde <- function(x, par, s, log = FALSE){
        gam <- par[1]; mu <- par[2]; delta <- par[3]
        cdf <- 0
        pdf <- 0
        if(delta <= 0){
                if(log == FALSE) return(pdf)
                if(log == TRUE) return(log(pdf))
        }
        else{
                z <- (x - mu)/delta
                #Define 0 neighbourhood
                eps <- .Machine$double.eps^.3
                #the smallest positive floating-point
                #number x such that 1 + x != 1
                if(abs(gam) <= eps){
                        cdf <- (exp(-exp(-z)))^(1/s)
                        pdf <- (1/(s*delta))*cdf*exp(-z)
                }
                else{
                        gev.k <- (1 + gam*z)^(-1/gam)
                        gev.k2 <- (1 + gam*z)^(-1/gam -1)
                        cdf <- (exp(-gev.k))^(1/s)
                        pdf <- (1/(s*delta))*cdf*gev.k2
                        if(gam > 0 & ((1 + gam*z) <= 0)){
                                cdf <- 0
                                pdf <- 0
                        }
                        if(gam < 0 & ((1 + gam*z) <= 0)){
                                cdf <-  1
                                pdf <- 0
                        }
                }
                if(log == FALSE) return(pdf)
                if(log == TRUE) return(log(pdf))
        }
}
```

```
dgev.tilde.V <- Vectorize(dgev1.tilde, "x")

#Cumulative density function

pgev1.tilde <- function(x, par, s, log = FALSE){
        gam <- par[1]; mu <- par[2]; delta <- par[3]
        cdf <- 0
        if(delta <= 0){
                if(log == FALSE) return(cdf)
                if(log == TRUE) return(log(cdf))
        }
        else{
                z <- (x - mu)/delta
                #Define 0 neighbourhood
                eps <- .Machine$double.eps^.3
                #the smallest positive floating-point
                #number x such that 1 + x != 1
                if(abs(gam) <= eps){
                        cdf <- (exp(-exp(-z)))^(1/s)
                }
                else{
                        gev.k <- (1 + gam*z)^(-1/gam)
                        cdf <- (exp(-gev.k))^(1/s)
                        if(gam > 0 & ((1 + gam*z) <= 0))
                                cdf = 0
                        if(gam < 0 & ((1 + gam*z) <= 0))
                                cdf = 1
                }
                if(log == FALSE) return(cdf)
                if(log == TRUE) return(log(cdf))
        }
}
pgev.tilde.V <- Vectorize(pgev1.tilde, "x")
```

Appendix B

```
#GP distribution
#Probability density function

dgp1 <- function(x, par, log = FALSE){
        gam <- par[1]; delta <- par[2]
        pdf <- 0
        if(delta <= 0 | x <= 0){
                if(log == FALSE) return(pdf)
                if(log == TRUE) return(log(pdf))
        }
        else{
                z <- x/delta
                #Define 0 neighbourhood
                eps <- .Machine$double.eps^.3
                #the smallest positive floating-point
                # number x
                 such that 1 + x != 1
                if(abs(gam) <= eps){
                        pdf <- (1/delta)*exp(-z)
                }
                else{
                        if((1 + gam*z) > 0){
                                pdf <- (1/delta)*
                                (1 + gam*z)^(-(1/gam) - 1)
                        }
                }
                if(log == FALSE) return(pdf)
                if(log == TRUE) return(log(pdf))
        }
}
dgp.V <- Vectorize(dgp1, "x")

#Cumulative density function

pgp1 <- function(x, par, log = FALSE){
        gam <- par[1]; delta <- par[2]
        cdf <- 0
```

```
        if(delta <= 0 | x <= 0){
                if(log == FALSE) return(cdf)
                if(log == TRUE) return(log(cdf))
        }
        else{
                z <- x/delta
                #Define 0 neighbourhood
                eps <- .Machine$double.eps^.3
                #the smallest positive floating-point
                # number x such that 1 + x != 1
                if(abs(gam) <= eps){
                        cdf <- 1 - exp(-z)
                }
                else{
                        cdf <- 1 - (1 + gam*z)^(-1/gam)
                        if(gam > 0 & ((1 + gam*z) <= 0))
                                cdf <- 0
                        if(gam < 0 & ((1 + gam*z) <= 0))
                                cdf <- 1
                }
                if(log == FALSE) return(cdf)
                if(log == TRUE) return(log(cdf))
        }
}
pgp.V <- Vectorize(pgp1, "x")
```

Appendix C

```
#Censored GEV likelihood

gev.lik <- function(par, data, t = NULL, p = 0.95,
 log = FALSE,
#llik.type = Max-Gev-Cens/ Gev-Cens
llik.type = "Gev-Cens",
#param.type = tilde / bar
param.type = "tilde"){

gam <- par[1]; mu <- par[2]; delta <- par[3]

#Set the smallest value of the log-likelihood
Low <- -1e300
#Check the validity of the required likelihood
if(!any(llik.type == "Max-Gev-Cens",
 llik.type == "Gev-Cens")){
        stop("Must choose the likelihood between
        Max-Gev-Cens and Gev-Cens \n")
        }
        #Check the validity of the parameterization
        if(!any(param.type == "tilde",
         param.type == "bar")){
                stop("Must choose the parameterization
                 between tilde and bar \n")
        }
        #Check the validity of the data
        if(!is.vector(data))
                {stop("Data must be a vector \n")}
        #Check the validity of the parameters
        if(length(par) != 3)
                {stop("Wrong length of
                 parameter vector \n")}
        #Check the support of the scale parameter delta
        if(delta <= 0){
                if(log == FALSE) return(0)
                if(log == TRUE) return(Low)
        }
```

```
#Empirical threshold
#Use discontinuous sample 95% quantile: type = 3
if(is.null(t)){
        t <- as.numeric(quantile(data, probs = p,
         type = 3))
        #message("T set to 95% quantile by default \n")
}
#Indicator of censoring
cond <- data > t
#Exceedances
exceed <- data[cond]
#Sample size
n <- length(data)
#Effective sample size
k <- sum(cond)

if(llik.type == "Max-Gev-Cens"){
        #Likelihood (tilde parameterization)
        if(log == FALSE){
        lik.cens <- (pgev.V(t, par,
         log = FALSE)^(k/n))^(n - k)
        lik.obs <- prod((k/n)*(pgev.V(exceed, par,
         log = FALSE)^(k/n - 1))*
        dgev.V(exceed, par, log = FALSE))
        lik <- prod(lik.cens, lik.obs)
        if(is.infinite(lik)) return(0)
        return(lik)
        }

        #Log-likelihood (tilde parameterization)
        if(log == TRUE){
        llik.cens <- (n - k)*((k/n)*pgev.V(t, par,
         log = TRUE))
        llik.obs <- sum(log(k/n) + dgev.V(exceed, par,
         log = TRUE) +
        (k/n - 1)*pgev.V(exceed, par, log = TRUE))
        llik <- sum(llik.cens, llik.obs)
        if(is.infinite(llik)) return(Low)
```

```
                return(llik)
                }
        }

        if(llik.type == "Gev-Cens"){
        #Define the new parameterization
        #(GEV parameterization)
        if(param.type == "tilde"){
                delta.new <- delta*(k/n)^(gam)
                mu.new <- mu - delta * (1 - (k/n)^gam)/gam
                par.new <- c(gam, mu.new, delta.new)

                #Likelihood (tilde parameterization)
                if(log == FALSE){
                        lik.cens <- (pgev.V(t,
                         par.new, log = FALSE))^(n - k)
                        lik.obs <- prod(dgev.V(exceed,
                         par.new, log = FALSE))
                        lik <- prod(lik.cens, lik.obs)
                        if(is.infinite(lik)) return(0)
                        return(lik)
                        }

                #Log-likelihood (tilde parameterization)
                if(log == TRUE){
                        llik.cens <- (n - k)*(pgev.V(t,
                         par.new, log = TRUE))
                        llik.obs <- sum(dgev.V(exceed,
                         par.new, log = TRUE))
                        llik <- sum(llik.cens, llik.obs)
                        if(is.infinite(llik)) return(Low)
                        return(llik)
                        }
                }
        if(param.type == "bar"){

                #Likelihood (bar parameterization)
                if(log == FALSE){
                        lik.cens <- (pgev.V(t, par,
```

```
                                log = FALSE))^(n - k)
                                lik.obs <- prod(dgev.V(exceed,
                                 par, log = FALSE))
                                lik <- prod(lik.cens, lik.obs)
                                if(is.infinite(lik)) return(0)
                                return(lik)
                                }

                        #Log-likelihood (bar parameterization)
                        if(log == TRUE){
                                llik.cens <- (n - k)*(pgev.V(t, par,
                                 log = TRUE))
                                llik.obs <- sum(dgev.V(exceed, par,
                                 log = TRUE))
                                llik <- sum(llik.cens, llik.obs)
                                if(is.infinite(llik)) return(Low)
                                return(llik)
                                }
                        }
                }
        }

        #Profile likelihood
        #(gamma, mu, \hat{delta}_{(gamma, mu)})
        gev.lik.gammu <- function(par, data, t = NULL, p = 0.95,
         log = FALSE,
        llik.type = "Gev-Cens", param.type = "tilde"){
                par0 <- 5
                prof.lik <- -nlminb(par0,
                 function(x) -gev.lik(par = c(par[1], par[2], x),
                data = data, t = t, p = t.prob,
                log = log, llik.type = llik.type,
                param.type = param.type),
                lower = 1e-10, upper = Inf)$objective
                return(prof.lik)
        }
        #gev.lik.gammu <- Vectorize(gev.lik.gammu, "par")

        #Profile likelihood
```

```
#(gamma, \hat{mu}_{(gamma, delta)}, delta)
gev.lik.gamdel <- function(par, data, t = NULL, p = 0.95,
 log = FALSE,
llik.type = "Gev-Cens", param.type = "tilde"){
        par0 <- 1
        prof.lik <- -nlminb(par0,
         function(x) -gev.lik(par = c(par[1], x, par[2]),
        data = data, t = t, p = t.prob,
        log = log, llik.type = llik.type,
        param.type = param.type),
        lower = -Inf, upper = Inf)$objective
        return(prof.lik)
}
#gev.lik.gamdel <- Vectorize(gev.lik.gamdel, "par")

#Profile likelihood
#(\hat{gamma}_{(mu, delta)}, mu, delta)
gev.lik.mudel <- function(par, data, t = NULL, p = 0.95,
 log = FALSE,
llik.type = "Gev-Cens", param.type = "tilde"){
        par0 <- 1
        prof.lik <- -nlminb(par0,
         function(x) -gev.lik(par = c(x, par[1], par[2]),
        data = data, t = t, p = t.prob,
        log = log, llik.type = llik.type,
        param.type = param.type),
        lower = -Inf, upper = Inf)$objective
        return(prof.lik)
}
#gev.lik.mudel <- Vectorize(gev.lik.mudel, "par")
```

Appendix D

```
#Uniform prior for (gamma, mu, delta)

gev.unif.prior <- function(par, log = FALSE){

        #Same prior for tilde and bar parameterization
        #Jeffreys' prior in scale and location families
        #Jeffreys' prior is parameterization invariant
        delta <- par[3]
        pi.prior <- 1/delta
        if(log == FALSE) return(pi.prior)
        if(log == TRUE) return(log(pi.prior))
}

#Empirical Bayes prior distribution for (gamma, mu, delta)

gev.emp.prior <- function(par, hyp,
#param.type = tilde / bar
param.type = "tilde",
#s = n/k
s.frac = NULL,
log = FALSE){

        #Parameters
        gam <- par[1]; mu <- par[2]; delta <- par[3]
        #Hyperparameters (hatmu, hatdelta)
        bk <- hyp[1]; ak <- hyp[2]

        if(param.type == "tilde"){
                #gamma prior
                pi.gam <- (1 - pt(-1, df = 1))^(-1)*dt(gam,
                 df = 1)*I(gam > -1)
                #Mu prior
                pi.mu <- (1/ak)*dnorm((mu - bk)/ak)
                #Delta prior
                pi.del <- (1/ak)*dgamma(delta, shape = 1,
                 scale = ak)*I(delta > 0)
```

```
                #Prior
                if(log == FALSE)
                        return(prod(pi.gam, pi.mu, pi.del))
                if(log == TRUE)
                        return(sum(log(pi.gam),
                         log(pi.mu), log(pi.del)))
        }

        if(param.type == "bar"){

                #Set s
                if(is.null(s.frac)){
                        stop("Set the inverse of the
                         proportion of exceedances s \n")
                }

                #Define 0 neighbourhood
                eps <- .Machine$double.eps^.3
                #the smallest positive floating-point
                #number x such that 1 + x != 1
                if(abs(gam) <= eps){
                        #Prior components
                        pi.1 <- (1 - pt(-1, df = 1))^(-1)
                        *dt(gam, df = 1)*I(gam > -1)
                        pi.2 <- (1/ak)*dnorm((mu - bk +
                        (delta*log(s.frac)))/ak)
                        pi.3 <- (1/ak^2)*exp(-(1/ak)*delta)
                        *I(delta > 0)
                }
                else{
                        #Prior components
                        pi.1 <- (s.frac^(gam))*
                        (1 - pt(-1, df = 1))^(-1)*
                        dt(gam, df = 1)*I(gam > -1)
                        pi.2 <- (1/ak)*dnorm((mu - bk +
                         ((s.frac^(gam))*delta*
                         (1 - s.frac^(-gam)))/gam)/ak)
                        pi.3 <- (1/ak^2)*
                        exp(-(1/ak)*(s.frac^(gam))*
```

```
                        delta)*I(delta > 0)
                }

                #Prior
                if(log == FALSE) return(prod(pi.1, pi.2, pi.3))
                if(log == TRUE) return(sum(log(pi.1),
                 log(pi.2), log(pi.3)))
        }
}

#Posterior distribution for (gamma, mu, delta)

gev.post <- function(par, data, hyp, t = NULL, p = 0.95,
 log = FALSE,
#llik.type = Max-Gev-Cens/ Gev-Cens
llik.type = "Gev-Cens",
#param.type = tilde / bar
param.type = "tilde",
#prior = "uniform" / "empirical"
prior = "empirical"){
        if(prior == "uniform"){
                if(log == FALSE)
                        return(prod(gev.lik(par, data, t = t,
                        p = p, log = FALSE,
                        llik.type = llik.type,
                        param.type = param.type),
                        gev.unif.prior(par, log = FALSE)))
                if(log == TRUE)
                        return(sum(gev.lik(par, data, t = t,
                        p = p, log = TRUE,
                        llik.type = llik.type,
                        param.type = param.type),
                        gev.unif.prior(par, log = TRUE)))
        }
        if(prior == "empirical"){

                #Empirical threshold
                #Use discontinuous sample 95% quantile: type = 3
                if(is.null(t)){
```

```
                    t <- as.numeric(quantile(data,
                     probs = p, type = 3))
                }
                #Indicator of censoring
                cond <- data > t
                #Sample size
                n <- length(data)
                #Effective sample size
                k <- sum(cond)
                #s = n/k
                s.frac <- n/k

                if(log == FALSE)
                     return(prod(gev.lik(par, data,
                     t = t, p = p, log = FALSE,
                    llik.type = llik.type,
                    param.type = param.type),
                    gev.emp.prior(par, hyp,
                    param.type = param.type,
                    s.frac = s.frac,
                    log = FALSE)))
                if(log == TRUE)
                     return(sum(gev.lik(par, data,
                     t = t, p = p, log = TRUE,
                    llik.type = llik.type,
                    param.type = param.type),
                    gev.emp.prior(par, hyp,
                    param.type = param.type,
                    s.frac = s.frac,
                    log = TRUE)))
            }
        }
```

Appendix E

```
#Adaptive Gaussian Random Walk Metropolis-Hastings Markov
#Chain Monte Carlo

adamh <- function(R, data, hyp, par0, k0, t = NULL, p = p,
#llik.type = Max-Gev-Cens/ Gev-Cens
llik.type = "Gev-Cens",
#param.type = tilde / bar
param.type = "tilde",
#prior = "uniform" / "empirical"
prior = "empirical",
#etastar: desired overall sampler acceptance probability
etastar = 0.234){

        #Required packages
        require(mvtnorm)

        #Set parameters constraints
        parcheck <- function(para){
                res <- any(para[3] <= 0,
                (para[1] > 0 &
                 (para[2] > min(data) + para[3]/para[1])),
                (para[1] < 0 &
                 any(para[2] <= data + para[3]/para[1])))
                return(res)
        }
        n <- length(data)
        d <- length(par0)
        if(!any(d == 3)){stop("Wrong length of parameter vector")}
        #Correct specification of the prior
        if(!any(prior == "uniform", prior == "empirical")){
                stop("Specify a prior type within
                 uniform and empirical")
        }

        #Quantities related to the adaptive update of k
        zeta0 <- -qnorm(etastar/2)
        #Steplength constant
```

```
#Suitable overstimate of steplength constant
#in multivariate framework
a <- (1 - 1/d)*(sqrt(2*pi))*
exp((zeta0^2)/2)/(2*zeta0) +
1/(d*etastar*(1 - etastar))

out <- array(dim = c(R, d))
#Acceptance vector
accepted <- rep(0, R)
#Acceptance probabilities vector
accepted.prob <- rep(0, R)
#Automatic rejection counter
#(proposed parameters that don't respect
#the constraints)
straight.reject <- rep(0, R)
#Vector of scaling parameters k
k.vec <- k0
#Monitor the adequate values of k
k.start <- k0
k.restart <- k0
#Number of iterations before the beginning
#of RM search
n0 <- round(5/(etastar*(1 - etastar)))
#Max number of iterations before the last restart
iMax <- 100
Numbig <- 0
Numsmall <- 0
#Counter of the iteration number at each restart
numRS <- 1
#Message about the outcome of the algorithm
msg <- "MCMC run without errors"
#Initialization
sigma0 <- diag(d)
par <- par0; sigma <- sigma0; k <- k0

#Progress bar
pb <- txtProgressBar(min = 0, max = R,
 initial = 0, style = 3)
```

```
for(i in 1:R){
        #Adequacy of the covariance matrix
        s <- try(eigen((k^2)*sigma,
         symmetric = TRUE),
         silent = TRUE)
        if(class(s) == "try-error"){
                msg <- "MCMC failed"
                return(list(values = out,
                accepted.vec = accepted,
                accepted = mean(accepted),
                accepted.prob = accepted.prob,
                straight.reject = straight.reject,
                k.vec = k.vec,
                k.restart = k.restart,
                msg = msg))
        }

        #Proposal
        if(d == 1) pars <- rnorm(1, sigma = k)
        #pars <- par + rmvnorm(1, sigma = k*sigma)
        else pars <- par +
         rmvnorm(1, sigma = (k^2)*sigma)
        numRS <- numRS + 1
        #Check for NA proposed values
        if(any(is.na(pars))){
                straight.reject[i] <- 1
                accepted.prob[i] <- 0
        }
        #Check basis condition
        else{
        if(parcheck(pars)){
                straight.reject[i] <- 1
                accepted.prob[i] <- 0
                }
        #Compute the acceptance probability
        else{
                eta <- min(1, exp(gev.post(pars,
                data, hyp, t = t, p = p, log = TRUE,
                llik.type = llik.type,
```

```
                        param.type = param.type,
                        prior = prior) -
                        gev.post(par, data, hyp,t = t,
                        p = p,log = TRUE,
                        llik.type = llik.type,
                        param.type = param.type,
                        prior = prior)))
                        #Check for unsuccessful
                        #probability computation
                        if(is.na(eta)){
                        straight.reject[i] <- 1
                        accepted.prob[i] <- 0
                        eta <- 0
                                }
                        else{
                        accepted.prob[i] <- eta
                        #Accepatance/rejection step
                        if(runif(1) < eta){
                                par <- pars
                                accepted[i] <- 1
                                }
                        }
                }
        }
        out[i,] <- par

        #Adaptive covariance matrix update
        #(Haario et al. 2001)
        if(i > 100 && d > 1){
                if(i == 101){
                #Update partial covariance matrix
                sigMat <- cov(out[1:i,])
                #Update partial mean vector
                thetaM <- apply(out[1:i,], 2, mean)
                sigma <- sigMat
                                + diag(d)/i
                }
                else{
                #Recursive update of partial
```

```
                #mean vector
                thetaM2 <- (thetaM*(i - 1) + out[i,])/i
                #Recursive update of partial
                #covariance matrix
                sigMat <- (i - 2)/(i - 1)*sigMat +
                        thetaM%*%t(thetaM) -
                (i)/(i - 1)*thetaM2%*%t(thetaM2)
                + 1/(i - 1)*out[i,]%*%t(out[i,])
                sigma <- sigMat
                + diag(d)/i
                #Recursive update of partial
                #mean vector
                thetaM <- thetaM2
                }
        }

        #Adaptive k update via Robbins-Monro process
        #(Garthwaite et al. 2016)
        if(i > n0){
        if(d == 1)
                kstar <- exp(log(k) + a*(eta - etastar)/i)
                else
                kstar <- exp(log(k) +
                         a*(eta - etastar)/max(200, i/d))
                k <- kstar
                k.vec <- c(k.vec, k)
                if ((i <= (iMax + n0)) &&
                         (Numbig < 5 || Numsmall < 5)) {
                        Toobig <- (k > (3*k.start))
                        Toosmall <- (k < (k.start/3))
                        if (Toobig || Toosmall) {
                                #Restart the algorithm
                                message("\n", "Restart the
                                program at ", numRS,
                                        "th iteration")
                                #k.restart
                                k.restart <- c(k.restart, k)
                                Numbig <- Numbig + Toobig
                                Numsmall <- Numsmall +
```

```
                                    Toosmall
                                    i <- n0
                                    k.start <- k
                                    }
                            }
                    }

            #Progress bar
            print.i <- seq(0, R, by = 100)
            if(i %in% print.i) setTxtProgressBar(pb, i)

            }
            #Return object
            return(list(values = out,
            accepted.vec = accepted,
            accepted = mean(accepted),
            accepted.prob = accepted.prob,
            straight.reject = straight.reject,
            k.vec = k.vec,
            k.restart = k.restart,
            msg = msg))
    }
```

Appendix F

```
#Censored POT frequentist/Empirical Bayes inference

fit.gev.inference <- function(data, t = NULL, t.prob = 0.95,
llik.type = "Gev-Cens", param.type = "tilde",
T.ret = 50, p = 1/length(data),
par0 = NULL, optim.meth = "Nelder-Mead", control = NULL,
hessian = FALSE,
#inf.type = c("Frequent", "Bayes")
inf.type = NULL,
#prior = "uniform" / "empirical"
prior = "empirical",
#Posterior chain size
R = NULL,
#Burn-in
burn = NULL,
#Starting RW steplength k
k = 1,
#Desired overall sampler acceptance probability
etastar = 0.234,
#Plot about the MCMC algorithm
val.show = FALSE,
...){

#Choose one of the possible types of inference
if(is.null(inf.type) || (inf.type != "Frequent" &&
inf.type != "Bayes")){
                stop("Need to specify the type of inference
                 between: Frequent or Bayes")
        }
#Setting of the starting value of the optimizator
if(is.null(par0))
         {stop("Need to specify a starting value
                 for the optimization")}
#Empirical threshold
#Use discontinuous sample 95% quantile: type = 3
if(is.null(t)){
        t <- as.numeric(quantile(data, probs = t.prob,
```

```
          type = 3))
        message(paste0("T set to the ", round(100*t.prob, 0),
         "% quantile by default"))
}
#Optimization method:
#c("Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN", "Brent")
if(is.null(optim.meth))
        {stop("Need to specify an existing
                 optimisation method!")}
#Flow control parameters
if(!("control" %in% names(sys.call()))){
        #Maximization
        control <- list(fnscale = -1, maxit = 8e+5)
}
else{
        if(!("fnscale" %in% names(control))){
                #Maximization
                control[["fnscale"]] <- -1
                control[["maxit"]] <- 8e+5
        }
}

#Sample size
n <- length(data)
#Effective sample size
n.eff <- sum(data > t)
#Define the proportion of exceedances
#exc.prop <- k/n
exc.prop <- n.eff/n
names(exc.prop) <- c("Excesses proportion")
#s = Reciprocal of proportion of exceedances
#s = (n/k)
s <- 1/exc.prop

#Outcome setting
optim.msg <- "Something went wrong"
optimfun <- function(par){
        return(gev.lik(par = par, data = data, t = t,
        p = t.prob, log = TRUE,
```

```
        llik.type = llik.type, param.type = param.type))
}
#Compute the maximum consered likelihood estimates
#Hessian = TRUE preferred for minimizations
mle <- optim(par0, optimfun, method = optim.meth,
 control = control, hessian = hessian)

#Define the estimated GEV parameters
names(mle$par) <- c("Shape", "Location", "Scale")

#Frequentist inference setting
if(inf.type == "Frequent"){

        optim.msg <- mle$convergence
        names(optim.msg) <- c("Optimization message")

        #Define the extreme quantiles
        Q.ext <- mle$par[2] + mle$par[3] *
         ((s*p)^(-mle$par[1]) - 1) / mle$par[1]
        names(Q.ext) <- c("Extreme-Quantile")
        #Return level associated with the return period T.ex
        R.lev <- mle$par[2] + mle$par[3] *
         ((-log(1 - 1/T.ret))^(-mle$par[1]) - 1) / mle$par[1]
        names(R.lev) <- c("Return-Level")

        #Return object
        if(hessian == FALSE){
                return(list(inference = inf.type,
                                        mle = mle$par,
                                        max.lik = mle$value,
                                        hessian = mle$hessian,
                                        Q.extreme = Q.ext,
                                        R.level = R.lev,
                                        optim.msg = optim.msg,
                                        exc.prop = exc.prop))
        }
        else{
                #Preferred for minimizations
                return(list(inference = inf.type,
```

```
                                        mle = mle$par,
                                        max.lik = mle$value,
                                        hessian = mle$hessian,
                                        Q.extreme = Q.ext,
                                        R.level = R.lev,
                                        optim.msg = optim.msg,
                                        exc.prop = exc.prop))
        }
}

#Bayesian inference setting
if(inf.type == "Bayes"){
        #Outcome setting
        optim.msg <- "Something went wrong"
        #Check for the number of posterior samples
        if(is.null(R))
                {stop("Missing number of replications
                        for MCMC")}
        #Set the number of quantiles to extrapolate
        nQuant <- length(p)
        #Correct specification of the prior
        if(!any(prior == "uniform", prior == "empirical")){
                stop("Specify a prior type within
                 uniform and empirical")
        }

        #Empirical Bayes
        #Maximum likelihood estimates for norming constants
        #in the tilde parameterization
        optimfun <- function(par){
                return(gev.lik(par = par, data = data,
                 t = t, p = t.prob, log = TRUE,
                llik.type = "Gev-Cens",
                param.type = "tilde"))
        }
        #Compute the maximum consered likelihood
        #of the tilde parameters
        mle.tilde <- optim(par0, optimfun,
         method = optim.meth, control = control,
```

```
hessian = hessian)
#Prior hyperparameters
hyp <- mle.tilde$par[2:3]
hyp.ret <- mle.tilde$par[2:3]
names(hyp.ret) <- c("b(n/k)", "a(n/k)")
#Optimization
optimfun <- function(par){
        return(gev.post(par = par, data = data,
        hyp = hyp,
        t = t, p = t.prob,
        llik.type = llik.type,
        param.type = param.type,
        prior = prior, log = TRUE))
}
#Compute the posterior mode estimates
#Hessian = TRUE preferred for minimizations
mode <- optim(par0, optimfun, method = optim.meth,
 control = control,
hessian = hessian)
optim.msg <- mode$convergence
names(optim.msg) <- c("Optimization message")
#Posterior mode GEV parameters
names(mode$par) <- c("Shape", "Location", "Scale")
#Posterior chain
post.mcmc <- adamh(R = R, data = data, hyp = hyp,
                        par0 = par0, k0 = k,
                        t = t, p = p, prior = prior,
                        llik.type = llik.type,
                        param.type = param.type,
                        etastar = etastar)
#Acceptance rate
post.acc <- post.mcmc$accepted
mean.acc <- rep(NA, length(post.mcmc$accepted.prob))
for(j in c(1:length(post.mcmc$accepted.prob))){
mean.acc[j] <- mean(post.mcmc$accepted.prob[round(j/2) :j])
}
index <- c(1:2000, seq(2001, length(post.mcmc$k.vec),
 by = 100))
#Plots of steplength and acceptance probability
```

```
if(val.show){
        par(mfrow=c(1, 2))
        plot(cbind(index, post.mcmc$k.vec[index]^2),
         type = "l", col = 3,
        ylim = c(0, max(post.mcmc$k.vec^2)),
        ylab = expression(kappa),
        xlab = "Iterations", lwd = 2)
        abline(h = 0, lwd = 2)
        plot(cbind(index, mean.acc[index]),
         type = "l", col = 2, ylim = c(0,1),
                ylab = "Acceptance Probability",
                xlab = "Iterations", lwd = 2)
        abline(h = etastar, lwd = 2)
        par(mfrow = c(1, 1))
        #Deduce approximate samples from the
        #parameter's posterior distributions
}
#Posterior distribution of the original parameters
#(after burn-in)
if(is.null(burn)){
        burn <- round(R/4)
        message(paste0("\n Burn-in set to the 25% of
         ", R, " by default"))
}
post.sam <- post.mcmc$values[(burn + 1):R,]
scale <- post.sam[,3]
loc <- post.sam[,2]
colnames(post.sam) <- c("Shape", "Location", "Scale")
#Extreme quantile chain
Q.ext <- matrix(NaN, nrow = nrow(post.sam),
 ncol= nQuant)
for(j in 1:nQuant){
        Q.ext[,j] <- loc + scale *
         ((exc.prop/p[j])^post.sam[,1] - 1)/
                post.sam[,1]
}
#Return level
R.lev <- loc + scale *
        ((-log(1 - 1/T.ret))^(-post.sam[,1]) - 1)/
```

```
        post.sam[,1]

#Return object
if(hessian == FALSE){
        return(list(inference = inf.type,
                    mle = mle$par,
                    mode = mode$par,
                    max.post = mode$value,
                    optim.msg = optim.msg,
                    emp.bayes.hyp = hyp.ret,
                    mcmc.acc = post.acc,
                    parameters = post.sam,
                    Q.extreme = Q.ext,
                    R.level = R.lev,
                    straight.reject =
                post.mcmc$straight.reject[(burn + 1):R],
                    k.vec = post.mcmc$k.vec,
                    accept.prob =
                     cbind(index, mean.acc[index]),
                    msg = post.mcmc$msg,
                    exc.prop = exc.prop))
}
else{
        return(list(inference = inf.type,
                    mle = mle$par,
                    mode = mode$par,
                    hessian = mode$hessian,
                    max.post = mode$value,
                    optim.msg = optim.msg,
                    emp.bayes.hyp = hyp.ret,
                    mcmc.acc = post.acc,
                    parameters = post.sam,
                    Q.extreme = Q.ext,
                    R.level = R.lev,
                    straight.reject =
                post.mcmc$straight.reject[(burn + 1):R],
                    k.vec = post.mcmc$k.vec,
                    accept.prob =
                     cbind(index, mean.acc[index]),
```

```
                                                msg = post.mcmc$msg,
                                                exc.prop = exc.prop))
                    }
          }
}
```

Bibliography

Beranger, B., Padoan, S. A. & Sisson, S. A. (2021). Estimation and uncertainty quantification for extreme quantile regions. *Extremes* **24**, 349–375.

Birkhoff, G. D. (1931). Proof of the Ergodic Theorem. *Proceedings of the National Academy of Sciences* **17**, 656–660.

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

Craiu, R. V., Rosenthal, J. & Yang, C. (2009). Learn from thy neighbor: Parallel-chain and regional adaptive mcmc. *Journal of the American Statistical Association* **104**, 1454–1466.

Davison, A. C. (2003). *Statistical Models*. Cambridge University Press.

De Haan, L. & Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer.

Dombry, C. & Ferreira, A. (2019). Maximum likelihood estimators based on the block maxima method. *Bernoulli* **25**, 1690–1723.

Fisher, R. A. & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society* **24**, 180–190.

Garthwaite, P. H. (1996). Confidence intervals from randomization tests. *Biometrics* **52**, 1387–1393.

Garthwaite, P. H., Fan, Y. & Sisson, S. A. (2016). Adaptive optimal scaling of Metropolis–Hastings algorithms using the Robbins–Monro process. *Communications in Statistics-Theory and Methods* **45**, 5098–5111.

Gelman, A., Gilks, W. R. & Roberts, G. O. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability* **7**, 110–120.

Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457–472.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics* **44**, 423–453.

Haario, H., Saksman, E. & Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* **7**, 223–242.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

Hosking, J. R. M. & Wallis, J. R. (1997). *Regional Frequency Analysis: An Approach based on L Moments*. Cambridge University Press.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* **21**, 1087–1092.

Padoan, S. A. & Rizzelli, S. (2022). Empirical Bayes inference for the block maxima method. *arXiv preprint arXiv:2204.04981* .

Plummer, M., Best, N., Cowles, K. & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News* **6**, 7–11.

Prescott, P. & Walden, A. (1983). Maximum likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution from censored samples. *Journal of Statistical Computation and Simulation* **16**, 241–250.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reid, N. (1996). Likelihood and Bayesian approximation methods. *Bayesian Statistics* **5**, 349–366.

Robbins, H. & Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics* **22**, 400–407.

Robbins, H. E. (1956). An empirical Bayes approach to statistics. *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability* **1**, 351–368.

Robert, C. P. & Casella (2010). *Introducing Monte Carlo Methods with R*. Springer.

Roberts, G. O. & Rosenthal, J. S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science* **16**, 351–367.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72**, 67–90.

Smith, R. L. (1994). Multivariate threshold methods. In *Extreme Value Theory and Applications*. Springer, pp. 225–248.

Stephenson, A. G. (2002). evd: Extreme value distributions. *R News* **2**, 0.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press.

Ypma, T. J. (1995). Historical development of the Newton–Raphson method. *SIAM review* **37**, 531–551.