

# mev package tutorial

*Léo Belzile*

*2019-06-30*



# Contents

<b>Preliminary remarks</b>	<b>5</b>
<b>1 Likelihood based inference for univariate extremes</b>	<b>7</b>
1.1 Basic theory . . . . .	7
1.2 Numerical example . . . . .	9
1.3 Your turn . . . . .	13
<b>2 Threshold selection</b>	<b>15</b>
2.1 Threshold stability plots . . . . .	16
2.2 Wadsworth's diagnostics: white noise process and simultaneous threshold stability plots . . . . .	16
2.3 Changepoint tests based on penultimate approximation . . . . .	17
2.4 Extended generalized Pareto . . . . .	17
2.5 Robust selection . . . . .	17
2.6 Your turn . . . . .	21
<b>3 Penultimate approximation</b>	<b>23</b>
3.1 Theory . . . . .	23
3.2 Your turn . . . . .	27
<b>4 Simulation</b>	<b>29</b>
4.1 Max-stable processes . . . . .	29
4.2 $\mathcal{R}$ -Pareto processes and generalizations . . . . .	29
4.3 Algorithms . . . . .	30
4.4 Extremal coefficient . . . . .	34
4.5 Generalized $\mathcal{R}$ -Pareto processes for functional threshold exceedances . . . . .	36
4.6 Your turn . . . . .	41



# Preliminary remarks

The `mev` package bundles routine for - likelihood-based inference for univariate extremes (including distributions, densities, score, information matrix and higher order asymptotics). The four basic likelihood (generalized extreme value, generalized Pareto, non-homogeneous Poisson process and  $r$ -largest) are implemented. For the first two distributions, additional higher order inference tools are available, including profile likelihoods in most parametrizations of interest. Many threshold selection diagnostics that have recently been proposed in the literature are implemented.

The second includes multivariate goodness-of-fit diagnostics and tests (max-stability test, information matrix test, coefficient of tail dependence, extremal coefficient, estimators of the extremal index, extremogram, nonparametric estimation of the angular measure, etc.) Pairwise composite likelihood for general models will be included in a future update.

Last, but not least, the package includes functions to simulate from multivariate max-stable vectors in dimension  $d \geq 2$ , including extensions of most of the families in the `evd` package. The package also includes simulation algorithms for  $R$ -Pareto processes using accept-reject and composition sampling, and accept-reject for generalized  $R$ -Pareto processes.

You can install the latest version of the package (v. 1.12) directly from CRAN via

```
install.packages("mev")
```



# Chapter 1

## Likelihood based inference for univariate extremes

### 1.1 Basic theory

Let  $\ell(\mathbf{y}; \boldsymbol{\theta})$  denotes the log-likelihood of an  $n$  sample with a  $p$ -dimensional parameter  $\boldsymbol{\theta}$ . The score vector is  $U(\boldsymbol{\theta}) = \partial \ell / \partial \boldsymbol{\theta}$ , while the Fisher information is  $i(\boldsymbol{\theta}) = E\{U(\boldsymbol{\theta})U(\boldsymbol{\theta})^\top\}$ . Under regularity conditions, we also have  $i(\boldsymbol{\theta}) = -E(\partial^2 \ell / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top)$ . The observed information is the negative Hessian  $-\partial^2 \ell / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$ , evaluated at the maximum likelihood estimator  $\hat{\boldsymbol{\theta}}$ .

By definition, the maximum likelihood estimator solves the score equation, i.e.  $U(\hat{\boldsymbol{\theta}}) = \mathbf{0}_p$ . If the maximum likelihood estimator is not available in closed-form, its solution is found numerically and this property can be used to verify that the optimization routine has converged or for gradient-based maximization algorithms.

#### 1.1.1 Likelihoods

There are four basic likelihoods for univariate extremes: the likelihood of the generalized extreme value (GEV) distribution for block maxima, the likelihood for the generalized Pareto distribution and that of the non-homogeneous Poisson process (NHPP) for exceedances above a threshold  $u$  and lastly the likelihood of the  $r$ -largest observations.

The generalized extreme value (GEV) distribution with location parameter  $\mu \in \mathbb{R}$ , scale parameter  $\sigma \in \mathbb{R}_+$  and shape parameter  $\xi \in \mathbb{R}$  is

$$G(x) = \begin{cases} \exp\left\{-\left(1 + \xi \frac{x-\mu}{\sigma}\right)^{-1/\xi}\right\}, & \xi \neq 0, \\ \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}, & \xi = 0, \end{cases}$$

defined on  $\{x \in \mathbb{R} : \xi(x - \mu)/\sigma > -1\}$  where  $x_+ = \max\{0, x\}$ . The case  $\xi = 0$  is commonly known as the Gumbel distribution. We denote the distribution by  $\text{GEV}(\mu, \sigma, \xi)$ .

The generalized Pareto (GP) distribution with scale  $\sigma \in \mathbb{R}_+$  and shape  $\xi \in \mathbb{R}$  is

$$G(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\sigma}\right)_+^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(-\frac{x}{\sigma}\right), & \xi = 0. \end{cases}$$

The range of the generalized Pareto distribution is  $[0, -\sigma/\xi]$  if  $\xi < 0$  and is  $\mathbb{R}_+$  otherwise. We denote the distribution by  $\text{GP}(\sigma, \xi)$ .

Let  $Y_{(1)} \geq \dots \geq Y_{(r)}$  denote the  $r$  largest observations from a sample. The likelihood of the limiting distribution of the point process for the  $r$ -largest observations is, for  $\mu, \xi \in \mathbb{R}, \sigma > 0$ ,

$$\ell(\mu, \sigma, \xi; \mathbf{y}) \equiv -r \log(\sigma) - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^r \log \left(1 + \xi \frac{y_{(j)} - \mu}{\sigma}\right)_+ - \left(1 + \xi \frac{y_{(r)} - \mu}{\sigma}\right)_+^{-1/\xi}.$$

This likelihood can be used to model the  $r$ -largest observations per block or threshold exceedances where the threshold is the  $r$ th order statistic

Consider a sample of  $N$  observations, of which  $n_u$  exceed  $u$  and which we denote by  $y_1, \dots, y_{n_u}$ . The likelihood associated to the limiting distribution of threshold exceedances is, for  $\mu, \xi \in \mathbb{R}, \sigma > 0$ ,

$$L(\mu, \sigma, \xi; \mathbf{y}) = \exp \left[ -c \left\{ 1 + \xi \left( \frac{u - \mu}{\sigma} \right) \right\}_+^{-1/\xi} \right] (c\sigma)^{-n_u} \prod_{i=1}^{n_u} \left\{ 1 + \xi \left( \frac{y_i - \mu}{\sigma} \right) \right\}_+^{-1/\xi - 1}, \quad (1.1)$$

where  $(\cdot)_+ = \max\{0, \cdot\}$ . The quantity  $c$  is a tuning parameter whose role is described in §7.5 of Coles (2001). If we take  $c = N/m$ , the parameters of the point process likelihood correspond to those of the generalized extreme value distribution fitted to blocks of size  $m$ . The NHPP likelihood includes a contribution for the fraction of points that exceeds the threshold, whereas the generalized Pareto is a conditional distribution, whose third parameter is the normalizing constant  $\zeta_u = \Pr(Y > u)$ . Since the latter has a Bernoulli and  $\zeta_u$  is orthogonal to the pair  $(\sigma, \xi)$ , it is often omitted from further analyses and estimated as the proportion of samples above the threshold.

### 1.1.2 Test statistics and the profile likelihood

The three main type of test statistics for likelihood-based inference are the Wald, score and likelihood ratio tests. The three main classes of statistics for testing a simple null hypothesis  $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$  against the alternative  $H_a : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$  are the likelihood ratio, the score and the Wald statistics, defined respectively as

$$\begin{aligned} w &= 2 \{ \ell(\hat{\boldsymbol{\theta}}) - \ell(\boldsymbol{\theta}_0) \}, \\ w_{\text{score}} &= U^\top(\boldsymbol{\theta}_0) i^{-1}(\boldsymbol{\theta}_0) U(\boldsymbol{\theta}_0), \\ w_{\text{wald}} &= (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top i(\boldsymbol{\theta}_0) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \end{aligned}$$

where  $\hat{\boldsymbol{\theta}}$  is the maximum likelihood estimate under the alternative and  $\boldsymbol{\theta}_0$  is the null value of the parameter vector. The statistics  $w, w_{\text{score}}, w_{\text{wald}}$  are all first order equivalent and asymptotically follow a  $\chi_p^2$  distribution, where  $q$  is the difference between  $p$  and the number of parameters under the null hypothesis. Under the conditions of the Neyman–Pearson theorem, the likelihood ratio test is most powerful test of the lot. The score statistic  $w_{\text{score}}$  only requires calculation of the score and information under  $H_0$ , which can be useful in problems where calculations under the alternative are difficult to obtain. The Wald statistic  $w_{\text{wald}}$  is not parametrization-invariant and typically has poor coverage properties.



Oftentimes, we are interested in a functional of the parameter vector  $\theta$ . The profile likelihood  $\ell_p$ , a function of  $\psi$  alone, is obtained by maximizing the likelihood pointwise at each fixed value  $\psi = \psi_0$  over the nuisance vector  $\lambda_{\psi_0}$ ,

$$\ell_p(\psi) = \max_{\lambda} \ell(\psi, \lambda) = \ell(\psi, \hat{\lambda}_{\psi}).$$

We denote the restricted maximum likelihood estimator  $\hat{\theta}_{\psi} = (\psi, \hat{\lambda}_{\psi})$ .

We can define score and information in the usual fashion: for example, the observed profile information function is

$$j_p(\psi) = -\frac{\partial \ell_p(\psi)}{\partial \psi \partial \psi^T} = \{j^{\psi\psi}(\psi, \hat{\lambda}_{\psi})\}^{-1}.$$

The profile likelihood is not a genuine likelihood in the sense that it is not based on the density of a random variable.

We can turn tests and their asymptotic distribution into confidence intervals. For the hypothesis  $\psi = \psi_0$ , a  $(1 - \alpha)$  confidence interval based on the profile likelihood ratio test is  $\{\psi : 2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_{\psi})\} \leq \chi_1^2(0.95)\}$ .

Two typical questions in extreme values are: given the intensity of an extreme event, what is its recurrence period? and what is a typical worst-case scenario over a given period of time? For the latter, suppose for simplicity that the daily observations are blocked into years, so that inference is based on  $N$  points for the  $N$  years during which the data were recorded. The *return level* is a quantile of the underlying distribution corresponding to an event of probability  $p = 1 - 1/T$  for an annual maximum, which is interpreted as “the level exceeded by an annual maximum on average every  $T$  years”. If observations are independent and identically distributed, then we can approximate the probability that a return level is exceeded  $l$  times over a  $T$  year period using a binomial distribution with probability of success  $1 - 1/T$  and  $T$  trials. For  $T$  large, the return level is exceeded  $l = 0, 1, 2, 3, 4$  times within any  $T$ -years period with approximate probabilities 36.8%, 36.8%, 18.4%, 6.1% and 1.5%. The probability that the maximum observation over  $T$  years is exceeded with a given probability is readily obtained from the distribution of the  $T$ -year maximum, leading (Cox et al., 2002, §3(b)) to advocate its use over return levels, among other quantities of interest such as the number of times a threshold  $u$  will be exceeded in  $T$  years or the average number of years before a threshold  $u$  is exceeded.

**Quantiles, mean and return levels of  $T$ -maxima:** consider the distribution  $H(x) = G^T(x)$  of the maximum of  $T$  independent and identically distributed generalized extreme value variates with parameters  $(\mu, \sigma, \xi)$  and distribution function  $G$ . By max-stability, the parameters of  $H(x)$  are  $\mu_T = \mu - \sigma(1 - T^{\xi})/\xi$  and  $\sigma_T = \sigma T^{\xi}$  when  $\xi \neq 0$ . We denote the expectation of the  $T$ -observation maximum by  $e_T$ , the  $p$  quantile of the  $T$ -observation maximum by  $q_p = H^{-1}(p)$  and the associated return level by  $z_{1/T} = G^{-1}(1 - 1/T)$ . Then, any of these three quantities can be written as

$$\begin{cases} \mu - \frac{\sigma}{\xi} \{1 - \kappa_{\xi}\}, & \xi < 1, \xi \neq 0, \\ \mu + \sigma \kappa_0, & \xi = 0, \end{cases}$$

where  $\kappa_{\xi} = T^{\xi} \Gamma(1 - \xi)$  for  $e_T$ ,  $\kappa_{\xi} = T^{\xi} \log(1/p)^{-\xi}$  for  $q_p$  and  $\kappa_{\xi} = \{-\log(1 - 1/T)\}^{-\xi}$  for  $z_{1/T}$ . In the Gumbel case, we have  $\kappa_0 = \log(T) + \gamma_e$  for  $e_T$ ,  $\kappa_0 = \log(T) - \log\{-\log(p)\}$  for  $q_p$  and  $\kappa_0 = -\log\{-\log(1 - 1/T)\}$  for  $z_{1/T}$ .

## 1.2 Numerical example

This example illustrates some of the functions used in peaks-over-threshold analysis based on fitting a generalized Pareto distribution to threshold exceedances. We use the Venezuelan rainfall data, a time series of daily rainfall

precipitations at Maiquetia airport in Venezuela, for the purpose of illustration.

```
library(mev)
data("maiquetia", package = "mev")
day <- seq.Date(from = as.Date("1961-01-01"),
               to = as.Date("1999-12-31"), by = "day")
# Keep non-zero rainfall, exclude 1999 observations
nzrain <- maiquetia[substr(day, 3, 4) < 99 & maiquetia > 0]
gpdf <- fit.gpd(nzrain, threshold = 20)
print(gpdf)
```

```
## Method: Grimshaw
## Log-likelihood: -832.629
##
## Threshold: 20
## Number Above: 216
## Proportion Above: 0.0604
##
## Estimates
##   scale   shape
## 15.5800  0.1088
##
## Standard Errors
##   scale   shape
## 1.60673  0.07785
##
## Optimization Information
##   Convergence: successful
```

We will ignore temporal dependence and stationarity, but these should be considered. The first step in our analysis is to choose a threshold. For the time being, we set the latter to 20 and consider threshold selection in the next section.

The default optimization routine for the generalized Pareto distribution is Grimshaw's method, which profiles out the likelihood. The method has theoretical convergence guarantees for convergence. Because of non-regularity, the maximum likelihood estimator for  $\xi < -1$  does not solve the score equation and leads to infinite log-likelihood, hence the maximum returned lies on the boundary of the parameter space. The standard errors are based on the inverse observed information matrix and provided only if  $\xi > -1/2$ . We can verify that our maximum likelihood estimate is indeed a maximum by checking if it solves the score equation if  $\hat{\xi} > -1$ .

```
isTRUE(all.equal(
  gpd.score(gpdf$estimate, dat = gpdf$exceedances),
  c(0,0), tolerance = 1e-5))
```

```
## [1] TRUE
```

If the sample is small, maximum likelihood estimators are biased for the generalized Pareto distribution (the shape

parameter is negatively biased, regardless of the true value for  $\xi$ ). Bias correction methods includes the modified score of Firth, but the default method is the implicit correction (subtract), which solves the implicit equation

$$\tilde{\theta} = \hat{\theta} - b(\hat{\theta}). \quad (1.2)$$

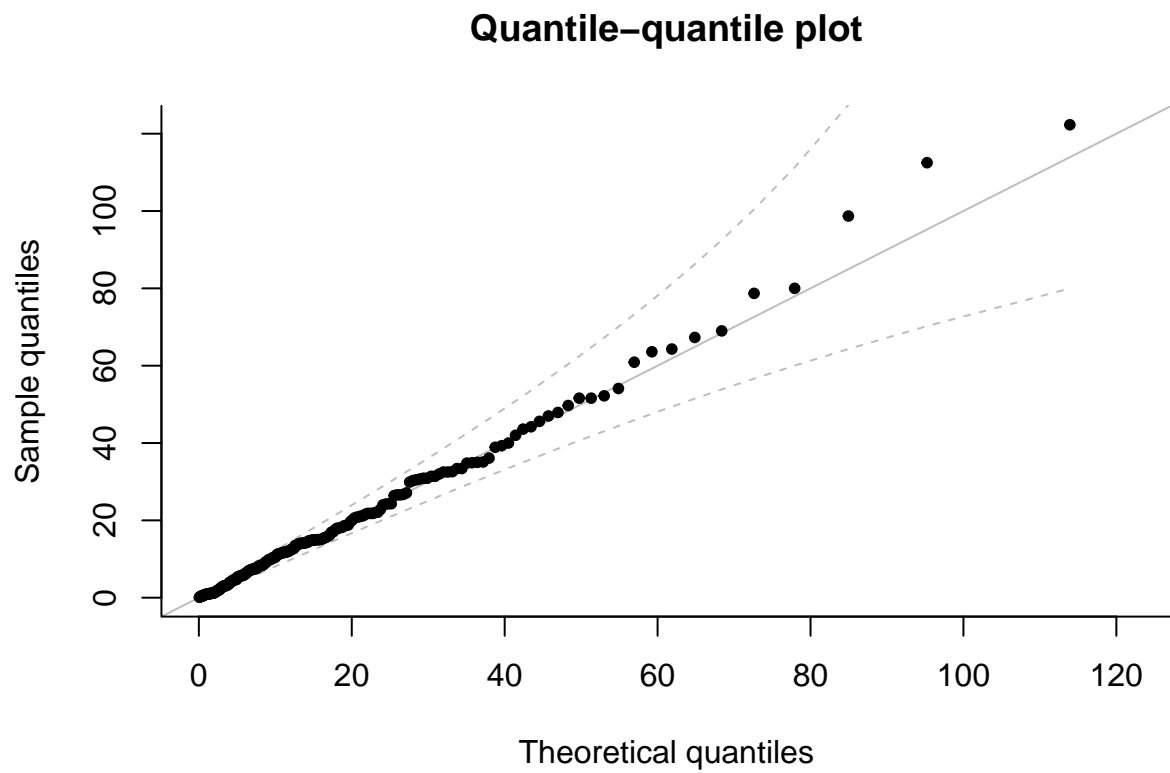
The point estimate  $\tilde{\theta}$  is obtained numerically as the root of this nonlinear system of equations. In the present case, the sample size is large and hence the first-order correction, derived through asymptotic arguments from the generalized Pareto distribution likelihood, is small. Note that the bias correction requires  $\xi > -1/3$ , since it is based on third-order cumulants of the distribution.

```
gpdbcor <- gpd.bcor(dat = gpdf$exceedances, par = gpdf$estimate)
#print the differences between MLE and bias-corrected estimates
gpdf$estimate - gpdbcor
```

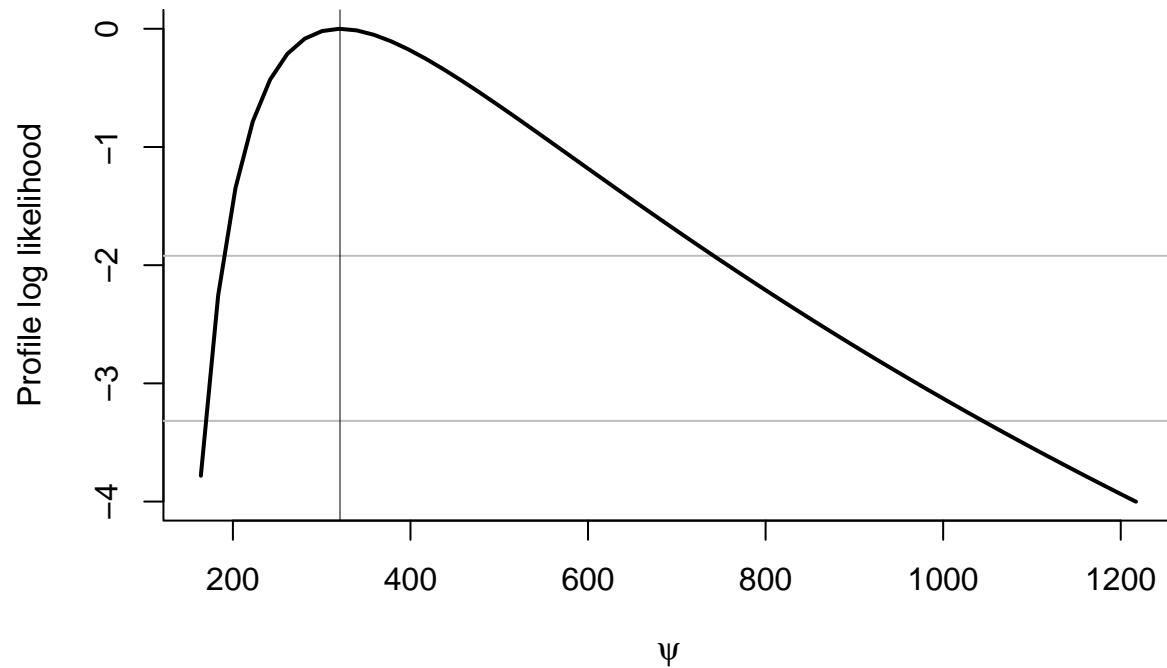
```
##      scale      shape
## 0.1915284 -0.0118876
```

The package includes some default diagnostic plots (probability-probability plots and quantile-quantile plots), which include approximate confidence intervals based on order statistics. We can also get profile likelihood and profile-based confidence intervals for most quantities of interest (parameters of the generalized Pareto distribution, excepted shortfall, return levels,  $N$ -observation maxima mean and quantiles). The example below gives the estimated profile for the median of the centennial maximum distribution conditional on exceeding 15, along with 95% confidence intervals.

```
plot(gpdf, which = 2) #Q-Q plot
```



```
# Profile of median of maxima of 100 years
profile <- gpd.pll(param = "Nquant", dat = nzrain, threshold = 15,
  N = length(nzrain)/(1998-1961)*100, q = 0.5)
```



```
# 95% confidence intervals
conf <- confint(profile, print = TRUE)

## Point estimate for the parameter of interest psi:
## Maximum likelihood      : 320.622
##
## Confidence intervals, levels : 0.025 0.975
## Wald intervals           : 118.72 522.524
## Profile likelihood        : 177.214 741.869
```

### 1.3 Your turn

1. Simulate 200 observations from the  $r$ -largest likelihood using `rrlarg` with shape parameter  $\xi = -0.2$  and  $r = 5$ .
2. Test the hypothesis  $H_0 : \xi = \xi_0$  using a score test and derive a 90% confidence interval for  $\xi$ . You can obtain the maximum likelihood estimator by calling `fit.rlarg` and the score and information matrix are implemented under `rlarg.score` and `rlarg.informat`. Recall that the score statistic  $w_{\text{score}} \equiv U(\hat{\theta}_{\xi_0})^\top i^{-1}(\hat{\theta}_{\xi_0}) U(\hat{\theta}_{\xi_0}) \sim$

$$\chi_1^2.$$

## Chapter 2

# Threshold selection

The generalized Pareto (GP) distribution with scale  $\sigma \in \mathbb{R}_+$  and shape  $\xi \in \mathbb{R}$  is

$$G(x) = \begin{cases} 1 - \{1 + \xi (\frac{x}{\sigma})\}_+^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp(-x\sigma), & \xi = 0. \end{cases}$$

The range of the generalized Pareto distribution is  $[0, -\sigma/\xi]$  if  $\xi < 0$  and  $\mathbb{R}_+$  otherwise.

Why is the generalized Pareto distribution so central to peaks-over-threshold analysis? The conditional distribution of exceedances over a threshold  $u < x^*$  converges to a generalized Pareto distribution as  $u$  converges to the endpoint of the distribution,

$$\lim_{u \rightarrow x^*} \frac{1 - F(xa_u + u)}{1 - F(u)} = 1 - H(x),$$

where  $H(x)$  is the distribution function of  $\text{GP}(1, \xi)$ .

If  $X \sim \text{GP}(\sigma, \xi)$ , straightforward calculations show that  $X - u \mid X > u \sim \text{GP}(\sigma + \xi u, \xi)$  for any  $u \in \mathbb{R}$  such that  $\sigma + \xi u > 0$ , meaning that conditional exceedances above a threshold  $u$  also follow a generalized Pareto distribution. This property is termed threshold-stability.

The limiting distribution of threshold exceedances is generalized Pareto as  $u \rightarrow x^*$  but, in practice, we must choose a finite threshold in order to have enough exceedances to draw inference. Since the scaling constant  $a_u$  is unknown, we have  $X \mid X > u \stackrel{\sim}{\sim} \text{GP}(\sigma_u, \xi)$ . The term  $1 - F(u)$  in the denominator is the fraction of points exceeding the threshold, which has a binomial distribution.

Threshold selection is subtle and it is common to select a high percentile of the data, say the 95% value, as the threshold, even if this is asymptotically incorrect, as in this case  $k/n \not\rightarrow 0$  as  $n \rightarrow \infty$ . Most approaches for threshold selection rely on properties of the generalized Pareto distribution (moments, threshold-stability) to determine a region within which the asymptotic distribution fits the data well and the parameter estimates are stable.

Below, we focus on recent graphical selection tools for likelihood based inference; mixture models are reviewed in Scarrott & MacDonald (2012) and are available in the package `evmix`.

## 2.1 Threshold stability plots

Consider a sequence of ordered candidate thresholds  $u_1 < \dots < u_k$ ; one of the most widely used tools for threshold selection is the threshold-stability plots of Davison & Smith (1990). These show the point estimates of the shape  $\xi$  and the modified scale  $\sigma_{u_i} - \xi u_i$ , which should be constant for any threshold  $u_j > u_i$  assuming that the generalized Pareto above  $u_i$  holds exactly. In addition to the point estimates, the asymptotic pointwise 95% Wald confidence intervals are displayed; the standard errors are obtained from the observed information matrix. While these are displayed by many packages, notably `extRemes`, `ismev` and `evd`, `mev` allows you to use profile likelihood based confidence intervals, which typically offer better coverage and capture some of the asymmetry of the distribution. The `mev` package functions for producing threshold stability plots are `tstab.gpd` and `tstab.pp` for respectively the generalized Pareto and Poisson process likelihoods.

Parameter stability plots can be difficult to interpret because the confidence intervals are pointwise rather than simultaneous (each fit likewise uses an overlapping portion of the data). The plots also ignore changes in the estimated parameters due to the penultimate approximation.

## 2.2 Wadsworth's diagnostics: white noise process and simultaneous threshold stability plots

The problem with the threshold stability plots lies in the point-wise nature of the estimate. Assuming a superposition of  $k$  Poisson processes, Wadsworth (2016) derives the limiting distribution of the maximum likelihood estimators from the Poisson process for overlapping windows as the number of windows  $k \rightarrow \infty$  and  $n_k \rightarrow \infty$ . The joint asymptotic Gaussian distribution allows Wadsworth (2016) to propose two additional diagnostics: a white noise sequence of differences in estimates of the shape, standardized to have unit variance. The variables  $\xi_i^* = (\hat{\xi}_{u_{i+1}} - \hat{\xi}_{u_i}) / \{(I_{u_{i+1}}^{-1} - I_{u_i}^{-1})_{\xi, \xi}^{1/2}\}$ , where  $I_{u_i}$  is the Fisher information of the Poisson process likelihood for exceedances above  $u_i$ , should form a white-noise sequence of independent variables centered around the origin; systematic deviations are indicative of inadequacy. To formally test the hypothesis, a likelihood ratio test can be used assuming a simple alternative, namely a single change point at threshold  $u_j$ . The null hypothesis is  $H_0 : \xi_i^* \stackrel{\text{iid}}{\sim} \text{No}(0, 1)$  for  $i = 1, \dots, k-1$  against the alternative  $H_a : \xi_i^* \sim \text{No}(\beta, \sigma) (i = 1, \dots, j-1) \text{ and } \xi_i^* \sim \text{No}(0, 1)$  for  $j, \dots, k-1$ . This alternative is motivated by results on model misspecification (White, 1982), which suggest that the asymptotic distribution may still be Gaussian, but with a different mean and variance. This can be used to automate threshold selection, by picking the smallest threshold for which the  $P$ -value is above the level  $\alpha$ . The function `W.diag` returns diagnostics plots (for the likelihood ratio statistic path, the white noise process and threshold stability along with asymptotic simultaneous confidence intervals) for non-homogeneous Poisson process model and for the bivariate exponential and the over a sequence of thresholds, specified using `q1`, `q2` and `k`. The argument `M` is a tuning parameter that can be chosen in a way such that the parameters of the non-homogeneous Poisson process likelihood coincide with those of the generalized extreme value distribution for blocks of size  $m$ ; see Coles (2001) to this effect.

A main criticism of the proposals of Wadsworth (2016) is their lack of robustness. For the asymptotic result to be approximately valid, the number of thresholds must be large, which implicitly requires large samples for each superposed point process. Moreover, the estimated difference in Fisher information matrices often fails to be positive definite in practice. The procedure is highly sensitive to the choice of  $k$ . Changing the set of thresholds  $\mathbf{u}$  under consideration leads to potentially completely different parameter estimates being chosen by the automated procedure.



## 2.3 Changepoint tests based on penultimate approximation

Let  $F(x)$  denote a thrice-differentiable distribution function with endpoint  $x^*$  and density  $f(x)$ . Define the reciprocal hazard function  $r(x) = \{1 - F(x)\}/f(x)$ . The existence of the limit  $\xi_\infty = \lim_{n \rightarrow \infty} s'\{b_n\}$  is necessary and sufficient for convergence to an extreme value distribution and Smith (1987) shows that there exists  $y$  such that

$$\frac{1 - F\{u + xr(u)\}}{1 - F(u)} = \{1 + xr'(y)\}_+^{-1/r'(y)}, \quad u < y < u + xr(u),$$

unless  $r'(x)$  is constant. The penultimate shape parameter for the generalized Pareto distribution is  $r'(u)$ , but the true shape parameter lies between  $r'(u)$  and  $\xi_\infty$ .

When we fit the limiting parametric models to finite samples, maximum likelihood estimates of the shape parameter will be closer to their penultimate counterparts than to the limiting value and we can expect them to vary as we increase the threshold.

Northrop & Coleman (2014) adapt the idea of Wadsworth & Tawn (2012) and fit a generalized Pareto model with piecewise constant shape to  $k$  different thresholds; continuity constraints at the thresholds impose  $k - 1$  restrictions on scale parameters, so the model only has  $k + 1$  parameters. A score test can be used to test the hypothesis of equal shape and it only requires evaluation of the model under the null hypothesis that a generalized Pareto distribution is valid above all thresholds. A diagnostic plot is obtained by plotting  $P$ -values against threshold. One can then choose to take, e.g., (a) the lowest threshold at which the  $P$ -value is non-significant, or (b) the lowest threshold at which the  $P$ -values for all higher thresholds are non-significant: under the null hypothesis, there is an  $\alpha\%$  probability of rejection at any given threshold. The function `NC.diag` computes the  $P$ -value of the score test as a function of the threshold.

## 2.4 Extended generalized Pareto

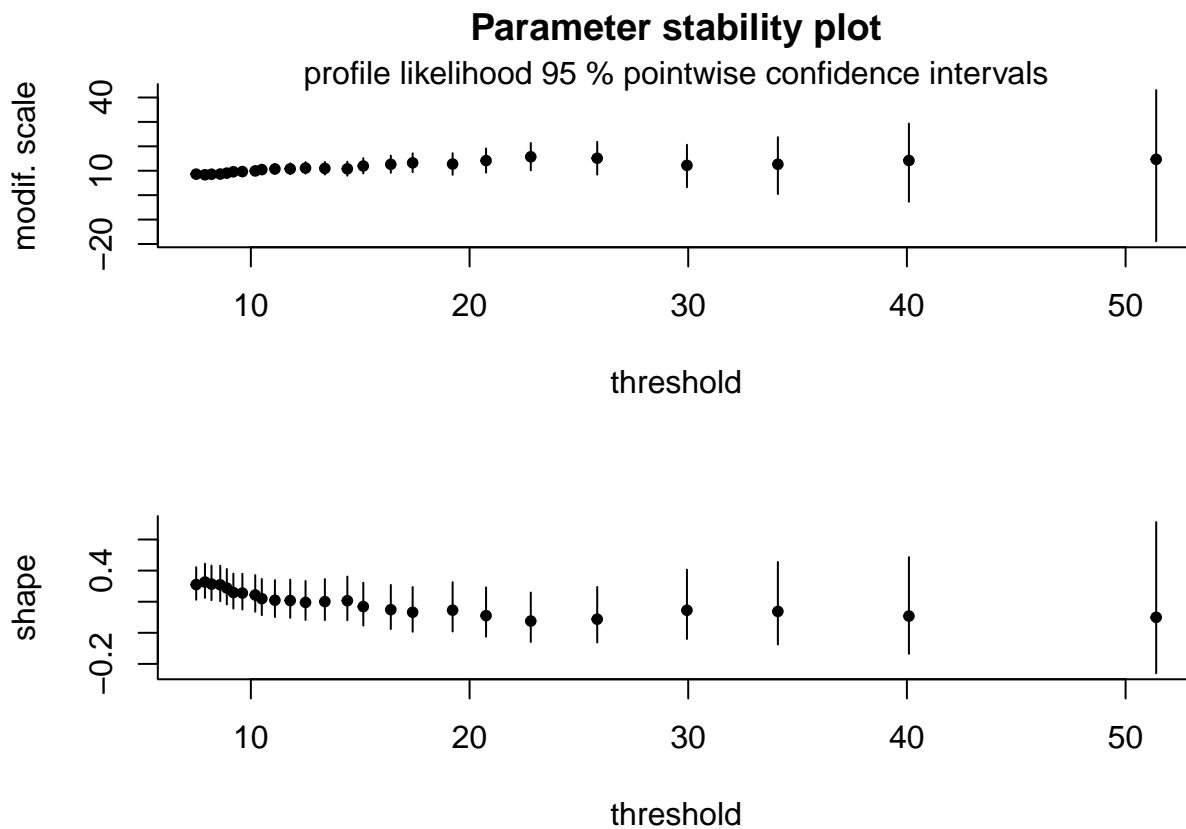
Papastathopoulos & Tawn (2013) propose three extended generalized Pareto distributions: for example, the third extended generalized Pareto model has distribution function  $\{1 - (1 + \xi x/\sigma)_+^{-1/\xi}\}^\kappa$  for  $x > 0$  and  $\kappa > 0$ . Each family reduces to the generalized Pareto when the additional parameter  $\kappa = 1$  and share the same tail index  $\xi$ , the extended generalized Pareto provide more flexibility for modelling departures from the limiting form. Standard parameter stability plots can be used to find a region in which  $\kappa \approx 1$  and the shape parameter stabilizes. The additional parameter gives flexibility for modelling departures from the limiting distribution and the hope is one can fit to exceedances over lower threshold and increase the number of points to which the distribution is fitted. The stability plots, obtained through `tstab.egp`, suffer from the same caveats as classical diagnostics.

## 2.5 Robust selection

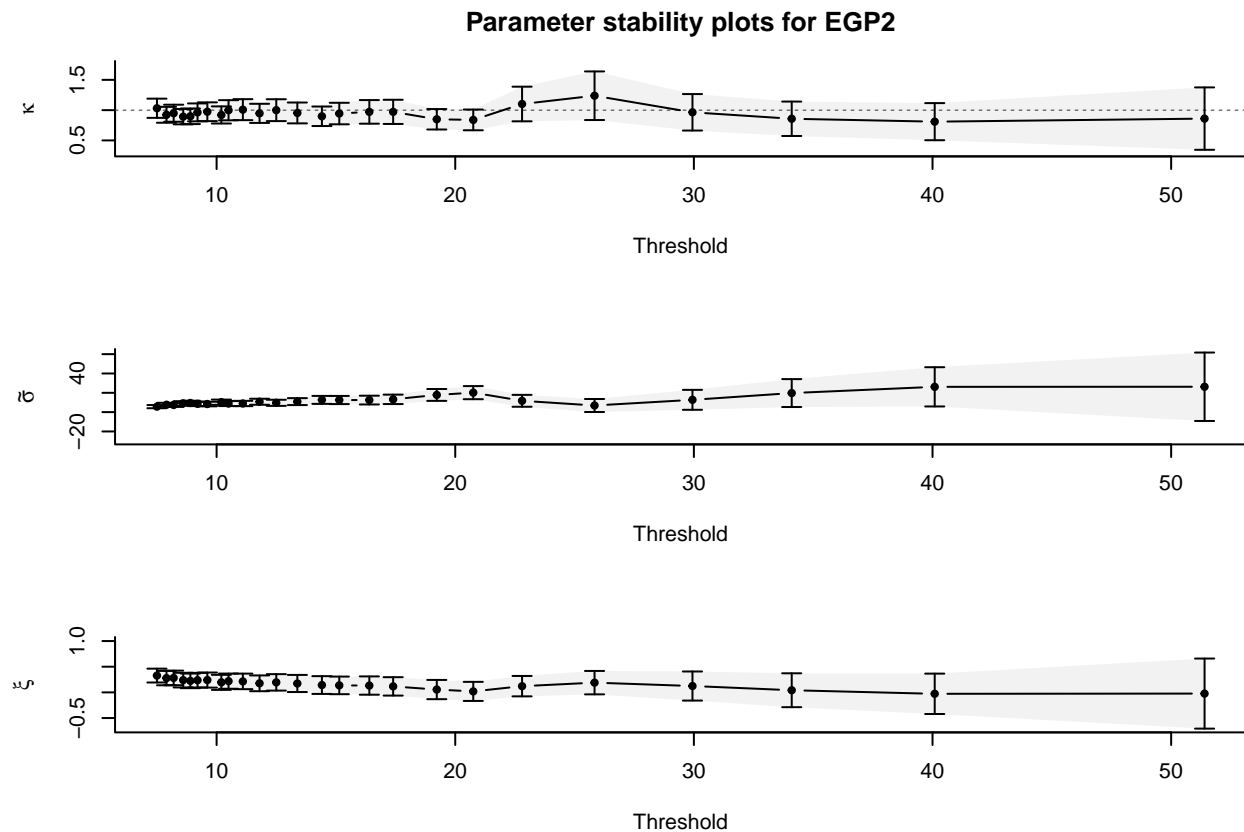
The extreme value distributions have unbounded influence functions and outliers can strongly affect the estimate of the shape. Dupuis (1999) proposes an optimal  $B$ -robust estimator of the generalized Pareto parameters. Points that are outlying or for which the fit is poor are downweighted; if the weights for the largest observations are very low, this suggests that the threshold is too low. While there is no guarantee that observations that were simulated from a generalized Pareto distributed would not be downweighted, systematic downweighting of the largest exceedances may be indicative of poor fit.

We are now ready to go back to the Maiquetia data and try to select a reasonable threshold. To this effect, we select a sequence of candidate thresholds and run the procedures described above.

```
# Keep non-zero rainfall, exclude 1999 observations
nzrain <- maiquetia[substr(day, 3, 4) < 99 & maiquetia > 0]
thcan <- quantile(nzrain, seq(0.8, 0.99, length.out = 25))
tstab.gpd(dat = nzrain, thresh = thcan, method = "profile")
```



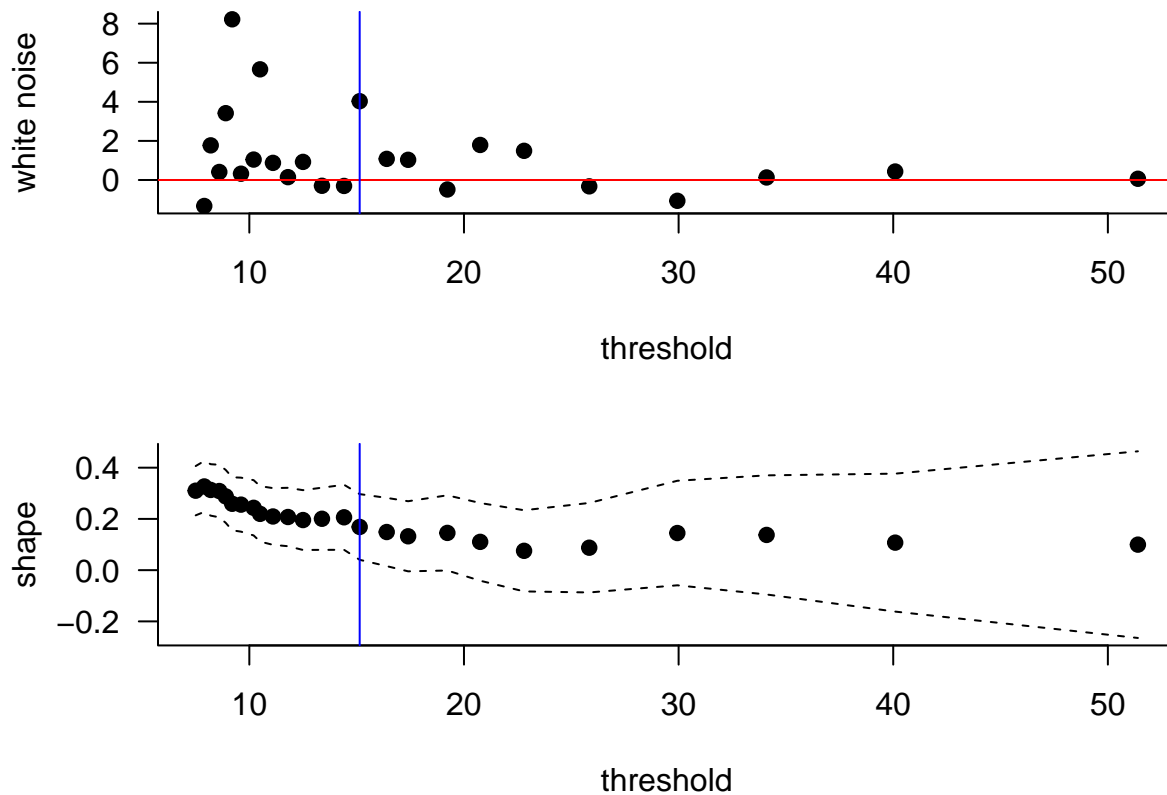
```
tstab.egp(xdat = nzrain, thresh = thcan, model = "egp2")
```



What do you notice on the parameter stability plots?

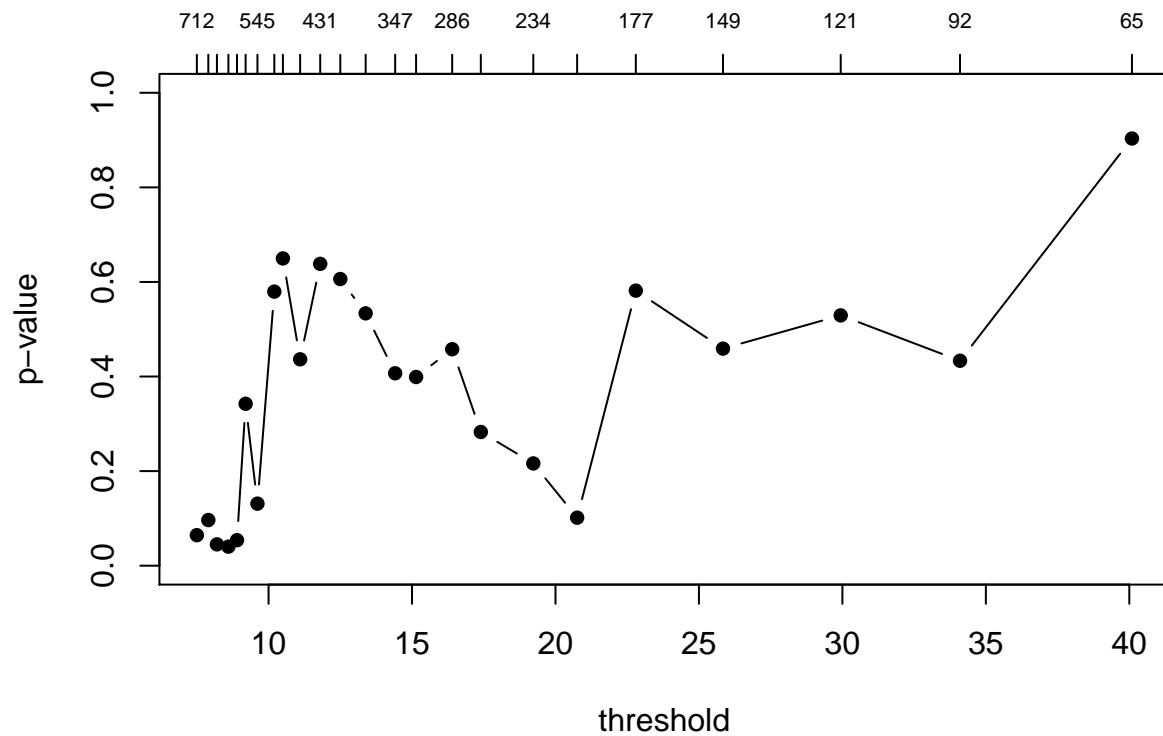
We can plot simultaneous parameter stability plots for the shape, as well as the white-noise sequence. The `W.diag` function, which implements the tests of Wadsworth (2016), returns the white noise process and calculates by simulation the null distribution, selecting the lowest threshold at which the hypothesis of changepoint is not rejected for all higher thresholds. For the Maiquetia rainfall, a threshold around 15.15mm is returned.

```
fWdiag <- W.diag(xdat = nzrain, model = "nhpp", u = thcan, plots = c("WN", "PS"))
```



Another option is the Northrop & Coleman (2014) penultimate test, which considers a sequence of threshold and tests whether a single shape fits the data against the alternative where the shape parameter differs at each candidate threshold following a step function. The `NC.diag` function plots the  $p$ -value path against the candidate thresholds; the top axis also displays the number of threshold exceedances. The extended generalized Pareto distribution can also be used to produce threshold stability plot with 95% confidence intervals. The parameter  $\kappa = 1$  if the model is generalized Pareto. Although there are three potential families, the family `egp2` exhibits the fastest convergence for

```
fNCdiag <- NC.diag(x = nzrain, u = thcan)
```



## 2.6 Your turn

1. Try using the threshold diagnostic tests on the nidd river flow data. Which threshold would you choose based on the conclusions of the test? Repeat the analysis on your favorite dataset.
2. Compare the inference obtained by using the different threshold (by comparing, for example, return levels or quantiles of a distribution of interest). Are your results sensitive to the choice of threshold?



## Chapter 3

# Penultimate approximation

### 3.1 Theory

The generalized extreme value (GEV) distribution arises as the non-degenerate limiting distribution of maxima. A key property of the generalized extreme value distribution is max-stability, meaning that the distribution of maxima of  $T \in \mathbb{N}$  is the same up to a location-scale transform: if  $F \sim \text{GEV}$ , then there exist location and scale constants  $b \in \mathbb{R}$  and  $a > 0$  such that  $F^T(ax + b) = F(x)$ . The parameters of the new distribution are easily derived: if  $X_i \stackrel{\text{iid}}{\sim} \text{GEV}(\mu, \sigma, \xi)$ , then  $\max\{X_1, \dots, X_T\} \sim \text{GEV}(\mu_T, \sigma_T, \xi)$  with  $\mu_T = \mu - \sigma(1 - T^\xi)/\xi$  and  $\sigma_T = \sigma T^\xi$  for  $\xi \neq 0$ , or  $\mu_T = \mu + \sigma \log(T)$  and  $\sigma_T = \sigma$  if  $\xi = 0$ .

In practice, the GEV is fitted to data that are partitioned into  $k$  blocks of sizes  $m$ , assumed equal for simplicity, so that  $n = km$ . The quality of the generalized extreme value approximation increases with the block size  $m$ . For fixed block sizes, the estimated shape parameter generally differs from its asymptotic counterpart and this discrepancy usually introduces bias if one extrapolates far beyond the range of the observed data.

Let  $F(x)$  denote a thrice-differentiable distribution function with endpoint  $x^*$  and density  $f(x)$  and define  $s(x) = -F(x) \log\{F(x)\} / f(x)$ . The existence of the limit  $\xi_\infty = \lim_{n \rightarrow \infty} s'(b_n)$  is necessary and sufficient for convergence to an extreme value distribution  $G$ ,

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \exp \left\{ -(1 + \xi_\infty x)^{-1/\xi_\infty} \right\} = G(x).$$

Smith (1987) shows that, for any  $x \in \{y : 1 + \xi_\infty y > 0\}$ , there exists  $z$  such that

$$\frac{-\log[F\{v + xs(v)\}]}{-\log\{F(v)\}} = \{1 + xs'(z)\}^{-1/s'(z)}, \quad v < z < v + xs(v).$$

For each  $n \geq 1$ , setting  $v = b_n$  and  $a_n = s(b_n)$  yields

$$F^n(a_n x + b_n) = \exp \left[ -\{1 + s'(z)x\}^{-1/s'(z)} \right] + O(n^{-1})$$

for  $z \in [\min(a_n x + b_n, b_n), \max(a_n x + b_n, b_n)]$ , depending on the support of  $F$ . The ultimate approximation replaces  $s'(z)$  by  $s'(x^*) = \xi_\infty$ , whereas Smith (1987) suggests instead taking  $s'(b_n)$ , which is closer to  $s'(a_n x + b_n)$  than is  $s'(x^*)$ .

The maximum likelihood estimate of the shape should change as the threshold or the block size increases. Consider for simplicity yearly maxima arising from blocking  $m = n_y$  observations per year. We are interested in the distribution of the maximum of  $N$  years and thus the target has roughly shape  $\xi_{Nn_y}$ , but our estimates will instead give approximately  $\xi_{n_y}$ ; an extrapolation error arises from this mismatch between the shape values, which would be constant if the observations were truly max-stable. The curvature of  $s'$  determines how stable the estimates of  $\xi_t$  are when the extrapolation window increases.

We illustrate the previous discussion by looking at block maximum from a log-normal distribution.

```
library(mev)
set.seed(123)
x <- seq(1, 30, length=200)
fitted <- t(replicate(n = 1000,
                     fit.gev(apply(matrix(rlnorm(30*100), ncol= 30), 1, max),
                                   method = "BFGS")$est))
penult.bm.lnorm <- mev::smith.penult(model = "bm", m = (m <- 30), family = "lnorm")

par(mfrow = c(1,2), mar = c(5,5,1,1))
hist(fitted[,3], probability = TRUE, breaks = 20,
     xlab = "Estimated shape", main = "")
segments(y0 = -0.5, y1 = 0, x0 = penult.bm.lnorm$shape, col = "red", lwd = 2)
segments(y0 = -0.5, y1 = 0, x0 = 0, col = "black", lwd = 2)

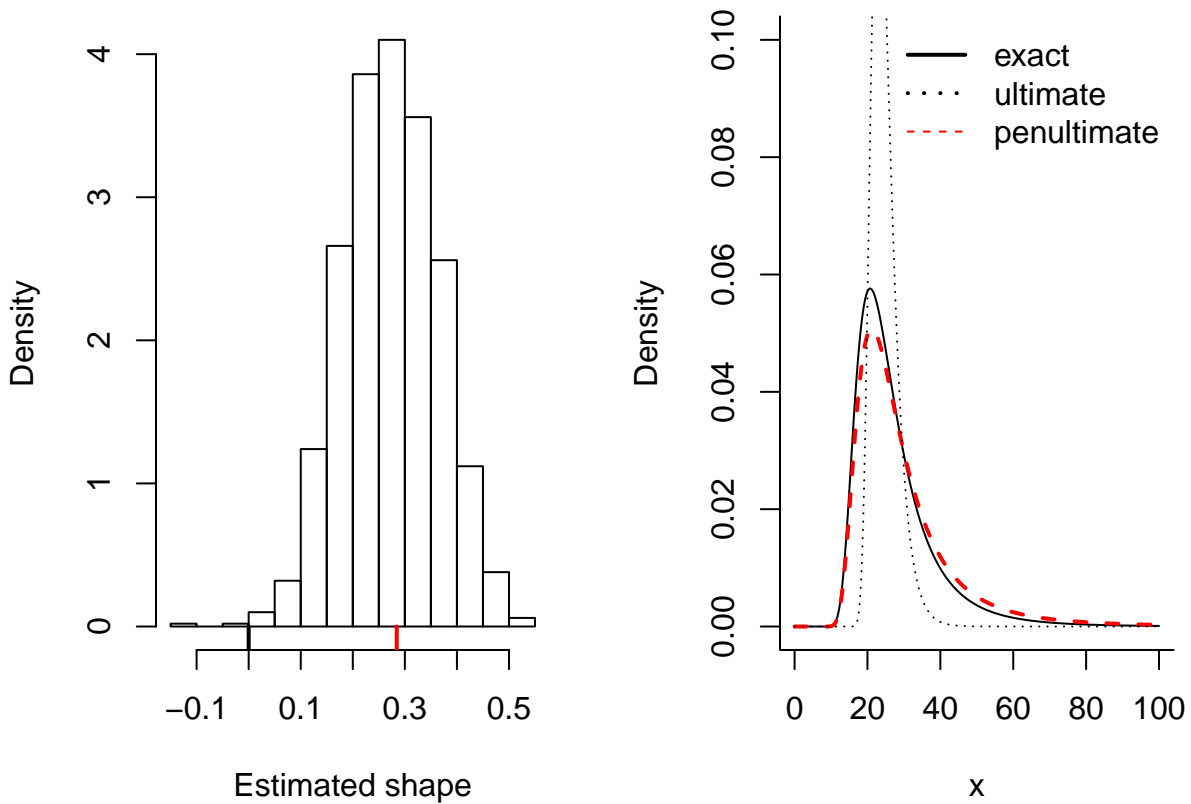
p30 <- penult.bm.lnorm
x = seq(0,100, length = 400)
N <- 1000; N30 = N/30
plot(x, N*exp((N-1)*plnorm(x, log.p=TRUE))*
     dlnorm(x),
     type="l", bty = "l", ylim = c(0,0.1),
     xlab="x", ylab="Density")
# Get parameters of maximum of N GEV
maxstabp <- function(loc, scale, shape, N){
  if(!isTRUE(all.equal(shape, 0, check.attributes = FALSE))){
    mut <- loc - scale*(1-N^shape)/shape
    sigmat = scale*N^shape
    return(c(loc = mut, scale = sigmat, shape = shape))
  } else{
    mut <- loc + scale*log(N)
    return(c(loc = mut, scale = scale, shape = shape))
  }
}
p30e <- maxstabp(loc = p30$loc, scale = p30$scale, shape = p30$shape, N = N30)
lines(x, evd::dgev(x, loc = p30e['loc'],
                  scale = p30e['scale'],
                  shape = p30e['shape']),
      col = "red", lwd = 2, lty = 2)
```



```

p301 <- maxstabp(loc = p30$loc, scale = p30$scale, shape = 0, N = N30)
lines(x, evd::dgev(x,
  loc = p30e['loc'],
  scale = p301['scale'],
  shape = 0), col = 1, lty=3, lwd=1)
legend(x = "topright",
  legend = c("exact", "ultimate", "penultimate"),
  col = c(1,1,2),
  lwd = c(2,2,1), bty = "n",
  lty = c(1,3,2))

```



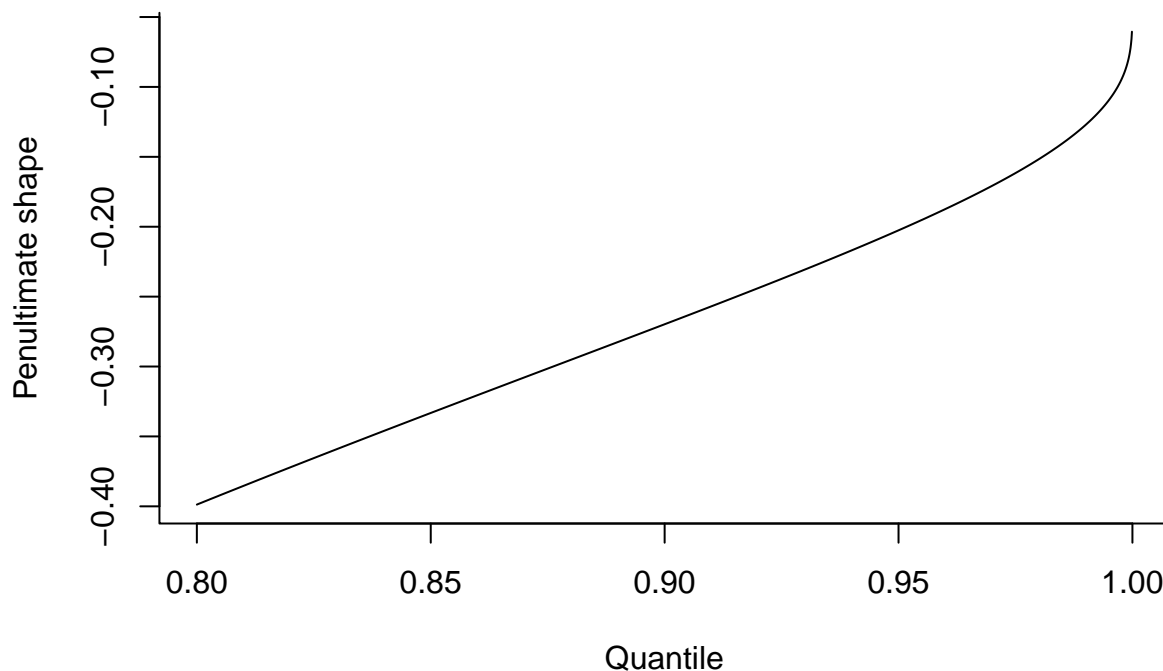
The left panel illustrates how maximum likelihood estimates of the parameters for repeated samples from a log-normal distribution are closer to the penultimate approximation than to the limiting tail index  $\xi_\infty = 0$ . Using max-stability, we could get an estimate of the distribution of the maximum of  $Nn_y$  observations. In this case, the penultimate shape for  $m = 1000$  is approximately  $\xi_{1000} = 0.215$ , compared to  $\xi_{30} \approx 0.284$ , which is far from the limiting value  $\xi_\infty = 0$ . The extrapolated density estimate is based on  $F_1^{1000/30}$ , the generalized extreme value density associated to the distribution function of the penultimate approximation  $F_1^{1000/30}$ , with  $F_1 \sim \text{GEV}(a_{30}, b_{30}, \xi_{30})$ . The

penultimate approximation  $F_2$  is more accurate than the ultimate approximation  $\text{GEV}(a_{30}, b_{30}, \xi_\infty)$ , which is too short tailed.

The function `smith.penult` computes the penultimate approximation for parameter models in **R** for the distribution assuming a model family with arguments `dfamily`, `qfamily` and `pfamily` exist. It returns the penultimate parameters for a given block size or quantile in case of threshold models. These functions are particularly useful for simulation studies in which one wants to obtain an approximation for  $F^N$  for large  $N$ , to get an approximation to the asymptotic distribution of a test statistic for finite  $N$ ; the GEV penultimate approximation is more numerically stable for certain models and its moments and cumulants are easily derived.

The following plot illustrates the penultimate shape parameter for the normal distribution based on threshold exceedances. In finite samples, the upper tail is bounded, but the quality of the approximation by the Gumbel distribution increases as  $u \rightarrow \infty$ , albeit very slowly.

```
penult <- smith.penult(family = "norm", method = "pot",
                      u = qnorm(u <- seq(0.8, 0.9999, by = 0.0001)))
plot(u, penult$shape, bty = "l", type='l', xlab='Quantile', ylab='Penultimate shape')
```



While the above discussion has been restricted to block maxima and the generalized extreme value distribution,

similar results exists for peaks over threshold using a generalized Pareto distribution. For threshold exceedances, we start with the reciprocal hazard  $r(x) = \{1 - F(x)\}/f(x)$  in place of  $s(x)$ . The normalizing constants are  $(u, r(u))$  and the penultimate shape parameter for the generalized Pareto distribution is  $r'(u)$ .

We can relate threshold exceedances to the distribution of maxima as follows: suppose a generalized Pareto distribution  $F$  is fitted to threshold exceedances above a threshold  $u$ , with parameters  $(\zeta_u, \sigma_u, \xi)$ , where  $\zeta_u$  denotes the unknown proportion of points above the threshold  $u$ . If there are  $n_y$  observations per year on average, then the  $T$ -year return level is  $z_{1/T} = u + \sigma_u/\xi\{(Tn_y\zeta_u)^\xi - 1\}$ . We take  $F^{\zeta_u T n_y}$  as the approximation to the  $T$ -year maxima distribution, conditional on exceeding  $u$ .

## 3.2 Your turn

1. Sample observations from the standard Gaussian distribution  $\text{No}(0, 1)$  for  $m = 25, 50, 100, 1000$  and  $k = 1000$  replications.
2. Fit a generalized extreme value distribution to block maxima of size  $m$ . Obtain an estimate of the distribution of  $m = 1000000$  observations from a standard Gaussian using max-stability (the functions `fit.gev` and `gev.mle` can be useful).
3. Plot the density of the resulting approximations and compare them to the density of the true distribution  $N\phi(x)\Phi^{N-1}(x)$ .
4. Superpose the density of the penultimate approximation for  $m = 1000000$  obtained using the Smith penultimate approximation (`smith.penult`).



## Chapter 4

# Simulation

### 4.1 Max-stable processes

Let  $\mathcal{S}$  be a compact subset of  $\mathbb{R}^d$ . The de Haan spectral representation (de Haan, 1984) of simple max-stable processes is (Schlather & Tawn, 2003)

$$Z(\mathbf{s}) = \max_{i \in \mathbb{N}} \zeta_i W_i(\mathbf{s}), \quad \mathbf{s} \in \mathcal{S},$$

where  $\mathbf{W}$  is a stochastic process satisfying  $E\{W(\mathbf{s})_+\} = 1$  for any  $\mathbf{s} \in \mathcal{S}$  and  $\{\zeta_i\}_{i \in \mathbb{N}}$  is a Poisson point process with intensity function  $\zeta^{-2} d\zeta$ . Max-stable processes arise as the pointwise maximum of an infinite collection of random functions  $\varphi_i = \zeta_i W_i$ .

The distribution function of a  $D$ -dimensional max-stable vector can be defined in terms of a measure  $\Lambda$ , termed exponent measure, defined on  $\mathbb{R}^D \setminus \{\mathbf{0}_D\}$  viz.

$$\Pr(\mathbf{Z} \leq \mathbf{z}) = \exp\{\Lambda([-\infty, \mathbf{z}]^c)\}, \quad \mathbf{z} > \mathbf{0}_D.$$

It follows from max-stability that the exponent measure is homogeneous of order  $-1$ .

### 4.2 $\mathcal{R}$ -Pareto processes and generalizations

Let  $\mathcal{F}^+$  denote the set of continuous non-negative non-null functions. A risk functional  $\mathcal{R} : \mathcal{F}^+ \rightarrow [0, \infty)$  is a continuous functional which is homogeneous of order 1, i.e.,  $\mathcal{R}(tf) = t\mathcal{R}(f)$  for  $f \in \mathcal{F}^+$  and  $t > 0$ . Examples include maxima,  $\mathcal{R}(f) = \max_{j=1}^D f(\mathbf{s}_j)$ , minima,  $\mathcal{R}(f) = \min_{j=1}^D f(\mathbf{s}_j)$ , sums  $\mathcal{R}(f) = \sum_{j=1}^D f(\mathbf{s}_j)$  and projections  $\mathcal{R}(f) = X(\mathbf{s}_j)$ . A  $\mathcal{R}$ -Pareto process admits the stochastic representation (Ferreira & de Haan, 2014)(Dombry & Ribatet, 2015)

$$X = P \frac{S}{\mathcal{R}(S)}, \tag{4.1}$$

for  $P$  a unit Pareto random variable with survival function  $\Pr(P > x) = x^{-\alpha}$  for  $x \geq 1$  and a stochastic process  $S$ , independent of  $P$ , with sample paths in  $\mathcal{S}_{\text{ang}} = \{f \in \mathcal{F}^+ \setminus \mathcal{C}_{\mathcal{R}} : \|f\|_{\text{ang}} = 1\}$  and where  $\mathcal{C}_{\mathcal{R}} = \{g \in \mathcal{F}^+ : \mathcal{R}(g) = 0\}$  and  $\|\cdot\|_{\text{ang}}$  is a norm.  $\mathcal{R}$ -Pareto processes come about when looking at the limit distribution of  $\mathcal{R}$ -exceedances,

i.e.  $\{f : \mathcal{R}(f) > u\}$  for a large threshold  $u$ , assuming unit Frechet margins. This can be achieved in practice by standardizing the margins, but with the drawback that the exceedances are defined on the transformed scale.

de Fondeville (2018) lifts this assumption and define generalized  $\mathcal{R}$ -Pareto processes, allowing for different location and scale marginal parameters (but with a common shape parameter). The risk region in which extreme observations lie is

$$\mathcal{A}_u = \left\{ f \in \mathcal{F}^+ \setminus \mathcal{C}_{\mathcal{R}} : \mathcal{R} \left( \tau \frac{f^\xi - 1}{\xi} + \eta \right) \geq u \right\},$$

where  $\eta$  is a location function and  $\tau$  is a scale function. The probability measure of  $Z^*$  over  $\mathcal{A}_u$  is  $\Lambda(\cdot)/\Lambda\{\mathcal{A}_u\}$ . The stochastic representation of the generalized  $\mathcal{R}$ -Pareto vector is

$$Z(\mathbf{s}) = \tau(\mathbf{s}) \frac{X(\mathbf{s})^\xi - 1}{\xi} + \eta(\mathbf{s}), \quad (4.2)$$

where  $X(\mathbf{s})$  is a  $\mathcal{R}$ -Pareto process. When  $\mathcal{R}$  is max, the finite-dimensional distributions are multivariate generalized Pareto (Rootzén et al., 2018).

### 4.3 Algorithms

The original goal of the `mev` package was to implement the algorithm of Dombry et al. (2016), in order to perform exact simulation from max-stable vectors by sampling each extremal function in turn. Since then, other algorithms for simulating from standard  $\mathcal{R}$ -Pareto processes and generalized  $\mathcal{R}$ -Pareto processes have been added to the package. These are based on accept-reject method or composition sampling and can be slow in high dimensions. Note that packages such as `RandomFields` and `SpatialExtremes` are tailored for spatial processes and include more efficient algorithms to sample the (log-)Gaussian components appearing in the Brown–Resnick and extremal Student models, which makes them typically faster. The vignette provides details about the parametrization of the models implemented in `mev`.

The various functions of interest for simulation include

- `rmev`, which is the workhorse for simulating max-stable random vectors;
- `rmevspec` returns samples from the angular distribution defined with respect to the  $l_1$  norm;
- `rparp` for simulating  $\mathcal{R}$ -Pareto processes through accept-reject;
- `rpapcs` for simulating  $\mathcal{R}$ -Pareto processes using composition sampling, e.g. Ho & Dombry (2019)
- `rgparp` for simulating generalized  $\mathcal{R}$ -Pareto processes using accept-reject.

This approach is only available for Brown–Resnick and extremal Student processes and is more computationally intensive since it requires computing numerically mixture weights. However, these need only be computed once and this approach is more efficient for simulating large samples, particularly in cases where the acceptance rate is very low (notably if  $R(\mathbf{X}) = \min_{j=1}^D X_j$ ).

Let's start by simulating from a simple max-stable random vector from the negative logistic model. The margins are unit Frechet, and we can verify this by pooling the observations and fitting a generalized extreme value distribution to the sample. The parameter estimates are correct, and a probability-probability plot shows that the .

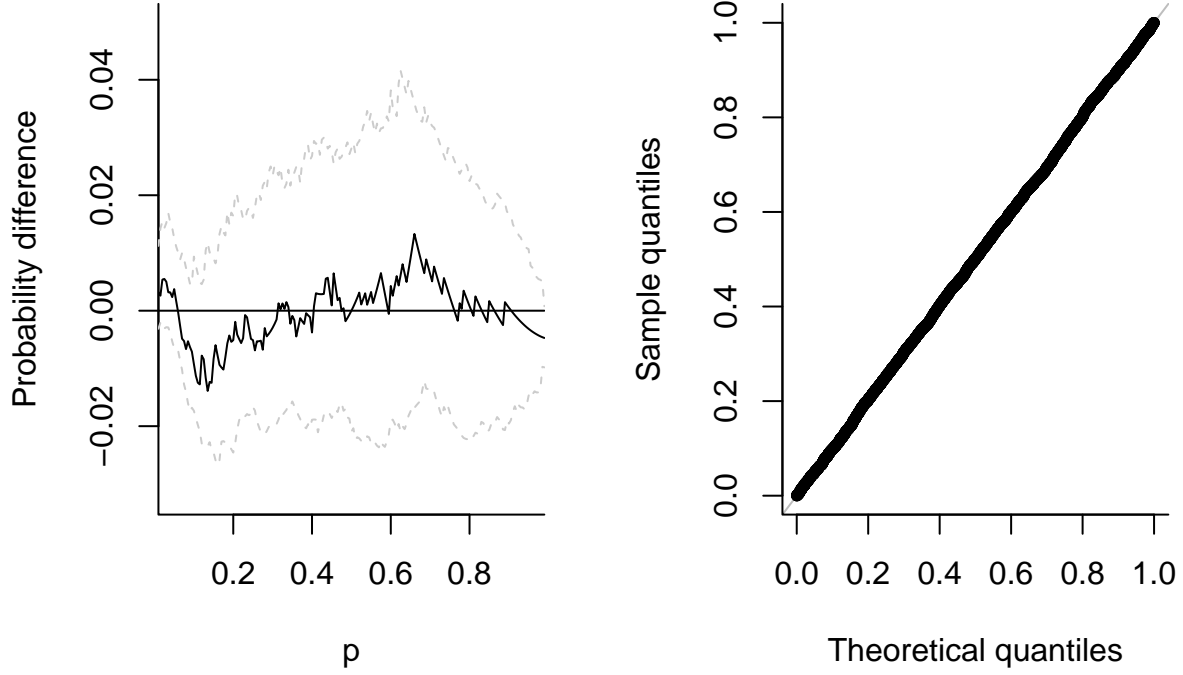
If  $\mathbf{Z}$  is a max-stable vector of size  $D$  with standard Gumbel margins, with distribution function  $G(\mathbf{z}) = \exp[-V\{\exp(\mathbf{z})\}]$ , then for any subset of size  $J \subseteq \{1, \dots, D\}$ , the variable  $Z_J = \max_{j \in J} Z(\mathbf{s}_j)$  has distribution

function  $H_J(z) = \exp\{-\exp(z - \mu_J)\}$ , where  $\mu_J = \log\{V_J(\mathbf{1})_{|J|}\}$  and  $0 \leq \mu_J \leq \log(|J|)$ . This follows from the homogeneity of the exponent measure. Gabda et al. (2012) propose a probability-probability plot based on fitting  $\mu_J$  through maximum likelihood with the parameter constraints for each set of  $|J|$  stations. The data are obtained by pooling replications of the max-stable field and selecting all subsets of size  $|J|$  if  $\binom{D}{|J|}$  is small, or else a limited number of stations among those that display the higher dependence so as to maximize the power of the test (independence being a special case of max-stability). Uncertainty quantification is performed using a nonparametric bootstrap. The function `maxstabtest` implements this for the special case  $J = D$ .

```
library(mev)
set.seed(0)
samp <- rmev(n = 1000, d = 5, param = 0.1, model = "neglog")
fgev <- fit.gev(c(samp), show = FALSE)
fgev$estimate
```

```
##          loc      scale      shape
## 0.9954939 0.9919429 0.9867670
```

```
par(mfrow = c(1, 2))
# Test of max-stability
maxstabtest(dat = samp)
# Probability-probability plot
plot(fgev, which = 1, main = "")
```



The likelihood of max-stable processes is untractable, so inference is typically performed using composite likelihood. The package `evd` includes routines for doing this in the bivariate setting (see `fbvpot`).

Consider a collection of independent and identically distributed vectors  $\mathbf{X}$  with continuous marginal distributions that are in the max-domain of attraction of a max-stable distribution with limit measure  $\Lambda$ . One can transform the observations  $\mathbf{X}_i$  into pseudo-uniform samples using the empirical distribution function or else the semi-parametric estimator,

$$\check{F}_j(x) = \begin{cases} I\{i : y_{ij} \leq x\}/(n+1), & x \leq u_j, \\ 1 - \{1 - \check{F}_j(u_j)\} \left\{1 + \hat{\xi} \left( \frac{x - u_j}{\hat{\tau}_j} \right)\right\}_+^{-1/\hat{\xi}_j}, & x > u_j. \end{cases} \quad (4.3)$$

(see `spunif`). We can back-transform the pseudo-uniform samples  $u_{ij} = \check{F}_j(y_{ij})$  to unit Pareto using the quantile transform,  $z_{ij}^p = 1/(1 - u_{ij})$ . If we map the random vector  $\mathbf{Z}^p$  to pseudo-polar coordinates,  $\mathbf{Z}^p \mapsto (R, \boldsymbol{\Omega})$  from  $\mathbb{R}_+^D \rightarrow (0, \infty) \times \mathbb{S}_1$ , where  $\mathbb{S}_1 = \{\boldsymbol{\omega} \in \mathbb{R}_+^D : \|\boldsymbol{\omega}\|_1 = 1\}$  is the  $\ell_1$ -simplex. The measure  $\Lambda$  factorizes as a product measure  $\Lambda(d\mathbf{z}) = D\zeta^{-2}d\zeta\rho_1(d\boldsymbol{\omega})$  with angular measure  $\rho_1$ , a probability measure satisfying the moment constraint  $D \int_{\mathbb{S}_1} \omega_j \rho_1(d\boldsymbol{\omega}) = 1$  for  $j = 1, \dots, D$ ; this moment constraint holds for any  $\rho_1$ , whereas it will be measure dependent if the radial measure is not  $\|\cdot\|_1$  (Einmahl & Segers, 2009). We can thus create an approximate sample  $\boldsymbol{\omega}$  from the



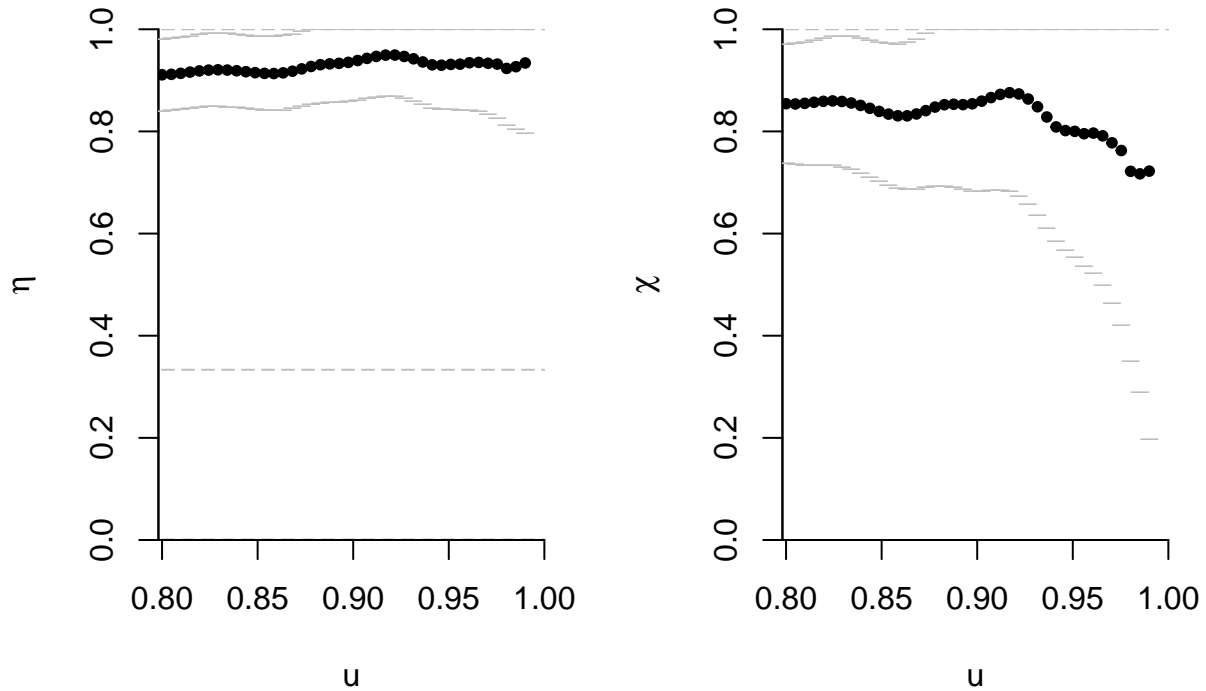
angular measure  $\rho_1$  and estimate its distribution non-parametrically as

$$\hat{H}(w) = \sum_{i=1}^N \hat{p}_i I\{\omega_i \leq w\},$$

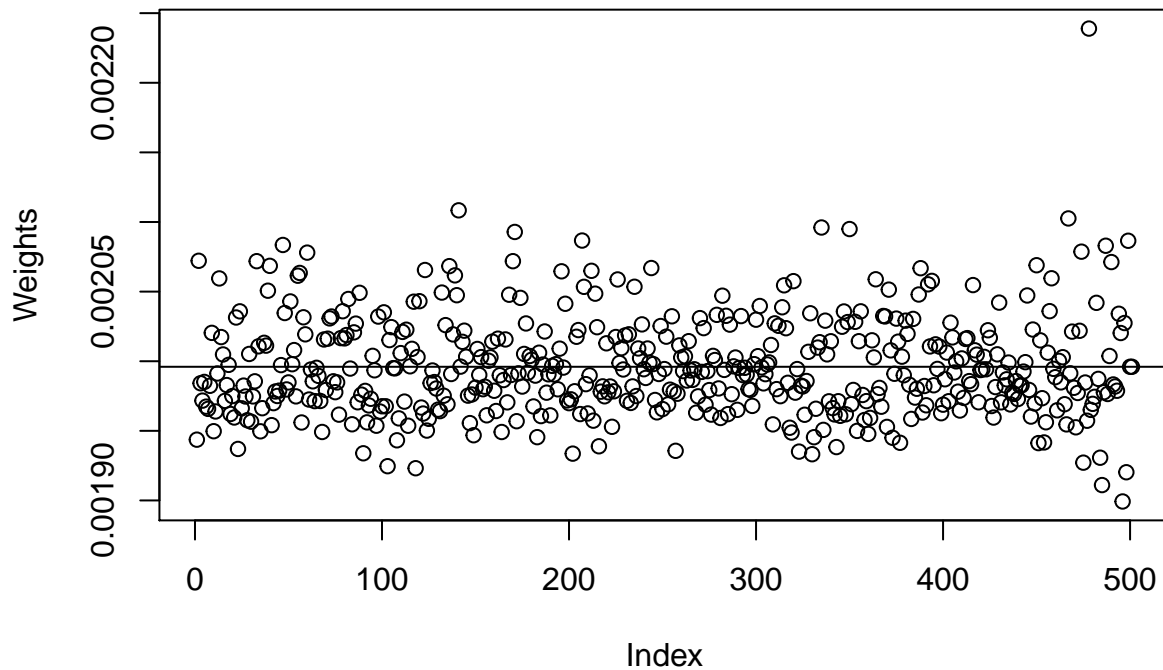
where the weights  $\hat{p}_i$  are obtained as the solution of either the empirical or Euclidean likelihood problems with a mean constraint (de Carvalho et al., 2013, Einmahl & Segers (2009)). Since the resulting spectral distribution is discrete (which is problematic in simulations), de Carvalho et al. (2013) suggest fitting a Dirichlet kernel to observations, with parameters  $\nu \mathbf{w}_i (i = 1, \dots, D)$  subject to the constraint  $\|\nu \mathbf{w}_i\|_1 = 1$ . The 'bandwidth' tuning parameter  $\nu$  is chosen via cross-validation. The function `angmeasdir` fits Euclidean likelihood and returns a list with the weights vector `wts`, the cross-validation parameter `nu` and a  $n \times D$  matrix of parameters for the Dirichlet distribution.

Alternatively, we can estimate the limiting spectral measure from the pseudo-observations enforcing the mean constraint using empirical likelihood, as proposed by Einmahl & Segers (2009). The function `angmeas` computes the weights associated to each observations.

```
samp <- rmev(n = 1000, d = 3, param = c(0.4, 0.6, 2.9, 0.1), model = "sdir")
taildep(samp, method = list(eta = "betacop", chi = "betacop"))
```



```
nparangmeas <- angmeas(samp, th = 0.5)
# Plot the probability weights and compute the column mean
plot(nparangmeas$wts, ylab = "Weights"); abline(h = 1/nrow(nparangmeas$ang))
```



```
colSums(nparangmeas$wts * nparangmeas$ang)
```

```
## [1] 0.3333333 0.3333333
```

```
dirangmeas <- angmeasdir(samp, th = 0.5)
```

## 4.4 Extremal coefficient

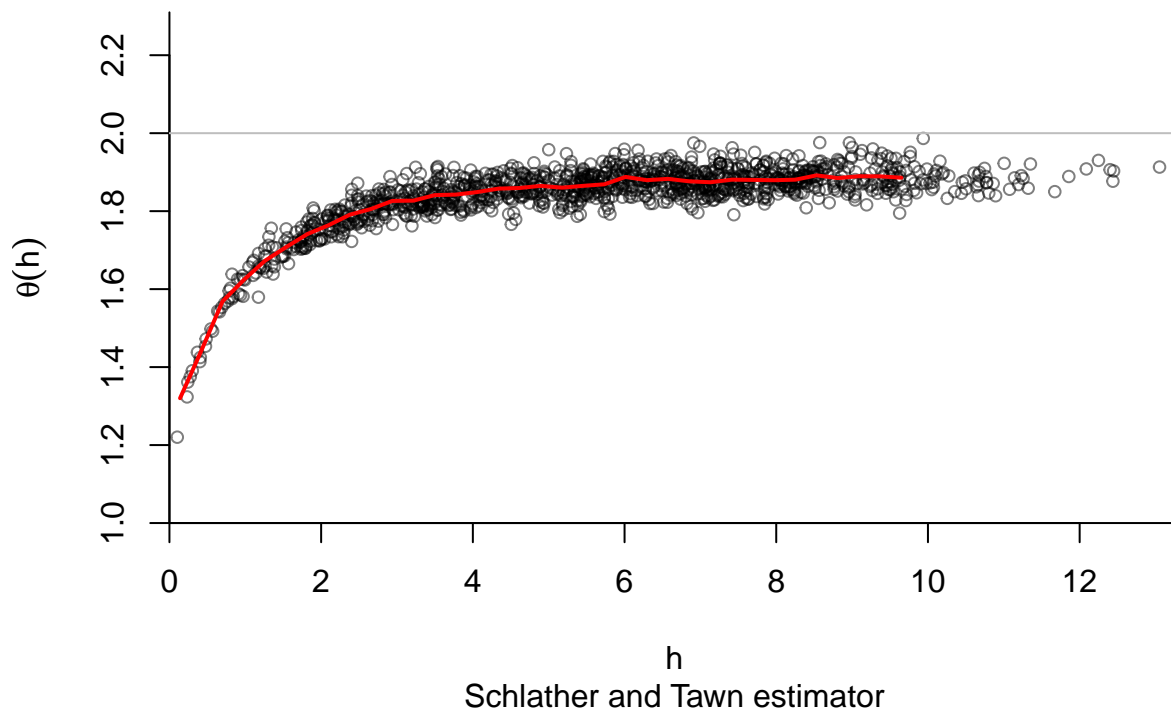
As in `SpatialExtremes`, we can use the nonparametric estimates of the extremal coefficient as goodness-of-fit diagnostic. The following code samples data

```

coord <- 10*cbind(runif(50), runif(50))
di <- as.matrix(dist(coord))
dat <- rmv(n = 1000, d = 100, param = 3, sigma = exp(-di/2), model = 'xstud')
res <- extcoef(dat = dat, coord = coord)

```

### Extremal coefficient



```

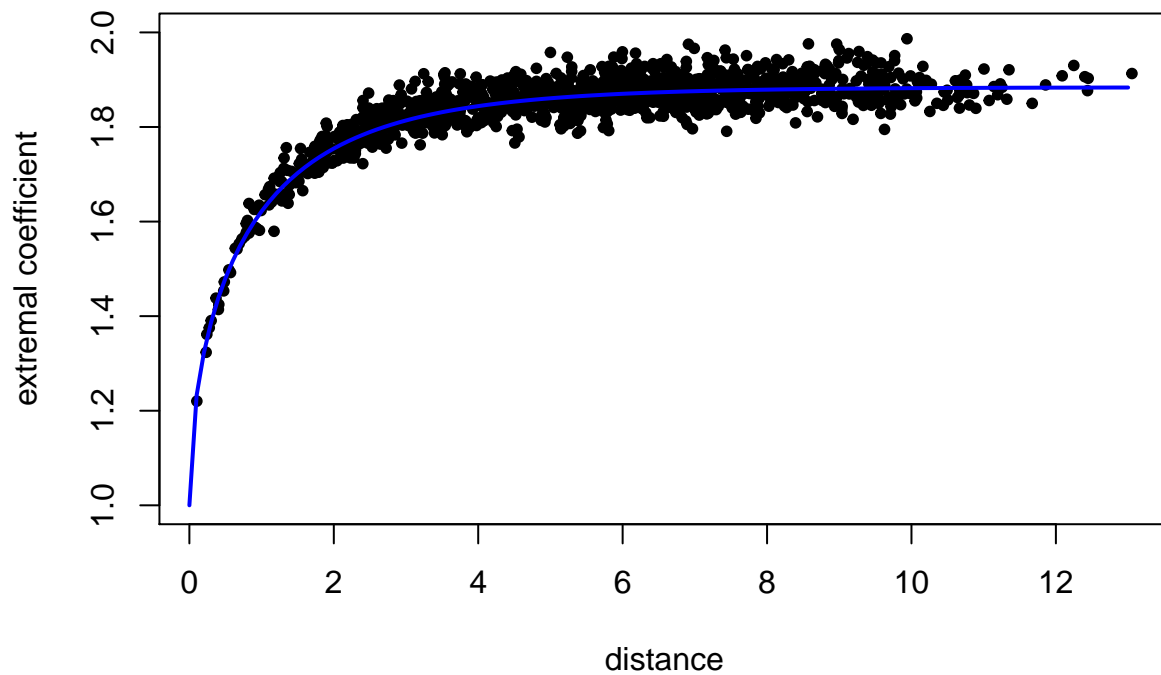
# Extremal Student extremal coefficient function
XT.extcoeffun <- function(h, nu, corrfun, ...){
  if(!is.function(corrfun)){
    stop('Invalid function `corrfun`.')
  }
  h <- unique(as.vector(h))
  rhoh <- sapply(h, corrfun, ...)
  cbind(h = h, extcoef = 2*pt(sqrt((nu+1)*(1-rhoh)/(1+rhoh)), nu+1))
}
#This time, only one graph with theoretical extremal coef
plot(res$dist, res$extcoef, ylim = c(1,2), pch = 20,

```

```

ylab = "extremal coefficient", xlab = "distance")
extcoefxt <- XT.extcoeffun(seq(0, max(res$dist), by = 0.1), nu = 3,
                          corrfun = function(x){exp(-x/2)})
lines(extcoefxt[, 'h'], extcoefxt[, 'extcoef'], type = 'l', col = 'blue', lwd = 2)

```



## 4.5 Generalized $\mathcal{R}$ -Pareto processes for functional threshold exceedances

We simulate a spatial generalized  $\mathcal{R}$ -Pareto process.

```

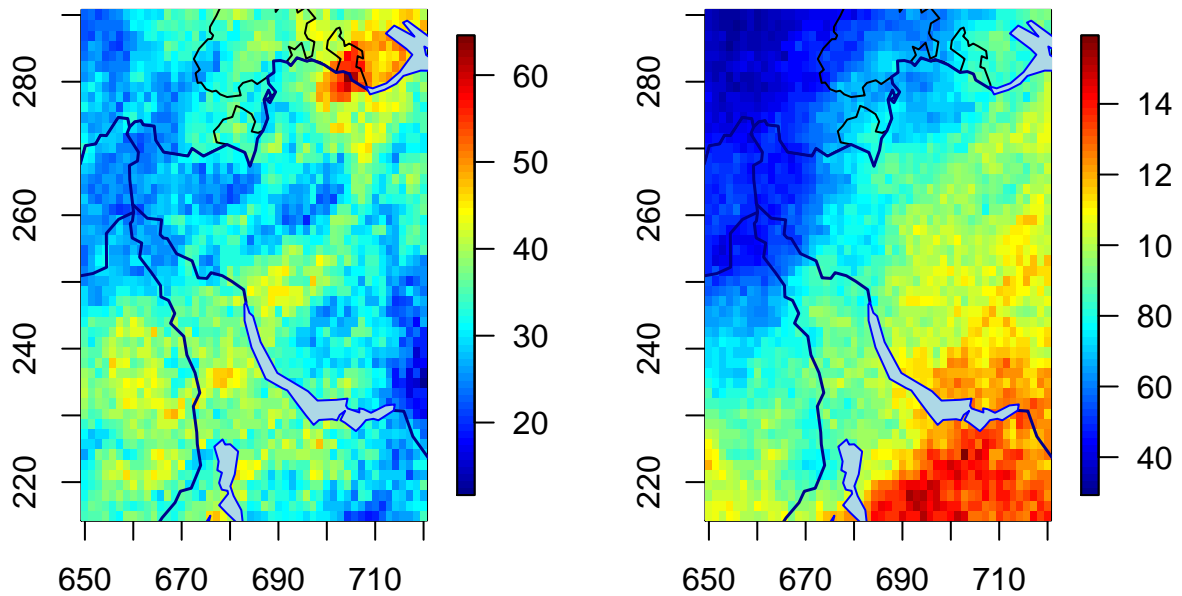
lon <- seq(650, 720, length = 50)
lat <- seq(215, 290, length = 50)
# Create a grid
grid <- expand.grid(lon, lat)
coord <- as.matrix(grid)
dianiso <- distg(coord, 1.5, 0.5)
sgrid <- scale(grid, scale = FALSE)

```

```

# Specify marginal parameters `loc` and `scale` over grid
eta <- 26 + 0.05*sgrid[,1] - 0.16*sgrid[,2]
tau <- 9 + 0.05*sgrid[,1] - 0.04*sgrid[,2]
# Parameter matrix of Huesler--Reiss
# associated to power variogram
Lambda <- ((dianiso/30)^0.7)/4
# Simulate generalized max-Pareto field above u=50
set.seed(345)
simu1 <- rgparp(n = 1, thresh = 50, shape = 0.1, riskf = "max",
               scale = tau, loc = eta, sigma = Lambda, model = "hr")
# The same, but conditional on an exceedance at a site
simu2 <- rgparp(n = 1, thresh = 50, shape = 0.1, riskf = "site",
               siteindex = 1225, scale = tau, loc = eta,
               sigma = Lambda, model = "hr")
#Plot the generalized max-Pareto field
par(mfrow = c(1,2))
fields::quilt.plot(grid[,1], grid[,2], simu1, nx = 50, ny = 50)
SpatialExtremes::swiss(add = TRUE)
fields::quilt.plot(grid[,1], grid[,2], simu2, nx = 50, ny = 50)
SpatialExtremes::swiss(add = TRUE)

```



```
# Value at conditioning coordinate should be greater than 50
simu2[1225]
```

```
## [1] 84.81661
```

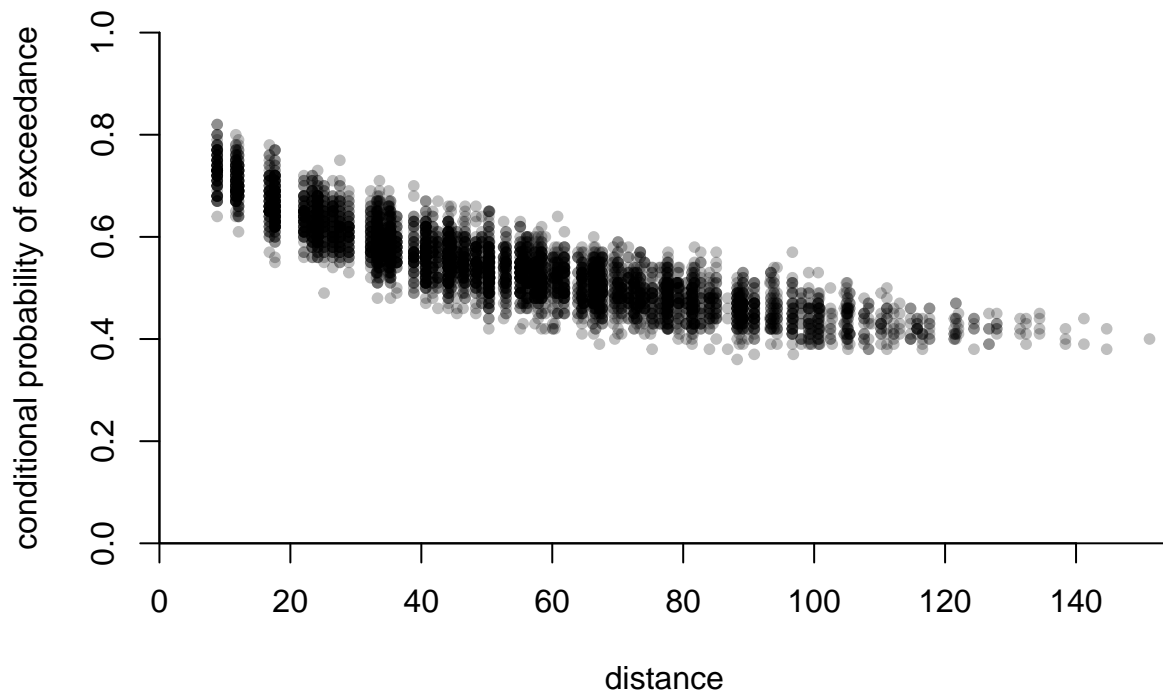
The code snippet below fits a Brown–Resnick model with power variogram to simulated data from the same model (based on more than one replicate). The parameters are estimated by minimizing the squared distance between empirical cloud of pairwise conditional probability of exceedance and the theoretical one. We include geometric anisotropy in the analysis.

```
lon <- seq(650, 720, length = 10)
lat <- seq(215, 290, length = 10)
# Create a grid
grid <- expand.grid(lon,lat)
coord <- as.matrix(grid)
dianiso <- distg(coord, 1.5, 0.5)
sgrid <- scale(grid, scale = FALSE)
# Specify marginal parameters `loc` and `scale` over grid
```

```

eta <- 26 + 0.05*sgrid[,1] - 0.16*sgrid[,2]
tau <- 9 + 0.05*sgrid[,1] - 0.04*sgrid[,2]
# Parameter matrix of Huesler--Reiss
# associated to power variogram
Lambda <- ((dianiso/30)^0.7)/4
# Simulate generalized max-Pareto field above u=50
set.seed(345)
simul <- rgparp(n = 1000, thresh = 50, shape = 0.1, riskf = "max",
               scale = tau, loc = eta, sigma = Lambda, model = "hr")
extdat <- extremo(dat = simul, margp = 0.9, coord = coord,
                 scale = 1.5, rho = 0.5, plot = TRUE)

```



```

# Constrained optimization
# Minimize distance between extremal coefficient from fitted variogram
mindistpvario <- function(par, emp, coord){
alpha <- par[1]; if(!isTRUE(all(alpha > 0, alpha < 2))){return(1e10)}

```

```

scale <- par[2]; if(scale <= 0){return(1e10)}
a <- par[3]; if(a<1){return(1e10)}
rho <- par[4]; if(abs(rho) >= pi/2){return(1e10)}
semivariomat <- mgp::power.vario(distg(coord, a, rho), alpha = alpha, scale = scale)
  sum((2*(1-pnorm(sqrt(semivariomat[lower.tri(semivariomat)]/2))) - emp)^2)
}
# constrained optimization for the parameters
hin <- function(par, ...){
  c(1.99-par[1], -1e-5 + par[1],
    -1e-5 + par[2],
    par[3]-1,
    pi/2 - par[4],
    par[4]+pi/2)
}
opt <- alabama::auglag(par = c(0.5, 30, 1.5, 0.5),
  hin = hin, control.optim = list(parscale = c(0.5, 30, 1.5, 0.5)),
  fn = function(par){
    mindistpvario(par, emp = extdat[, 'prob'], coord = coord)})

```

```

## Min(hin): 0.49999
## Outer iteration: 1
## Min(hin): 0.49999
## par: 0.5 30 1.5 0.5
## fval = 77.37
##
## Outer iteration: 2
## Min(hin): 0.3992261
## par: 0.668296 74.8862 1.39923 0.556808
## fval = 6.52
##
## Outer iteration: 3
## Min(hin): 0.3992502
## par: 0.668285 74.8855 1.39925 0.556807
## fval = 6.52
##

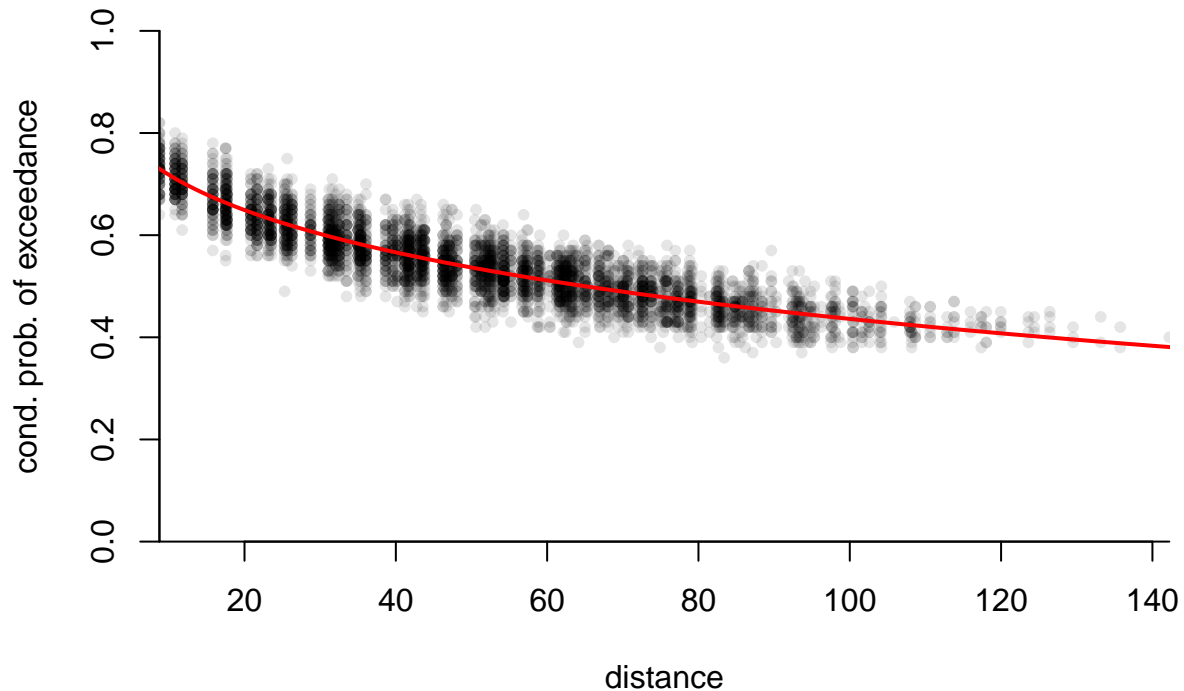
```

```

stopifnot(opt$kk1, opt$kk2)
# Plotting the extremogram in the deformed space
distfa <- distg(loc = coord, opt$par[3], opt$par[4])
plot(c(distfa[lower.tri(distfa)]), extdat[,2], pch = 20,
col = scales::alpha(1,0.1), yaxs = "i", xaxs = "i", bty = 'l',
  xlab = "distance", ylab= "cond. prob. of exceedance", ylim = c(0,1))
lines(x = (distvec <- seq(0,200, length = 1000)), col = 2, lwd = 2,
  2*(1-pnorm(sqrt(power.vario(distvec, alpha = opt$par[1], scale = opt$par[2])/2))))

```





## 4.6 Your turn

1. Simulate Gaussian random vector (using e.g. the function `mvrnorm` or through the package `RandomFields`) with independent observations and vectors with covariance matrix  $\alpha \mathbf{1}_D \mathbf{1}_D^\top + (1 - \alpha) \text{diag}(\mathbf{1}_D)$  for  $\alpha \in \{0, 0.25, 0.5, 0.75\}$ . Use `maxstabtest` on your simulated data and comment on the diagnostic plots.
2. Simulate 100 realizations from a  $\mathcal{R}$ -Pareto process with risk function max and min for the Brown–Resnick model with power variogram function `power.vario` with shape  $\alpha = 1$  and scale = 2 on a grid  $\{1, 2, \dots, 4\}^2$  using `rparp` and `rparpcs`. Benchmark the two methods: which is fastest and what is the acceptance rate of the accept-reject? Some sample code for doing this is found below

```
grid <- as.matrix(expand.grid(1:5, 1:5))
depmat <- power.vario(h = distg(grid, scale = 1, rho = 0), alpha = 1, scale = 2)/4
# This is where composition sampling shines!
samp <- rparpcs(n = 1000, shape = 0.1, riskf = "min",
               Lambda = depmat, model = "br")
#rparp is computationally intensive with "min" - only for
samp2 <- rparp(n = 1000, shape = 0.1, riskf = "max",
```

```
sigma = depmat, model = "hr")
```

# Bibliography

- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer-Verlag.
- Cox, D. R., Isham, V. S., & Northrop, P. J. (2002). Floods: some probabilistic and statistical approaches. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 360(1796), 1389–1408.
- Davison, A. C. & Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society. Series B. (Methodological)*, 52(3), 393–442.
- de Carvalho, M., Oumow, B., Segers, J., & Warchoř, M. (2013). A Euclidean likelihood estimator for bivariate tail dependence. *Communications in Statistics - Theory and Methods*, 42(7), 1176–1192.
- de Fondeville, R. (2018). *Functional Peaks-Over-Threshold Analysis for Complex Extreme Events*. PhD thesis, EPFL, Lausanne.
- de Haan, L. (1984). A spectral representation for max-stable processes. *Ann. Probab.*, 12(4), 1194–1204.
- Dombry, C., Engelke, S., & Oesting, M. (2016). Exact simulation of max-stable processes. *Biometrika*, 103(2), 303–317.
- Dombry, C. & Ribatet, M. (2015). Functional regular variations, Pareto processes and peaks over threshold. *Statistics and Its Interface*, 8(1), 9–17.
- Dupuis, D. (1999). Exceedances over high thresholds: A guide to threshold selection. *Extremes*, 1(3), 251–261.
- Einmahl, J. H. J. & Segers, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Ann. Statist.*, 37(5B), 2953–2989.
- Ferreira, A. & de Haan, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli*, 20(4), 1717–1737.
- Gabda, D., Towe, R., Wadsworth, J., & Tawn, J. (2012). Discussion of “Statistical modeling of spatial extremes” by A. C. Davison, S. A. Padoan and M. Ribatet. *Statistical Science*, 27(2), 189–192.
- Ho, Z. W. O. & Dombry, C. (2019). Simple models for multivariate regular variation and the Hüsler–Reiss Pareto distribution. *Journal of Multivariate Analysis*, 173, 525–550.
- Northrop, P. J. & Coleman, C. L. (2014). Improved threshold diagnostic plots for extreme value analyses. *Extremes*, 17(2), 289–303.

- Papastathopoulos, I. & Tawn, J. A. (2013). Extended generalised Pareto models for tail estimation. *Journal of Statistical Planning and Inference*, 143(1), 131–143.
- Rootzén, H., Segers, J., & Wadsworth, J. L. (2018). Multivariate peaks over thresholds models. *Extremes*, 21(1), 115–145.
- Scarrott, C. & MacDonald, A. (2012). A review of extreme-value threshold estimation and uncertainty quantification. *REVSTAT – Statistical Journal*, 10(1), 33–60.
- Schlather, M. & Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, 90(1), 139–156.
- Smith, R. L. (1987). Approximations in extreme value theory. Technical Report 205.
- Wadsworth, J. L. (2016). Exploiting structure of maximum likelihood estimators for extreme value threshold selection. *Technometrics*, 58(1), 116–126.
- Wadsworth, J. L. & Tawn, J. A. (2012). Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3), 543–567.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1), 1–25.