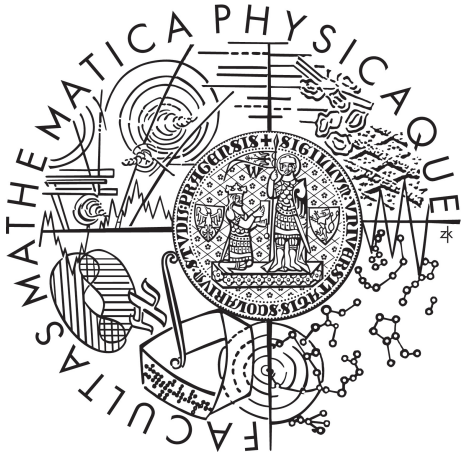
Charles University in Prague
Faculty of Mathematics and Physics

MASTER THESIS



Lukáš Drápal

Multivariate extreme value models and their application in hydrology

Department of Probability and Mathematical Statistics

Supervisor of the master thesis:	Prof. RNDr. Daniela Jarušková, CSc.
Study programme:	Mathematics
Specialization:	Probability, Mathematical Statistics and Econometrics

Prague 2014

Děkuji paní profesorce Jaruškové za množství času, který do konzultací vložila. Díky Vám jsem mohl poznat, jak je nutné při výzkumu věcem do detailu rozumět.

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In ..... on ..... Author signature

Název práce: Multivariate extreme value models and their application in hydrology

Autor: Lukáš Drápal

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Prof. RNDr. Daniela Jarušková, CSc. Katedra matematiky, stavební fakulta, České vysoké učení technické

Abstrakt: Práce se zabývá mnohorozměrnou extremální statistikou. Nejprve jsou zopakovány metody modelování v jednorozměrné statistice: pomocí blokového maxima či přesahu nad prahem. Pro modelovaní závislosti v mnohorozměrné statistice je použit přístup využívající bodového procesu. Závislost je tak modelována pomocí spekrální hustoty či pomocí exponentové funkce. Jsou probrány známé modely pro asymptoticky závislé proměnné. Podrobně rozebrán je konstrukční princip tvorby spekrální hustoty z Ballani and Schlather (2011). Je navrhnuta nová varianta modelu pairwise beta Cooley et al. (2010), která umožňuje větší flexibilitu tohoto modelu. Modely jsou využity na hydrologická data z devíti stanic ze severní Moravy, které již dříve byly analyzovány v Jarušková (2009). Nový pairwise beta model ukázal výrazně vyšší věrohodnost. Na data rovněž byla aplikována Bayesovská selekce modelu publikovaná v Sabourin et al. (2013).

Klíčová slova: Extremální statistika, pairwise beta model, spektrální hustota.

Title: Multivariate extreme value models and their application in hydrology

Author: Lukáš Drápal

Department: Department of Probability and Mathematical Statistics

Supervisor of the master thesis: Prof. RNDr. Daniela Jarušková, CSc. Katedra matematiky, stavební fakulta, České vysoké učení technické

Abstract: Present thesis deals with the multivariate extreme value theory. First, concepts of modelling block maxima and threshold excesses in the univariate case are reviewed. In the multivariate setting the point process approach is chosen to model dependence. The dependence structure of multivariate extremes is provided by a spectral measure or an exponent function. Models for asymptotically dependent variables are provided. A construction principle from Ballani and Schlather (2011) is discussed. Based on this discussion the pairwise beta model introduced by Cooley et al. (2010) is modified to provide higher flexibility. Models are applied to data from nine hydrological stations from northern Moravia previously analysed by Jarušková (2009). Usage of the new pairwise beta model is justified as it brought a substantial improvement of log-likelihood. Models are also compared with Bayesian model selection introduced by Sabourin et al. (2013).

Keywords: Extreme value theory, pairwise beta model, spectral density.

Contents

Introduction

Present thesis deals with the multivariate extreme value theory. The extreme value theory is a hot topic. As climate changes, extreme floods, droughts or heat waves are more frequent. Governments want to know risks of these in order that they could protect their citizens in a relevant matter. Insurance companies need to calculate probabilities of such events so that they could price insurance policies. Use of the extreme value theory is not limited to geographical applications. For example, financial institutions need to manage a risk of a market crash as they could lose in one day profit that has been acquired in years. The field has even received attention of the public thanks to popular science books like *The Black Swan* by *Nassim Taleb*.

The multivariate extreme value theory models occurrence of multiple extreme events at a time. This is a challenging task as it is difficult to provide a flexible dependence structure. One of common approaches is to model dependence with *copulas*. In this thesis, we use a point process approach and we model dependence with a spectral density or an exponent measure. However, a connecting link with the approach via copulas is provided.

The thesis is structured as follows. The first chapter deals with the univariate extreme value theory. Two approaches are introduced. The first models the *block maxima* by the generalized extreme value distribution. The second models *threshold excesses* by the generalized Pareto distribution. Then it is shown how the task can be alternatively expressed in the point processes notion.

In Chapter 2 the multivariate extreme value theory is introduced. The chapter starts with an introduction of *copulas* approach. Then the multivariate version of convergence of *block maxima* is presented. It provides models of dependence based on *exponent measure* or *spectral density*. A general principle of Coles and Tawn (1991) how to provide a spectral density on a simplex is introduced. The chapter ends with a description of asymptotically independent and asymptotically dependent variables.

Chapter 3 contains the bulk of the thesis. It describes known parameter models for asymptotically dependent variables. There are introduced three known families of models – the logistic, Dirichlet and the pairwise beta. The principle of constructing spectral measures introduced by Coles and Tawn (1991) is of a limited practical use. Ballani and Schlather (2011) suggested a principle how to construct a spectral density on a low dimensional simplex and extend it to higher dimensions. This result is discussed in detail. Based on this result a modification of the pairwise beta model is suggested. The new model provides a higher flexibility as it provides a dependence structure that can more precisely reflect the nature of data. The chapter ends with a description of Bayesian model averaging technique. This provides a way how to combine spectral densities or

can serve as a selection tool.

The gathered theory is applied in the last chapter. Data from nine hydrological stations from northern Moravia are used for modelling. It is discussed whether data should be fitted with models for asymptotically dependent and asymptotically independent variables. Models parameters are estimated via maximum likelihood technique. Models are compared via log-likelihood and Bayesian model selection is applied. New variants of the pairwise beta model bring a substantial improvement. The chapter ends with calculating exceedance probabilities.

Analysis was performed in statistical software `R`. Several packages for extreme value analysis have been used. Spectral densities of new models are provided along with functions that allow to generate random samples from these. In total over 1000 lines of code can be find on the attached CD or downloaded from `http://artax.karlin.mff.cuni.cz/~drapall/multivariate/`.

Chapter 1

Univariate extreme value theory

Before we shall deal with the multivariate extreme value theory we shall review the basic concepts of the univariate extreme value theory.

Let us suppose through out this chapter that $X_1, \ldots, X_n, n \in \mathbb{N}$ are independent and identically distributed random samples from distribution F, with expected values μ and finite variances σ^2.

Behaviour of sample mean $\overline{X}_n$ is described by laws of large numbers. The central limit theorem states the limiting distribution of difference between expected value and sample mean.

Now suppose that the object of the interest is not the sample mean but maximum of $X_1, \ldots, X_n$. Does this maximum converge almost surely as described for sample mean in Strong law of large numbers? Is there an equivalent to Central limit theorem – can we describe its distribution? *The univariate extreme value theory* deals with these kinds of questions.

1.1 Block maxima

We shall answer the question about a distribution of maxima. Let

$$M_n = \max(X_1, \ldots, X_n).$$

It holds

$$P(M_n \leq z) = P(X_1 \leq z, \ldots, X_n \leq z).$$

As $X_1, \ldots, X_n$ are independent it follows

$$P(M_n \leq z) = P(X_1 \leq z) \cdots P(X_n \leq z) = (F(z))^n.$$

As n approaches infinity we get

$$\lim_{n\to\infty} F^n(z) = \begin{cases} 0 & \text{if } F(z) < 1, \\ 1 & \text{if } F(z) = 1. \end{cases}$$

This result has a very limited practical use. In 1928 Fisher and Tippett developed a technique how to compensate the increase of n. In 1943 Gnedenko further improved the result to the powerful form stated below.

Theorem 1 (Fisher-Tippett-Gnedenko theorem)**.** *If there exists a sequence of pairs* $(a_n > 0, b_n \in \mathbb{R})$ *and*

$$\lim_{n\to\infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F^n(a_n X + b_n) = G(x) \tag{1.1}$$

then distribution function G *is either degenerate or it belongs to one of the three types*

$$I: \quad \Lambda(x) = \exp\left\{-\exp\left(-\frac{x-b}{a}\right)\right\}, x \in \mathbb{R},$$

$$II: \quad \Phi_\alpha(x) = \begin{cases} 0 & \text{if } x < b, \\ \exp\{-\left(\frac{x-b}{a}\right)^{-\alpha}\} & \text{if } x \geq b. \end{cases}$$

$$III: \quad \Psi_\alpha(x) = \begin{cases} \exp\{-(-\left(\frac{x-b}{a}\right)^{\alpha})\} & \text{if } x < b, \\ 1 & \text{if } x \geq b. \end{cases}$$

where $\alpha > 0,\ a > 0$ *and* $b \in \mathbb{R}$.

Proof. See Gnedenko (1943). ❑

Example 1. Let $X_1, X_2, \ldots$ be a sequence of independent unit Fréchet variables with a distribution function

$$F(x) = \exp\{-1/x\}, \quad x > 0.$$

Put $a_n = n, b_n = 0$ and fix z, then

$$\begin{aligned} \lim_{n\to\infty} P\left(\frac{M_n - b_n}{a_n} \leq z\right) &= \lim_{n\to\infty} P\left(M_n \leq nz\right) \\ &= \lim_{n\to\infty} F^n(nz) \\ &= \lim_{n\to\infty} \left[\exp\{-1/(nz)\}\right]^n \\ &= \exp(-1/z). \end{aligned}$$

Hence, the limiting distribution is again the unit Fréchet.

The three types from the Theorem 1 are called *Gumbel*, *Fréchet* and *Weibull*, respectively. The three types can be combined into a single *generalized extreme value* (GEV) distribution.

Definition 1. *A distribution function*

$$G(x) = \exp\left\{-\left(1 + \xi\frac{x-\mu}{\sigma}\right)_+^{-1/\xi}\right\},$$

where

$$(z)_+ = \max(z, 0)$$

is the generalized extreme value *distribution function. A parameter* μ *is called a location parameter,* $\sigma > 0$ *is a scale parameter and* ξ *is a shape parameter.*

How does GEV distribution relate to the three types? As $\xi \to 0$ we get the *Gumbel* distribution. For

$$\xi > 0, \alpha = \frac{1}{\xi}$$

we get the *Fréchet* distribution and for

$$\xi < 0, \alpha = -\frac{1}{\xi}$$

we get the *Weibull* distribution. We shall denote G_ξ a GEV distribution with a given shape parameter ξ.

If the normalizing constants a_n, b_n from Theorem 1 and the limit distribution G in (1.1) is not degenerate we say that F is in the *domain of attraction* of G, written as $F \in D(G)$. Sufficient conditions for a distribution function F to belong into the domain of attraction of the GEV distribution is formulated in Theorem 2.

Theorem 2 (von Mises' conditions)**.** *Let F be a non-degenerate distribution function and let*

$$x_* := \sup\{x; F(x) < 1\}.$$

(i) If

1. $x_* < \infty$
2. $F''(x)$ *exists*
3. $\forall x \leq x_* : F'(x) > 0$
4. $$\lim_{x \uparrow x_*} \left(\frac{1 - F(x)}{F'(x)} \right)' = \xi,$$

then F is in the domain of attraction G_ξ.
(ii) If

1. $x_* = \infty$
2. $F'(x)$ *exists*
3. $$\lim_{x \to \infty} \frac{xF'(x)}{1 - F(x)} = \frac{1}{\xi}, \ \xi > 0,$$

then F is in the domain of attraction G_ξ.

Proof. See de Haan and Ferreira (2006, p. 15). □

Necessary and sufficient conditions for domain attraction are formulated in the following theorem.

Theorem 3. *Let F be a non-degenerate distribution function and let*

$$x_* := \sup\{x; F(x) < 1\}.$$

Distribution function F is in the domain of attraction of G_ξ if and only if

(i) for $\xi = 0$:

$$\lim_{t \uparrow x_*} \frac{1 - F(t + xh(t))}{1 - F(t)} = \exp(-x),\ x \in R,$$

where h is a suitable positive function.
(ii) for $\xi > 0$: $x_* = \infty$ and

$$\lim_{t \to \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\xi},\ x > 0.$$

(iii) for $\xi < 0$: $x_* < \infty$ and

$$\lim_{t \downarrow 0} \frac{1 - F(x_* - tx)}{1 - F(x_* - t)} = x^{-1/\xi},\ x > 0.$$

Proof. See de Haan and Ferreira (2006, Chapter 1). □

There exist various versions of mentioned necessary and sufficient conditions for domain attraction. As univariate extreme value theory is not the primary concern of this thesis we shall not include them. Interested reader can find them in de Haan and Ferreira (2006, Chapter 1) along with other examples of normalizing constants a_n, b_n for selected distributions.

An important property of extreme value distributions is *max-stability*.

Definition 2. *A distribution function F is* max-stable *if there are coefficients $(a_n, b_n), n > 1$ such that*

$$P\left(\max(X_1, \ldots, X_n) < a_n x + b_n\right) = F^n(a_n x + b_n) = F(x).$$

This means that a max-stable function is invariant to taking a normalized maximum over independent sample. The relationship with GEV distribution functions is stated in the following theorem.

Theorem 4. *The class of max-stable functions coincide with the class of extreme value distribution functions.*

Proof. See de Haan and Ferreira (2006, Remark 1.1.5). □

The gathered extreme value theory can be used as follows. Let us say that we have data of daily precipitation. We can take the annual maxima and use them to estimate parameters of GEV distribution. This can be done via maximum likelihood technique. Another approach is to estimate just ξ with Hill or Pickands estimator – see de Haan and Ferreira (2006, Chapter 3). Given the parameters of GEV distribution we can estimate return levels. In our case the $\frac{1}{\beta}$-year return level equals to the $1 - \beta$ quantile of GEV distribution.

The described technique is called the *block maxima* approach. The choice of number of variables over which a maximum is computed exhibits a bias-variance

trade-off. If a small number of observations is chosen, convergence to GEV distribution might not hold. This would result in bias estimation. On the other hand, if too large number of observations is chosen, estimation of GEV parameters would depend on too few observations of maxima resulting in high variance.

The *block maxima* approach has a disadvantage – it uses only few values. Could for example the second largest event also be relevant for parameter estimation? Alternative approach via *Peaks over thresholds (POT)* takes into account all observations above a given threshold. It also offers a way to calculate return levels in case when observation time is not recorded and thus *block maxima* approach is not possible.

1.2 Peaks over thresholds (POT)

The core idea of the *Peaks over thresholds* method is summarized in the following theorem.

Theorem 5. *Let distribution function F be in the domain of attraction of G_ξ and*

$$x_* = \sup\{x; F(x) < 1\}.$$

As a high enough threshold t approaches x_* conditional exceedances

$$(X_i - t | X_i > t)$$

follow distribution function

$$H_\xi(x, \sigma, \xi) = \begin{cases} \left(1 - \left(1 + \xi \frac{x}{\sigma}\right)^{-1/\xi}\right)_+, \ x > 0, & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\sigma}\right), \ x > 0, & \text{if } \xi = 0. \end{cases} \tag{1.2}$$

where $\sigma > 0$ and $\xi \in R$ are respectively scale and shape parameters.

Proof. Only an outline of the proof is provided. For a more precise argument see for example Pickands (1975). By the assumption of Theorem 1 for a large n it holds

$$F^n(x) \approx \exp\left\{-\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi}\right\},$$

where $\mu, \sigma > 0$ and $\xi \in \mathbb{R}$ are parameters of GEV distribution. Applying logarithm to both sides gives

$$n \log\left(F(x)\right) \approx -\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi}. \tag{1.3}$$

For a large value x is $F(x)$ sufficiently close to 1 and Taylor expansion

$$\log\left(F(x)\right) \approx F(x) - 1$$

holds. Substitution to (1.3) yields

$$1 - F(u) \approx \frac{1}{n}\left(1 + \xi \frac{u - \mu}{\sigma}\right)^{-1/\xi}$$

for a large u. Clearly, for $y > 0$,

$$1 - F(u+y) \approx \frac{1}{n}\left(1 + \xi\frac{u+y-\mu}{\sigma}\right)^{-1/\xi}. \tag{1.4}$$

Hence,

$$\begin{aligned} P(X > u+y | X > u) &= \frac{n^{-1}\left(1+\xi\frac{u+y-\mu}{\sigma}\right)^{-1/\xi}}{n^{-1}\left(1+\xi\frac{u-\mu}{\sigma}\right)^{-1/\xi}} \\ &= \left(\frac{\sigma+\xi(u+y-\mu)}{\sigma+\xi(u-\mu)}\right)^{-1/\xi} \\ &= \left(1+\frac{\xi y}{\sigma+\xi(u-\mu)}\right)^{-1/\xi}. \end{aligned}$$

Computing the limit

$$\lim_{\xi\to 0}\left(1+\frac{\xi y}{\sigma+\xi(u-\mu)}\right)^{-1/\xi} = \exp\left\{-\frac{y}{\sigma}\right\}$$

hints the shape of the distribution function for $\xi = 0$.

□

The distribution function (1.2) is called the *generalized Pareto distribution*. The shape parameter ξ is the same as for GEV distribution in the block maxima approach. As the proof suggests the relationship of scale parameter σ in the two techniques is

$$\sigma_{POT} = \sigma_{GEV} + \xi(t-\mu).$$

Once a threshold is selected, parameters can be estimated via maximum likelihood technique. These estimates are consistent if $\xi > -1$. Asymptotic normality holds for $\xi > -0,5$.

The *Peaks over threshold* technique can be used to represent a semi-parametric extremal model. The excesses above a high threshold t follow *generalized Pareto distribution* while empirical distribution function $\hat{F}$ is used for values below u. More precisely, supposing that t is high enough that Theorem 5 holds, extremal model is the distribution function

$$\tilde{F} = \begin{cases} \hat{F}(x), & \text{for} \quad x \le t \\ (1-p_t) + p_t\left(1 - \left(1+\frac{\hat{\xi}}{\hat{\sigma}}(x-t)^{-1/\hat{\xi}}\right)\right), & \text{for} \quad x > t, \end{cases}$$

where $p_t = P(X_1 > t)$.

The choice of a threshold t exhibits a bias-variance trade-off. Choosing t too high results in estimation based on a small amount of data. This means that estimated parameters shall have a high variance. On the other hand, choosing t too low causes that the asymptotic distribution is not truly the *generalized Pareto*. Hence, it is recommended to choose the lowest threshold for which the Pareto hypothesis seems reliable.

One of the diagnostic tools used to detect where Pareto hypothesis is reliable is the *mean residual life plot*. Its purpose is to observe whether there is a linear

relationship between mean excesses and thresholds. To understand why this should hold, suppose that excesses of X_1 above t_0 are Pareto with parameters ξ and σ. Then, for $t > t_0$:

$$\begin{aligned} P(X_1 - t > x | X_1 > t)) &= \frac{P(X_1 > x + t | X_1 > t_0)}{P(X_1 > t | X_1 > t_0)} \\ &= \frac{(1 + \sigma^{-1}\xi(x - t_0 + t))^{-1/\xi}}{(1 + \sigma^{-1}\xi(t - t_0))^{-1/\xi}} \\ &= \left(1 + \frac{\xi x}{\sigma + \xi(t - t_0)}\right)^{-1/\xi}. \end{aligned}$$

This means that excesses above $t > t_0$ have generalized Pareto distribution with parameters $\sigma + \xi(t - t_0)$ and ξ. The mean of generalized Pareto distribution with parameters σ and ξ is

$$\frac{\sigma}{1 + \xi}$$

for $\xi < 1$, otherwise it is infinite. Thus, for any $t > t_0$:

$$E(X_1 - t | X_1 > t) = \frac{\sigma + \xi(t - t_0)}{1 + \xi},$$

which is a linear function of t. This explains why mean excesses should depend roughly linearly on thresholds when the distribution is truly the generalized Pareto.

Estimation of return levels of Pareto distribution can be done as follows. Let

$$p_t = P(X_1 > t)$$

for a threshold t. Suppose that for variable of our interest

$$\beta = P(X_1 > x_\beta)$$

holds $x_\beta > t$. This means

$$\beta = P(X_1 > x_\beta | X_1 \leq t)(1 - p_t) + P(X_1 > x_\beta | X_1 > t)p_t = P(X_1 > x_\beta | X_1 > t)p_t.$$

Consequently,

$$\frac{\beta}{p_t} = P(X_1 - t > x_\beta - t | X_1 > t).$$

The return level is

$$x_\beta = y_\beta + t,$$

where y_β is the $\frac{\beta}{p_t}$-quantile of generalized Pareto distribution with the given parameters.

1.3 Point processes

Two ways of computation of return levels – via the block maxima and threshold excesses were shown. We shall now provide a way of characterising these two derived from point process theory. This is useful as it shows the extreme

value theory from a different point of view. Moreover, the point process representation will be used in the multivariate setting. As formal treatment of point process theory is beyond the scope of this thesis we shall provide an informal development.

Definition 3. *Let $\mathcal{A}$ be a set and let $N(A)$ be a non-negative integer-valued random variables for each $A \subset \mathcal{A}$ such that $N(A)$ is the number of points in the set A. Let*

$$\Lambda(A) = \mathbb{E}N(A)$$

be called the intensity measure *of the process. Assuming that A is of shape*

$$A = [a_1, x_1] \times \cdots \times [a_k, x_k] \subset \mathbb{R}^k$$

and given that it exists, the derivative function

$$\lambda(x) = \frac{\partial \Lambda(A)}{\partial x_1 \cdots \partial x_k}$$

is called the intensity (density) function *of the process.*

In the modelling context, often $\mathcal{A} = \mathbb{R}$ is time, $N(A)$ then represent number of occurrences of meteorological event of interest – thunderstorms for example.

A basic example of point process is the one-dimensional homogeneous Poisson process.

Definition 4. Homogeneous Poisson process *with parameter $\lambda > 0$ on $\mathcal{A} \subset \mathbb{R}$ satisfies:*

(i) $\forall A = [t_1, t_2] \subset \mathcal{A} : N(A) \sim$ *Poisson* $(\lambda(t_2 - t_1))$;
(ii) $\forall A, B \subset \mathcal{A},\ A \cap B = \emptyset : N(A)$ *and* $N(B)$ *are independent random variables.*

The corresponding intensity measure can be easily computed as a mean of Poisson random variable

$$\Lambda([t_1, t_2]) = \lambda(t_2 - t_1).$$

Deriving yields intensity density function

$$\lambda(t) = \lambda.$$

A one-dimensional homogeneous Poisson process can be further generalized to a multi-dimensional non-homogeneous Poisson process.

Definition 5. *A point process on $\mathcal{A} \subset \mathbb{R}^k$ is said to be a k-dimensional* non-homogeneous Poisson process *with intensity density function $\lambda(\cdot)$ if*

(i) $\forall A \subset \mathcal{A} : N(A) \sim$ *Poisson* $(\Lambda(A))$, *where*
$\Lambda(A) = \int_A \lambda(\mathbf{x})\, \mathrm{d}\mathbf{x}$;
(ii) $\forall A, B \subset \mathcal{A},\ A \cap B = \emptyset : N(A)$ *and* $N(B)$ *are independent random variables.*

To represent the extreme value behaviour via point processes a notion of convergence is needed.

Definition 6. *Let $N_1, N_2, \ldots$ be a sequence of point processes on $\mathcal{A}$. The sequence* converges in distribution *to a point process N denoted*

$$N_n \xrightarrow{d} N,$$

if for all m and all bounded sets $A_1, \ldots, A_m$ such that for their boundaries $B(A_j),\ j = 1, \ldots, m$ holds

$$P\{N(B(A_j)) = 0\} = 1,\ j = 1, \ldots, m,$$

the joint distribution of

$$(N_n(A_1), \ldots, N_n(A_m))$$

converges in distribution to

$$(N(A_1), \ldots, N(A_m)).$$

Theorem 6. *Let $X_1, X_2, \ldots$ be a series of independent and identically distributed random variables for which exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that*

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) \to G(z),$$

where G(z) is the GEV distribution function. Let

$$z_* := \sup\{x; G(x) < 1\},\ z_- := \inf\{x; G(x) > 0\}.$$

The sequence of point processes

$$N_n = \left\{\frac{i}{n+1}, \frac{X_i - b_n}{a_n} : i = 1, \ldots, n\right\}$$

converges on regions of the form $(0,1) \times [u, \infty)$, for any $u > z_-$, to a Poisson process with intensity measure

$$\Lambda(A) = (t_2 - t_1)\left(1 + \xi\frac{z - \mu}{\sigma}\right)^{-1/\xi}$$

on $A = [t_1, t_2] \times [z, z_*)$.

Proof. Only an outline of the proof is provided. Let us consider a region $A = [0,1] \times (u, \infty)$ for a sufficiently large value u. Then the probability that a point from N_n falls into region A is

$$p := P\left(\frac{X_1 - b_n}{a_n} > u\right) \approx \frac{1}{n}\left(1 + \xi\frac{u - \mu}{\sigma}\right)^{-1/\xi},$$

where the approximation holds by (1.4). Since $X_i, i = 1, \ldots, n$ are mutually independent, $N_n(A)$ has binomial distribution with parameters (n, p). As $n \to \infty$ it holds $p \to 0$ and

$$np = n\frac{1}{n}\left(1 + \xi\frac{u - \mu}{\sigma}\right)^{-1/\xi} = \left(1 + \xi\frac{u - \mu}{\sigma}\right)^{-1/\xi}.$$

Thus the limiting distribution of $N_n(A)$ as $n \to \infty$ is Poisson($\Lambda'(A)$) with

$$\Lambda'(A) = \left(1 + \xi \frac{u - \mu}{\sigma}\right)^{-1/\xi}.$$

The process is homogeneous in time direction which implies that for any region of the form $A = [t_1, t_2] \times (u, \infty)$, where $[t_1, t_2] \subset [0, 1]$, the limiting distribution of $N_n(A)$ is also Poisson($\Lambda(A)$) with

$$\Lambda(A) = (t_2 - t_1)\left(1 + \xi \frac{u - \mu}{\sigma}\right)^{-1/\xi}.$$

The fact that the distributions of the $N(A)$ on non-overlapping sets are independent by construction completes the argument. □

Chapter 2

Multivariate extreme value theory

In Chapter 1 we have summarized the basics of the univariate extreme value theory. In this chapter, the theory is expanded into multiple dimensions. The focus is on the dependence structure between extremes in different components as isolating a single component would bring as back to the univariate extreme value theory. Also, in some cases only the joint occurrence of extremes is the subject of our interest. The event of interest does not have to be daily precipitation in an individual hydrological station but whether a sum of daily precipitation in several stations exceeds a certain level. Another example are air pollutant measurements: it is likely that compound effects of high levels of multiple pollutants have more severe impact on human health then effects resulting from high levels of the of the individual pollutants. The issue of temporal dependence is not covered, the dependence is understood as cross-sectional only.

Let

$$\mathbf{X}_i = (X_{i1}, \ldots, X_{id}),\ i = 1, \ldots, n,$$

be independent and identically distributed random vectors of dimension d. Consider the sample maximum for each component $j = 1, \ldots, d$ and create a new random vector

$$\mathbf{M}_n = (M_{n1}, \ldots, M_{nd}),\ \text{where } M_{nj} = \max(X_{1j}, \ldots, X_{nj}).$$

Vector M_n is called the *component-wise maxima*. Note that the component-wise maxima is usually not a part of the sample as the maxima in different components typically occur at different times.

Weak convergence of a sequence of random vectors implies weak convergence of each of its components. As in the univariate extreme value theory, if normalizing constants for each components can be found, then Theorem 1 implies convergence

$$\lim_{n\to\infty} P\left(\frac{M_{nj} - b_{nj}}{a_{nj}} < x\right) = G_j(x),\ j = 1, \ldots, d, \tag{2.1}$$

where $a_{nj} > 0$ and G_j are GEV distributions functions.

However, weak convergence of each of d components in (2.1) is strictly weaker than the joint convergence of the vector of normalized maxima. What is needed in addition is a condition of the dependence structure of the common joint distribution of vectors $\mathbf{X}_i$. One way to describe this dependence structure is via

copulas. Approach via *copulas* will not be used later in the thesis. However, it is widely used and it is desirable to show its connection with the point processes approach.

2.1 Copulas

Definition 7. *Let* $(Y_1, \ldots, Y_d)$ *be a random vector with continuous margins. Then the random vector*

$$(U_1, \ldots, U_d) = (F(Y_1), \ldots, F(Y_d))$$

has uniformly distributed marginals. The copula *of* $(Y_1, \ldots, Y_d)$ *is the joint distribution function of* $(U_1, \ldots, U_d)$*:*

$$C(u_1, \ldots, u_d) = P(U_1 \leq u_1, \ldots, U_d \leq u_d).$$

Let F be the joint distribution function of the random vectors $\mathbf{X}_i$ and C_1 be its copula. Then from definition of copula:

$$P(\mathbf{X}_i \leq \mathbf{x}) = F(\mathbf{x}) = C_1(F_1(x_1), \ldots, F_d(x_d)). \tag{2.2}$$

Let us assume throughout this section that $F_1, \ldots, F_d$ are continuous. This provides uniqueness of copula C_1 from equation (2.2) and it can be obtained as the joint distribution function of random vector $(F_1(X_{i1}), \ldots, F_d(X_{id}))$.

Theorem 7. *Let*

$$H^1, H^2, \ldots$$

be a sequence of d-variate distribution functions with marginal distributions

$$(H_1^1, \ldots, H_d^1), (H_1^2, \ldots, H_d^2), \ldots.$$

Let

$$L_1, L_2, \ldots$$

be a sequence of its copulas. Then the following statements are equivalent

(i) $\forall \mathbf{x} \in \mathbb{R}^d : \lim_{n\to\infty} H^n \xrightarrow{d} H$,

where H *has continuous marginal distributions* $H_1, \ldots, H_d$

(ii) $\forall \mathbf{x} \in \mathbb{R}^d : \lim_{n\to\infty} H_i^n \xrightarrow{d} H_i,\ i = 1, \ldots, d$ *and* $\lim_{n\to\infty} L_n = L$,

where L *is the copula corresponding to a distribution function* H*.*

Proof. See Galambos (1987). □

The intuition behind Theorem 7 is the following. Let C_n be the copula of the vector of component-wise maxima $\mathbf{M}_n$. Following (2.2) it holds

$$P(\mathbf{M}_n \leq \mathbf{x}) = F^n(\mathbf{x}) = (C_1(F_1(x_1), \ldots, F_d(x_d)))^n. \tag{2.3}$$

Marginal distribution of $F^n(\mathbf{x})$ are $F_j^n(x_j), j = 1, \ldots, d$, thus from definition of C_n if follows

$$P(\mathbf{M}_n \leq \mathbf{x}) = F^n(\mathbf{x}) = C_n(F_1^n(x_1), \ldots, F_d^n(x_d)). \tag{2.4}$$

Comparing (2.3) and (2.4) yields

$$\left(C_1\left(u_1^{1/n},\dots,u_d^{1/n}\right)\right)^n = C_n(u_1,\dots,u_d),\ (u_1,\dots,u_d)\in[0,1]^d.$$

This provides explanation why not only a sequence of marginal distributions should weakly converge but also a sequence of copulas should converge in order that joint distribution weakly converges.

Copulas that arise as weak limits of C_n are called *extreme-value* copulas.

Definition 8. *A copula C is called an* extreme-value *copula if there exists a copula C_1 such that*

$$\lim_{n\to\infty}\left(C_1\left(u_1^{1/n},\dots,u_d^{1/n}\right)\right)^n = C(u_1,\dots,u_d). \tag{2.5}$$

A copula C_1 is said to be in the domain of attraction *of C.*

Extreme-value copulas arise as the class of possible limit copulas of vectors $\mathbf{M}_n$.

Convergence of copulas share some similarities with convergence of normalized distribution functions in the univariate extreme value theory. Namely, the class of extreme-value copulas coincides with the class of *max-stable copulas*.

Definition 9. *A copula C is* max-stable *if for all $\mathbf{u}\in[0,1]^d$ and $k=1,2,\dots$*

$$C(\mathbf{u}) = \left(C(u_1^{1/k},\dots,u_d^{1/k})\right)^k.$$

Obviously, a max-stable copula is also an extreme-value copula, being in its own domain of attraction. On the other hand, putting $n=mk$ in equation (2.5) for a fixed k yields

$$\begin{aligned} C(u_1,\dots,u_d) &= \lim_{m\to\infty}\left(C_1\left(\left(u_1^{1/k}\right)^{1/m},\dots,\left(u_d^{1/k}\right)^{1/m}\right)^m\right)^k \\ &= \left(C(u_1^{1/k},\dots,u_d^{1/k})\right)^k. \end{aligned}$$

Now we shall proceed to provide a bridge between copulas and other dependence functions. Taking $u_i=\exp\{-x_i\}$ puts (2.5) into a form

$$\lim_{n\to\infty}\left(C_1\left(\exp\left\{-\frac{x_1}{n}\right\},\dots,\exp\left\{-\frac{x_d}{n}\right\}\right)\right)^n = C(\exp\{-x_1\},\dots,\exp\{-x_d\}).$$

Applying logarithm to both sides then yields

$$\lim_{n\to\infty} n\log\left(C_1\left(\exp\left\{-\frac{x_1}{n}\right\},\dots,\exp\left\{-\frac{x_d}{n}\right\}\right)\right) = \log\left(C(\exp\{-x_1\},\dots,\exp\{-x_d\})\right).$$

Via Taylor expansion it is derived

$$\lim_{n\to\infty} n\left(-1+C_1\left(1-\frac{x_1}{n},\dots,1-\frac{x_d}{n}\right)\right) = \log\left(C(\exp\{-x_1\},\dots,\exp\{-x_d\})\right).$$

Multiplying by -1 gives

$$\lim_{n\to\infty} n\left(1-C_1\left(1-\frac{x_1}{n},\dots,1-\frac{x_d}{n}\right)\right) = -\log\left(C(\exp\{-x_1\},\dots,\exp\{-x_d\})\right).$$

Definition 10. *Limit*

$$\ell(\mathbf{x}) = \lim_{n\to\infty} n\left(1 - C_1\left(1 - \frac{x_1}{n}, \ldots, 1 - \frac{x_d}{n}\right)\right), \mathbf{x} \in [0,\infty)^d$$

is called the stable tail dependence function *of C.*

The stable tail dependence function contains probability that at least one among the components exceeds its high quantile:

$$\begin{aligned}\ell(\mathbf{x}) &= \lim_{n\to\infty} n\left(1 - C_1\left(1 - \frac{x_1}{n}, \ldots, 1 - \frac{x_d}{n}\right)\right) \\ &= \lim_{n\to\infty} n\, P\left(F_1(X_1) > 1 - \frac{x_1}{n} \text{ or } \ldots \text{ or } F_d(X_d) > 1 - \frac{x_d}{n}\right). \qquad (2.6)\end{aligned}$$

The probability in (2.6) contains the union of events $F_j(X_j) > 1 - \frac{x_j}{n}$, each with probability x_j/n, given $0 \le x_j \le n$. Complete dependence and complete independence between the events provide bounds

$$\begin{aligned}\max\left(\frac{x_1}{n}, \ldots, \frac{x_d}{n}\right) &\le P\left(F_1(X_1) > 1 - \frac{x_1}{n} \text{ or } \ldots \text{ or } F_d(X_d) > 1 - \frac{x_d}{n}\right) \\ &\le \frac{x_1}{n} + \cdots + \frac{x_d}{n}.\end{aligned}$$

This leads to

$$\max(x_1, \ldots, x_d) \le \ell(x_1, \ldots, x_d) \le x_1 + \cdots + x_d.$$

The stable tail dependence function ℓ is homogeneous of order 1:

$$\begin{aligned}\ell(a\mathbf{x}) &= \lim_{n\to\infty} n\left(1 - C_1\left(1 - \frac{ax_1}{n}, \ldots, 1 - \frac{ax_d}{n}\right)\right) \\ &= \lim_{m\to\infty} am\left(1 - C_1\left(1 - \frac{x_1}{m}, \ldots, 1 - \frac{x_d}{m}\right)\right) \\ &= a\ell(\mathbf{x}),\ a > 0,\ \mathbf{x} \in [0,\infty)^d.\end{aligned}$$

Therefore it is sufficient to consider the restriction to the unit simplex

$$\mathcal{S}_{d-1} = \left\{(\omega_1, \ldots, \omega_d) \in [0,1]^d : \omega_1 + \cdots + \omega_d = 1\right\}.$$

Definition 11. *The restriction of ℓ to $\mathcal{S}_{d-1}$ is called the* Pickands dependence function *D.*

By homogenity,

$$\ell(\mathbf{x}) = (x_1 + \cdots + x_d)D(\omega_1, \ldots, \omega_d), \quad \omega_j = \frac{x_j}{x_1 + \cdots + x_d}.$$

Relationship between a stable tail dependence function and a copula from Definition 10 can be rewritten as

$$C(u_1, \ldots, u_d) = \exp\{-\ell(-\log(u_1), \ldots, -\log(u_d)\}, \mathbf{u} \in (0,1]^d.$$

In the extreme value theory, it is often more useful to standardize to a different distribution then the uniform. A frequent choice is the unit Fréchet distribution. Then the relationship between a copula and a stable tail dependence function is then of a form

$$C(e^{-1/x_1}, \ldots, e^{-1/x_d}) = \exp\left\{-\ell(1/x_1, \ldots, 1/x_d)\right\},\ \mathbf{x} \in (0,\infty)^d. \qquad (2.7)$$

2.2 Block maxima

Now, let us abandon the approach via copulas for a while and let us go back to the problem of block maxima. Suppose now that

$$X_{ij}, i = 1, \ldots, n$$

are unit Fréchet random variables with distribution function F for each $j = 1, \ldots, d$. This assumption is not restrictive as a suitable transformation can be applied otherwise. From Example 1 we know that

$$\lim_{n\to\infty} P\left(\frac{M_n}{n} \leq z\right) = \exp(-1/z).$$

The result is even stronger: it holds not only for the limiting distribution but also for all n due to max-stability:

$$P\left(\frac{M_n}{n} \leq z\right) = P\left(M_n \leq nz\right) = F^n(nz) = F(z).$$

Hence, to obtain standard univariate results for margins, the rescaled vector

$$\mathbf{M}_n^* = (M_{n1}^*, \ldots, M_{nd}^*), \text{ where } M_{nj}^* = \max(X_{1j}/n, \ldots, X_{nj}/n). \tag{2.8}$$

should be considered. The following theorem provides a multivariate version of Theorem 1.

Theorem 8. *Let*

$$X_{ij}, \quad i = 1, \ldots, n$$

be independent vectors with unit Fréchet marginal distributions for each $j = 1, \ldots, d$. Then following notation from (2.8) if

$$P(M_{n1}^* \leq x_1, \ldots, M_{nd}^* \leq x_d) \xrightarrow{d} G(x_1, \ldots, x_d). \tag{2.9}$$

where G is a non-degenerate distribution function, G has the form

$$G(x_1, \ldots, x_d) = \exp\{-V(x_1, \ldots, x_d)\}, \quad x_1 > 0, \ldots, x_d > 0 \tag{2.10}$$

where

$$V(x_1, \ldots, x_d) = d \int_{S_{d-1}} \max_{1\leq j\leq d} \left(\frac{\omega_j}{x_j}\right) \, \mathrm{d}H(\boldsymbol{\omega}) \tag{2.11}$$

and H is a positive finite measure on S_{d-1} satisfying the mean constraint

$$\int_{S_{d-1}} \omega_j \, \mathrm{d}H(\boldsymbol{\omega}) = \frac{1}{d}, \quad j = 1, \ldots, d. \tag{2.12}$$

Proof. See Resnick (1987). □

Function V from Theorem 8 is called an *exponent measure* and a measure H is called a *spectral measure*. Note that the only constraint on a spectral measure H is the condition (2.12).

Exponent measure has a close connection with a copula with unit Fréchet margins as in (2.7). From (2.7) and (2.10) it is noted that

$$V(x_1, \ldots, x_n) = \ell(1/x_1, \ldots, 1/x_d).$$

To understand the outline of proof of Theorem 8 observe that exponent measure V is *homogeneous of order* -1: from (2.11) it follows:

$$V(a^{-1}\mathbf{x}) = aV(\mathbf{x});$$

this is also a consequence of connection with ℓ. Using this property in (2.10) gives

$$G^n(\mathbf{x}) = (\exp\{-V(\mathbf{x})\})^n = \exp\{-nV(\mathbf{x})\} = \exp\{-V(\mathbf{x}/n)\} = G(\mathbf{x}/n).$$

This means that G possesses the multivariate version of max-stability property. Then it is shown that the limit distribution in (2.9) must be max-stable and that distributions of form (2.10) are the only ones that possess this property. To understand why (2.12) must hold, let us introduce the point processes approach.

2.3 Point processes approach

Theorem 9. *Let*

$$X_{ij}, \quad i = 1, \ldots, n$$

be independent vectors with unit Fréchet marginal distributions for each $j = 1, \ldots, d$ *that satisfy the convergence for component-wise maxima*

$$P(M_{n1}^* \leq x_1, \ldots, M_{nd}^* \leq x_d) \xrightarrow{d} G(x_1, \ldots, x_d).$$

Let $\{N_n\}$ *be a sequence of point processes defined by*

$$N_n = \{(X_{11}/n, \ldots, X_{1d}/n), \ldots, (X_{n1}/n, \ldots, X_{nd}/n)\},\ n \in \mathbb{N}. \tag{2.13}$$

Then

$$N_n \xrightarrow{d} N$$

on regions bounded from the origin $\mathbf{0}$*, where* N *is a non-homogenous Poisson process on* $\mathbb{R}_+^d \backslash \mathbf{0}$*. Moreover, the intensity function of* N *is*

$$\lambda(\mathrm{d}r, \mathrm{d}\boldsymbol{\omega}) = \frac{d}{r^2}\,\mathrm{d}r \times \mathrm{d}H(\boldsymbol{\omega}),$$

where

$$r = \sum_{i=1}^{d} x_i, \quad \omega_i = \frac{x_i}{r}, \quad i = 1, \ldots, d$$

and H *is related to* G *through (2.11) and (2.10).*

Proof. See Resnick (1987). □

In Theorem 9 a transformation from Cartesian to pseudo-polar coordinates was performed, in which r stands for "distance" from the origin and $\boldsymbol{\omega}$ provides angle through point on simplex $\mathcal{S}_{d-1}$. The spectral measure then determines the angular spread of points in the limit Poisson process. When spectral measure H is differentiable with a spectral density h, interpretation of the role that H plays can be made explicit as follows. When extremes are near-independent then large values occur typically only in one dimension, $h(\cdot)$ being concentrated near simplex vertices. In case when dependence is very strong, $h(\cdot)$ is concentrated around the simplex centre $(1/d, \ldots, 1/d)$.

Note also that a choice of divisor n in (2.13) is motivated by the normalizing constant of unit Fréchet variables.

Now we shall proceed with finishing the outline of proof of Theorem 8. Note that

$$P(M_{n1}^* \leq x_1, \ldots, M_{nd}^* \leq x_d) = P(N_n(A) = 0),$$

where N_n is the point process defined in (2.13) and

$$A = \mathbb{R}_+^d \backslash \{(0, x_1) \times \cdots \times (0, x_d)\}. \tag{2.14}$$

Thus, by Poisson processes limit

$$\lim_{n\to\infty} P(M_{n1}^* \leq x_1, \ldots, M_{nd}^* \leq x_d) = P(N(A) = 0) = \exp\{-\Lambda(A)\}, \tag{2.15}$$

where

$$\begin{aligned}
\Lambda(A) &= \int_A \frac{d}{r^2}\,\mathrm{d}r \times \mathrm{d}H(\boldsymbol{\omega}) \\
&= \int_{S_d}\int_{r\,=\,\min_{1\leq j\leq d}\left(\frac{x_j}{\omega_j}\right)}^{\infty} \frac{d}{r^2}\,\mathrm{d}r \times \mathrm{d}H(\boldsymbol{\omega}) \\
&= d\int_{S_d} \max_{1\leq j\leq d}\left(\frac{\omega_j}{x_j}\right)\,\mathrm{d}H(\boldsymbol{\omega}).
\end{aligned} \tag{2.16}$$

Putting (2.15) and (2.16) together yields (2.12). Note also that for a given A in (2.14) the intensity measure $\Lambda(A)$ coincides with the exponent measure $V(x_1, \ldots, x_d)$, hence following (2.16) exponent measure V can be written as

$$V(x_1, \ldots, x_d) = d\int_{S_d} \max_{1\leq j\leq d}\left(\frac{\omega_j}{x_j}\right)\,\mathrm{d}H(\boldsymbol{\omega}). \tag{2.17}$$

Now let us further describe the relationship between a spectral measure H and an exponent measure V.

Theorem 10. *If a spectral density $h(\boldsymbol{\omega})$ of spectral measure $H(\boldsymbol{\omega})$ exists then*

$$h(\omega_1, \ldots, \omega_d) = -\frac{1}{d}\frac{\partial}{\partial x_1 \ldots \partial x_d} V(x_1, \ldots, x_d)|_{x_1=\omega_1,\ldots,x_d=\omega_d}$$

for $\omega_1 + \cdots + \omega_d = 1$.

Proof. See Coles and Tawn (1991). □

Theorem 10 has its more powerful form that states spectral densities on subspaces of S_{d-1}. Let

$$S_{j-1,c} = \{\boldsymbol{\omega} \in S_{d-1} : \omega_k = 0,\, k \notin c\},$$

where

$$c = \{i_1, \ldots, i_j\}, \quad 1 \leq j \leq d.$$

This means that $S_{j-1,c}$ is isomorphic to simplex S_{j-1}. If V is differentiable, H has density $h_{j,c}$ on each of the subspaces $S_{j-1,c}$. The density $h_{j,c}$ describes the dependence structure for events that are extreme only in components $c = \{i_1, \ldots, i_j\}$.

Theorem 11. *Let V be a differentiable exponent measure and H corresponding spectral measure with densities $h_{j,c}$ as above. Then for*

$$c = \{i_1, \ldots, i_m\}, \quad b = \sum_{j=1}^{m} x_{i_j},\ 1 \leq m \leq d$$

it holds

$$\frac{\partial V(x_1, \ldots, x_d)}{\partial x_{i_1} \ldots \partial x_{i_m}} = -d\, b^{-m-1}\, h_{m,c}\left(\frac{x_{i_1}}{b}, \ldots, \frac{x_{i_m}}{b}\right).$$

on $\{\mathbf{x} \in \mathbb{R}_+^d : x_r = 0$ if $r \notin c\}$.

Proof. See Coles and Tawn (1991). □

It has been noted that the only constraint on spectral measure H is (2.12). This is the reason why no finite parametrization exists for this measure. The following theorem summarizes how to obtain a spectral density from a wide range of functions defined on a simplex.

Theorem 12. *If h^* is a positive function on S_{d-1} with finite first moments, then*

$$h(\boldsymbol{\omega}) := d(\mathbf{m} \cdot \boldsymbol{\omega})^{-d-1} \prod_{j=1}^{d} m_j\, h^*\left(\frac{m_1\omega_1}{\mathbf{m} \cdot \boldsymbol{\omega}}, \ldots, \frac{m_d\omega_d}{\mathbf{m} \cdot \boldsymbol{\omega}}\right), \tag{2.18}$$

where

$$m_j = \int_{S_{d-1}} u_j\, h^*(\mathbf{u})\ \mathrm{d}\mathbf{u}, j = 1, \ldots, d \tag{2.19}$$

satisfies constraint (2.12) and thus $h(\boldsymbol{\omega})$ is a spectral density of a valid measure H.

Proof. See Coles and Tawn (1991). □

Overview of known parametric models given by either exponent measure $V(\mathbf{x})$ or spectral density $h(\boldsymbol{\omega})$ is provided in Chapter 3.

2.4 Asymptotic dependence measures

Let H be a spectral measure that concentrates $\frac{1}{d}$ of its mass on each of simplex S_{d-1} vertices i.e on

$$\boldsymbol{\omega} = \mathbf{e_j}, \quad j = 1, \ldots, d,$$

where $\mathbf{e_j}$ is a d-dimensional unit vector that has 1 in j-th dimension. Hence, (2.12) is clearly satisfied and (2.6) gives

$$V(x_1, \ldots, x_d) = \frac{1}{x_1} + \cdots + \frac{1}{x_d}.$$

and the corresponding extreme value distribution is

$$\begin{aligned} G(x_1, \ldots, x_d) &= \exp\left\{\frac{1}{x_1} + \cdots + \frac{1}{x_d}\right\} \\ &= \exp\left\{\frac{1}{x_1}\right\} \times \cdots \times \exp\left\{\frac{1}{x_d}\right\}. \end{aligned}$$

This distribution corresponds to case when

$$X_{ij},\ 1 \leq i \leq n,\ 1 \leq j \leq d$$

are independent variables. Of course, the fact that the limiting distribution is the distribution corresponding to independent margins does not imply that the margins are truly independent. For example, let $(X_1, \ldots, X_d)$ be a d-variate normal random vector where each pair of variables

$$X_{ij},\ 1 \leq i < j \leq d$$

has correlation coefficient $\rho_{ij} < 1$. Then it can be shown (see Sibuya (1959)) that the limiting distribution is the one corresponding to independent variables. However, data even at moderately extreme levels are likely to exhibit a strong dependence, especially for ρ close to 1. Hence, models fitted to data are likely to overestimate dependence in unobserved extreme levels.

The provided example shows the importance of determining whether the limiting distribution is asymptotically independent or whether models for asymptotic dependence as those in Chapter 3 should be used. Several measures have been proposed to this purpose. First, Sibuya (1959) proposed a quite intuitive measure for bivariate distributions:

$$\chi = \lim_{u \to 1} P(F_{X_1}(X_1) > u | F_{X_2}(X_2) > u).$$

Defining also

$$\chi(u) = 2 - \frac{\log P(F_X(X) < u, F_Y(Y) < u)}{\log(P(F_X(X) < u)}, \quad 0 < u < 1. \tag{2.20}$$

It is straightforward to show that

$$\chi = \lim_{u \to 1} \chi(u).$$

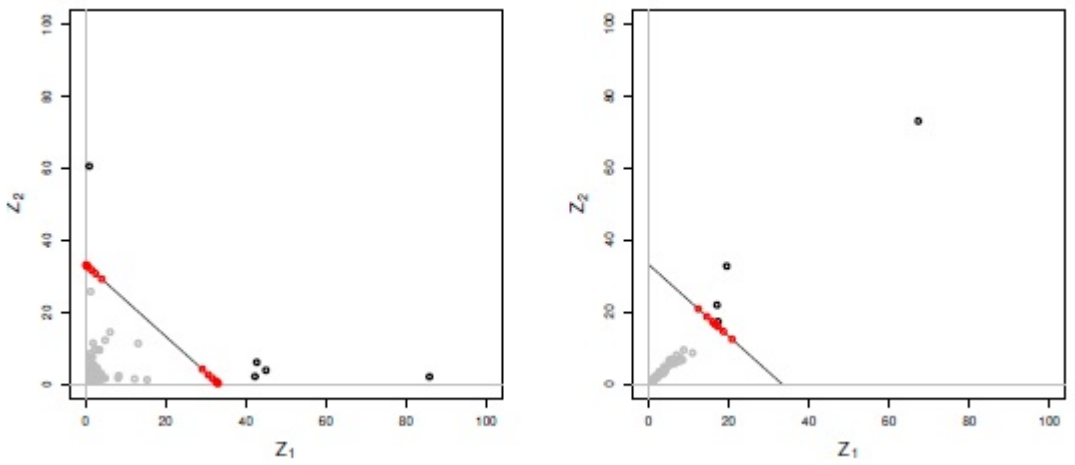


Figure 2.1: Asympotic indeependence (left) versus asymptotic perfect dependence (right).

In general χ varies between 0 and 1. If $\chi = 1$, data are perfectly asymptotically depend. If $\chi = 0$, data are asymptotically independent and difficulties described above arise. The difference between these two is depicted on Figure 2.1 from Segers (2012).

The function $\chi(u)$ can be empirically estimated for increasing u which provides an exploratory tool to analyse strength of dependence; see Coles, Heffernan, and Tawn (1999).

Dependence measure χ does not provide any measure of discrimination for asymptotically independent variables. Another measure

$$\bar{\chi} = \frac{2\log(P(F_X(X) > u))}{\log(P(F_X(X) > u, F_Y(Y) > u))} - 1$$

has been propose for these. Details can be found for example in Coles (2001, Chapter 8.4).

As data that are analysed in Chapter 4 are asymptotically dependent we shall restrict ourselves to this class.

Measure χ is suitable for analysing bivariate dependence. There is another measure aiming to summarize the overall dependence in higher dimensions.

Definition 12. *Suppose that Y is a d-variate extreme value random variable with unit Fréchet marginal distributions and*

$$Y^{(j)}, j = 1, \ldots, n$$

are independent replicates of Y. Then there exists a real number $1 \leq \theta_A \leq A$ for all non-empty $A \subseteq \{1, \ldots, d\}$ such that

$$\begin{aligned}\lim_{n\to\infty} P\left(\max_{i\in A}\max_{j=1,\ldots,n} Y_i^{(j)}/n \leq y\right) &= \lim_{n\to\infty}\left(P\left(\max_{j=1,\ldots,n} Y_i^{(j)}/n \leq y\right)\right)^{\theta_A} \\ &= \exp\{-\theta_A/y\};\end{aligned}$$

the parameter θ_A is called extremal coefficient of the process for the set A.

The extremal coefficient can be thought of as the effective number of independent unit Fréchet random variables in the set $\{Y_d, d \in A\}$. If $\theta_A = 1$ then variables corresponding to set A are perfectly asymptotically dependent. On the other hand, $\theta_A = |A|$ implies asymptotic independence. Detailed properties of extremal coefficient can be found in Schlather and Tawn (2003).

Chapter 3

Parametric models

In this chapter known parametric models for asymptotically dependent variables are summarized. These are either given via an exponent function $V(\mathbf{x})$ or through a spectral density $h(\boldsymbol{\omega})$. For models that are used for modelling in Chapter 4 an algorithm how to generate random samples is also provided. Emphasis is given on recent result by Ballani and Schlather (2011).

3.1 Logistic family

Logistic model described by Gumbel (1960) defines the model through exponent function

$$V(\mathbf{x}) = \left(\sum_{i=1}^{d} (x_i)^{-1/\alpha} \right)^{\alpha}, \quad 0 \leq \alpha \leq 1.$$

The model was further extended to model asymmetrical behaviours of subsets of variables by Coles and Tawn (1991).

The *asymmetric logistic model* is defined through

$$V(\mathbf{x}) = \sum_{c \in C} \left\{ \sum_{i \in c} \left(\frac{\theta_{i,c}}{x_i} \right)^{r_c} \right\}^{1/r_c},$$

where C is the set of all non-empty subsets of $\{1, \ldots, d\}$ and the parameters are constrained by

$$\forall c \in C : r_c \geq 1; \text{ if } i \notin c, \theta_{i,c} = 0 \text{ else } \theta_{i,c} \geq 0,\ i = 1, \ldots, d \text{ and } \sum_{c \in C} \theta_{i,c} = 1.$$

Spectral densities $h_{j,c}$ can be calculated through Theorem 11. Following its notation it gives for $\boldsymbol{\omega} \in S_{j,c}$

$$h_{j,c}(\boldsymbol{\omega}) = \frac{1}{d} \left\{ \prod_{k=1}^{j-1} (kr_c - 1) \right\} \left(\prod_{i \in c} \theta_{i,c} \right)^{r_c} \left(\prod_{i \in c} \omega_i \right)^{-r_c - 1} \left\{ \sum_{i \in c} \left(\frac{\theta_{i,c}}{\omega_i} \right)^{r_c} \right\}^{1/r_c - j}. \tag{3.1}$$

Negative asymmetric logistic model was described by Joe (1990). Its exponent function is

$$V(\mathbf{x}) = \sum_{i=1}^{d} \frac{1}{x_i} - \sum_{c \in C : |c| \geq 2} (-1)^{|c|} \left\{ \sum_{i \in c} \left(\frac{\theta_{i,c}}{x_i} \right)^{r_c} \right\}^{1/r_c}$$

with parameter constraints given by

$$\forall c \in C : r_c \leq 0;\ \text{if } i \notin c, \theta_{i,c} = 0 \text{ else } \theta_{i,c} \geq 0 \text{ and } \sum_{c \in C} (-1)^{|c|} \theta_{i,c} \leq 1.$$

Again by Theorem 11 we get

$$h_{j,c}(\boldsymbol{\omega}) = \sum_{d \in C : c \subset d} \frac{(-1)^{|d|}}{d} \left\{ \prod_{k=1}^{j-1} (kr_c - 1) \right\} \left(\prod_{i \in c} \theta_{i,c} \right)^{r_c} \left(\prod_{i \in c} \omega_i \right)^{-r_c - 1} \left\{ \sum_{i \in c} \left(\frac{\theta_{i,c}}{\omega_i} \right)^{r_c} \right\}^{1/r_c - j}.$$

The similarity of spectral density with the one of asymmetric logistic model (3.1) is clear.

A parametric family of multivariate extreme value distribution is *closed* if its exponent measure $V(\mathbf{x})$ is closed. The family of asymmetric logistic models is not closed (Tawn (1990)). On the other hand, the family of negative asymmetric logistic models is closed, see Joe (1990).

Nested logistic model further generalizes the asymmetric logistic model. The family of nested logistic models allows many kinds of parametrization, for details see Tawn (1990). In Chapter 4 the following trivariate nested logistic model is used for modelling:

$$V(\mathbf{x}) = 2^{-\alpha_0} \left[\left(x_1^{\frac{-1}{\alpha_0 \alpha_{12}}} + x_2^{\frac{-1}{\alpha_0 \alpha_{12}}} \right)^{\alpha_{12}} + \left(x_1^{\frac{-1}{\alpha_0 \alpha_{13}}} + x_3^{\frac{-1}{\alpha_0 \alpha_{13}}} \right)^{\alpha_{13}} + \left(x_2^{\frac{-1}{\alpha_0 \alpha_{23}}} + x_3^{\frac{-1}{\alpha_0 \alpha_{23}}} \right)^{\alpha_{23}} \right]^{\alpha_0} \tag{3.2}$$

with

$$0 < \alpha_0,\ \alpha_{12}, \alpha_{23}, \alpha_{13} < 1.$$

Theorem 10 can be applied to obtain spectral density $h(\boldsymbol{\omega})$ - see Sabourin et al. (2013) for steps and exact form.

In logistic family, the dependence between a coordinates subset is a decreasing function of the corresponding parameter. Here, α_0 sets the overall dependence, α_{ij} characterizes the additional pairwise dependence.

To understand how nested logistic model (3.2) can be generated let us review the concept of stable distributions.

Definition 13. *Let Z be a random variable with its independent copies $Z^{(1)}, Z^{(2)}$. Then Z is said to be* stable *if for any constants $a > 0, b > 0$ the random variable*

$$aZ^{(1)} + bZ^{(2)}$$

has the same distribution as

$$cZ + d,\ c > 0, d \in \mathbb{R}.$$

If $d = 0$ for all choices of a, b the distribution of Z is said to be strictly stable.

The characteristic function of a stable random variable can be written as

$$\varphi(t; \mu, c, \alpha, \beta) = \exp\left\{ it\mu - |ct|^{\alpha} \left(1 - i\beta \operatorname{sgn}(t) \Phi\right) \right\},$$

where

$$\Phi = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right), & \alpha \neq 1 \\ -\frac{2}{\pi} \log|t|, & \alpha = 1 \end{cases}$$

and $0 < \alpha \leq 2$ is the *stability* parameter, $-1 \leq \beta \leq 1$ is the *skewness* parameter, $0 < c < \infty$ is the *scale* parameter and $-\infty < \mu < \infty$ is the *location* parameter. When $\beta = 0$ the distribution is symmetric around μ and is referred to as a *symmetric alpha-stable* distribution. A standardized stable random variable is understood a such with $c = 1$ and $\mu = 0$.

An example of a stable distribution is the normal distribution ($\alpha = 2, \beta = 0$). Another example is the Cauchy distribution ($\alpha = 1, \beta = 0$).

The following algorithm generates random samples from model (3.2). It was proposed by Stephenson (2003) and adapted by Sabourin et al. (2013).

1. Generate independently standardized symmetric α-stable variables $S, S_{12}, S_{13}, S_{23}$ with respective index $\alpha_0, \alpha_{12}, \alpha_{13}, \alpha_{23} \in (0, 1)$ and take absolute value from these. Chambers, Mallows, and Struck (1976) provides a following method for generating standardized symmetric α-stable variables. Let Θ and W be independent with Θ uniformly distributed on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, W being standard exponential. Then,

$$S = \frac{\sin(\alpha_0\Theta)}{(\cos(\Theta))^{1/\alpha_0}} \left[\frac{\cos\left((\alpha_0 - 1)\Theta\right)}{W}\right]^{(1-\alpha_0)/\alpha_0}$$

   is a standardized symmetric α_0-stable variable.

2. For $i \in \{1, 2, 3\}$:

   (a) Simulate $E_{i,ij}, E_{i,ik}$ from the standard exponential distribution.

   (b) Set

$$X_{i,ij} := \left[\left(\frac{S}{2}\right)^{1/\alpha_{ij}} \frac{S_{ij}}{E_{i,ij}}\right]^{\alpha_{ij}\alpha_0}, \quad X_{i,ik} := \left[\left(\frac{S}{2}\right)^{1/\alpha_{ik}} \frac{S_{ik}}{E_{i,ik}}\right]^{\alpha_{ik}\alpha_0}. \tag{3.3}$$

   (c) Set

$$X_i = \max(X_{i,ij}, X_{i,ik}).$$

3. Now

$$\mathbf{X} = (X_1, X_2, X_3)$$

   has unit Fréchet margins and a multivariate distribution according to (3.2).

Now let us reason why the described algorithm works. Observe from (3.3) that

$$X_{i,ij} | (S = s, S_{ij} = s_{ij})$$

are independent with distribution

$$P(X_{i,ij} \leq x_i | s, s_{ij}) = \exp\left\{-s_{ij}\left(\frac{s}{2}\right)^{1/\alpha_{ij}}\left(\frac{1}{x_i}\right)^{1/(\alpha_0\alpha_{ij})}\right\}.$$

Therefore, $\mathbf{X}$ has conditional distribution

$$P(\mathbf{X} \leq \mathbf{x} | s, s_{12}, s_{13}, s_{23}) = \exp\left\{-\sum_{1 \leq i < j \leq 3} s_{ij}\left(\frac{s}{2}\right)^{1/\alpha_{ij}}\left(\left(\frac{1}{x_i}\right)^{1/(\alpha_0\alpha_{ij})} + \left(\frac{1}{x_j}\right)^{1/(\alpha_0\alpha_{ij})}\right)\right\}. \tag{3.4}$$

From a basic property of conditional expectation we get

$$P(\mathbf{X} \leq \mathbf{x}) = \mathbf{E}\, P(\mathbf{X} \leq \mathbf{x} | s, s_{12}, s_{13}, s_{23}). \tag{3.5}$$

As S_0, S_{12}, S_{13} and S_{23} are independent combining (3.5) and (3.4) yields

$$P(\mathbf{X} \leq \mathbf{x}) = I_1 \times I_2 \times I_3,$$

where

$$I_1 = \mathbf{E} \exp\left\{ -s_{12} \left(\frac{s}{2}\right)^{1/\alpha_{12}} \left(\left(\frac{1}{x_1}\right)^{1/(\alpha_0\alpha_{12})} + \left(\frac{1}{x_2}\right)^{1/(\alpha_0\alpha_{12})} \right) \right\}.$$

and I_2, I_3 correspond to pairs $\{1,3\}, \{2,3\}$, respectively. Now is used a Laplace transform property of positive α-stable variables:

$$\mathbf{E} \exp\{-tS_{12}\} = \exp\{-t^{\alpha_{12}}\}.$$

Concretely, we get

$$I_1 = \exp\left\{ -\frac{s}{2} \left(\left(\frac{1}{x_1}\right)^{1/(\alpha_0\alpha_{12})} + \left(\frac{1}{x_2}\right)^{1/(\alpha_0\alpha_{12})} \right)^{\alpha_{12}} \right\}.$$

Similar results for I_2 and I_3 give

$$P(\mathbf{X} \leq \mathbf{x}) = \exp\left\{ -\frac{s}{2} \sum_{1 \leq i < j \leq 3} \left(\left(\frac{1}{x_i}\right)^{1/(\alpha_0\alpha_{ij})} + \left(\frac{1}{x_j}\right)^{1/(\alpha_0\alpha_{ij})} \right)^{\alpha_{ij}} \right\}.$$

Applying the Laplace transform property for s yields the desired distribution function.

The angular components

$$\omega_i = \frac{X_i}{X_1 + X_2 + X_3},\ 1 \leq i \leq 3$$

follow immediately. By having a high threshold r_0 and keeping only the angular points for which $X_1 + X_2 + X_3 > r_0$ one obtains a sample on the simplex, approximately following angular distribution with the spectral density corresponding to exponent function (3.2).

For completeness, let us mention that Boldi (2004) proposed a version of asymmetric logistic model that coincides with the classical one for $d = 2$ but differs in higher dimensions. Its spectral density for the symmetric case is

$$h(\omega) = \frac{\Gamma(d-\alpha)}{\Gamma(1-\alpha)} \frac{\alpha^{d-1}}{d} \prod_{i=1}^{d} d\omega_i^{-1/\alpha - 1} \left(\sum_{i=1}^{d} \omega_i^{-1/\alpha} \right)^{\alpha - d}$$

and simulation is done via ratio of independent Gamma variables.

3.2 Dirichlet family

Dirichlet model was developed in Coles and Tawn (1991) via Theorem 12. Take

$$h^*(\boldsymbol{\omega}) = \left\{ \prod_{j=1}^{d} \Gamma(\alpha_j) \right\}^{-1} \Gamma\left(\sum_{i=1}^{d} \alpha_j \right) \prod_{j=1}^{d} \omega_j^{\alpha_j - 1},\ \alpha_j > 0,\ j = 1, \ldots, d,\ \boldsymbol{\omega} \in S_{d-1}. \tag{3.6}$$

Calculating

$$m_j = \frac{\alpha_j}{\sum_{j=1}^{d} \alpha_j}$$

from (2.19) and substituting to (2.18) yields

$$h(\boldsymbol{\omega}) = d \prod_{j=1}^{d} \frac{\alpha_j}{\Gamma(\alpha_j)} \frac{\Gamma\left(\sum_{j=1}^{d} \alpha_j + 1 \right)}{\left(\sum_{j=1}^{d} \alpha_j \omega_j \right)^{d+1}} \prod_{j=1}^{d} \left(\frac{\alpha_j \omega_j}{\sum_{j=1}^{d} \alpha_j \omega_j} \right)^{\alpha_j - 1}, \boldsymbol{\omega} \in S_{d-1}$$

which is known as the *Dirichlet model*. The exponent measure V has no closed form and has to be obtained by numerical integration of (2.17).

Dirichlet model allows for asymmetry between variables. The symmetric version with

$$\alpha = \alpha_1 = \cdots = \alpha_d$$

allows total asymptotic independence by taking $\alpha \to 0$ and perfect asymptotic dependence via $\alpha \to \infty$. This hints the interpretation of parameters: α_i measures dependence between variable X_i and the remaining variables, higher values meaning larger dependence.

Jarušková (2009) proposed that it can be generated from the Dirichlet model as follows. First, a sample from Dirichlet probability distribution with the same values of parameters as in the considered spectral density is generated. Angular component is then obtained via the acceptance-rejection method.

Boldi and Davison (2007) explored a model where a mixture of Dirichlet densities (3.6) is considered instead of a single one. They used the result of Dalal and Hall Jr (1980) that a distribution of any random probability vector can be approximated in the weak sence by a sequence of finite mixtures of Dirichlet distribution. They found that this property carries over to distribution functions satisfying the mean constraint (2.12), implying that any spectral distribution function may be weakly approximated by a mixture of Dirichlet distributions that satisfies (2.12). EM algorithm is used to fit mixture densities, for details see Boldi and Davison (2007) and Boldi (2004).

A re-parametrization of Dirichlet mixture model has been done recently by Sabourin and Naveau (2014) in order that the Bayesian model averaging procedure that will be introduced later in this chapter could be performed.

3.3 The pairwise beta model

The *pairwise beta model* was described in Cooley, Davis, and Naveau (2010). It is given via spectral density

$$h(\boldsymbol{\omega}|\beta_0, \{\beta_{i,j}\}_{1\le i<j\le d}) = \sum_{1\le i<j\le d} h_{ij}(\boldsymbol{\omega}|\beta_0, \beta_{i,j}), \quad \beta_0, \beta_{i,j} > 0, \tag{3.7}$$

where each spectral density h_{ij} is the following beta function:

$$h_{ij}(\boldsymbol{\omega}|\beta_0, \beta_{i,j}) = K_d(\beta_0)\, \omega_{ij}^{2\beta_0-1}\, (1-\omega_{ij})^{(d-2)\beta_0-d+2}\, \frac{\Gamma(2\beta_{ij})}{\Gamma^2(\beta_{i,j})}\, \omega_{i/ij}^{\beta_{ij}-1}\, \omega_{j/ij}^{\beta_{ij}-1},$$

where

$$\omega_{ij} = \omega_i + \omega_j,\ \omega_{i/ij} = \frac{\omega_i}{\omega_i+\omega_j} \text{ and } K_d(\beta_0) = \frac{2(d-3)!}{d(d-1)} \frac{\Gamma(\beta_0 d+1)}{\Gamma(2\beta_0+1)\Gamma(\beta_0(d-2))}.$$

The exponent function $V(\mathbf{x})$ is obtained via numerical integration of (2.17).

Parameter β_0 characterizes the overall dependence between variables, larger values meaning higher association. Parameter β_{ij} measures dependence between i-th and j-th coordinates, again larger values mean higher association.

The pairwise beta model does not exhibit the exponent measure closure, see Cooley et al. (2010).

Cooley et al. (2010) generated random samples from the model via acceptance-rejection method. Jarušková (2009) generated samples through conditional beta distribution. More precisely, when T has a beta distribution with parameters γ and δ and the conditional distribution of X given $T = t$ is a symmetric beta distribution on $[0, t]$ given by density

$$f(x|T=t) = \frac{\Gamma(2\beta_{ij})}{\Gamma(\beta_{ij})} \left(\frac{x}{t}\right)^{\beta_{ij}-1} \left(1-\frac{x}{t}\right)^{\beta_{ij}-1} \frac{1}{t},\ x \in [0,t]$$

then the joint density of $(X, Y = T - X)$ has a form

$$f(x,y) = \frac{\Gamma(2\beta_{ij})}{\Gamma(\beta_{ij})} \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \left(\frac{x}{x+y}\right)^{\beta_{ij}-1} \left(\frac{y}{x+y}\right)^{\beta_{ij}-1} (x+y)^{\gamma-2}(1-x-y)^{\delta-1}$$

for

$$0 \le x \le 1,\ 0 \le y \le 1,\ 0 \le x+y \le 1.$$

The simulation algorithm used in data analysis in Chapter 4 is described in Sabourin et al. (2013) and is based on the same principle as the approach of Jarušková (2009). The notation from the definition of pairwise beta model is used, additionally it is defined

$$\mathbf{s}_{ij} = \frac{\boldsymbol{\omega}_{[-(i,j)]}}{1-\omega_{ij}} \text{ with } \boldsymbol{\omega}_{[-(i,j)]} = (\omega_1, \dots, \omega_{i-1}, \omega_{i+1}, \dots, \omega_{j-1}, \omega_{j+1}, \dots, \omega_d).$$

Transformation

$$\omega \to (\omega_{ij}, \omega_{i/ij}, \mathbf{s}_{ij})$$

is made. Its Jacobian is

$$J_{ij} = \frac{1}{\omega_{ij}(1-\omega_{ij})^{d-3}}.$$

Each beta function

$$h_{ij}((\omega_{ij}, \omega_{i/ij}, \mathbf{s}_{ij})|\beta_0, \beta_{i,j})$$

can be expressed in new coordinates as

$$h_{ij}((\omega_{ij}, \omega_{i/ij}, \mathbf{s}_{ij})|\beta_0, \beta_{i,j}) \propto \omega_{ij}^{2\beta_0}(1-\omega_{ij})^{(d-2)\beta_0-1}\omega_{i/ij}^{\beta_{ij}-1}(1-\omega_{i/ij})^{\beta_{ij}-1}J_{ij}$$

which corresponds to two beta distributions and one uniform distribution on the unit simplex of dimension $d-3$. Hence, the algorithm can be summarized as follows.

1. Choose uniformly a pair $i < j$.
2. Generate independently vector R_{ij} according to beta distribution $Be(2\beta_0 + 1, (d-2)\beta_0)$, vector $W_{i/ij}$ according to $Be(\beta_{ij}, \beta_{ij})$ and vector S_{ij} according to uniform Dirichlet distribution $Dir_{d-2}(1, \ldots, 1)$.
3. Define

$$\begin{aligned} \omega_i &:= R_{ij}W_{i/ij} \\ \omega_j &:= R_{ij}(1-W_{i/ij}) \\ \boldsymbol{\omega}_{[-(i,j)]} &:= (1-R_{ij})S_{ij}. \end{aligned}$$

Figure 3.1 helps to visualize the pairwise beta model. It depicts a single h_{12} spectral density on a simplex S_2. Parameters of the density are $\beta_0 = 1.4$ and $\beta_{12} = 1.6$. Parameter β_0 sets the shape of density on the line from the vertex to the middle of the opposite site. Parameter β_{12} sets the shape of lines parallel to this site. The entire pairwise beta model is then build as a normalized sum of these densities.

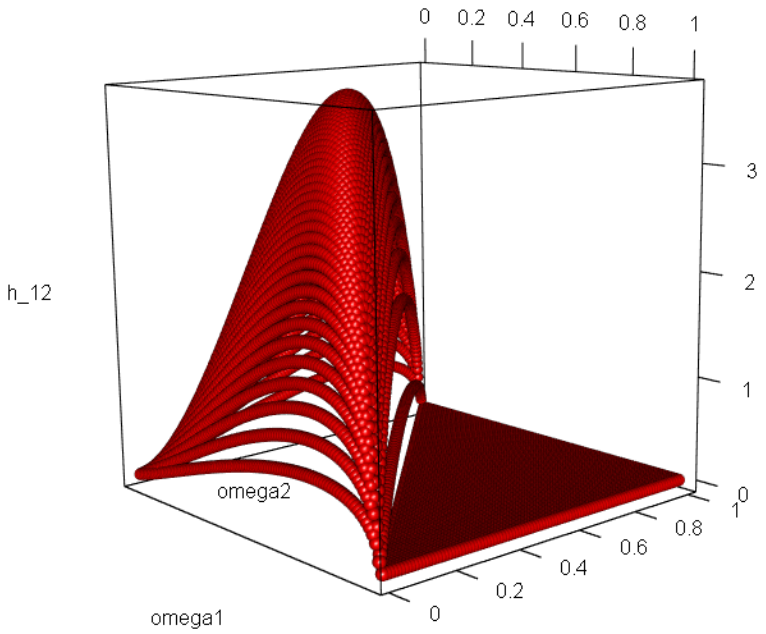


Figure 3.1: Scatter plot of pairwise beta density h_{12} with parameter $\beta_0 = 1.4$, $\beta_{12} = 1.6$

3.4 Construction principle

Theorem 12 is a powerful result that allows to construct a spectral density from a wide range of functions on a simplex. However, it is very difficult to propose a new meaningful class of functions that would model dependence of variables on a simplex. Especially, if one wants this new class of functions to be applicable to a simplex of a general dimension. This is the reason why only three types of models – logistic family, Dirichlet family and pairwise beta model – have been been introduced so far.

It seems easier to propose a function on a simplex of a low dimension and then extend it to a simplex of a high dimension. This is actually what pairwise beta model does – its symmetric beta density with parameter $\beta_{i,j}$ is a function on two dimensional unit simplex: interval $(0, 1)$. The following Theorem of Ballani and Schlather (2011) provides conditions for extending a density on a simplex to a simplex of a higher dimension. It also states necessary conditions for these lower-dimensional densities. Note that in the pairwise beta model the only used property of a symmetric beta density was its expectation 0.5.

Theorem 13. *Let*

$$2 \leq k \leq d-2,\ 1 \leq i_1 < \cdots < i_k \leq d \text{ and } \{i_{k+1}, \ldots, i_d\} = \{1, \ldots, d\} \backslash \{i_1, \ldots, i_k\}$$

and $\mathbf{m_k}$ *denote the vector*

$$\{1/k\}^{k-1} = (1/k, \ldots, 1/k)$$

of length $k-1$*. Let* p *be a probability density function on the parameter space*

$$\widetilde{S_{k-1}} = \{\boldsymbol{\omega} \in [0, \infty)^{k-1} : \|\boldsymbol{\omega}\|_1 \leq 1\}$$

with centre of mass $\mathbf{m_k}$ *and* $\widetilde{p}$ *be a probability density function on the parameter space* $\widetilde{S_{d-1}}$ *with centre of mass* $\mathbf{m_{d-k}}$*. Furthermore, let* q *be a density function on* $[0, 1]$ *such that*

$$\frac{M_q(k, d-k-1)}{M_q(k-1, d-k-1)} = \frac{k}{d}, \tag{3.8}$$

where

$$M_q(k, l) = \int_0^1 q(r)\, r^k\, (1-r)^l \, \mathrm{d}r.$$

Then

$$h_{i_1,\ldots,i_k}(\boldsymbol{\omega}, p, \tilde{p}, q) = \\ = \frac{1}{\sqrt{d}} \frac{q(\omega_{i_1} + \cdots + \omega_{i_k})}{M_q(k-1, d-k-1)} p\left\{\frac{(\omega_{i_1}, \ldots, \omega_{i_{k-1}})}{\sum_{j=1}^k \omega_{i_j}}\right\} \tilde{p}\left\{\frac{(\omega_{i_{k+1}}, \ldots, \omega_{i_{d-1}})}{1 - \sum_{j=1}^k \omega_{i_j}}\right\}$$

is a valid spectral density on S_{d-1}*.*

Proof. Without loss of generality suppose

$$\{i_1, \ldots, i_k\} = \{1, \ldots, k\}.$$

Let

$$r = \sum_{i=1}^{k} \omega_i.$$

The change of variables

$$\theta_i = \frac{\omega_i}{r},\ i = 1, \ldots, k-1,$$
$$\varphi_j = \frac{\omega_{k+j}}{1-r},\ j = 1, \ldots, d-k-1$$

has Jacobian

$$r^{k-1}(1-r)^{d-k-1}.$$

Then

$$\begin{aligned}
&\int_{S_{d-1}} h_{1,\ldots,k}(\boldsymbol{\omega}, p, \tilde{p}, q)\ \mathrm{d}\boldsymbol{\omega} = \\
&= \int_{\widetilde{S_{d-1}}} \frac{q(\omega_1 + \cdots \omega_k)}{M_q(k-1, d-k-1)} p\left\{\frac{(\omega_{i_1}, \ldots, \omega_{i_{k-1}})}{\sum_{j=1}^{k} \omega_{i_j}}\right\} \tilde{p}\left\{\frac{(\omega_{i_{k+1}}, \ldots, \omega_{i_{d-1}})}{1 - \sum_{j=1}^{k} \omega_{i_j}}\right\}\ \mathrm{d}\boldsymbol{\omega} \\
&= \int_0^1 \int_{\widetilde{S_{k-1}}} \int_{\widetilde{S_{d-k-1}}} \frac{q(r) r^{k-1}(1-r)^{d-k-1}}{M_q(k-1, d-k-1)} p(\theta)\tilde{p}(\varphi)\ \mathrm{d}\varphi \mathrm{d}\theta \mathrm{d}r \\
&= \int_0^1 \frac{q(r) r^{k-1}(1-r)^{d-k-1}}{M_q(k-1, d-k-1)}\ \mathrm{d}r \\
&= \frac{M_q(k-1, d-k-1)}{M_q(k-1, d-k-1)} = 1
\end{aligned}$$

Now, let us verify the constraint (2.12). For $i = 1, \ldots, k-1$ using the fact that p has the centre of mass at $\mathbf{m_k}$ we get

$$\begin{aligned}
&\int_{S_{d-1}} \omega_i\, h_{1,\ldots,k}(\boldsymbol{\omega}, p, \tilde{p}, q)\ \mathrm{d}\boldsymbol{\omega} = \\
&= \int_{\widetilde{S_{d-1}}} \frac{\omega_i\, q(\omega_1 + \cdots \omega_k)}{M_q(k-1, d-k-1)} p\left\{\frac{(\omega_{i_1}, \ldots, \omega_{i_{k-1}})}{\sum_{j=1}^{k} \omega_{i_j}}\right\} \tilde{p}\left\{\frac{(\omega_{i_{k+1}}, \ldots, \omega_{i_{d-1}})}{1 - \sum_{j=1}^{k} \omega_{i_j}}\right\}\ \mathrm{d}\boldsymbol{\omega} \\
&= \int_0^1 \int_{\widetilde{S_{k-1}}} \int_{\widetilde{S_{d-k-1}}} \frac{q(r) r^{k}(1-r)^{d-k-1}}{M_q(k-1, d-k-1)} \theta_i p(\theta)\tilde{p}(\varphi)\ \mathrm{d}\varphi \mathrm{d}\theta \mathrm{d}r \\
&= \int_0^1 \frac{q(r) r^{k}(1-r)^{d-k-1}}{M_q(k-1, d-k-1)} \frac{1}{k}\ \mathrm{d}r \\
&= \frac{M_q(k, d-k-1)}{M_q(k-1, d-k-1)} = \frac{k}{d}\frac{1}{k} = \frac{1}{d}
\end{aligned}$$

where we have used (3.8). For $i = k$ realize that

$$\begin{aligned}
\omega_k &= r - \sum_{i=1}^{k-1} \omega_i \\
&= r\left(1 - \sum_{j=1}^{k-1} \theta_i\right)
\end{aligned}$$

after change of variables. Then the same approach applies.

Similarly for $i = k+1 \dots, d-1$ we get

$$\begin{aligned}\int_{S_{d-1}} \omega_i \, h_{1,\dots,k}(\boldsymbol{\omega}, p, \tilde{p}, q) \, \mathrm{d}\boldsymbol{\omega} &= \int_0^1 \frac{q(r) r^{k-1}(1-r)^{d-k}}{M_q(k-1, d-k-1)} \frac{1}{d-k} \, \mathrm{d}r \\ &= \frac{M_q(k-1, d-k)}{M_q(k-1, d-k-1)} \frac{1}{d-k} \\ &= \frac{M_q(k-1, d-k-1) - M_q(k, d-k-1)}{M_q(k-1, d-k-1)} \frac{1}{d-k} \\ &= \left(1 - \frac{k}{d}\right) \frac{1}{d-k} = \frac{1}{d}.\end{aligned}$$

For $i = d$ it holds

$$\begin{aligned}\omega_k &= 1 - r - \sum_{i=k+1}^{d-1} \omega_i \\ &= (1-r) \left(1 - \sum_{j=1}^{k-1} \theta_i\right).\end{aligned}$$

and again we get

$$\int_{S_{d-1}} \omega_i \, h_{1,\dots,k}(\boldsymbol{\omega}, p, \tilde{p}, q) \, \mathrm{d}\boldsymbol{\omega} = \frac{1}{d}$$

□

Although the proof was quite technical it points out a way how to obtain an intuitive understanding of the construction principle. First, compute the distance

$$r = \sum_{i=1}^{k} \omega_i.$$

In three dimensional simplex – triangle – this would be a distance from one of the vertices. Density $q(r)$ captures the dependence of spectral density on the distance r. Densities p and $\tilde{p}$ then set dependence structures for fixed r.

Theorem 13 provides a way how to make a valid spectral density on a high dimensional simplex via lower dimensional densities. However, as in the pairwise beta model it is desirable to make a mixture of these in order that all lower-dimensional simplexes of a given size are included. This is made explicit in the following Corollary.

Corollary 1. Let $2 \le k \le d-2$. For all

$$1 \le i_1 < \dots < i_k \le d$$

let $p_{i_1,\dots,i_k}$ be a probability density function on the parameter space $\widetilde{S_{k-1}}$ with the centre of mass $\mathbf{m_k}$ and

$$\tilde{p}_{i_{k+1},\dots,i_d}, \ \{i_{k+1}, \dots, i_d\} = \{1, \dots, d\} \backslash \{i_1, \dots, i_k\}$$

be a probability density function on the parameter space $\widetilde{S_{d-k-1}}$ with the centre of mass $\mathbf{m_{d-k}}$. Let $q_{i_1,\ldots,i_k}$ be a density function on $[0,1]$ such that (3.8) is satisfied. Then

$$h(\boldsymbol{\omega}, p_{i_1,\ldots,i_k}, \tilde{p}_{i_{k+1},\ldots,i_d}, q_{i_1,\ldots,i_k}) =$$
$$= \frac{1}{\sqrt{d}} \frac{1}{\binom{d}{k}} \sum_{1\le i_1<\ldots<i_k\le d} \frac{q_{i_1,\ldots,i_k}(\omega_{i_1}+\cdots+\omega_{i_k})}{M_{q_{i_1,\ldots,i_k}}(k-1,d-k-1)} p_{i_1,\ldots,i_k} \left\{ \frac{(\omega_{i_1},\ldots,\omega_{i_{k-1}})}{\sum_{j=1}^{k}\omega_{i_j}} \right\} \times$$
$$\times \tilde{p}_{i_{k+1},\ldots,i_d} \left\{ \frac{(\omega_{i_{k+1}},\ldots,\omega_{i_{d-1}})}{1-\sum_{j=1}^{k}\omega_{i_j}} \right\}$$

is a valid spectral density on S_{d-1}. The constraint (3.8) is not needed when q does not depend on indices $i_1,\ldots,i_d$.

Proof. Let us verify that the constraint (2.12) is satisfied.

If constraint (3.8) holds it follows from the proof of Theorem 13 that

$$\int_{S_{d-1}} \omega_j \, h(\boldsymbol{\omega}, p_{i_1,\ldots,i_k}, \tilde{p}_{i_{k+1},\ldots,i_d}, q_{i_1,\ldots,i_k}) = \frac{1}{d}, \ j \in \{i_1,\ldots,i_d\}$$

and hence the result clearly holds.

If constraint (3.8) does not hold and q does not depend on indices $i_1,\ldots,i_k$ then following proof of Theorem 13 we have

$$\int_{S_{d-1}} \omega_j \, h(\boldsymbol{\omega}, p_{i_1,\ldots,i_k}, \tilde{p}_{i_{k+1},\ldots,i_d}, q) = \frac{M_q(k,d-k-1)}{M_q(k-1,d-k-1)} \frac{1}{k}, \ j \in \{i_1,\ldots,i_k\}.$$

Similarly, for $j \in \{i_{k+1},\ldots,i_d\}$

$$\int_{S_{d-1}} \omega_j \, h(\boldsymbol{\omega}, p_{i_1,\ldots,i_k}, \tilde{p}_{i_{k+1},\ldots,i_d}, q) = \left(1 - \frac{M_q(k,d-k-1)}{M_q(k-1,d-k-1)}\right) \frac{1}{d-k}$$

Now calculating the sum via counting the cases when j is or is not included in $\{i_1,\ldots,i_k\}$ gives

$$\frac{1}{\binom{d}{k}} \sum_{1\le i_1<\ldots<i_k\le d} \omega_j \, h(\boldsymbol{\omega}, p_{i_1,\ldots,i_k}, \tilde{p}_{i_{k+1},\ldots,i_d}, q) = \frac{1}{\binom{d}{k}} \times$$
$$\times \left[\binom{d-1}{k-1} \frac{1}{k} \frac{M_q(k,d-k-1)}{M_q(k-1,d-k-1)} + \binom{d-1}{k} \frac{1}{d-k} \left(1 - \frac{M_q(k,d-k-1)}{M_q(k-1,d-k-1)}\right) \right]$$
$$= \frac{1}{\binom{d}{k}} \left[\frac{(d-1)!}{k!(d-k)!} \frac{M_q(k,d-k-1)}{M_q(k-1,d-k-1)} + \frac{(d-1)!}{k!(d-k)!} \left(1 - \frac{M_q(k,d-k-1)}{M_q(k-1,d-k-1)}\right) \right]$$
$$= \frac{1}{d}$$

and the proof is complete.

□

Note. Defining $p_{i_1} \equiv 1, \tilde{p}_{i_d} \equiv 1$, Corollary 1 allows for $k = 1$ and $k = d-1$, respectively.

Corollary 1 is easily interpretable. If condition (3.8) holds then the centre of each spectral density is such that it is satisfies the constraint (2.12). However, if q does not depend on indices $i_1, \ldots, i_k$ each individual centre of mass does not have to satisfy (2.12) as summing over all densities averages this out. Result in Corollary 1 can be made even little stronger: q can depend on indices $i_1, \ldots, i_k$ as long as

$$\frac{M_q(k, d-k-1)}{M_q(k-1, d-k-1)}$$

remains fixed.

Let us illustrate Theorem 13 on an example.

Example 2. One may obtain a model similar to the pairwise beta model in the following fashion. Let $q(r)$ be of a form

$$q(r) = \frac{1}{\mathrm{B}(a,b)}\, r^{a-1}(1-r)^{b-1},\ 0 \leq r \leq 1.$$

Then we have

$$M_q(k, d-k-1) = \int_0^1 r^{a+k}(1-r)^{b+d-k-1}\ \mathrm{d}r = \frac{\Gamma(a+k)\Gamma(b+d-k-1)}{\Gamma(a+b+d-1)}.$$

In order that (3.8) holds we need

$$\begin{aligned} \frac{M_q(k, d-k-1)}{M_q(k-1, d-k-1)} &= \frac{k}{d} \\ \frac{a+k-1}{a+b+d-2} &= \frac{k}{d}. \end{aligned}$$

This implies

$$b = \frac{(d-k)a - d + 2k}{k}.$$

Choosing

$$a = k(\alpha - 1) + 1$$

gives

$$b = (d-k)(\alpha-1) + 1.$$

This gives

$$q(r) = q(r;\alpha) = \frac{1}{\mathrm{B}(k\alpha - k + 1, \alpha(d-k) - d + k + 1)}\, r^{k(\alpha-1)}(1-r)^{(d-k)(\alpha-1)}$$

which for $k = 2$ gives exponents different but quite similar to those of the pairwise beta model. In the pairwise beta model density q is

$$q(r) = \frac{1}{\mathrm{B}(2\alpha, (\alpha-1)(d-2)+1)}\, r^{2\alpha-1}(1-r)^{(d-2)(\alpha-1)}$$

and the constraint (3.8) is not satisfied. Also, density q does not depend on the indices i, j. Corollary 1 justifies why pairwise beta model provides a valid spectral density.

Ballani and Schlather (2011) provide also new models constructed via Theorem 13. One of them is the *pairwise exponential model*.

Example 3. Let

$$f_u(t;b) = \frac{1}{c_u(b)} \left(\frac{1}{2} - \left| \frac{1}{2} - t \right| \right)^{u-1} \exp\left\{ b \left(\frac{1}{2} - \left| \frac{1}{2} - t \right| \right) \right\} 1_{[0,1]}(t),$$

where $c_u(b)$ is a normalizing constant such that $f_u(t;b)$ is a probability density function on $[0,1]$. This is possible for all $u > 0$ and $b \in \mathbb{R}$. If $u \in \mathbb{N}$ then the constant $c_u(b)$ even has a closed form. For instance

$$c_1(b) = 2\frac{\exp\{b/2\} - 1}{b}.$$

As $f_u(t;b)$ is symmetric around $1/2$ it has expectation $1/2$ and hence is a possible choice for p_{i_1,i_2} in case $k = 2$. The pairwise beta density from the pairwise beta model can be replaced with the density $f_1(t;b)$. Using Corollary 1 and keeping

$$q(r) = q(r;\alpha) = \frac{1}{\mathrm{B}(k\alpha - k + 1, \alpha(d-k) - d + k + 1)} r^{k(\alpha-1)}(1-r)^{(d-k)(\alpha-1)}$$

and $\tilde{p}$ the uniform distribution provides model

$$h(\boldsymbol{\omega}, \alpha, (b_{ij})_{1\le i<j\le d}) = \frac{(d-3)!}{\binom{d}{2}} \frac{\Gamma(\alpha d + 1)}{\Gamma(2\alpha+1)\Gamma(\alpha(d-2))} \times \\ \times \sum_{1\le i<j\le d} (\omega_i + \omega_j)^{2\alpha-1} (1 - \omega_i - \omega_j)^{(\alpha-1)(d-2)} f_1\left(\frac{\omega_i}{\omega_i + \omega_j}; b_{ij} \right).$$

The parameter α measures the overall dependence between variables, b_{ij} measures dependence between variables i, j. As in pairwise beta model, larger parameter values indicate higher association.

This model is also modified into a *weighted exponential model*. In similar fashion is modified Dirichlet model into a new *weighted Dirichlet model*. For details see Ballani and Schlather (2011).

3.5 Variants of the pairwise beta model

In this section more properties of the pairwise model are discussed and variants of the model are suggested.

A disadvantage of the pairwise beta model is that parameters β_{ij} do not work independently: adjusting a single β_{ij} parameter affects the level of dependence among all pairs of components. The reason is simply the additive nature of the model. This property is shared also by all models constructed via Corollary 1 as these are all additive models.

A model that allows to capture various dependence structures is desirable. Beta distribution provides a wide range of dependence possibilities and thus, it is a good choice. However, there are restrictions on model parameters that decrease this flexibility. The restriction that the distribution modelling the relationship

between i-th and j-th variable has expectation 0.5 implying symmetric beta distribution is not overly demanding – a symmetric beta distribution can still model both the case when variables show a high and very low level of association at extreme levels.

Density q from Theorem 13 is in the pairwise beta model the beta distribution with parameters $(2\beta_0, (d-2)\beta_0 - d + 3)$. The influence of density q is tremendous – it captures the overall dependence of parameters and hence, it sets the overall shape of the model. However, the imposition of relationship between the two shape parameters decreases a lot of the flexibility of the beta distribution. This is illustrated on Figure 3.2 for case when $d = 3$.

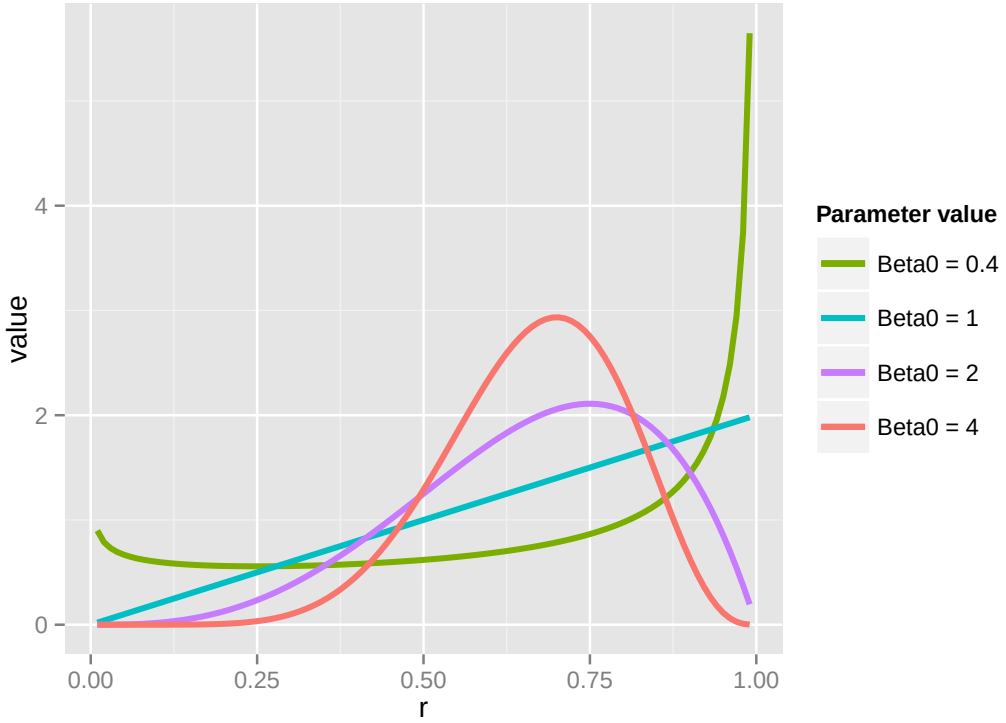


Figure 3.2: Beta density with parameters $(2\beta_0, \beta_0)$

From the Figure 3.2 it is clear that in such a setting beta distribution looses a lot of its flexibility. Its mean is $2/3$ and thus it cannot provide a model where observations are located mostly near a simplex vertex – in a situation where the overall dependence is low. This requirement was probably originally meant in order that each spectral density h_{ij} satisfies the limit constraint (2.12). However, due to the Corollary 1 this is not needed.

In order that (3.8) holds, Ballani and Schlather (2011) suggest for $k = 2, d = 3$ a beta distribution with parameters $(2\beta_0 - 1, \beta_0)$. This limits the parameter space to $\beta_0 > 0.5$. As Figure 3.3 shows such a distribution is able to put emphasis on observations near vertex. However, it cannot capture well a linearly descending dependence.

We suggest three variants of the pairwise beta model. First, as density $q(r)$ has a great influence on the model structure we suggest to model both parameters (α_1, α_2) of beta distribution. We called this the *double pairwise beta model*. The model is more flexible at the cost of increased number of parameters. To provide a

model with the same number of parameters we suggest to fix one of the parameters at a constant value. The model with $q(r)$ as beta distribution with parameter $(\alpha_1, 1)$ we call the *left pairwise beta model*. This model with parameter $(1, \alpha_2)$ we call the *right pairwise beta model*.

The Figure 3.4 shows possible beta distribution densities of the *right pairwise beta model* for different parameter choices. Observe that the mean value is not limited and that it allows a linearly descending densities.

For *double pairwise beta model* following Corollary 1 we have

$$\begin{aligned} M_q(k-1, d-k-1) &= \int_0^1 \frac{1}{\mathrm{B}(\alpha_1, \alpha_2)} x^{\alpha_1 - 1 + k - 1} (1-x)^{\alpha_2 - 1 + d - k - 1} \\ &= \frac{\mathrm{B}(\alpha_1 + k - 1, \alpha_2 + d - k - 1)}{\mathrm{B}(\alpha_1, \alpha_2)}. \end{aligned}$$

Combining with

$$q(\omega_i + \omega_j) = \frac{1}{\mathrm{B}(\alpha_1, \alpha_2)} (\omega_i + \omega_j)^{\alpha_1 - 1} (1 - \omega_i - \omega_j)^{\alpha_2 - 1}$$

we get

$$\frac{q(\omega_i + \omega_j)}{M_q(k-1, d-k-1)} = \frac{(1 - \omega_i - \omega_j)^{\alpha_2 - 1}}{\mathrm{B}(\alpha_1 + k - 1, \alpha_2 + d - k - 1)} (\omega_i + \omega_j)^{\alpha_1 - 1}.$$

As $k = 2$ and p is the symmetric pairwise beta distribution and $\tilde{p}$ is the uniform distribution we get:

$$h(\boldsymbol{\omega} | \alpha_1, \alpha_2, \{\beta_{ij}\}_{1 \le i < j \le d}) = \sum_{1 \le i < j \le d} h_{ij}(\boldsymbol{\omega} | \alpha_1, \alpha_2, \beta_{ij}), \quad \alpha_1, \alpha_2, \beta_{ij} > 0, \tag{3.9}$$

where each spectral density h_{ij} is:

$$h_{ij}(\boldsymbol{\omega} | \alpha_1, \alpha_2, \beta_{i,j}) = K_d(\alpha_1, \alpha_2) \frac{2(d-3)!}{d(d-1)} \omega_{ij}^{\alpha_1 - 1} (1 - \omega_{ij})^{\alpha_2 - 1} \frac{\Gamma(2\beta_{ij})}{\Gamma^2(\beta_{ij})} \omega_{i/ij}^{\beta_{ij} - 1} \omega_{j/ij}^{\beta_{ij} - 1},$$

where

$$\omega_{ij} = \omega_i + \omega_j, \ \omega_{i/ij} = \frac{\omega_i}{\omega_i + \omega_j} \text{ and } K_d(\alpha_1, \alpha_2) = \frac{1}{\mathrm{B}(\alpha_1 + 1, \alpha_2 + d - 3)}.$$

Similarly for the *left pairwise beta model* we get sum as in (3.9), where

$$h_{ij}(\boldsymbol{\omega} | \alpha_1, \beta_{i,j}) = K_d(\alpha_1) \frac{2(d-3)!}{d(d-1)} \omega_{ij}^{\alpha_1 - 1} \frac{\Gamma(2\beta_{ij})}{\Gamma^2(\beta_{ij})} \omega_{i/ij}^{\beta_{ij} - 1} \omega_{j/ij}^{\beta_{ij} - 1}, \tag{3.10}$$

where

$$\omega_{ij} = \omega_i + \omega_j, \ \omega_{i/ij} = \frac{\omega_i}{\omega_i + \omega_j} \text{ and } K_d(\alpha_1) = \frac{1}{\mathrm{B}(\alpha_1 + 1, d - 2)}.$$

For the *right pairwise beta model* we have

$$h_{ij}(\boldsymbol{\omega} | \alpha_2, \beta_{i,j}) = K_d(\alpha_2) \frac{2(d-3)!}{d(d-1)} (1 - \omega_{ij})^{\alpha_2 - 1} \frac{\Gamma(2\beta_{ij})}{\Gamma^2(\beta_{ij})} \omega_{i/ij}^{\beta_{ij} - 1} \omega_{j/ij}^{\beta_{ij} - 1}, \tag{3.11}$$

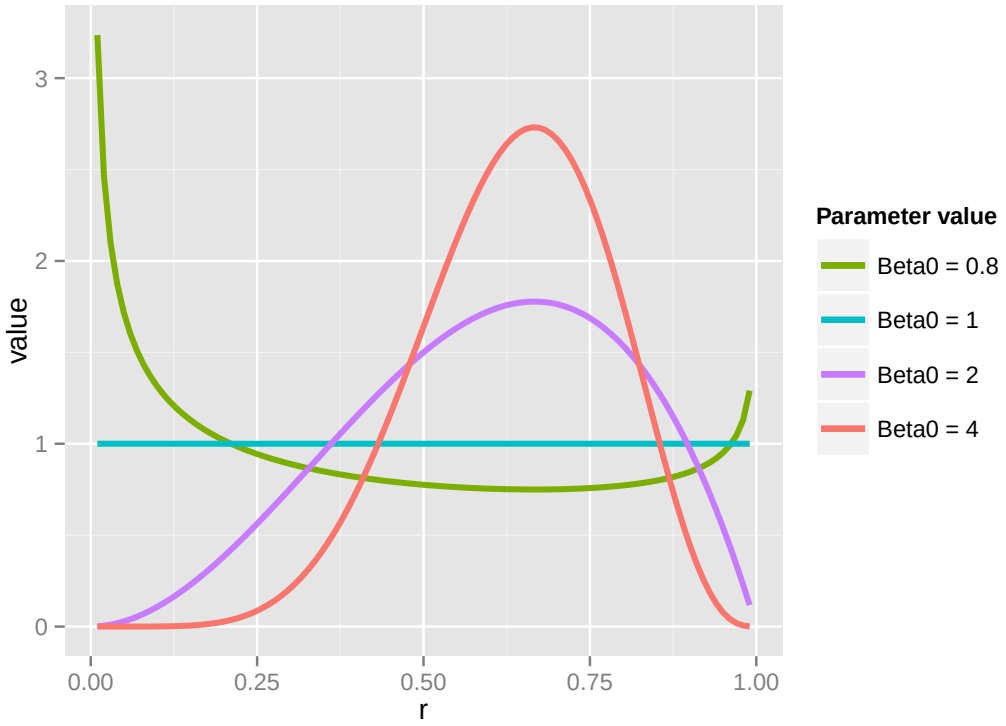


Figure 3.3: Beta density with parameters $(2\beta_0 - 1, \beta_0)$

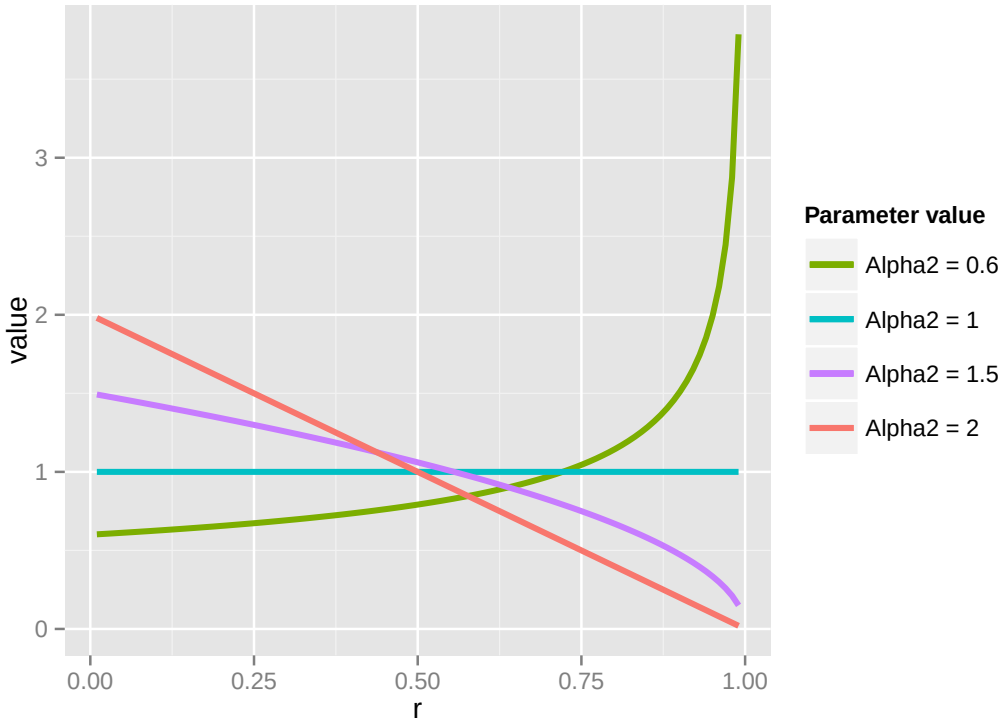


Figure 3.4: Beta density with parameters $(1, \alpha_2)$

where

$$\omega_{ij} = \omega_i + \omega_j,\ \omega_{i/ij} = \frac{\omega_i}{\omega_i + \omega_j} \text{ and } K_d(\alpha_2) = \frac{1}{\mathrm{B}(2, \alpha_2 + d - 3)}.$$

The double, the left and *the right* pairwise beta models share many properties with the original pairwise beta model. The exponent measure $V(\mathbf{x})$ can be obtained only via numerical integration of (2.17). Models do not exhibit the exponent measure closure; disadvantages of an additive model are still present. Simulation from these models can be done as in the original pairwise beta model; one just sets R_{ij} parameters according to the chosen model.

Clearly, the interpretation of parameters is different. In the original pairwise beta model the mean was constant. The choice of β_0 affected only variance. With a high variance the distribution was less distributed around the centre of mass of a simplex, meaning a lower overall dependence.

Now, the obtained distribution function more reflects the way how observations are truly distributed. The level of association is not that obvious as in the original model, but conclusions based on a value of the mean and the variance can provide in-depth diagnostics.

One can calculate the mean of the distribution as

$$\frac{\alpha_1}{\alpha_1 + \alpha_2}.$$

The closer is the mean to the simplex centre of mass k/d, the higher is the overall dependence between parameters. If the mean is higher than k/d this implies that there is a high overall dependence of a *subset* of variables, but extremes typically do not occur jointly in all stations. If the mean is lower then k/d it indicates that there is often only an extreme event only in one of stations.

Interpretation of variance

$$\frac{\alpha_1\alpha_2}{(\alpha_1 + \alpha_2)^2(\alpha_1 + \alpha_2 + 1)}$$

is the same as in the original model – higher variance indicates lower overall dependence.

During our research we have considered also other shapes of density q. For example, we considered $q(r)$ as a linearly increasing or decreasing up to $c \in (0, 1]$ and then remaining constant. The exact form of the spectral density was then obtained via Corollary 1. These models were able to outperform the original pairwise beta model. However, they seemed dull in comparison with the *double, left or right* pairwise beta model. As these models did not bring a fundamental improvement or idea and as we view the beta distribution with arbitrary parameters as a better and more flexible way to model dependence, we have not included these models.

3.6 Bayesian model selection

Several parametric models have been introduced in this chapter. All of them provide a way how to fulfil the constraint (2.12). As constraint (2.12) does not

imply any parametric form, the list of models is not complete. But even considering only models that were introduced so far, it seems like a Herculean task to provide guidelines how to choose between the models based on something else than their performance.

Provided models are predictive in their nature. In practice, the criterion is how well models fit available data. Methods that have the same number of parameters can be compared via maximum likelihood, i.e maximizing function

$$\sum_i \log(h_i(\boldsymbol{\omega})).$$

Comparing models with different number of parameters is more difficult, measures such as Akaike information criterion (AIC) or Bayesian information criterion (BIC) that penalize the number of parameters are used.

Each model fitted a different shape of the parameter function. Hence, it might be better not to choose one model, but to blend models together. Combining models not only provides a shape of a function that could not be obtained by each individual model, but blended models can also provide better predictions. What exactly is suggested to be blended?

If a random vector $\mathbf{M_j}$ follows a max-stable distribution $G_j(\mathbf{x}) = \exp(-V(\mathbf{x}))$ with unit Fréchet margins, then the averaged distribution

$$G(\mathbf{x}) = p_1 G_1(\mathbf{x}) + \cdots + p_J G_J(\mathbf{x}),\ \sum_{j=1}^{J} p_j = 1$$

in general is not max-stable anymore. However, averaging spectral measures

$$p_1 H_1(\cdot) + \cdots + p_J H_J(\cdot),\ \sum_{j=1}^{J} p_j = 1$$

proves a valid spectral measure. Clearly, as each spectral density satisfies the constraint (2.12), the constraint (2.12) holds also for the averaged measure.

One would like to find a mixture of spectral measures that is based on their performance – in the Bayesian setting it is typically the Bayesian Information Criterion (BIC). A tool for that provides *Bayesian model averaging* applied by Sabourin et al. (2013). Bayesian model averaging can recommend a mixture of spectral densities or if enough data is provided and differences between models are high, it can serve as a selection tool. Below is given a brief summary of Bayesian model averaging in the context of spectral densities. For general information about Bayesian model averaging the reader is refered to Hoeting et al. (1999) or Marin and Robert (2007), the latter concentrated more on Bayesian decision making.

Now, suppose that we have M spectral density models such that each model

$$\mathcal{M}_m = \{h_m(\cdot|\theta_m), \theta_m \in \Theta_m\}$$

has a finite dimensional parameter space Θ_m. Let the statistical model $\tilde{\mathcal{M}}$ be a disjoint union of individual models. This means that the parameter space $\tilde{\Theta}$ is the disjoint union

$$\tilde{\Theta} = \bigsqcup_{m=1}^{M} \Theta_m.$$

Each parameter space Θ_m is endowed with a *prior* probability measure π_m. Measure π_m expresses the statistician's beliefs about θ_m e.g. arising from expert knowledge. In the Bayesian model averaging framework there is also needed an *a priori weight* p_m of each model $\mathcal{M}_m$. The set of weights

$$(p_1, \ldots, p_M), \ \sum_{i=1}^{M} p_i = 1$$

is typically set uniform

$$p_m = \frac{1}{m}, \ m = 1, \ldots, M$$

for the lack of expert knowledge. Putting this together, on $\tilde{\mathcal{M}}$ we get prior distribution

$$\tilde{\pi}\left(\bigsqcup_{m=1}^{M} B_m\right) = \sum_{m=1}^{M} p_m \pi_m(B_m),$$

for any collection of measurable sets

$$(B_1, \ldots, B_M), \ B_m \subset \Theta_m, \ m = 1, \ldots, M.$$

In Chapter 4 it is described how a sample of excesses

$$\mathbf{W} = (\mathbf{W_1}, \ldots, \mathbf{W_n})$$

is obtained. Common density estimator, a *Posterior predictive density* averages density estimates produced in separate bayesian models. In the disjoint union model it is defined as

$$\tilde{h}(\boldsymbol{\omega}|\mathbf{W}) = \sum_{m=1}^{M} p_m(\mathbf{W}) \int_{\Theta_m} h_m(\boldsymbol{\omega}|\theta_m) \ \mathrm{d}(\pi_m|\mathbf{W})(\theta_m), \tag{3.12}$$

where

$$\pi_m|\mathbf{W}$$

is the posterior distribution restricted to $\mathcal{M}_m$ and $p_m(\mathbf{W})$ is the posterior weight of $\mathcal{M}_m$.

A posterior weight $p_m(\mathbf{W})$ is proportional to a *marginal* likelihood $\mathcal{L}_m(\mathbf{W})$ of an observed angular sample in each model $\mathcal{M}_m$ multiplied by the corresponding prior model weight:

$$p_m(\mathbf{W}) = \frac{p_m \mathcal{L}_m(\mathbf{W})}{p_1 \mathcal{L}_1(\mathbf{W}) + \cdots + p_M \mathcal{L}_M(\mathbf{W})},$$

where

$$\mathcal{L}_m(\mathbf{W}) = \int_{\Theta_m} h_m(\mathbf{W}|\theta_m) \ \mathrm{d}(\pi_m)(\theta_m). \tag{3.13}$$

In practice, for high dimensional parameter spaces, the main hurdle lies in estimating the parameter (3.13). It is done either by Monte-Carlo methods or asymptotic approximations from which the BIC is derived, see Kass and Raftery (1995).

Inside each single model, a Metropolis-Hastings algorithm is used to produce a posterior sample $\theta_{m,1}, \ldots, \theta_{m,N}$. This sample is used to approximate each term

$$\tilde{h}_m(\boldsymbol{\omega}) = \int_{\Theta_m} h_m(\boldsymbol{\omega}|\theta_m) \, \mathrm{d}(\pi_m|\mathbf{W})(\theta_m)$$

from (3.12) through

$$\tilde{h}_m(\boldsymbol{\omega}) = \frac{1}{N} \sum_{t=1}^{N} h_m(\omega|\theta_{m,t}).$$

In the Chapter 4 it is made explicit how Bayesian model averaging was used for modelling.

Chapter 4

Hydrological data application

In this chapter the gathered theory is used to build multivariate extreme value models for hydrological data. First, the modelling technique is described. Then data description follows along with models comparison. All models were build using the statistical software R and can be find on the attached CD or downloaded from `http://artax.karlin.mff.cuni.cz/~drapall/multivariate/`.

Data used for modelling consist of daily precipitation values from nine hydrological stations in northern Moravia, Czech Republic.

4.1 Modelling

The first step in the modelling is the estimation of marginal distribution in order that it can be transformed to unit Fréchet. Above a subjectively chosen threshold it is assumed that the inference from Theorem 5 holds and that an excess can be modelled by the generalized Pareto distribution. For observations below this threshold it is assumed that the empirical distribution function provides a good estimate. This means that observations

$$X_{ij},\ i = 1, \ldots, n,\ j = 1, \ldots, d$$

are transformed to unit Fréchet for each $j = 1, \ldots, d$ via

$$Z_{ij} = \begin{cases} \left[\log\left(\tilde{F}_j(X_{ij})\right)\right]^{-1} & \text{for } X_{ij} \leq q_j(0.95), \\ \left[\log\left(1 - 0.95\left(1 + \hat{\xi}_j \dfrac{X_{ij} - q_j(0.95)}{\hat{\sigma}_j}\right)_+^{-1/\hat{\xi}_j}\right)\right]^{-1} & \text{for } X_{ij} \geq q_j(0.95), \end{cases} \tag{4.1}$$

where $\tilde{F}_j$ is the empirical distribution function of

$$X_{ij},\ i = 1, \ldots, n,$$

$q_j(0.95)$ is its 95% quantile and $\hat{\xi}_j, \hat{\sigma}_j$ are respectively estimated scale and shape parameter of the generalized Pareto distribution estimated from top 5% of $X_{ij},\ i = 1, \ldots, n$ via the maximum likelihood technique.

In our model we picked a subset of stations of size $p = 3$ to be used for modelling. The corresponding unit Fréchet variables

$$Z_{ij},\ j = 1, \ldots, p$$

are transformed into a radial component R_i and angular components ω_{ij}:

$$R_i = \sum_{j=1}^{p} Z_{ij},\ \omega_{ij} = \frac{Z_{ij}}{R_i}, \quad j = 1, \ldots, p,\ i = 1, \ldots n.$$

Only those angular components ω_{ij} for which the corresponding radial component R_i exceeds a certain threshold r_0 are kept. The mean residual excess plot is used to choose threshold r_0. From these data coefficients of models were estimated via maximum likelihood approach.

There are two families of models – the pairwise beta family and the nested logistic model (3.2) – for which the Bayesian model averaging was applied. None of these models is part of the exponential family. Hence, there are no obvious uninformative or invariant prior choices. As we would like to use the normal distribution as a prior, the parameter space is transformed to obtain an unconstrained one. For the pairwise beta model the log transform is used

$$\left(\beta_0^{'}, \{\beta_{ij}\}_{1\leq i<j\leq p}^{'}\right) = \left(\log(\beta_0), \{\log(\beta_{ij})\}_{1\leq i<j\leq p}\right).$$

For the nested logistic model the logit transform was chosen

$$\left(\alpha_0^{'}, \{\alpha_{ij}\}_{1\leq i<j\leq p}^{'}\right) = \left(\log\left(\frac{\alpha_0}{1-\alpha_0}\right), \left\{\log\left(\frac{\alpha_{ij}}{1-\alpha_{ij}}\right)\right\}_{1\leq i<j\leq p}\right).$$

Then, parameters in each model are assumed to be normally distributed with mean 0 and standard deviation 3 and *a priori* mutually independent. Sabourin et al. (2013) showed on simulated data that this prior specification does not introduce a strong bias in estimates.

4.2 Exploratory analysis

The data of daily precipitation from nine hydrological stations were provided by Czech Hydro-Meteorolgoical Institute and used previously by Jarušková (2009) whose approach is followed. The nine stations are Heřmanovice (HE), Karlovice (KA), Krnov (KR), Lichnov (LI), Opava (OP), Praděd (PR), Rejvíz (RE), Vidly (VI) and Albrechtice - Žáry (ZI) and their location is depicted on Figure 4.1 from Jarušková (2009).

Data were measured from Janurary 1, 1960 to February 6, 2005. Records have missing values, ranging from over a month to almost 7 and half year in total. This is more than in the dataset used by Jarušková (2009) – unfortunately, we were unable to obtain the exact dataset.

To get a basic idea about the data, let us count the length of records for each site, number of days with positive precipitation, quantiles of these and attained maximum, see table 4.1.

A basic measure of dependence is correlation. The table 4.2 shows correlation coefficients between all pairs of sites for days with positive precipitation. As for each pair were considered days with positive precipitation on these two sites, correlation coefficients are based on different sets of data. Sites with correlation coefficient greater than 0.7 are connected with a line on Figure 4.1.

Station	No. of observations	No. of positive observations	0.90 quantile	0.95 quantile	0.99 quantile	Maximum
HE	15221	6911	13.2	20.3	41.2	196.5
KA	16407	6547	12.7	17.8	31.0	124.2
KR	16194	6304	11.3	16.8	30.2	59.2
LI	16437	6581	11.2	16.3	28.8	110.0
OP	16108	6419	11.0	16.2	31.1	62.0
PR	13758	7368	14.1	20.2	37.8	139.4
RE	14730	6459	15.8	23.6	43.3	214.2
VI	15098	7934	15.2	21.4	40.5	199.3
ZY	16347	7962	11.2	16.8	30.4	125.0

Table 4.1: Number of all observations, number of positive observations, 0.90, 0.95 and 0.99 quantile of positive observations and maximum for each site

	KA	KR	LI	OP	PR	RE	VI	ZY
HE	0.70	0.59	0.57	0.56	0.68	0.83	0.75	0.74
KA		0.66	0.65	0.59	0.68	0.70	0.73	0.78
KR			0.76	0.74	0.55	0.61	0.57	0.78
LI				0.72	0.50	0.56	0.55	0.70
OP					0.50	0.57	0.55	0.67
PR						0.69	0.81	0.68
RE							0.73	0.76
VI								0.72

Table 4.2: Correlation coefficients between pairs of sites for days with positive precipitation.

A subject of our interest is the dependence between extreme events which may or may not coincide with the general dependence. Hence, we are interested in extremal dependence indices $\chi(u)$ from (2.20). Table 4.3 shows indices $\chi(0.975)$ calculated for each pair of stations using function `taildep` from `R` package `extRemes` available from CRAN. Quantile 0.975 was chosen as it is high enough and it still provides enough data to calculate from. However, it still measures association at moderate extreme level not the limiting behaviour itself.

In Chapter 2 it has been mentioned that plotting $\chi(u)$ for values of u close to 1 can serve as a diagnostic tool of asymptotic dependence. Later on, stations HE, ZY and LI will be included into a single model. Therefore we have chosen to plot $\chi(u)$ for these three stations. Figure 4.2 depicts values of $\chi(u)$ for different quantiles for stations HE and ZY. The value of $\chi(u)$ is decreasing. However, it seems that as $u \to 1$ it would still be above 0. A similar behaviour exhibits the pair HE and LI.

On the other hand, the pair LI and ZY exhibits quite different behaviour as shown on Figure 4.3. Here, $\chi(u)$ reaches 0 which corresponds to asymptotic independence. It is shown later that this pair of stations exhibits the highest association of the three stations.

Estimator $\chi(u)$ has the disadvantage that only values of u very close to 1

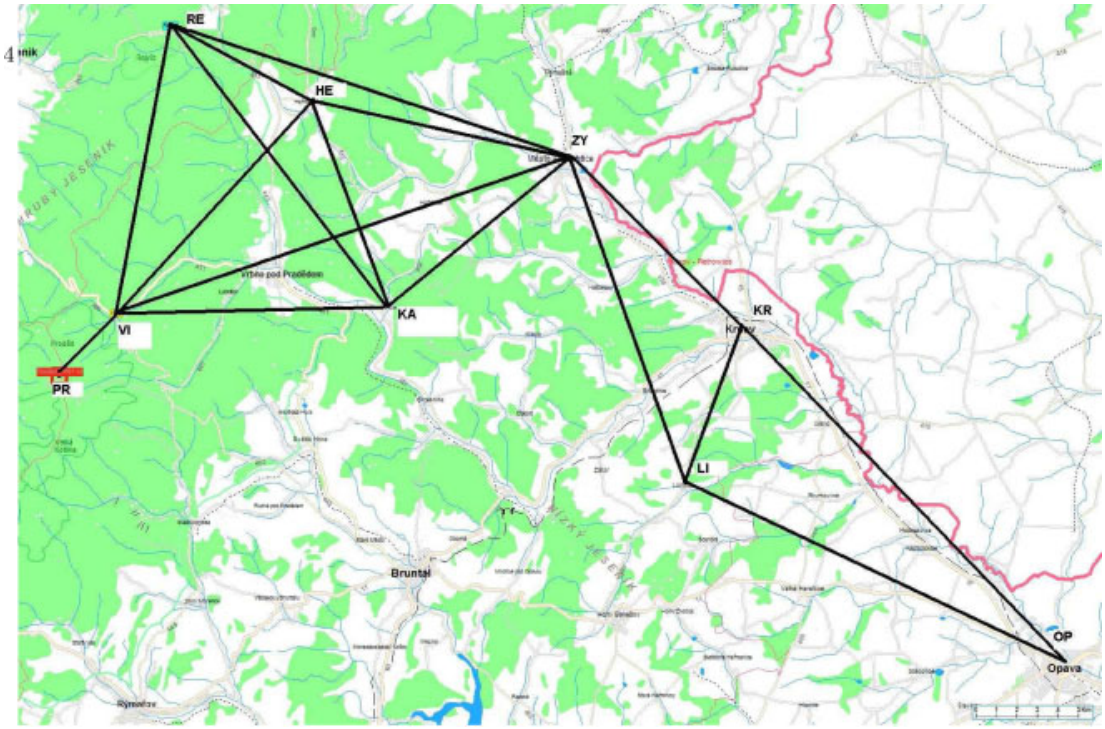


Figure 4.1: The location of the nine Moravian hydrological stations.

	KA	KR	LI	OP	PR	RE	VI	ZY
HE	0.52	0.44	0.40	0.42	0.48	0.67	0.55	0.56
KA		0.48	0.49	0.43	0.46	0.51	0.52	0.60
KR			0.63	0.57	0.38	0.46	0.42	0.60
LI				0.59	0.35	0.42	0.41	0.55
OP					0.36	0.41	0.38	0.51
PR						0.47	0.57	0.46
RE							0.50	0.57
VI								0.51

Table 4.3: Dependence parameter $\chi(0.975)$ for each of pair of sites

determine the asymptotic dependence. For such u the estimate is based on little data. The fact that two pairs of stations suggest different asymptotic behaviour means that it is not clear which models should be chosen – a reasoning both for the asymptotic dependence and the asymptotic independence can be found. Due to the described characteristics of estimator $\chi(u)$ this is quite common in the extreme value analysis. For example, for Leeds air pollution data (dataset `winterdat` in `R` package `BMAmevt`) Cooley et al. (2010) and Boldi and Davison (2007) suggested models of asymptotic dependence while Heffernan and Tawn (2004) study focused on conditional distribution that allows also independence at extreme levels.

It seems that the data could be a mixture of both asymptotically dependent and independent data. This could arise as a consequence of two types of extreme precipitation: one that occurs just locally and the other associated with a weather front. In this work models of asymptotic dependence are used as both approaches

seem applicable and the focus of the thesis are models of asymptotic dependence.

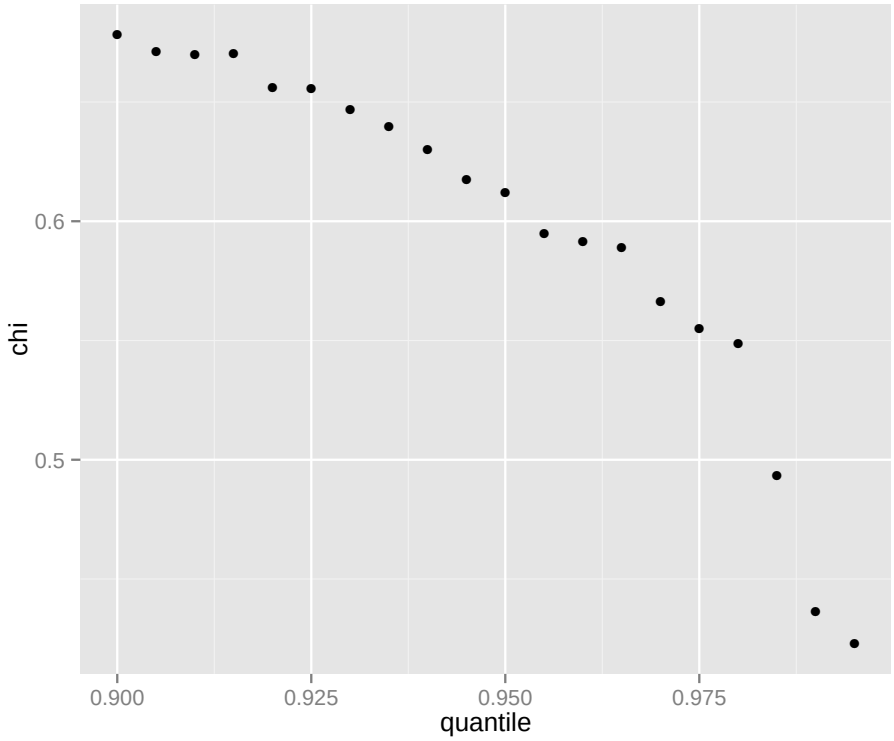


Figure 4.2: Estimator $\chi(u)$ at different quantiles for station HE and ZY

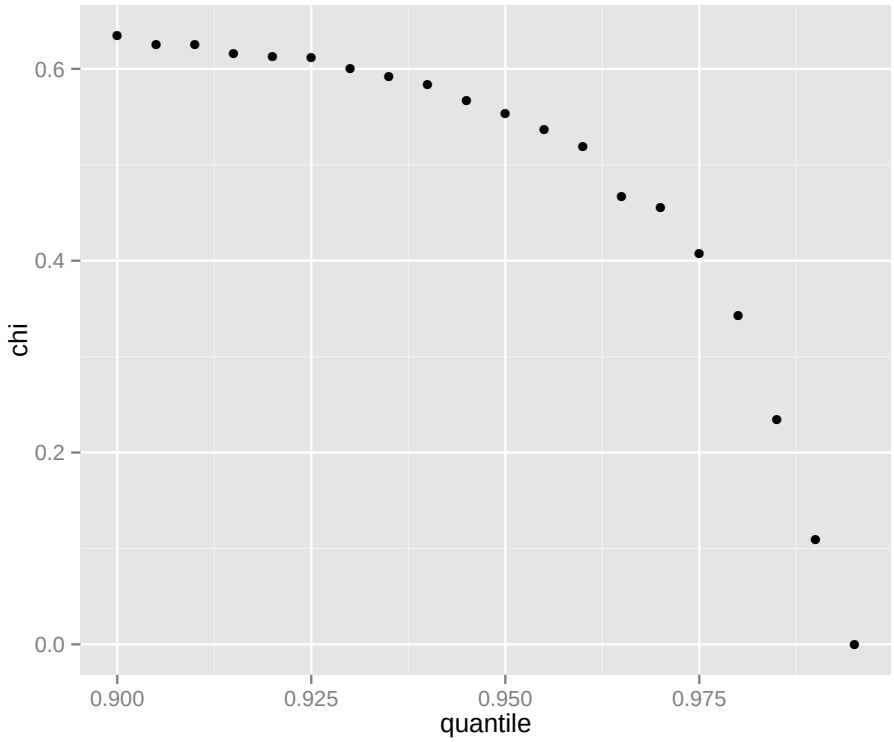


Figure 4.3: Estimator $\chi(u)$ at different quantiles for station LI and ZY

4.3 Transformed data

Estimated coefficients of Generalized Pareto distribution via maximum likelihood are shown in Table 4.4.

Station	Threshold	Scale	Shape
HE	12.40	9.18	0.24
KA	11.40	6.61	0.18
KR	9.70	7.78	-0.05
LI	9.50	7.49	0.04
OP	9.30	7.80	0.02
PR	14.61	8.43	0.22
RE	14.40	10.12	0.21
VI	15.60	8.80	0.20
ZY	11.00	7.90	0.08

Table 4.4: Threshold values and parameters of GPD fitted to excesses above threshold for each site

A trivariate model has been built for stations HE-ZY-LI as these represent the region: HE is in the highlands, LI is in the lowlands and ZY is located in the slope.

To chose a radial threshold, a mean residual excess plot was computed – see Figure 4.4. Function `mrl.plot` from `R` package `ismev` was used to produced the plot. The lines were added manually – red line shows the linear relationship between threshold excess and threshold, blue line states the threshold. This way threshold exp{4.2} is chosen as a minimum value for which a linear relationship holds.

The angular components for those observations whose radial component is higher then the threshold are shown in Figure 4.5. The scatter plot 4.5 shows that there is a strong tail dependence between stations ZY and LI and a little weaker tail dependence between stations HE and ZY. It also seems that there is almost none tail association between stations HE and LI. The observed pair dependences corresponds to conclusions based on the location of stations.

4.4 Models comparison

There were two families of models fitted to data from stations HE, ZY and LI – nested logistic (3.2) and pairwise beta with its original version (3.7), double (3.9), right (3.11) and left (3.10) modifications. All of these models allow Bayesian model selection.

Table 4.5 shows obtained parameters for models from the pairwise beta family via the maximum likelihood technique and obtained log-likelihood

$$\sum_i \log(h_i(\boldsymbol{\omega})).$$

Function `maxLikelihood` from package `BMAmevt` was used to compute these. It calls function `optim` from `stats` package with method `"L-BFGS-B"` which is a

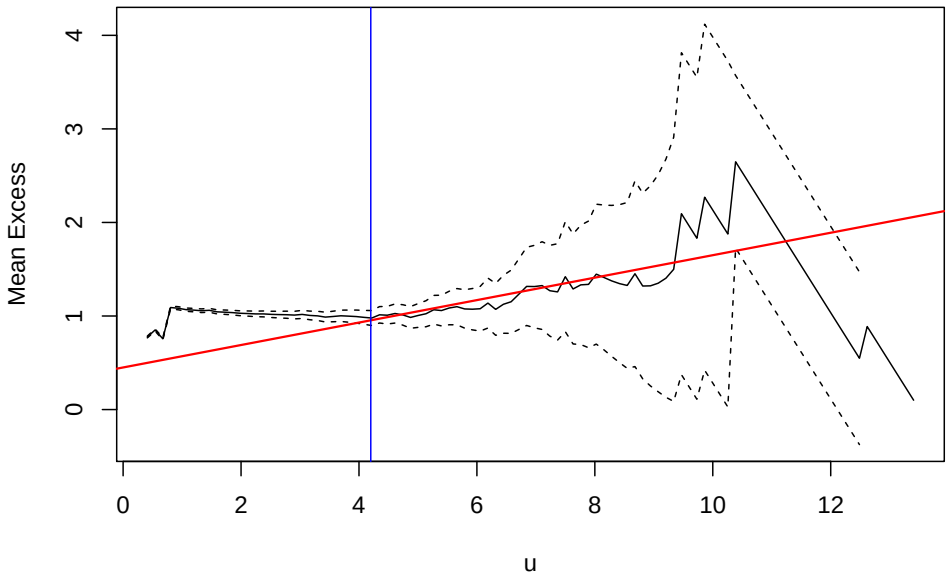


Figure 4.4: Mean residual plot of excesses for radial component of stations HE, ZY and LI on logarithmic scale

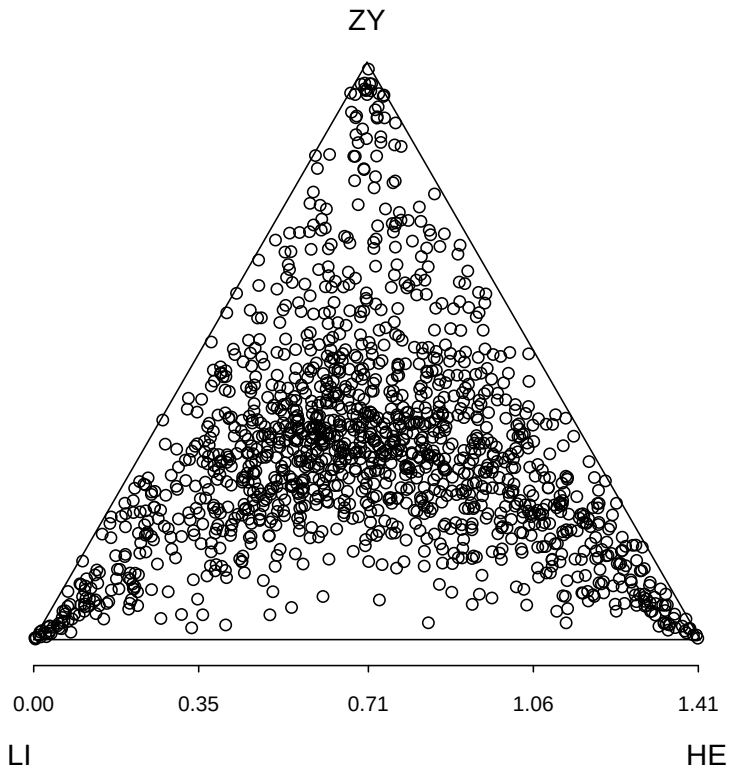


Figure 4.5: Angular components of excesses above threshold for stations HE, ZY and LI

quasi-Newton method. As initial parameters can affect obtained parameters, multiple runs were performed to obtain truly the best set of parameters. Table 4.6 shows obtained parameters and likelihood for the nested logistic model (3.2).

model	α_1	α_2	β_0	$\beta_{HE,ZY}$	$\beta_{HE,LI}$	$\beta_{LI,ZY}$	log-likelihood
Original	—	—	1.03	1.06	0.56	5.04	546.7
Double	0.20	1.11	—	3.76	0.92	4.30	644.5
Left	0.14	1	—	3.84	0.93	4.40	643.4
Right	1	2.15	—	3.10	0.93	3.47	597.9

Table 4.5: Parameters of variants of pairwise beta model and their log-likelihood

model	α_0	$\alpha_{HE,ZY}$	$\alpha_{HE,LI}$	$\alpha_{LI,ZY}$	log-likelihood
Nested Logistic	0.60	0.79	1.00	0.77	664.2

Table 4.6: Nested logistic model parameters and its log-likelihood

Comparing obtained log-likelihoods from table 4.5 proofs that new pairwise beta models bring a substantial improvement. The log-likelihood of the nested logistic model from table 4.6 makes it still preferable to pairwise beta models. However, unlike likelihood of the original pairwise beta model, the log-likelihood of new models is comparable to the one of nested logistic model.

It is not only the log-likelihood that makes new models preferable to the original pairwise beta model. The original beta distribution's beta density $q(r)$ depending on β_0 is not flexible enough to reflect that lots of observations are located near a simplex vertex. In order to compensate for this, it estimates coefficient $\beta_{HE,LI}$ very low so that it can allocate enough mass to these points. Thus, it underestimates the dependence between HE and LI. On the other hand, it overestimates the coefficient $\beta_{LI,ZY}$ compared to the coefficient $\beta_{HE,ZY}$. The dependence between LI and ZY is bigger than between HE and ZY. However, as Figure 4.5 shows and estimated coefficients of the nested logistic model suggest, dependence between these two pairs of stations is of similar order.

New pairwise beta models are able to capture that dependence between LI, ZY is only a little higher compared to dependence between HE and ZY. Coefficient $\beta_{HE,LI}$ does not overly underestimate dependence between HE and LI. Recall that an estimate of a parameter depends on other parameter values. To conclude: set of parameters of new pairwise beta models seem to better reflect dependence between stations.

If one insists that a four parameter model should be used, picking the better of the right and the left pairwise beta model could be an option measured merely by the log-likelihood. However, as we will see later, for making predictions it is better to use the double pairwise beta model. The left pairwise beta model's log-likelihood is very close to the one of the double pairwise beta model as estimated α_2 in double pairwise beta model is close to 1. Although $\alpha_1 = 1$ is not an ideal choice in this problem, the right beta model still clearly outperforms the original one.

On Figures 4.6, 4.7, 4.8, 4.9 and 4.10 are shown contour plots of the original, the double, the left and the right pairwise beta model and of the nested logistic model, respectively. The inflexibility of the original pairwise beta model is obvious. Plots for the left and the double pairwise beta model are very similar as so they are their parameters. The right pairwise beta model exhibits the same structure as these two.

To compare models the Bayesian model selection procedure was applied. Using function `posteriorWeights` from `R` package `BMAmevt` we have compared nested logistic separately with pairwise beta models that have the same number of parameters: with the left, the right and the original model. In all cases almost entire weight was assigned to the nested logistic model. The order assigned to a pairwise beta model differed: ranging from 10^{-9} for the left pairwise beta model to 10^{-54} for the original pairwise beta model. We conclude that the nested logistic model should be used to model this dataset.

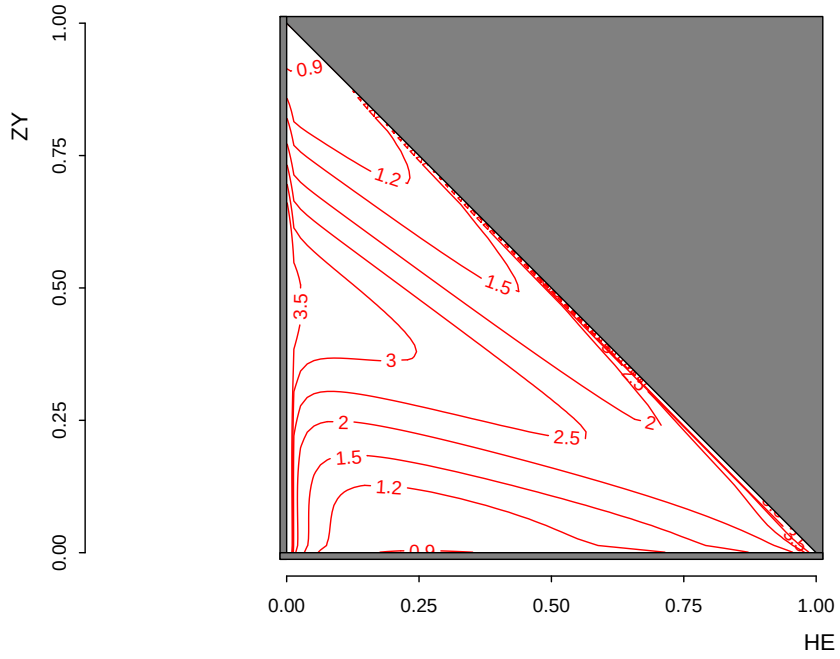


Figure 4.6: Contour plot of original pairwise beta model for stations HE, ZY and LI

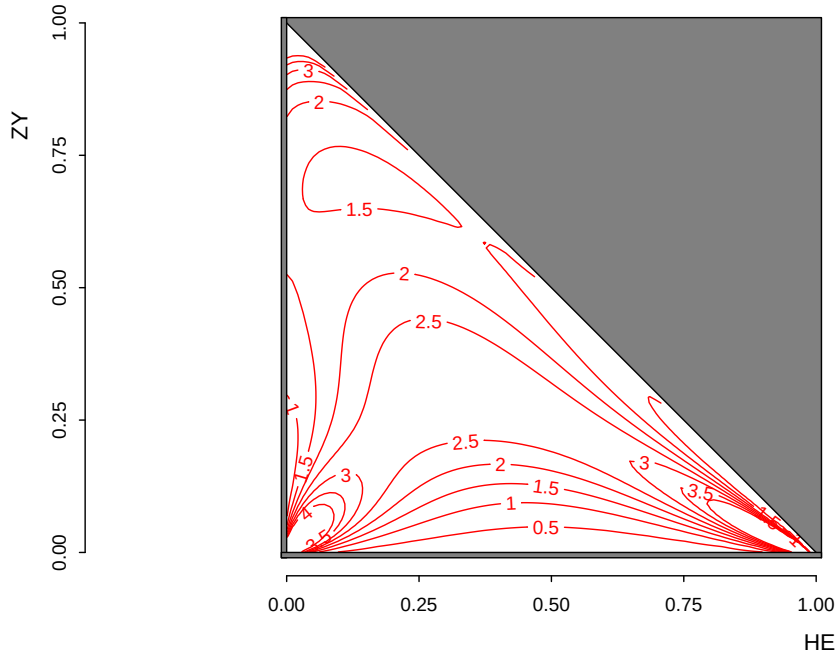


Figure 4.7: Contour plot of double pairwise beta model for stations HE, ZY and LI

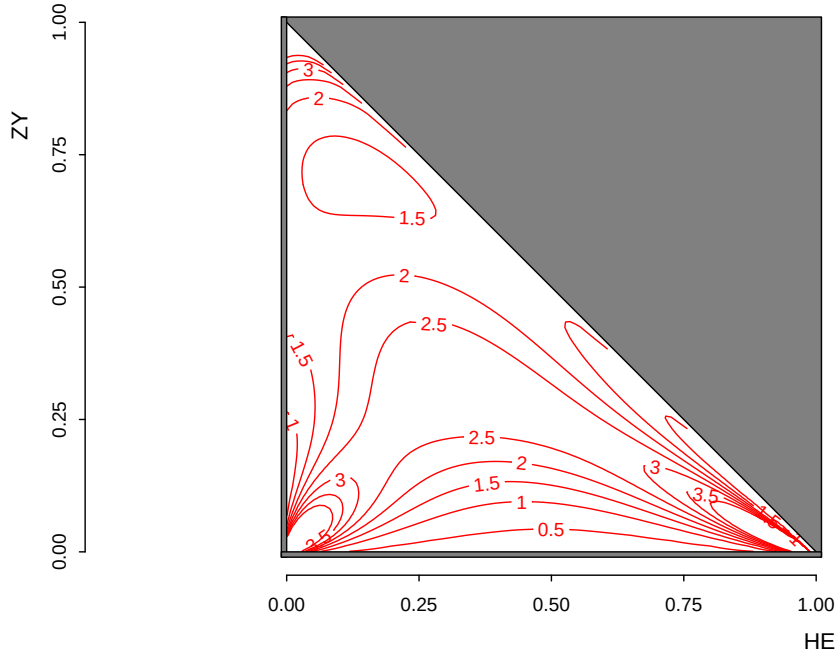


Figure 4.8: Contour plot of left pairwise beta model for stations HE, ZY and LI

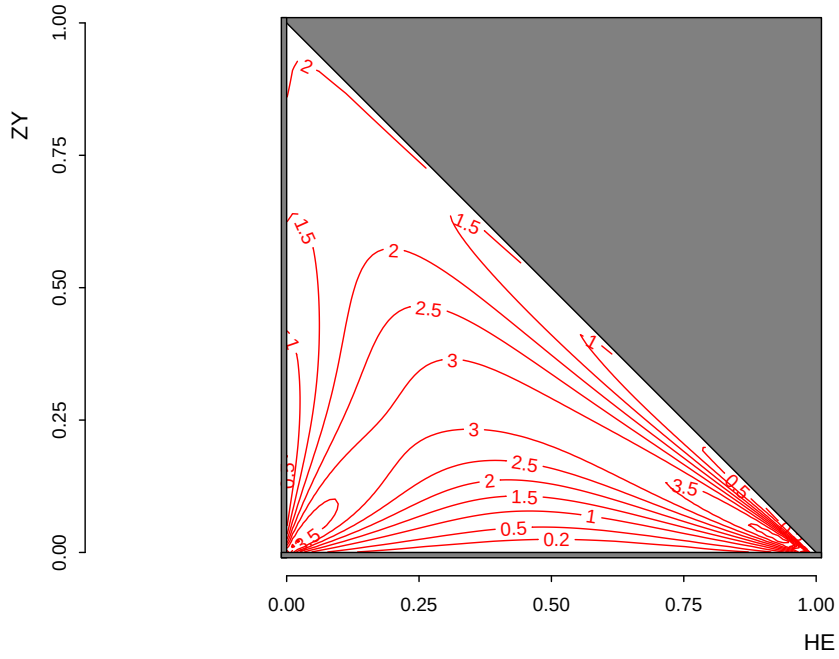


Figure 4.9: Contour plot of right pairwise beta model for stations HE, ZY and LI

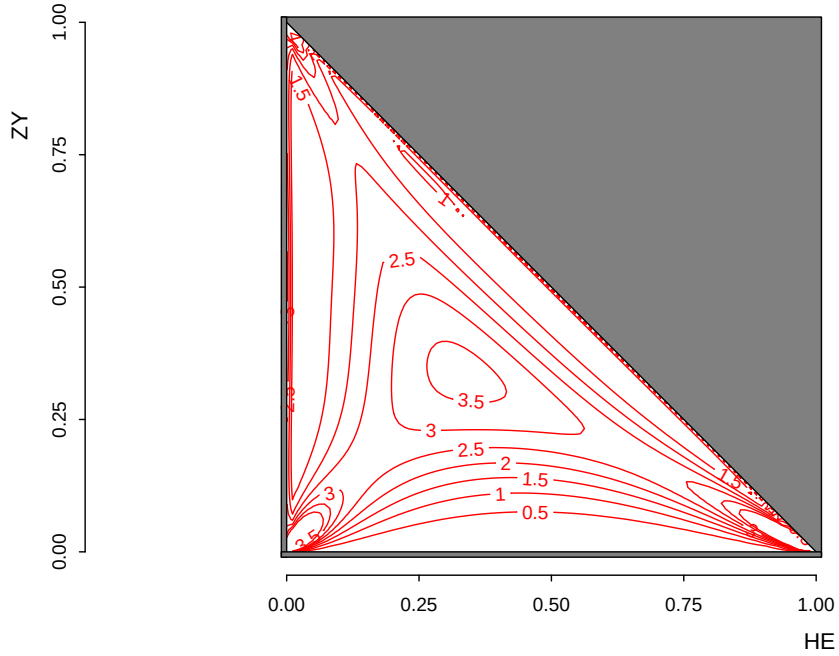


Figure 4.10: Contour plot of nested logistic model for stations HE, ZY and LI

Suppose that we would like to estimate exceedance probabilities, i.e. values of survival function, for large arguments x_1, x_2, x_3. For X_1, X_2 and X_3 transformed to unit Fréchet variables Z_1, Z_2, Z_3 via (4.1) it approximately holds

$$P(X_1 < x_1 \cap X_2 < x_2 \cap X_3 < x_3) \approx P(Z_1 < z_1 \cap Z_2 < z_2 \cap Z_3 < z_3)$$

where x_1, x_2, x_3 were transformed to z_1, z_2, z_3 via (4.1) as well. The nested logistic model has a closed form of the exponent measure $V(\mathbf{x})$. Thus $V(\mathbf{x})$ is used to estimate exceedance probabilities. The pairwise beta family does not share such a property. Hence, we have generated samples from a pairwise beta model and used Monte Carlo approximation. Table 4.7 shows calculated exceedance probabilities. Values x_1, x_2, x_3 have been chosen the same as in Jarušková (2009). However, data used in our analysis are slightly different. Also a different radial threshold was picked. There has been added a three parameters pairwise beta distribution from Jarušková (2009) for reference. This is the original pairwise beta with fixed $\beta_0 = 1$. The three parameters are those estimated by Jarušková (2009).

x_1	x_2	x_3	Nested logistic	Double PB	Left PB	Right PB	Original PB	Three parameters PB
75.00	83.30	34.50	48.72	21.25	16.97	41.77	48.57	49.47
108.90	50.90	51.10	141.23	46.47	37.29	104.57	127.55	139.45
196.50	125.00	51.90	3.15	1.25	1.02	2.63	3.13	3.25
133.30	57.00	20.00	104.54	37.46	29.90	81.16	100.85	112.68
80.90	57.70	110.00	1.35	0.92	0.76	1.35	1.36	1.35
53.70	25.50	61.00	114.07	57.16	45.04	107.36	113.32	116.49

Table 4.7: Three-dimensional survival function for nested logistic and pairwise beta models for HE-ZY-LI. All probabilities have to be multiplied by 10^{-6}

As the three parameters pairwise beta model gives similar predictions, it seems that the choice of the threshold has not much affected parameters estimation. All models of Jarušková (2009) gave similar predictions. However, this is not the case here. None of those models was flexible enough to assign weight both to observations near the centre of mass of the simplex and to those near a simplex vertex. New pairwise beta models do this which results in decreased estimates of exceedance probabilities. However, the question whether the double pairwise beta model underestimates or other models overestimate excess probability is open. Exceedance probabilities of the left, the right and the double pairwise beta model differ quite a bit. Imposition of fixed parameter can have a great influence on the final structure. Thus, we recommend to use the double pairwise beta model.

It has been noted that data may be a mixture of asymptotically independent and asymptotically dependent data. To obtain the most reliable estimate of excess probabilities, we would suggest to model the data as such a mixture.

Bibliography

F. Ballani and M. Schlather. A construction principle for multivariate extreme value distributions. *Biometrika*, 98(3):633–645, 2011.

M. Boldi. *Mixture models for multivariate extremes*. PhD thesis, Lausanne, 2004.

M. Boldi and A. C. Davison. A mixture model for multivariate extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):217–229, 2007.

J. Chambers, C. Mallows, and B. Struck. A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(354):340–344, 1976.

S. Coles. *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.

S. Coles, J. Heffernan, and J. Tawn. Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365, 1999.

S. G. Coles and J. A. Tawn. Modelling extreme multivariate events. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 377–392, 1991.

D. Cooley, R. A. Davis, and P. Naveau. The pairwise beta distribution: A flexible parametric multivariate model for extremes. *Journal of Multivariate Analysis*, 101(9):2103–2117, 2010.

S. R. Dalal and G. J. Hall Jr. On approximating parametric bayes models by nonparametric bayes models. *The Annals of Statistics*, pages 664–672, 1980.

L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer, 2006. ISBN 9780387239460.

J. Galambos. *The asymptotic theory of extreme order statistics*. R.E. Krieger Pub. Co., 1987. ISBN 9780898749571.

B. Gnedenko. Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *The Annals of Mathematics*, 44(3):423+, 1943.

E. J. Gumbel. Bivariate exponential distributions. *Journal of the American Statistical Association*, 55(292):698–707, 1960.

J. E. Heffernan and J. A. Tawn. A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):497–546, 2004.

J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. Bayesian model averaging: a tutorial. *Statistical science*, pages 382–401, 1999.

D. Jarušková. Modeling multivariate extremes of precipitation series in northern moravia. *Environmetrics*, 20(6):751–775, 2009.

H. Joe. Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistics & probability letters*, 9(1):75–81, 1990.

R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.

J. M. Marin and C. Robert. *Bayesian Core: A Practical Approach to Computational Bayesian Statistics: A Practical Approach to Computational Bayesian Statistics*. Springer, 2007.

J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131, 1975.

S. I. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Springer, 1987. ISBN 9783540964810.

A. Sabourin and P. Naveau. Bayesian dirichlet mixture model for multivariate extremes: A re-parametrization. *Computational Statistics & Data Analysis*, 71:542–567, 2014.

A. Sabourin, P. Naveau, and A. Fougres. Bayesian model averaging for multivariate extremes. *Extremes*, 16(3):325–350, 2013.

M. Schlather and J. A. Tawn. A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, 90(1):139–156, 2003.

J. Segers. Max-stable models for extremal dependence. *RevStat Statistical Journal*, 10:61–82, 2012.

M. Sibuya. Bivariate extreme statistics, i. *Annals of the Institute of Statistical Mathematics*, 11(2):195–210, 1959.

A. Stephenson. Simulating multivariate extreme value distributions of logistic type. *Extremes*, 6(1):49–59, 2003.

J. A. Tawn. Modelling multivariate extreme value distributions. *Biometrika*, 77 (2):245–253, 1990.