Classifying and Exploiting Structure in Multivariate Extremes

Emma Siobhan Simpson,

M.Math.(Hons.), M.Res.

Submitted for the degree of Doctor of Philosophy at Lancaster University.

May 2019

STOR-i
excellence with impact

Abstract

The aim of this thesis is to present novel contributions in multivariate extreme value analysis, with focus on extremal dependence properties. A set of random variables is often categorized as being asymptotically dependent or asymptotically independent, based on whether the largest values occur concurrently or separately across all members of the set. However, there may be a more complicated structure to the extremal dependence that is not fully described by this classification, with different subsets of the variables potentially taking their largest values simultaneously while the others are of smaller order. Knowledge of this detailed structure is essential, and aids efficient statistical inference for rare event modelling.

We propose a new set of indices, based on a regular variation assumption, that describe the extremal dependence structure, and present and compare a variety of inferential approaches that can be used to determine the structure in practice. The first approach involves truncation of the variables, while in the second we study the joint tail behaviour when subsets of variables decay at different rates. We also consider variables in terms of their radial-angular components, presenting one method based on a partition of the angular simplex, alongside two soft-thresholding approaches that incorporate the use of weights. The resulting estimated extremal dependence structures can be used for dimension reduction, and aid the choice or construction of appropriate extreme value models.

We also present an extensive analysis of the multivariate extremal dependence properties of vine copulas. These models are constructed from a series of bivariate copulas according to an underlying graphical structure, making them highly flexible and useful in moderate or even high dimensions. We focus our study on the coefficient of tail dependence, which we calculate for a variety of vine copula classes by applying and extending an existing geometric approach involving gauge functions. We offer new insights by presenting results for trivariate vine copulas constructed from bivariate extreme value and inverted extreme value copulas. We also present new theory for a class of higher dimensional vine copulas.

An approach for predicting precipitation extremes is presented, resulting from participation in a challenge at the 2017 EVA conference. We propose using a Bayesian hierarchical model with inference via Markov chain Monte Carlo methods and spatial interpolation.

Acknowledgements

First and foremost, I want to say a massive thank you to my PhD supervisors, Jenny Wadsworth and Jonathan Tawn. I couldn't have asked for better guidance over the last few years, and I've learnt so much from both of you. Thank you for your patience and wisdom, and all the time you've given to helping me develop as a researcher. You've really made the thesis-writing process a lot less daunting than it could have been.

Thank you to Ingrid Hobæk Haff for your input on the vine copula work. It was a pleasure to visit Oslo in the second year of my PhD and to have the opportunity to work with you.

I gratefully acknowledge funding from EPSRC through the STOR-i centre for doctoral training. I count myself as extremely lucky to have done my PhD as part of STOR-i. It's a really wonderful environment full of supportive people and friends. So thank you to everyone in the CDT, especially my cohort, who has inserted some fun into the last few years of hard work. Tuesday mornings wouldn't have been the same without extremes (and spatial statistics) coffee; thank you to everyone who contributed to the EVA paper, and provided interesting discussion (and snacks) over the last few years.

A special thank you to Emily for being a wonderful desk neighbour and friend,

and for all those calming words over cups of tea; to Lucy for your friendship and support since day one of the STOR-i internship; and to Kathryn, Emma and Chrissy for helping me to switch-off and relax with some crochet from time to time! Outside of STOR-i, I also want to thank Georgia, Jess and Tom for your years of friendship, and many weekends of fun and laughter; they're so often just what I need.

Thank you to my parents, Julia and Paul, for always encouraging curiosity and learning. Five-year-old Emma dreamt of doing 'lots of maths' when she grew up, and I think this thesis might count! That wouldn't have happened without all your support, for which I'm forever grateful. Also thank you to my sister Rebecca, for always being there for me, and letting me follow you to Lancaster in the first place!

Last, but by no means least, Ollie, thank you for being there every step of the way, to share the ups and downs of the last few years. I'm grateful to have had you by my side through all of this, and massively appreciate your ability to always put a smile on my face, even when life throws us challenges. I'm so lucky to be part of our team.

Declaration

I declare that the work in this thesis has been done by myself and has not been submitted elsewhere for the award of any other degree.

A version of Chapter 3 has been submitted for publication as Simpson, E. S., Wadsworth J. L. and Tawn J. A. (2019). Determining the dependence structure of multivariate extremes.

Chapter 6 is the result of a STOR-i collaboration for a competition at the 2017 conference on Extreme Value Analysis in Delft. It has been published as Barlow, A. M., Rohrbeck, C., Sharkey, P., Shooter, R. and Simpson, E. S. (2018). A Bayesian spatio-temporal model for precipitation extremes - STOR team contribution to the EVA2017 challenge. *Extremes*, 21(3):431-439.

The word count for this thesis is approximately 68,000.

Emma S. Simpson

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

There are many situations where it may be necessary to understand and model extreme values. Areas of interest range from environmental applications, such as high river flows or extreme weather events, to financial ones, like stock market movements. The occurrence of extreme events can have a huge impact, and a better understanding of the types of extremes we may expect can help us to mitigate their effect.

One of the main issues faced in modelling extreme values is the often limited data available; the events of interest may have never been observed, so standard statistical modelling techniques may not apply. We therefore require a range of asymptotically-motivated tools, which come from extreme value theory. A wide range of results are available for modelling extremes, including classical approaches for univariate and multivariate cases, as well as a variety of more recent advancements. There is increasing interest in more complicated situations involving multivariate extremes, and there is still work to be done to develop the theory and methodology necessary for these circumstances; we focus on two such cases in this thesis.

Many existing methods for modelling multivariate extremes are only applicable either if all of the variables take their largest values simultaneously, or if they occur separately in each one. However, in some scenarios it is possible for certain subsets of the variables to take their largest values simultaneously while the others are small, and models for the extreme values should reflect this. We aim to develop methods to determine these possibly complex extremal dependence structures, with the ultimate aim being to aid model selection.

We also consider the case where there is a specific underlying dependence structure among the variables. The example we focus on is vine copulas (Joe, 1996; Bedford and Cooke, 2001, 2002), which are a class of multivariate model constructed from a series of bivariate copulas, and whose dependence structure can be represented graphically. Owing to their flexibility, vine copulas have grown in popularity in recent years, and have the potential to be used for modelling multivariate extremes. We investigate how imposing the graphical dependence structure of vine copulas influences the tail dependence properties that can be captured by such models.

1.2 Outline of thesis

The overall aim of this thesis is to develop methods for assessing and modelling dependence in multivariate extremes. There already exists a rich literature in this area; we aim to build on this work with some novel approaches, which are outlined in this section.

Chapter 2 gives an overview of some of the techniques that can be used in extreme value modelling. We begin with a brief introduction to some of the standard methods for modelling univariate extremes, before reviewing existing methods for multivariate extreme value analysis, and finally introducing the concept of vine copulas.

In Chapter 3, we introduce methodology for determining the dependence structure of multivariate extremes; that is, the subsets of variables that can take their largest values simultaneously, while the others are of smaller order. Under a regular variation assumption, we present a new set of indices that reveal aspects of the extremal dependence structure not available through any existing measures of dependence. We derive theoretical properties of these indices, demonstrate their value through a series of examples, and develop inferential methods that also allow us to estimate the proportion of extremal mass associated with each subset of variables. We apply the methods to UK river flows, estimating the probabilities of different subsets of sites being large simultaneously.

In Chapter 4, we present alternative methods for determining extremal dependence structures; this time in a radial-angular setting. The first of these is a simplex-partitioning approach based on a regular variation assumption. This setting also enables us to introduce a soft-thresholding approach that allows the information from each observation to be shared across multiple faces. This is achieved by assigning weights to points in the angular simplex based on their proximity to the various faces. We also implement this soft-thresholding technique to extend the approach of Goix et al. (2016). We compare the simplex-partitioning method and both these weighted approaches to the methods in Chapter 3 via a simulation study, using receiver operating characteristic curves to test their performance as classifiers, and comparing Hellinger distances to assess the estimation of the proportion of extremal mass assigned to each face of the angular simplex. In several cases in this simulation study, these radial-angular methods show improvement over those in Chapter 3.

In Chapter 5, we investigate some of the tail dependence properties of vine copulas by calculating the coefficient of tail dependence η (Ledford and Tawn, 1996) for

certain sub-classes of this model. We focus on the trivariate case, and vine copulas constructed from extreme value and inverted extreme value pair copulas. We follow the approach of Nolde (2014), who investigates the limiting shape of suitably-scaled sample clouds to give a geometric interpretation of η, which allows for calculation of this coefficient from a joint density. By extending this theory, we propose a way to calculate η for multivariate margins when the joint density can only be expressed analytically for a higher order multivariate distribution. We also consider using numerical approximation for cases where η cannot be found analytically.

Chapter 6 was written following entry of a STOR-i team to the EVA2017 challenge. The aim of the challenge was to predict extreme precipitation quantiles across several sites in the Netherlands. Our proposed method uses a Bayesian hierarchical structure, and a combination of Gamma and generalized Pareto distributions. We impose a spatio-temporal structure in the model parameters via an autoregressive prior, and propose estimating model parameters using Markov chain Monte Carlo techniques and spatial interpolation. Our approach was successful in the context of the challenge, providing reasonable improvements over the benchmark method in terms of the quantile loss function of Koenker (2005).

In Chapter 7, we summarize the contributions of this thesis, and discuss some possible avenues for future work. The latter includes an alternative parameter estimation method for the approaches of Chapter 3, and a way to deal with negligible mass assigned to faces of the angular simplex in the methods of Chapters 3 and 4. We also propose a mixture model based on the conditional extreme value modelling approach of Heffernan and Tawn (2004) that has the potential to capture different combinations of tail dependence features; we suggest using Hamiltonian Monte Carlo techniques for inference, and demonstrate the use of this model in the bivariate case.

Chapter 2

Literature Review

2.1 Introduction

When studying extreme events, we are concerned with the tails of a distribution, where the intrinsically small number of observations make statistical modelling a challenge. We may also be interested modelling events that have not been observed in the data, meaning that extrapolation is a key concern. Due to the usually limited amount of data, empirical methods and other standard statistical techniques may not be applicable; we instead require results from extreme value theory.

In this chapter, we introduce some of the methods currently available to model extreme events. We begin with an overview of classical results for modelling univariate extremes in Section 2.2, before discussing multivariate techniques in Section 2.3. In Section 2.4, we discuss existing ideas for modelling multivariate extremes where there is some underlying structure controlling the dependence between the variables. Finally, we introduce a class of multivariate models known as vine copulas in Section 2.5; these models also have a specific dependence structure which can be represented graphically, and could be exploited in multivariate extreme value modelling.

2.2 Univariate extreme value theory

2.2.1 Overview

A thorough review of standard methods for modelling univariate extremes can be found in Coles (2001). Here, we discuss some of the main results, including two common approaches for modelling univariate extremes, based on the generalized extreme value (GEV) and generalized Pareto (GP) distributions, discussed in Sections 2.2.2 and 2.2.3, respectively.

2.2.2 Generalized extreme value distribution

Let $X_1, \ldots, X_n$ be independent random variables with common distribution function F. The upper tail of distribution F can be modelled by considering

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

By the extremal types theorem of Leadbetter et al. (1983), if there exist series $a_n > 0$ and b_n such that

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

as $n \to \infty$, with G non-degenerate, then G belongs to the family of extreme value distributions, and the distribution function F of each of the X_i variables is said to lie in the domain of attraction of G.

The distribution function G corresponds to either a negative Weibull, Fréchet or Gumbel distribution, and these three cases can be combined into a single family of the form

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, \tag{2.2.1}$$

with location parameter $\mu \in (-\infty, \infty)$, scale parameter $\sigma \in (0, \infty)$ and shape parameter $\xi \in (-\infty, \infty)$, and where $x_+ := \max(x, 0)$. This model is known as the

generalized extreme value (GEV) distribution. For $\xi < 0$, this corresponds to the negative Weibull distribution; for $\xi > 0$, the distribution is Fréchet; and the GEV distribution with $\xi = 0$ (interpreted as $\xi \to 0$) corresponds to a Gumbel distribution.

Assuming that equation (2.2.1) holds exactly for large n, we have

$$\Pr(M_n \leq x) = G\left(\frac{x - b_n}{a_n}\right) = \tilde{G}(x),$$

where $\tilde{G}$ represents a GEV distribution with different location and scale parameters to G. This result allows us to use the GEV distribution to model maxima in practice. A common method in modelling univariate extremes is to use the block maxima approach, where data are separated into sections of equal length, and the maximum value observed in each block is considered to be a realization of a GEV random variable; this is demonstrated in the left panel of Figure 2.2.1. The maximum observed values from the different blocks are used for inference of the GEV parameters.

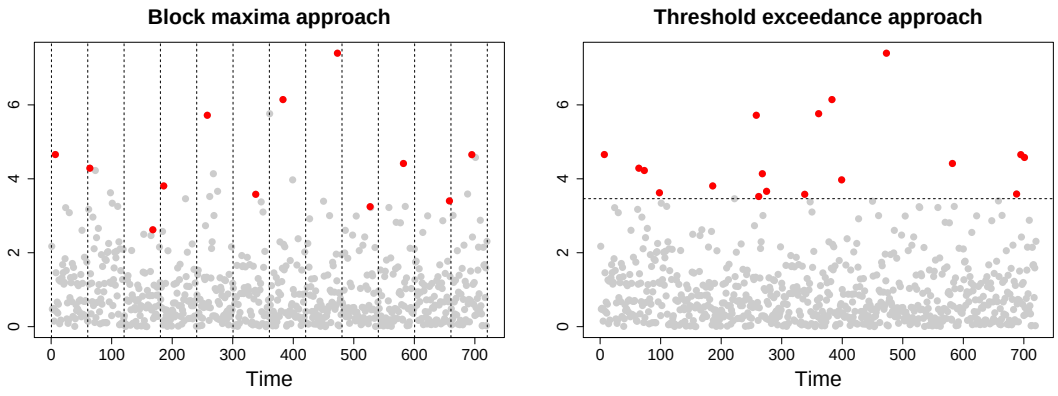

Figure 2.2.1: A demonstration of the data considered extreme in the block maxima and threshold exceedance methods.

Theory for modelling the lower tail of a distribution can be derived from the theory for the upper tail by exploiting the relation

$$\min(X_1, \ldots, X_n) = -\max(-X_1, \cdots - X_n).$$

It is therefore sufficient for us to consider only the latter case here.

2.2.3 Generalized Pareto distribution

A common criticism of the block maxima approach is that taking only one value per block may lead to some information about extreme events being disregarded. A typical alternative is to consider values above a high threshold, u, as being extreme, and to fit a model to the exceedances above that threshold. This method generally utilizes more of the extreme observations than the block maxima approach; an example is shown in the right panel of Figure 2.2.1.

A model for threshold exceedances can be motivated via the point process representation of Pickands (1971). Suppose we have independent and identically distributed random variables $X_1, \ldots, X_n$, whose common distribution function F has upper endpoint x^F, and that there exist $a_n > 0$ and b_n such that $\Pr\{(M_n - b_n)/a_n \leq x\} \to G(x)$, where G follows a GEV distribution. As $n \to \infty$, it can be shown that the point process

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

converges to a non-homogeneous Poisson process, P, with integrated intensity

$$\Lambda\big([c_1, c_2] \times [x, \infty)\big) = (c_2 - c_1) \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]_+^{-1/\xi}, \tag{2.2.2}$$

on $[0, 1] \times [x_G, \infty)$, for $x_G = \inf\{x : G(x) > 0\}$ and $0 < c_1 < c_2 < 1$. Coles (2001), for example, demonstrates that a model for threshold exceedances may be obtained from this limiting Poisson process result. First note that the integrated intensity in (2.2.2) may be written as

$$\Lambda\big([c_1, c_2] \times (x, \infty)\big) = \Lambda_1\big([c_1, c_2]\big) \times \Lambda_2\big([x, \infty)\big),$$

with

$$\Lambda_1\big([c_1, c_2]\big) = c_2 - c_1 \quad \text{and} \quad \Lambda_2\big([x, \infty)\big) = \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]_+^{-1/\xi}.$$

Then, for $u > x_G$, we have

$$\Pr\left(\frac{X_i - b_n}{a_n} \leq x + u \middle| \frac{X_i - b_n}{a_n} > u \right) = 1 - \Pr\left(\frac{X_i - b_n}{a_n} > x + u \middle| \frac{X_i - b_n}{a_n} > u \right)$$

$$
\begin{aligned}
&\to 1 - \frac{\Lambda_2\big([x+u,\infty)\big)}{\Lambda_2\big([u,\infty)\big)}, \quad \text{as } n \to \infty \\
&= 1 - \frac{\left[1+\xi\left(\frac{x+u-\mu}{\sigma}\right)\right]_+^{-1/\xi}}{\left[1+\xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi}} \\
&= 1 - \left[1+\frac{\xi x}{\sigma_u}\right]_+^{-1/\xi} =: H(x), \qquad (2.2.3)
\end{aligned}
$$

where $\sigma_u = \sigma + \xi(u-\mu)$. Distributions of the form $H(x)$ are termed generalized Pareto (GP) distributions.

By a similar argument as for the GEV distribution, treating (2.2.3) as an equality for large n, we may disregard the constants a_n and b_n for modelling purposes. To see this, first note that by (2.2.3), for large u, we have

$$
\Pr\left(X_i > a_n x + a_n u + b_n \mid X_i > a_n u + b_n\right) = \left[1+\frac{\xi x}{\sigma_u}\right]_+^{-1/\xi}.
$$

Now, setting $u_n = a_n u + b_n$, and $y = a_n x$,

$$
\Pr\left(X_i > y + u_n \mid X_i > u_n\right) = \left[1+\frac{\xi x}{\sigma_{u,n}}\right]_+^{-1/\xi},
$$

which corresponds to a generalized Pareto distribution with scale parameter $\sigma_{u,n} = a_n\sigma_u$. As the threshold u_n tends towards x^F, exceedances above u_n may therefore be modelled using the GP distribution.

We can fit both GEV and GP distributions numerically using standard maximum likelihood techniques. One of the main issues that arises when carrying out inference for the GP distribution is choosing a suitable value for the threshold. This must be small enough that we have sufficient data to fit the model, but large enough that the results provide a valid asymptotic approximation. One simple approach is to consider parameter stability plots, as outlined by Coles (2001), in which parameter estimates are observed over a range of threshold choices. An issue here is that stability plots corresponding to different parameters do not always lead to the same conclusions,

meaning they are often difficult to interpret. To overcome this, Wadsworth (2016) proposes an alternative threshold selection method based on the joint distribution of the estimated model parameters, and discusses a way to automate this method, avoiding the need to interpret plots by eye. There is a large amount of further research on the topic of threshold selection; a review of many of these techniques is given by Scarrott and MacDonald (2012), while more recent approaches include a Bayesian cross-validation technique proposed by Northrop et al. (2017).

2.3 Multivariate extreme value theory

2.3.1 Defining extreme events

Unlike in the univariate case where we consider maxima and minima, there is no clear, single way to define an extreme event for multiple variables. The most suitable definition of an extreme event will usually depend on context; Barnett (1976) suggests several possibilities. One of the most common ways of defining a multivariate extreme event is to consider componentwise maxima, i.e., the maximum of each variable, which may not correspond to an actual observation. Alternatives include defining a convex hull around the data, with points lying on or beyond this region deemed extreme; considering observations which contain the maximum of at least one variable; and defining some function of the data, which may allow transformation of the problem to the univariate setting, as considered by Coles and Tawn (1994).

2.3.2 Copula theory

In multivariate modelling, it is important to consider the dependence properties of the variables. Copulas provide a way to separate marginal modelling from dependence modelling, which is often useful in practice. Joe (1997) and Nelsen (2006) give textbook treatments of copula modelling, while Joe (2014) provides a review of more

recent developments in the area.

By Sklar's theorem (Sklar, 1959), if $\boldsymbol{X} = (X_1, \ldots, X_d)$ has joint distribution function F, and $X_i \sim F_i$, for $i = 1, \ldots, d$ and each F_i continuous, then there exists a unique copula C such that

$$F(x_1, \ldots, x_d) = C\{F_1(x_1), \ldots, F_d(x_d)\}.$$

The copula C is essentially a distribution function with Uniform$(0, 1)$ marginal distributions, and determines the dependence structure of the variables. This result, alongside the probability integral transform, allows for transformation between different marginal distributions while preserving the dependence properties of the variables.

If X_i is a continuous random variable with distribution function F_i and inverse distribution function F_i^{-1}, then by the probability integral transform, $U = F_i(X) \sim$ Uniform$(0, 1)$, and $F_i^{-1}(U) \sim F_i$. So for instance, we may obtain a copula C_F with standard Fréchet margins via

$$F(x_1, \ldots, x_d) = C_F \left\{ -\frac{1}{\log F_1(x_1)}, \ldots, -\frac{1}{\log F_d(x_d)} \right\},$$

a copula C_E with standard exponential margins is defined by

$$F(x_1, \ldots, x_d) = C_E \left[-\log\{1 - F_1(x_1)\}, \ldots, -\log\{1 - F_d(x_d)\} \right],$$

and for standard Gumbel margins, the copula C_G satisfies

$$F(x_1, \ldots, x_d) = C_G \left[-\log\{-\log F_1(x_1)\}, \ldots, -\log\{-\log F_d(x_d)\} \right].$$

Analogous results can be used for transformation to other marginal distributions.

These results are often used in multivariate extreme value theory, where transforming margins can highlight particular features of the extreme values. In the Fréchet case, the very largest values are accentuated, whilst exponential or Gumbel margins

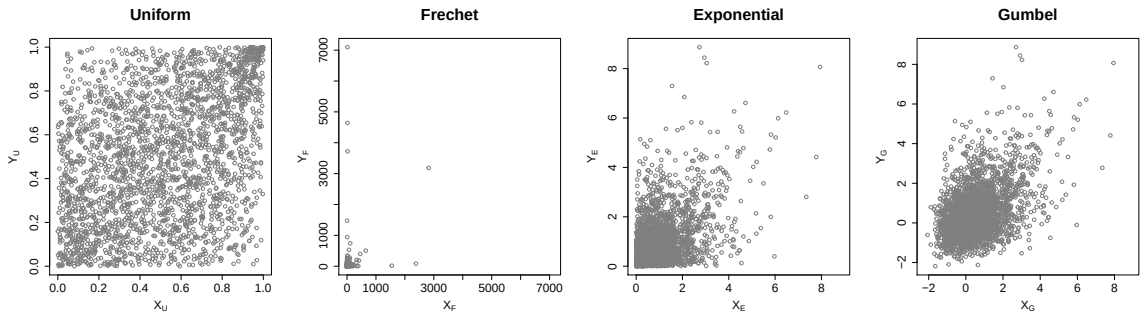

Figure 2.3.1: Example of data transformed to different marginal distributions.

are used in the conditional approach of Heffernan and Tawn (2004), discussed in Section 2.3.11, to focus on the behaviour of variables given that one of the variables is extreme. A demonstration of data transformed to different marginal distributions is given in Figure 2.3.1, for data simulated from a bivariate extreme value logistic distribution with dependence parameter $\alpha = 0.75$. This model will be discussed further in Section 2.3.3.

2.3.3 Componentwise maxima

Consider n independent d-dimensional random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ with common distribution function F. We denote the vector of componentwise maxima by $\boldsymbol{M}_n = (M_{n,1}, \ldots, M_{n,d})$, where $M_{n,i} = \max_{j \in \{1,\ldots,n\}} X_{j,i}$, for $i = 1, \ldots, d$, and $X_{j,i}$ denotes element i of the vector $\boldsymbol{X}_j$. In a similar way to the univariate case, suppose there exist constants $a_{n,i} > 0$ and $b_{n,i}$, for $i = 1, \ldots, d$, such that

$$\Pr\left(\frac{M_{n,1} - b_{n,1}}{a_{n,1}} \leq x_1, \ldots, \frac{M_{n,d} - b_{n,d}}{a_{n,d}} \leq x_d\right) \to G(x_1, \ldots, x_d),$$

as $n \to \infty$, for some limiting distribution function G that is non-degenerate in each margin. Unlike in the univariate case, the distribution G does not have a single parametric form, although each of the marginal distributions of G is a GEV distribution. Setting identical margins is often the easiest way in practice to study the properties of G, since this allows for focus on the dependence between the variables, as motivated

by the copula theory in Section 2.3.2. Choosing standard Fréchet marginal distributions emphasizes the largest values, as demonstrated in Figure 2.3.1; this is therefore the usual choice for modelling componentwise maxima. In particular, if X_i has a unit Fréchet distribution for each $i = 1, \ldots, d$, so that $\Pr(X_i < x) = \exp(-1/x)$ for $x > 0$, G may be written as

$$G(\boldsymbol{x}) = \exp\{-V(\boldsymbol{x})\}, \tag{2.3.1}$$

where $\boldsymbol{x} = (x_1, \ldots, x_d)$ and $x_i > 0$ for $i = 1, \ldots, d$. The function V is known as the exponent measure, and, for $\mathcal{S}_{d-1} = \left\{\boldsymbol{w} \in [0,1]^d : \sum_{i=1}^d w_i = 1\right\}$, takes the form

$$V(\boldsymbol{x}) = d \int_{\mathcal{S}_{d-1}} \max_{i=1,\ldots,d} \left(\frac{w_i}{x_i}\right) dH(\boldsymbol{w}), \tag{2.3.2}$$

where H is termed the spectral measure and, for $i = 1, \ldots, d$, satisfies

$$\int_{\mathcal{S}_{d-1}} w_i dH(\boldsymbol{w}) = \frac{1}{d}. \tag{2.3.3}$$

Special cases include independence between X_1 and X_2, where $V(x_1, x_2) = x_1^{-1} + x_2^{-1}$ and $H(\{0\}) = H(\{1\}) = 1/2$; and perfect dependence, which corresponds to $V(x_1, x_2) = \max\left(x_1^{-1}, x_2^{-1}\right)$ and $H(\{1/2\}) = 1$.

Functions $G(\boldsymbol{x})$ satisfying (2.3.1) are termed multivariate extreme value distributions. Several parametric models of this form have been proposed in the literature, including the logistic distribution of Gumbel (1960), for which the exponent measure is

$$V(\boldsymbol{x}) = \left(\sum_{i=1}^d x_i^{-1/\alpha}\right)^{\alpha}, \tag{2.3.4}$$

for $\alpha \in (0, 1]$. For this model, taking $\alpha = 1$ leads to independence between the variables, while as $\alpha \to 0$ we approach the case of complete dependence; in general, smaller values of α correspond to stronger dependence between the variables. Another parametric model is the asymmetric logistic distribution, proposed in the bivariate

case by Tawn (1988), and extended to the multivariate setting by Tawn (1990). The exponent measure for the asymmetric logistic distribution has the form

$$V(\boldsymbol{x}) = \sum_{C \in 2^D \setminus \emptyset} \left\{ \sum_{i \in C} \left(\frac{\theta_{i,C}}{x_i} \right)^{1/\alpha_C} \right\}^{\alpha_C}, \tag{2.3.5}$$

with $\alpha_C \in (0, 1]$, $\theta_{i,C} \in [0, 1]$, $\theta_{i,C} = 0$ if $i \notin C$, and $\sum_{C \in 2^D \setminus \emptyset} \theta_{i,C} = 1$ for all $i = 1, \ldots, d$, and $C \in 2^D \setminus \emptyset$, where 2^D denotes the power set of $D = \{1, \ldots, d\}$. The parameters α_C control dependence in a similar way to α in the logistic model, although this time just for the corresponding subsets of the variables. Other parametric models belonging to the class of multivariate extreme value distributions include the negative logistic model of Galambos (1975); the negative asymmetric logistic model of Joe (1990); and the Hüsler-Reiss distributions (Hüsler and Reiss, 1989).

Ledford and Tawn (1997) propose inverting models of the form (2.3.1) for the case where $d = 2$, so that the joint lower tail becomes the joint upper tail, and vice versa. The resulting class of models are known as inverted bivariate extreme values distributions, with copula of the form

$$C(u, v) = u + v - 1 + \exp\left[-V\left\{ \frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)} \right\} \right]. \tag{2.3.6}$$

These models exhibit different tail dependence properties to the corresponding extreme value distributions, and will be discussed further in Sections 2.3.7 and 2.3.10.

2.3.4 The linking between V and H

The exponent measure V and spectral measure H are related, due to the definition of V in (2.3.2), and Coles and Tawn (1991) show that these measures are linked via further relations. Specifically, in the bivariate case, they show that

$$h(w) = -\frac{(x_1 + x_2)^3}{2} \frac{\partial^2}{\partial x_1 \partial x_2} V(x_1, x_2), \quad \text{for } w \in (0, 1),$$
$$H(\{0\}) = -\frac{x_2^2}{2} \frac{\partial}{\partial x_2} V(x_1, x_2)\Big|_{x_1 = 0},$$

$$\text{and} \quad H(\{1\}) = -\frac{x_1^2}{2}\frac{\partial}{\partial x_1}V(x_1, x_2)\Big|_{x_2=0}.$$

A similar result for the d-dimensional case links the exponent measure V and spectral density h via

$$h\left(\frac{\boldsymbol{x}}{\sum_{i=1}^d x_i}\right) = -\frac{\left(\sum_{i=1}^d x_i\right)^{d+1}}{d}\frac{\partial^d}{\partial x_1, \ldots, \partial x_d}V(\boldsymbol{x}).$$

Moreover, for any subset of variables $\boldsymbol{X}_C = \{X_i : i \in C\}$ for $C \in 2^D \setminus \emptyset$ and $D = \{1, \ldots, d\}$, Coles and Tawn (1991) show that the spectral density h_C for events that are extreme only in $\boldsymbol{X}_C$ is given by

$$h_C\left(\frac{\boldsymbol{x}_C}{\sum_{i\in C} x_i}\right) = -\frac{\left(\sum_{i\in C} x_i\right)^{|C|+1}}{|C|}\left(\prod_{i\in C}\frac{\partial}{\partial x_i}\right)V(\boldsymbol{x}). \tag{2.3.7}$$

For instance, consider the asymmetric logistic model with exponent measure (2.3.5). In this case, applying (2.3.7), we find that the spectral density corresponding to the set C is given by

$$h_C(\boldsymbol{w}_C) = -\frac{\left(\sum_{i\in C} x_i\right)^{|C|+1}}{|C|}\left\{\prod_{i=0}^{|C|-1}\left(\frac{i-\alpha_C}{\alpha_C}\right)\right\}\left(\prod_{i\in C}\frac{\theta_{i,C}^{1/\alpha_C}}{x_i^{1+1/\alpha_C}}\right)\left\{\sum_{i\in C}\left(\frac{\theta_{i,C}}{x_i}\right)^{1/\alpha_C}\right\}^{\alpha_C-|C|},$$

for $\boldsymbol{w}_C = \boldsymbol{x}_C / \sum_{i\in C} x_i$.

2.3.5 Pickands' dependence function

In the bivariate case, again with standard Fréchet margins, the exponent measure $V(x_1, x_2)$ links to Pickands' dependence function $A(w)$ (Pickands, 1981) via the equation

$$V(x_1, x_2) = \left(\frac{1}{x_1} + \frac{1}{x_2}\right)A(w),$$

for $w = x_1/(x_1 + x_2)$, i.e., $0 \le w \le 1$. Pickands' dependence function is convex, and has the conditions that $A(0) = A(1) = 1$, and $\max(w, 1-w) \le A(w) \le 1$. The case of independence between X_1 and X_2 corresponds to $A(w) = 1$, while for perfect

dependence, $A(w) = \max(w, 1-w)$. These cases highlight a drawback of the componentwise maxima approach for modelling the joint upper tail of (X_1, X_2); since $A(w)$ is a convex function, with independence and perfect dependence as boundary cases, it is not possible to model negative dependence using these componentwise maxima results.

Pickands' dependence function also links to the spectral measure H by the relations

$$\frac{dA}{dw} = 2H(w) - 1 \quad \text{and} \quad \frac{d^2A}{dw^2} = 2\frac{dH}{dw} = 2h(w),$$

where $h(w)$ is the density corresponding to the spectral measure H.

2.3.6 Regular variation

An alternative tail characterization of $\boldsymbol{X}$ is in terms of its pseudo-polar coordinates $(R, \boldsymbol{W})$. Consider the random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$, with common marginal distributions satisfying $\Pr(X_i > x) \sim cx^{-1}$ for $i = 1, \ldots, d$ and some $c > 0$. For arbitrary norms, $\|\cdot\|_1$ and $\|\cdot\|_2$, the coordinates $(R, \boldsymbol{W})$ are defined as

$$R = \|\boldsymbol{X}\|_1 \quad \text{and} \quad \boldsymbol{W} = \boldsymbol{X}/\|\boldsymbol{X}\|_2,$$

with $R > 0$ and $\boldsymbol{W} \in \mathcal{S}_{d-1} = \left\{(w_1, \ldots, w_d) \in [0,1]^d : \sum_{i=1}^d w_i = 1\right\}$, the $(d-1)$-dimensional unit simplex.

Taking both norms to be the L_1 norm, i.e., $\|\boldsymbol{x}\|_1 = \|\boldsymbol{x}\|_2 = \sum_{i=1}^d x_i$, the assumption of multivariate regular variation states that

$$\lim_{t\to\infty} \Pr(R > tr, \boldsymbol{W} \in B \mid R > t) = H(B)r^{-1}, \tag{2.3.8}$$

for $r \geq 1$, spectral measure H satisfying (2.3.3) and B some measurable subset of $\mathcal{S}_{d-1}$. The right-hand side of (2.3.8) shows that under the regular variation assumption, we have limiting independence of the angular and radial components. For large radial

values, the position of angular mass on $\mathcal{S}_{d-1}$ links to the extremal dependence structure of $\boldsymbol{X}$, i.e., the subsets of $\boldsymbol{X}$ that can take their largest values simultaneously, which will be discussed further in Section 2.3.8.

2.3.7 Modelling asymptotic dependence

Many methods for modelling multivariate extremes are only applicable for variables with certain tail dependence properties, and an important consideration is whether or not the variables can take their largest values simultaneously. It is useful for model selection to have measures that allow us to categorize data as belonging to different tail dependence classes. Several such methods are discussed by Coles et al. (1999); we focus on one such measure here, denoted by χ.

In the bivariate case, consider variables X_1 and X_2 with respective distribution functions F_1 and F_2. The measure χ, taking values in $[0, 1]$, is defined via the limiting conditional survivor function

$$\chi = \lim_{u \to 1} \Pr\left\{F_2(X_2) > u \mid F_1(X_1) > u\right\} = \lim_{u \to 1} \frac{\Pr\left\{F_1(X_1) > u, F_2(X_2) > u\right\}}{1-u},$$

where the limit is assumed to exist. In this case, if $\chi = 0$, the variables are said to be asymptotically independent, i.e., X_1 and X_2 cannot take their largest values simultaneously, while if $\chi \in (0, 1]$, the variables may be simultaneously extreme and are said to exhibit asymptotic dependence.

For a bivariate extreme value distribution with exponent measure $V(x_1, x_2)$ and distribution function of the form (2.3.1), we have

$$\begin{aligned}
\chi &= \lim_{u \to 1} \Pr\left\{F_2(X_2) > u \mid F_1(X_1) > u\right\} \\
&= \lim_{u \to 1} \frac{1 - 2u + \exp\{-V(-1/\log u, -1/\log u)\}}{1-u} = \lim_{u \to 1} \frac{1 - 2u + u^{V(1,1)}}{1-u} \\
&= \lim_{u \to 1} \frac{1 - 2u + 1 + V(1,1)(u-1) + O\{(u-1)^2\}}{1-u} = 2 - V(1,1).
\end{aligned}$$

This implies that the bivariate logistic model with exponent measure (2.3.4) has $\chi = 2 - 2^{\alpha}$. That is, it exhibits asymptotic dependence for $\alpha \in (0, 1)$, and asymptotic independence for $\alpha = 1$, which corresponds to the independence case. The asymmetric logistic model with exponent measure (2.3.5) has

$$\chi = \theta_{1,\{1,2\}} + \theta_{2,\{1,2\}} - \left(\theta_{1,\{1,2\}}^{1/\alpha_{\{1,2\}}} + \theta_{2,\{1,2\}}^{1/\alpha_{\{1,2\}}}\right)^{\alpha_{\{1,2\}}}.$$

For this model, if $\alpha_{\{1,2\}} = 1$, or if $\theta_{1,\{1,2\}} = \theta_{2,\{1,2\}} = 0$, we are in the independence setting with $\chi = 0$. Otherwise, the model exhibits asymptotic dependence. Examples of models with $\chi = 0$, corresponding to asymptotic independence, include the bivariate Gaussian and inverted bivariate extreme value distributions in (2.3.6).

The idea of using a limiting conditional probability to assess tail dependence extends naturally to a multivariate setting. Suppose we are interested in the random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ with $X_i \sim F_i$ for each $i \in D = \{1, \ldots, d\}$. To investigate the tail dependence properties of a subset of variables $\boldsymbol{X}_C = \{X_i : i \in C\}$ for some $C \in 2^D$ with $|C| \geq 2$, a suitable measure of tail dependence is

$$\chi_C = \lim_{u \to 1} \frac{\Pr\{F_i(X_i) > u : i \in C\}}{1 - u}, \tag{2.3.9}$$

see for example Hua and Joe (2011) or Wadsworth and Tawn (2013). In this case, if $\chi_C \in (0, 1]$, the variables $\boldsymbol{X}_C$ are asymptotically dependent, i.e., all components can be large simultaneously. On the other hand, if $\chi_C = 0$, the variables in $\boldsymbol{X}_C$ cannot all take their largest values together, although any subset $\underline{C} \subset C$ with $|\underline{C}| \geq 2$ could have $\chi_{\underline{C}} > 0$, that is, $\boldsymbol{X}_{\underline{C}}$ could still exhibit asymptotic dependence.

As an example, consider the asymmetric logistic model with only pairwise dependence, i.e., with exponent measure

$$V(x_1, x_2, x_3) = \left\{\left(\frac{\theta_1}{x_1}\right)^{1/\alpha_1} + \left(\frac{\theta_2}{x_2}\right)^{1/\alpha_1}\right\}^{\alpha_1} + \left\{\left(\frac{1-\theta_1}{x_1}\right)^{1/\alpha_2} + \left(\frac{\theta_3}{x_3}\right)^{1/\alpha_2}\right\}^{\alpha_2} + \left\{\left(\frac{1-\theta_2}{x_2}\right)^{1/\alpha_3} + \left(\frac{1-\theta_3}{x_3}\right)^{1/\alpha_3}\right\}^{\alpha_3},$$

with $\alpha_1, \alpha_2, \alpha_3 \in (0, 1]$ and $\theta_1, \theta_2, \theta_3 \in [0, 1]$, where the parameter subscripts have been simplified from (2.3.5) to improve readability. In this case, it can be shown that $\chi_{\{1,2,3\}} = 0$, while the bivariate measures of dependence are

$$\chi_{\{1,2\}} = \theta_1 + \theta_2 - \left(\theta_1^{1/\alpha_1} + \theta_2^{1/\alpha_1}\right)^{\alpha_1}, \quad \chi_{\{1,3\}} = 1 - \theta_1 + \theta_3 - \left\{(1-\theta_1)^{1/\alpha_2} + \theta_3^{1/\alpha_2}\right\}^{\alpha_2},$$
$$\text{and} \quad \chi_{\{2,3\}} = 1 - \theta_2 + 1 - \theta_3 - \left\{(1-\theta_2)^{1/\alpha_3} + (1-\theta_3)^{1/\alpha_3}\right\}^{\alpha_3}.$$

That is, this particular model does not exhibit overall asymptotic dependence, but can still have asymptotic dependence in all pairs of variables if $\alpha_1, \alpha_2, \alpha_3 \neq 1$ and $\theta_1, \theta_2, \theta_3 \in (0, 1)$.

2.3.8 Extremal dependence structures

The definition of χ_C in the previous section demonstrates that extremal dependence in multivariate extremes can have a complicated structure, with only certain subsets of the variables being simultaneously large while other variables are of smaller order. In the radial-angular representation of Section 2.3.6, this corresponds to the spectral measure H placing extremal mass on various faces of the angular simplex $\mathcal{S}_{d-1}$. The extremal dependence properties exhibited by a particular set of data should be considered when selecting a model for its extreme values; we should aim to match the extremal dependence structure of the data to the structures that proposed models can capture. Many parametric models for multivariate extremes are only suitable for the asymptotic dependence or asymptotic independence cases, such as the logistic model with $\alpha \in (0, 1)$ and the multivariate Gaussian, respectively. However, some models, such as the asymmetric logistic model with exponent measure (2.3.5) allow for more complicated extremal dependence structures. In particular, if $\sum_{i \in C} \theta_{i,C} > 0$ for $C \in 2^D \setminus \emptyset$, the asymmetric logistic distribution places extremal mass on the face of $\mathcal{S}_{d-1}$ corresponding to the variables $\{X_i : i \in C\}$ being simultaneously large while $\{X_i : i \notin C\}$ are of smaller order.

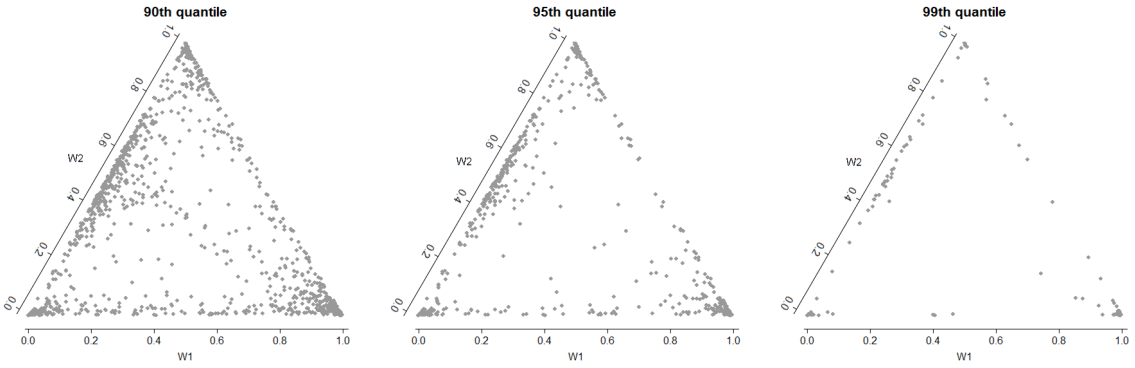

Figure 2.3.2: Example of a trivariate distribution with mass on five faces of the angular simplex. The data are generated using a multivariate extreme value distribution with an asymmetric logistic model.

Figure 2.3.2 shows an example of data simulated from the asymmetric logistic distribution with the dependence parameters $\alpha_C = 0.6$ for all $C \in 2^D \setminus \emptyset$, controlling how close to the centre of the faces with mass the extremal points lie. The dependence structure has been chosen to correspond to limiting mass on two of the vertices, two of the edges and the centre of the unit simplex. In particular, X_1 and X_3 can both take their largest values independently of the other variables, and the subsets $\{X_1, X_2\}$, $\{X_2, X_3\}$ and $\{X_1, X_2, X_3\}$ may be simultaneously large. Here, we have taken a sample of size $n = 10,000$, and the three plots show $(W_1, W_2) \mid R > r$, with r taken to be the observed 0.9, 0.95 and 0.99 radial quantiles, respectively. The data in Figure 2.3.2 are presented on the equilateral simplex for visual purposes.

As we increase the radial threshold, we observe that, although there are points close to the boundaries of the unit simplex, none of these points lie exactly on the boundary, so simply considering the proportion of points on each face above a high threshold does not reveal the extremal dependence structure. Moreover, there appears to be mass close to all three corners of the unit simplex, when in fact only two of the vertices have positive mass in the limit. To assess the underlying extremal de-

pendence structure of the variables using such data, conditioning on R being above a finite threshold, we need to determine which faces truly represent this limiting dependence structure, and which faces appearing to have mass do so as an artefact of lack of convergence at finite levels. We propose methods for determining this structure in Chapters 3 and 4, and introduce some existing methods in the remainder of this section.

Goix et al. (2016) propose a non-parametric simplex-partitioning method for estimating extremal dependence structures. In this approach, the different regions of the partition are chosen to estimate the various faces of the angular simplex $\mathcal{S}_{d-1}$, with extremal mass on each of these faces corresponding to a different subset of the variables being simultaneously extreme while the others are of smaller order. Conditioning on the radial component being above some high threshold, empirical estimates are obtained for the amount of extremal mass associated with each face, and a sparse representation of the extremal dependence structure is obtained by considering faces where this empirical estimate is sufficiently large. This method is shown to work well in practice, particularly when the asymptotic dependence between variables in the same subsets is strong.

In the method proposed by Goix et al. (2016), the aim is to obtain a representation of the extremal dependence structure that is sparse, i.e., the number of subsets of variables being simultaneously large should be small compared to the dimension of the problem. Chiapino and Sabourin (2017) point out that sparsity may not always be achieved by this method, as too many faces that are similar in some way could be detected. They therefore propose an algorithm that aims to group together nearby faces with extremal mass into feature clusters. This method exploits the graphical structure of clusters, and uses a measure of extremal dependence related to χ to group variables that are likely to take their largest values simultaneously. Chiapino et al.

(2019) extend this approach by using the coefficient of tail dependence of Ledford and Tawn (1996), discussed in Section 2.3.9, to assess the extremal dependence of groups of variables. Our approaches in Chapters 3 and 4 exploit a new set of parameters, related to this coefficient of tail dependence, that reveal additional information about the extremal dependence structure.

Extremal dependence structures have also been studied elsewhere in the literature. This includes the factor analysis approach of Klüppelberg et al. (2015) for elliptical copulas; the method of Chautru (2015) which incorporates principal component analysis and clustering techniques; and the Bayesian clustering approach of Vettori et al. (2018).

2.3.9 Modelling asymptotic independence

Our focus so far has mainly been on modelling asymptotic dependence between variables; we now consider asymptotic independence in more detail. The approach of Ledford and Tawn (1996) allows for the modelling of both asymptotic independence and asymptotic dependence, with the latter a boundary case. For random variables X_1 and X_2 with common marginal distribution F with an infinite upper endpoint, they propose a model for the limiting behaviour of the joint survivor function. This is given by

$$\Pr(X_1 > x, X_2 > x) \sim L\left[\{1 - F(x)\}^{-1}\right] \{1 - F(x)\}^{1/\eta}, \tag{2.3.10}$$

as $x \to \infty$, for $\eta \in (0, 1]$ and some function L that is slowly varying at infinity, that is $L(tx)/L(t) \to 1$ as $t \to \infty$ for all $x > 0$. This is often presented for (X_1, X_2) having standard Fréchet margins, so that as $x \to \infty$,

$$\Pr(X_1 > x, X_2 > x) \sim L(x)x^{-1/\eta}.$$

If $\eta = 1$ and $L(x) \not\to 0$ as $x \to \infty$, there is asymptotic dependence between the variables, and asymptotic independence otherwise. As such, the parameter η is often

used to determine the tail dependence properties of a given pair of variables. If we have $\eta < 1$, the particular value of η and the function L are indicative of the strength of asymptotic independence between the variables, with $\eta = 1/2$ corresponding to exact independence if $L(x) = 1$ or near independence for $L(x) \neq 1$, and the variables having some positive association for $\eta \in (1/2, 1)$.

A similar approach can be adopted in the multivariate case, as considered, for instance, by de Haan and Zhou (2011) and Eastoe and Tawn (2012). For the random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ with standard Fréchet margins, and any set $C \in 2^D \setminus \emptyset$ with $D = \{1, \ldots, d\}$, we can study

$$\Pr(X_i > x : i \in C) \sim L_C(x) x^{-1/\eta_C},$$

as $x \to \infty$, for some slowly varying function L_C, $\eta_C \in (0, 1]$. In this case, if $\eta_C = 1$, there is asymptotic dependence between the variables $\boldsymbol{X}_C = \{X_i : i \in C\}$. For $\eta_C < 1$, $\boldsymbol{X}_C$ exhibits asymptotic independence, although it could still be the case that certain subsets of the variables can be large together, relating to the more complicated extremal dependence structures discussed in Section 2.3.8.

Inference for the parameter η_C is not straightforward since estimation of the parameter is difficult, and the value corresponding to asymptotic dependence is a boundary case. One standard approach is to use the Hill estimator (Hill, 1975). Consider ordered realizations $t_{(1)} > t_{(2)} > \cdots > t_{(n)}$ of the variable $T = \min(X_i : i \in C)$. Assuming model (2.3.10) holds exactly for x above some high threshold u, the Hill estimate can be derived using maximum likelihood techniques as

$$\hat{\eta}_C = \frac{1}{n_u} \sum_{i=1}^{n_u} \log\left(\frac{t_{(i)}}{u}\right), \tag{2.3.11}$$

where n_u denotes the number of realizations above the threshold u. In practice, this can take values greater than 1, so we often take $\hat{\eta}_C$ to be the minimum of 1 and the value in (2.3.11). Beirlant et al. (2004) discuss some issues with estimating η

via the Hill estimate. In particular, they comment that uncertainty in the parameter estimation is often underestimated in practice, which could, for instance, lead to likelihood ratio tests incorrectly rejecting asymptotic dependence. In our studies, we found that the Hill estimate is often lower than the true value of η_C, particularly if this true value is 1 or close to 1, also leading to difficulty in distinguishing between asymptotic dependence and asymptotic independence. This is an issue we investigate further in Section 7.2.2 in the context of the methods we propose in Chapters 3 and 4.

A review of other techniques for estimating η_C can be found in Section 9.5.2 of Beirlant et al. (2004). These include procedures developed by Peng (1999) and Draisma et al. (2004), who suggest non-parametric estimates of η_C that do not depend on the marginal distributions of the variables, and Beirlant and Vandewalle (2002) whose approach is based on scaled log-ratios.

2.3.10 Calculating the coefficient of tail dependence from a density

The coefficient of tail dependence η_C is defined in terms of the survivor function of the variables $\boldsymbol{X}_C$. However, for certain distributions, it may not be possible to obtain this function analytically, and it may only be the joint density that is available. Nolde (2014) presents a strategy for determining η_C in this setting using a geometric approach, by considering the shape of scaled random samples from the joint distribution of $\boldsymbol{X}_C$.

Assuming standard exponential margins, and that the variables $\boldsymbol{X}_C$ have joint density $f(\boldsymbol{x})$, Nolde proposes studying the gauge function $g(\boldsymbol{x})$, which is homogeneous of order 1, such that

$$-\log f(t\boldsymbol{x}) \sim tg(\boldsymbol{x}),$$

as $t \to \infty$. Interest then lies with the set $G = \{\boldsymbol{x} \in \mathbb{R}^{|C|} : g(\boldsymbol{x}) = 1\}$, which forms the boundary of the scaled random sample $(\boldsymbol{X}_{C,1}/\log n, \ldots, \boldsymbol{X}_{C,n}/\log n)$ for large n, where the scaling function $\log n$ is chosen due to the exponential margins. In particular, note that for each variable X_i, $i \in C$, with $M_{i,n} = \max(X_{i,1}, \ldots, X_{i,n})$,

$$\Pr\left\{\frac{M_{i,n}}{\log n} < x\right\} = \left(1 - e^{-x\log n}\right)^n = \left(1 - \frac{1}{n^x}\right)^n \to \begin{cases} 0, & x < 1, \\ e^{-1}, & x = 1, \\ 1, & x > 1, \end{cases}$$

as $n \to \infty$, so that, asymptotically, $M_{i,n}/\log n = 1$. As such, the set G lies within the region $[0,1]^{|C|}$. The coefficient η_C corresponds to the smallest value of r such that $G \cap [r,\infty)^{|C|} = \emptyset$ (or the largest value of r such that $G \cap [r,\infty)^{|C|} \neq \emptyset$). We demonstrate this approach using a series of bivariate examples.

Let us first consider an inverted bivariate extreme value distribution. In exponential margins, this model has distribution function

$$F(x,y) = 1 - e^{-x} - e^{-y} + \exp\left\{-V\left(x^{-1}, y^{-1}\right)\right\},$$

for $x, y > 0$. By first differentiating with respect to both components to obtain the density $f(x,y)$, with V_1, V_2 and V_{12} denoting the derivatives of the exponent measure with respect to the first, second and both components, respectively, we have

$$\begin{aligned} -\log f(tx,ty) =& 2\log(tx) + 2\log(ty) + V\left\{(tx)^{-1}, (ty)^{-1}\right\} \\ & - \log\left[V_1\left\{(tx)^{-1}, (ty)^{-1}\right\} V_2\left\{(tx)^{-1}, (ty)^{-1}\right\} - V_{12}\left\{(tx)^{-1}, (ty)^{-1}\right\}\right] \\ =& 2\log(tx) + 2\log(ty) + tV\left(x^{-1}, y^{-1}\right) \\ & - \log\left\{t^4 V_1\left(x^{-1}, y^{-1}\right) V_2\left(x^{-1}, y^{-1}\right) - t^3 V_{12}\left(x^{-1}, y^{-1}\right)\right\} \\ =& tV\left(x^{-1}, y^{-1}\right) + O(\log t), \end{aligned}$$

as $t \to \infty$, by exploiting the homogeneity of the exponent measure. That is, the gauge function is given by $g(x,y) = V(x^{-1}, y^{-1})$.

For the bivariate inverted logistic example, this corresponds to the gauge function being $g(x, y) = (x^{1/\alpha} + y^{1/\alpha})^{\alpha}$. For the case where $\alpha = 0.5$, the grey points in the left panel of Figure 2.3.3 show a suitably-normalized sample from this distribution, while the set G where $g(x, y) = 1$ is shown by the red line. The blue region shows the set $[r, \infty)^2$, with the smallest value of r such that the two sets do not intersect being $2^{-0.5} = 1/2^{\alpha}$, occurring when $x = y$. This corresponds to the known value of $\eta_{\{1,2\}}$ for this copula.

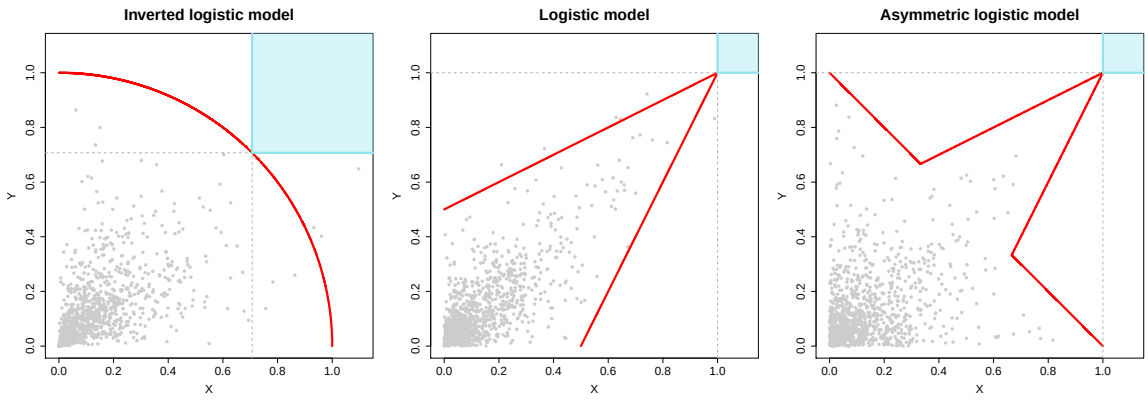

Figure 2.3.3: Scaled samples (grey) from an inverted logistic copula (left), a logistic copula (centre) and an asymmetric logistic copula (right) with $\alpha = 0.5$ and $\theta_{1,\{1\}} = \theta_{2,\{2\}} = \theta_{1,\{1,2\}} = \theta_{2,\{1,2\}} = 0.5$; the corresponding sets $G = \{(x, y) \in \mathbb{R}^2 : g(x, y) = 1\}$ (red); and the sets $[\eta_{\{1,2\}}, \infty)^2$ (blue).

To obtain similar results for the bivariate extreme value copula, we assume that the corresponding spectral density $h(w)$ places no mass on $\{0\}$ or $\{1\}$ and has regularly varying tails. Specifically, $h(w) \sim c_1(1 - w)^{s_1}$ as $w \to 1$, and $h(w) \sim c_2 w^{s_2}$ as $w \to 0$, for $c_1, c_2 \in \mathbb{R}$ and $s_1, s_2 > -1$. In this case, the gauge function is

$$g(x, y) = \left(2 + s_1 \mathbb{1}_{\{x \geq y\}} + s_2 \mathbb{1}_{\{x < y\}}\right) \max(x, y) - \left(1 + s_1 \mathbb{1}_{\{x \geq y\}} + s_2 \mathbb{1}_{\{x < y\}}\right) \min(x, y).$$

For the logistic model with dependence parameter $\alpha \in (0, 1)$, we have $s_1 = s_2 =$

$1/\alpha - 2$. Hence the gauge function is

$$g(x, y) = \frac{1}{\alpha} \max(x, y) + \left(1 - \frac{1}{\alpha}\right) \min(x, y).$$

In this case, the point $(x, y) = (1, 1)$ satisfies $g(x, y) = 1$, and since both variables are at most 1 in the set G, we must have $\eta_{\{1,2\}} = 1$. This is demonstrated in the centre panel of Figure 2.3.3 for $\alpha = 0.5$.

We now consider the asymmetric logistic model with exponent measure (2.3.5). This model does not satisfy the condition used when calculating the gauge function for bivariate extreme value copulas, that the spectral density places no mass on $\{0\}$ or $\{1\}$. However, calculating the gauge function for this model directly, we obtain

$$g(x, y) = \min\left\{(x + y); \frac{1}{\alpha} \max(x, y) + \left(1 - \frac{1}{\alpha}\right) \min(x, y)\right\},$$

for all parameters taking values in $(0, 1)$. This is demonstrated the right panel of Figure 2.3.3 for $\alpha = 0.5$, and we again note that since $g(1, 1) = 1$, the coefficient of tail dependence has value $\eta_{\{1,2\}} = 1$. The bivariate asymmetric logistic copula is essentially a mixture of independence and logistic models; this is reflected in the gauge function, which is the minimum of the gauge functions corresponding to the two mixture components.

Finally, we consider a bivariate Gaussian copula having covariance matrix Σ with $\Sigma_{1,1} = \Sigma_{2,2} = 1$ and $\Sigma_{1,2} = \Sigma_{2,1} = \rho \in [0, 1)$. Nolde (2014) shows that for exponential margins, this model has gauge function

$$g(x, y) = (1 - \rho^2)^{-1}\{x + y - 2\rho(xy)^{1/2}\}.$$

We demonstrate this in Figure 2.3.4 for $\rho = 0.5$. In this case, the smallest value of r such that the sets G and $[r, \infty)^2$ intersect is shown to be $0.75 = (1 + \rho)/2$, corresponding to the known value of $\eta_{\{1,2\}}$ in this case.

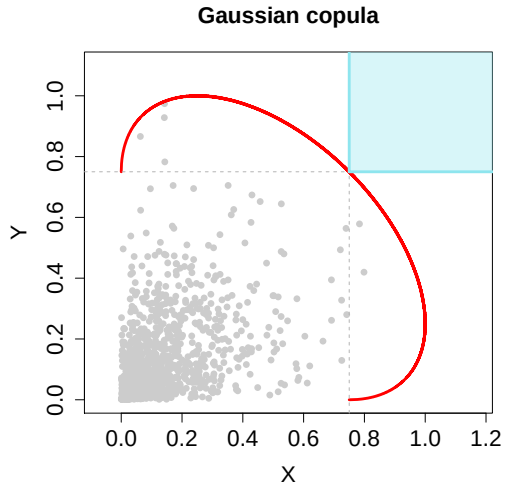

Figure 2.3.4: Scaled samples (grey) from a Gaussian copula with $\rho = 0.5$; the corresponding set $G = \{(x, y) \in \mathbb{R}^2 : g(x, y) = 1\}$ (red); and the set $[\eta_{\{1,2\}}, \infty)^2$ (blue).

In the bivariate examples we have studied here, the intersection of interest between the sets $[r, \infty)^2$ and G occurred when x and y were equal, but we note that this is not the case in general. We will adopt and extend this geometric approach of Nolde (2014) to study another class of models in Chapter 5.

2.3.11 Conditional extreme value modelling

An alternative approach to modelling multivariate extremes is to condition on one of the variables being extreme, and study the behaviour of the remaining variables. This is considered, for instance, by Heffernan and Resnick (2007); we focus on the statistical perspective of Heffernan and Tawn (2004). This approach can be used whether there is asymptotic dependence or asymptotic independence between the variables: an advantage over many other models for multivariate extremes.

In the bivariate case, consider variables X_1 and X_2 with standard Gumbel marginal distributions, although the results hold whenever the variables have a common, con-

tinuous marginal distribution with an exponential upper tail. Taking the conditioning variable to be X_1, Heffernan and Tawn consider the standardized residual

$$Z = \frac{X_2 - a(X_1)}{b(X_1)},$$

when X_1 is very large, for regularly varying functions a and b satisfying certain conditions, with $b(x) > 0$ for all $x \in \mathbb{R}$. Heffernan and Resnick (2007) consider the constraints on a and b, showing that their indices of regular variation should be 1 and less than 1, respectively. Under the assumption that

$$\Pr(Z \leq z \mid X_1 = x) \to G_{|1}(z),$$

as $x \to \infty$, for some non-degenerate limit distribution $G_{|1}$, Heffernan and Tawn show that for fixed z and $x > 0$,

$$\Pr\left(Z \leq z, X_1 - u > x \mid X_1 > u\right) \to G_{|1}(z)e^{-x}, \tag{2.3.12}$$

as $u \to \infty$. That is, for $X_1 > u$ and $u \to \infty$, the variables Z and $X_1 - u$ are asymptotically independent, with the latter following an exponential distribution. Common choices of the standardizing functions in this bivariate case are $a(x) = \alpha x$ and $b(x) = x^\beta$, for $\alpha \in (0, 1]$ and $\beta < 1$. There is asymptotic dependence between X_1 and X_2 if $\alpha = 1$ and $\beta = 0$, corresponding to $a(x) = x$ and $b(x) = 1$, and asymptotic independence otherwise. In the latter case, it is possible to have either independence or positive dependence between the variables in the current setting. Keef et al. (2013a) propose an extension which applies if there is negative dependence between the variables, with asymptotic independence in the tails.

For modelling purposes, we assume that limit (2.3.12) holds exactly for some large threshold u. Then, using the working assumption that $Z \sim N(\mu, \sigma^2)$, for $\mu \in \mathbb{R}$ and $\sigma > 0$, and writing

$$X_2 = \alpha X_1 + X_1^\beta Z,$$

for $X_1 > u$ and where Z is independent of X_1, one can use maximum likelihood techniques to estimate the parameters (α, β). Once these estimates, denoted by $(\hat{\alpha}, \hat{\beta})$, have been obtained, it is possible use simulation to extrapolate beyond the observed values. Suppose we have n observations, denoted $(x_{1,1}, x_{2,1}), \ldots, (x_{1,n}, x_{2,n})$. Now let $i_1, \ldots, i_{n_u}$ be the indices i in the set $\{1, \ldots, n\}$ corresponding to $x_{1,i} > u$. The residual associated with the i_lth observation is defined as

$$z_l = \frac{x_{2,i_l} - \hat{\alpha} x_{1,i_l}}{x_{1,i_l}^{\hat{\beta}}}$$

for $l = 1, \ldots, n_u$. We now drop the assumption that $Z \sim N(\mu, \sigma^2)$, and instead take Z to have an empirical distribution based on the observed residuals $\{z_1, \ldots, z_{n_u}\}$. A set of simulated values $(x^*_{1,1}, x^*_{2,1}), \ldots, (x^*_{1,m}, x^*_{2,m})$ can be obtained by the algorithm

1. set $x^*_{1,j} = u + e_j$, where $e_j \sim \text{Exp}(1)$,
2. sample z^*_j with replacement from the set of observed residuals $\{z_1, \ldots, z_{n_u}\}$, independently of $x^*_{1,j}$,
3. set $x^*_{2,j} = \hat{\alpha} x^*_{1,j} + z^*_j (x^*_{1,j})^{\hat{\beta}}$,

for $j = 1, \ldots, m$. Empirical probabilities can be obtained from these simulated values that, for large enough m, provide reliable estimates of the probabilities of extreme events. Suppose we want to estimate $\Pr\{(X_1, X_2) \in A\}$, where the set $A \in \mathbb{R}^2$ has values $X_1 > v$ for some $v > u$. Then, noting that $\Pr\{(X_1, X_2) \in A\} = \Pr\{(X_1, X_2) \in A \mid X_1 > v\} \Pr(X_1 > v)$, an estimate of this probability is given by

$$\widehat{\Pr}\{(X_1, X_2) \in A\} = \left\{ \frac{1}{m} \sum_{i=1}^{m} \mathbb{1}_{\{(x^*_{1,i}, x^*_{2,i}) \in A\}} \right\} \cdot e^{-v}.$$

Similar results hold in the multivariate case. Suppose we have variables $\boldsymbol{X} = (X_1, \ldots, X_d)$ with standard Gumbel marginal distributions, and we take the conditioning variable to be X_i for some $i \in \{1, \ldots, d\}$. The standardized residuals are defined by Heffernan and Tawn in this case as

$$\boldsymbol{Z}_{|i} = \frac{\boldsymbol{X}_{-i} - \boldsymbol{a}_{|i}(X_i)}{\boldsymbol{b}_{|i}(X_i)},$$

for normalizing functions $\boldsymbol{a}_{|i}$, $\boldsymbol{b}_{|i}$: $\mathbb{R} \to \mathbb{R}^{(d-1)}$, and $\boldsymbol{X}_{-i} = \{X_k : k \in \{1, \ldots, d\}, k \neq i\}$, and they assume that

$$\Pr(\boldsymbol{Z}_{|i} \leq \boldsymbol{z}_{|i} | X_i = x) \to G_{|i}(\boldsymbol{z}_{|i}),$$

as $x \to \infty$, where the marginal distributions of $G_{|i}(\boldsymbol{z}_{|i})$ are non-degenerate. From this assumption, for fixed $\boldsymbol{z}_{|i}$ and $x > 0$, they obtain the result

$$\Pr\left(\boldsymbol{Z}_{|i} \leq \boldsymbol{z}_{|i}, X_i - u > x \mid X_i > u\right) \to G_{|i}(\boldsymbol{z}_{|i}) e^{-x},$$

as $u \to \infty$. As in the bivariate case, this suggests that the vector of standardized variables $\boldsymbol{Z}_{|i}$ is asymptotically independent of $X_i - u$, for $X_i > u$ as $u \to \infty$. The normalizing functions are commonly set to be $\boldsymbol{a}_{|i}(x) = \{\alpha_{k|i} x : k \in \{1, \ldots, d\}, k \neq i\}$ and $\boldsymbol{b}_{|i}(x) = \{x^{\beta_{k|i}} : k \in \{1, \ldots, d\}, k \neq i\}$ with each $\alpha_{k|i} \in (0, 1]$ and each $\beta_{k|i} < 1$, analogously to the bivariate case. It is most straightforward to consider the tail dependence of each variable pairwise with the conditioning variable. In particular, if $\alpha_{k|i} = 1$ and $\beta_{k|i} = 0$, variables X_i and X_k are asymptotically dependent; if $\alpha_{k|i} \in (0, 1)$, X_i and X_k are asymptotically independent.

Heffernan and Tawn again propose a sampling procedure to obtain estimates of probabilities of extreme events using the above result. The algorithm is similar to the one in the bivariate case, although it should be noted that in step 2 the residuals must be sampled as full vectors of length $d - 1$ from the set of observed residuals $\{\boldsymbol{z}_{|i,1}, \ldots, \boldsymbol{z}_{|i,n_u}\}$ that occur with observations of $x_i > u$, although the simulated value in step 3 can be calculated separately for each variable.

2.4 Modelling extremes with underlying dependence structures

We have already discussed that variables can exhibit complex tail dependence properties, and that this is an important consideration when modelling multivariate ex-

tremes. Exploiting some underlying structure controlling the dependence of the variables can be useful in such cases. One example of this is in Markov models, which have been proposed to capture dependence structures when modelling time series extremes. This idea is studied from a probabilistic viewpoint by Rootzén (1988), Smith (1992) and Perfekt (1994), for example, while Smith et al. (1997) focus on the topic from a statistical perspective. In each of these instances, the focus is on asymptotic dependence between the variables; the case of asymptotic independence in Markov models is studied by Papastathopoulos et al. (2017).

Elsewhere in the literature, it has been proposed to exploit the structure of graphical models for extreme value modelling. Hitz and Evans (2016) consider graphical modelling of multivariate extremes by exploiting the assumption of regular variation, while Gissibl and Klüppelberg (2018) use a combination of directed acyclic graphs and max-linear models to impose structure on the extremal dependence properties of variables. Engelke and Hitz (2018) define conditional independence for multivariate extremes, allowing for the study of sparsity and graphical models in this setting. In each of these cases, the focus is on the asymptotic dependence setting, with this property being imposed on each link in the respective graphs.

In Chapter 5, we will study some of the tail properties of vine copulas, another class of models where the underlying dependence structure of the variables can be represented graphically. We introduce this class of models in Section 2.5.

2.5 Vine copulas

2.5.1 Introduction

In Section 2.3.2, we discussed copulas and their use in multivariate modelling. We now turn our attention to a particular class of these models, known as pair copula

constructions (PCCs). The idea behind these models is to exploit the wide range of bivariate copulas that are already known, to create models for higher dimensions, where there are fewer options available. This allows for the construction of flexible models with the possibility of capturing a wide range of dependence features. The idea was first proposed by Joe (1996); later developed by Bedford and Cooke (2001, 2002), who introduced the use of a type of graphical model called vines to aid the modelling procedure; hence the term 'vine copula'; and further studied by Aas et al. (2009). A thorough overview of these models is given in Kurowicka and Joe (2010). We focus on models for continuous variables here, although PCCs are also available for modelling discrete random variables.

2.5.2 Pair copula constructions

Recall the definition of a copula from Section 2.3.2. That is, the joint distribution function F of variables $\boldsymbol{X} = (X_1, \ldots, X_d)$ with $X_i \sim F_i$, for $i = 1, \ldots, d$, can be written in terms of a unique copula function C as

$$F(x_1, \ldots, x_d) = C\left\{F_1(x_1), \ldots, F_d(x_d)\right\}.$$

Differentiating this with respect to each variable gives the joint density function as

$$f(x_1, \ldots, x_d) = c\left\{F_1(x_1), \ldots, F_d(x_d)\right\} \prod_{i=1}^{d} f_i(x_i), \tag{2.5.1}$$

for $f_i(x_i)$, $i = 1, \ldots, d$, representing the marginal densities, and copula density

$$c(u_1, \ldots, u_d) = \partial^d C(u_1, \ldots, u_d) / \prod_{i=1}^{d} \partial u_i.$$

As outlined by Aas et al. (2009), the joint density can be decomposed as

$$\begin{aligned} f(x_1, \ldots, x_d) = {} & f_d(x_d) \cdot f_{d-1|d}(x_{d-1} \mid x_d) \\ & \cdot f_{d-2|d-1,d}(x_{d-2} \mid x_{d-1}, x_d) \ldots f_{1|2,\ldots,d}(x_1 \mid x_2, \ldots, x_d), \end{aligned} \tag{2.5.2}$$

and by repeatedly applying decomposition (2.5.1) to each term in the right-hand side of (2.5.2), it is possible to write the joint density of the variables $\boldsymbol{X} = (X_1, \ldots, X_d)$

in terms of only marginal and bivariate copula densities.

For instance, in the bivariate case, $f(x_1, x_2) = f_2(x_2) \cdot f_{1|2}(x_1 \mid x_2)$ from (2.5.2) and $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2) \cdot c_{12}\{F_1(x_1), F_2(x_2)\}$ from (2.5.1), so that

$$f_{1|2}(x_1 \mid x_2) = f_1(x_1) \cdot c_{12}\{F_1(x_1), F_2(x_2)\}.$$

Similarly, in the trivariate case,

$$\begin{aligned} f(x_1, x_2, x_3) &= f_3(x_3) \cdot f_{2|3}(x_2 \mid x_3) \cdot f_{1|23}(x_1 \mid x_2, x_3) \\ &= f_3(x_3) \cdot f_2(x_2) \cdot c_{23}\{F_2(x_2), F_3(x_3)\} \cdot f_{1|23}(x_1 \mid x_2, x_3). \end{aligned}$$

Again following Aas et al.,

$$\begin{aligned} f_{1|23}(x_1 \mid x_2, x_3) &= c_{13|2}\left\{F_{1|2}(x_1 \mid x_2), F_{3|2}(x_3 \mid x_2)\right\} \cdot f_{1|2}(x_1 \mid x_2) \\ &= c_{13|2}\left\{F_{1|2}(x_1 \mid x_2), F_{3|2}(x_3 \mid x_2)\right\} \cdot c_{12}\{F_1(x_1), F_2(x_2)\} \cdot f_1(x_1), \end{aligned}$$

so that a full decomposition of $f(x_1, x_2, x_3)$ is given by

$$\begin{aligned} f(x_1, x_2, x_3) =& f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \\ & \cdot c_{12}\{F_1(x_1), F_2(x_2)\} \cdot c_{23}\{F_2(x_2), F_3(x_3)\} \\ & \cdot c_{13|2}\left\{F_{1|2}(x_1 \mid x_2), F_{3|2}(x_3 \mid x_2)\right\}. \end{aligned}$$

For modelling purposes, different bivariate copula densities can be chosen for each of c_{12}, c_{23} and $c_{13|2}$, and different marginal distributions can be selected for each variable, i.e., F_1, F_2, and F_3, showing the flexibility in this class of model. A similar process can be applied to obtain models in higher than three dimensions in terms of bivariate copulas.

The decomposition of density f is not unique, as we have a choice about the conditioning variable used in each step of the decomposition. Bedford and Cooke (2001, 2002) proposed the use of regular vines, a class of graphical model, to represent the underlying structure of certain pair copula constructions and help to systematize the different possibilities.

2.5.3 Vine representations

An introduction to the vine representation of pair copula constructions is given in Chapter 3 of Kurowicka and Joe (2010), with formal definitions provided in Kurowicka and Cooke (2006); we give an overview here.

Suppose we are interested in modelling variables $\boldsymbol{X} = (X_1, \ldots, X_d)$. A regular vine corresponding to these d variables consists of $d-1$ connected trees labelled $T_1, \ldots, T_{d-1}$, with tree T_i having $d+1-i$ nodes and $d-i$ edges. The nodes in tree T_1 each have a different label in the set $\{1, \ldots, d\}$, and the edges are labelled according to the pair of nodes they connect. The labels of the nodes in tree T_{i+1} correspond to the labels of the edges in tree T_i, for $i = 1, \ldots, d-2$, creating a nested structure among the set of all trees. In tree T_i, $i \geq 2$, the pair of nodes connected by each edge will have $i-1$ variable labels in common; these become the conditioning variables in the corresponding edge label of T_i. Figure 2.5.1 gives an example of a regular vine for $d = 5$.

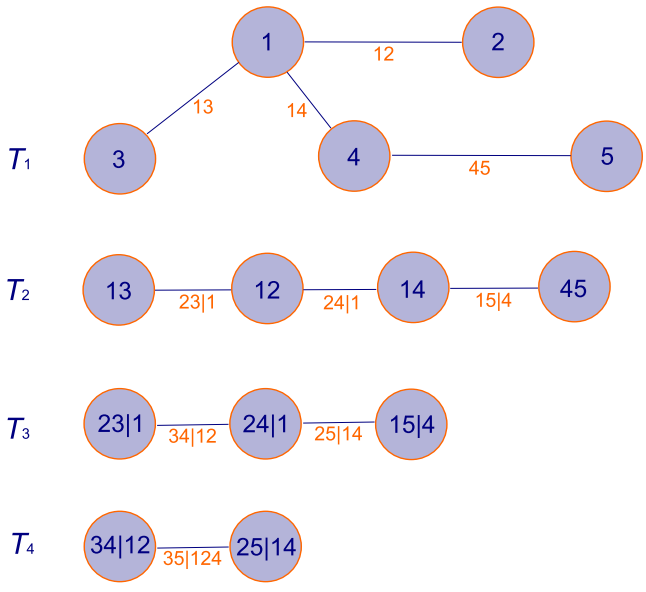

Figure 2.5.1: An example of a regular vine in five dimensions.

Each edge in a regular vine can be used represent one of the copula densities used in the pair copula construction. Not all PCCs can be represented in this way, but a large number can; these vine copulas are therefore a large subclass of PCCs. There are also certain subclasses of vine copula that are often of interest. These include D-vines, where each tree is a path, and C-vines, where each tree has exactly one node that is connected to all other nodes, with the latter class often being simpler to use in terms of inference. Figure 2.5.2 gives an example of these vine structures for $d = 4$. For the C-vine example, the corresponding decomposition of the density is

$$\begin{aligned} f(x_1, x_2, x_3, x_4) =& f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_4) \\ & \cdot c_{12}\left\{F_1(x_1), F_2(x_2)\right\} \cdot c_{13}\left\{F_1(x_1), F_3(x_3)\right\} \cdot c_{14}\left\{F_1(x_1), F_4(x_4)\right\} \\ & \cdot c_{23|1}\left\{F_{2|1}(x_2 \mid x_1), F_{3|1}(x_3 \mid x_1)\right\} \cdot c_{24|1}\left\{F_{2|1}(x_2 \mid x_1), F_{4|1}(x_4 \mid x_1)\right\} \\ & \cdot c_{34|12}\left\{F_{3|12}(x_3 \mid x_1, x_2), F_{4|12}(x_4 \mid x_1, x_2)\right\}, \end{aligned}$$

with the result for the D-vine found in a similar way.

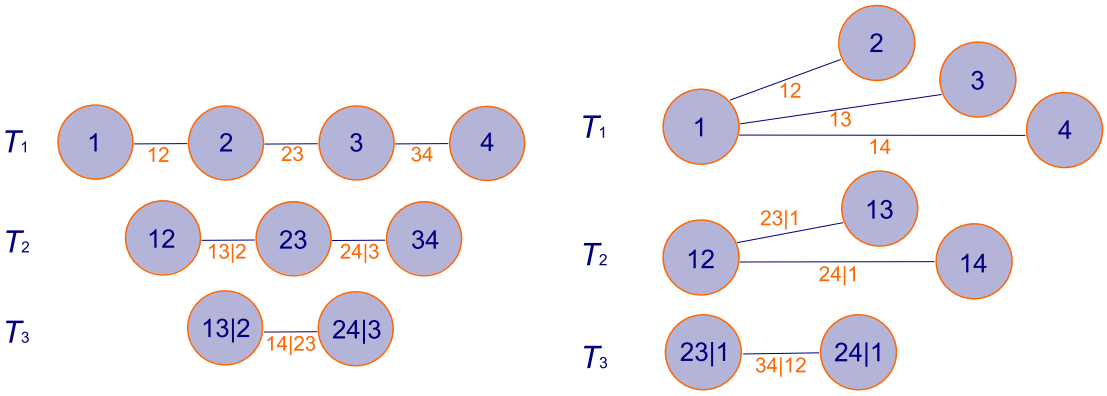

Figure 2.5.2: Four dimensional vine copula models; D-vine (left) and C-vine (right).

For modelling d variables, there are $d!/2$ possible D-vines, and the same number of possible C-vines (Aas et al., 2009). For $d = 3$, all vine structures are equivalent, with different decompositions only occurring with different labelling of the nodes. For $d = 4$, all possible structures fall into either the D-vine or C-vine category. For $d \geq 5$,

the more general regular vines provide a greater range of possible structures.

2.5.4 Tail behaviour of vine copulas

Joe et al. (2010) study some of the tail dependence properties of vine copulas by considering a range of multivariate and conditional tail dependence functions. They are particularly interested in the property of asymmetric dependence in the upper and lower tail, which can be captured by certain vine copulas if suitable linking copulas are chosen. Consider the vector of random variables $\boldsymbol{X} = (X_1, \ldots, X_d)$, with $X_i \sim F_i$ for each $i \in D = \{1, \ldots, d\}$. In equation (2.3.9), we gave the definition of the measure χ_C, which can be used to assess the upper tail dependence of variables $\boldsymbol{X}_C = \{X_i : i \in C\}$ with $C \in 2^D$ and $|C| \geq 2$. For the remainder of this section, we will denote this by $\chi_{U,C}$ in order to distinguish between upper and lower tail dependence measures. An equivalent measure to $\chi_{U,C}$ for the lower tail is $\chi_{L,C} \in [0, 1]$ defined by

$$\chi_{L,C} = \lim_{u \to 0} \frac{\Pr\{F_i(X_i) < u : i \in C\}}{u}.$$

In this case, if $\chi_{L,C} = 0$, the variables in $\boldsymbol{X}_C$ cannot all take their smallest values simultaneously and exhibit lower tail independence, while for $\chi_{L,C} \in (0, 1]$, the variables are lower tail dependent.

Joe et al. (2010) compare vine copulas to other standard multivariate models, including multivariate Gaussian and multivariate t copulas. The former exhibits no tail dependence, with $\chi_{U,C} = \chi_{L,C} = 0$ for all $C \in 2^D$ with $|C| \geq 2$; the latter has symmetry in the dependence of its upper and lower tail, $\chi_{U,C} = \chi_{L,C}$ for all C. These properties may not always be desirable in modelling multivariate extremes, as also discussed in Chapter 8 of Kurowicka and Joe (2010), where it is suggested that the flexibility of vine copulas means they can be used as approximations to a wide range of other copulas with varying tail dependence properties, i.e., vine copulas do not necessarily have $\chi_{U,C} = \chi_{L,C}$.

Joe et al. (2010) discuss the particular importance of the pair copulas in tree T_1. For a copula to have upper and lower tail dependence in each bivariate margin, each of the copulas in T_1 should have this property, with all other pair copulas having support on $(0,1)^2$. Moreover, if any pair copula in tree T_1 exhibits asymptotic independence in the upper or lower tail, then the overall vine copula will have asymptotic independence in that tail. For example, if we impose in T_1 that variables X_1 and X_2 are upper tail independent, with $\chi_{U,\{1,2\}} = 0$, then any subset of variables including both X_1 and X_2 cannot have all variables large at the same time, i.e., $\chi_{U,C} = 0$ for any set $C \supseteq \{1,2\}$. However, other subsets of variables could still exhibit upper tail dependence in this case, i.e., it is possible that $\chi_{U,C'} > 0$ for any $C' \not\supseteq \{1,2\}$.

In Chapter 5, we study the tail properties of vine copulas from a different perspective, by calculating the coefficient of tail dependence η_C of Ledford and Tawn (1996) for a variety of examples. As well as classifying the tail dependence of subsets of variables as being either asymptotically dependent or asymptotically independent, this allows us to also study any residual dependence that may be present in the latter case.

Chapter 3

Determining the Dependence Structure of Multivariate Extremes

3.1 Introduction

When constructing models in multivariate extreme value analysis, we often need to exploit extremal dependence features. Consider the random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$, with $X_i \sim F_i$, as well as a subset of these variables $\boldsymbol{X}_C = \{X_i : i \in C\}$, for some $C \in 2^D \setminus \emptyset$, i.e., C lies in the power set of $D = \{1, \ldots, d\}$ without the empty set. For any C with $|C| \geq 2$, extremal dependence within $\boldsymbol{X}_C$ can be summarized by

$$\chi_C = \lim_{u \to 1} \Pr\left\{F_i(X_i) > u : i \in C\right\} / (1 - u) \tag{3.1.1}$$

if the limit exists. In particular, if $\chi_C > 0$, the variables in $\boldsymbol{X}_C$ are asymptotically dependent, i.e., can take their largest values simultaneously. If $\chi_C = 0$, the variables in $\boldsymbol{X}_C$ cannot all take their largest values together, although it is possible that for some $\underline{C} \subset C$, $\chi_{\underline{C}} > 0$, see for example Hua and Joe (2011) or Wadsworth and Tawn (2013).

Many models for multivariate extremes are only applicable when data exhibit either full asymptotic dependence, entailing $\chi_C > 0$ for all $C \in 2^D \setminus \emptyset$ with $|C| \geq 2$, or

full asymptotic independence, i.e., $\chi_{i,j} = 0$ for all $i < j$ (Heffernan and Tawn, 2004). However, often some χ_C are positive whilst others are zero, i.e., only certain subsets of the variables take their largest values simultaneously, while the other variables are of smaller order. The extremal dependence between variables can thus have a complicated structure, which should be exploited when modelling. In this chapter, we present two methods for determining this structure.

The full extremal dependence structure is not completely captured by the $2^d - d - 1$ coefficients $\{\chi_C : C \in 2^D \setminus \emptyset, |C| \geq 2\}$ since we do not learn fully whether small values of some variables occur with large values of others, or whether individual variables can be extreme in isolation. This is revealed more clearly by decomposing the vector into radial and angular components, $(R, \boldsymbol{W})$, and examining their asymptotic structure. If the X_i follow a common heavy-tailed marginal distribution, usually achieved via a transformation, these pseudo-polar coordinates are defined as $R = \|\boldsymbol{X}\|_1$ and $\boldsymbol{W} = \boldsymbol{X} / \|\boldsymbol{X}\|_2$, for arbitrary norms $\|\cdot\|_1$ and $\|\cdot\|_2$. We take both to be the L_1 norm, and assume that $\boldsymbol{X}$ has standard Fréchet margins, so that $\Pr(X_i < x) = \exp(-1/x)$ for $x > 0$ and $i = 1, \ldots, d$. As such, the radial and angular components are $R = \sum_{i=1}^{d} X_i$ and $\boldsymbol{W} = \boldsymbol{X}/R$, respectively, with $R > 0$ and $\boldsymbol{W} \in \mathcal{S}_{d-1} = \left\{(w_1, \ldots, w_d) \in [0,1]^d : \sum_{i=1}^{d} w_i = 1\right\}$, the $(d-1)$-dimensional unit simplex. It follows that $\Pr(R > r) \sim a r^{-1}$ as $r \to \infty$, for $a \geq 1$, so all the information about extreme events is contained in $\boldsymbol{W}$, and in particular the distribution of $\boldsymbol{W}$ conditioned on $R > r$ as $r \to \infty$. Under the assumption of multivariate regular variation (Resnick, 2007, Chapter 6),

$$\lim_{t \to \infty} \Pr(R > tr, \boldsymbol{W} \in B \mid R > t) = H(B) r^{-1}, \quad r \geq 1, \tag{3.1.2}$$

for B a measurable subset of $\mathcal{S}_{d-1}$, where the limiting spectral measure H satisfies

$$\int_{\mathcal{S}_{d-1}} w_i \, dH(\boldsymbol{w}) = 1/d, \quad i = 1, \ldots, d. \tag{3.1.3}$$

As the radial component becomes large, the position of mass on $\mathcal{S}_{d-1}$ reveals the

extremal dependence structure of $\boldsymbol{X}$. We note the link between the dependence measure χ_C in (3.1.1), and the spectral measure H: if $\chi_C > 0$, then H places mass on at least one region $\mathcal{S}_{d-1}^{\overline{C}} = \left\{(w_1, \ldots, w_d) \in [0,1]^d : \sum_{i \in \overline{C}} w_i = 1\right\}$, with $C \subseteq \overline{C} \subseteq D$. This underlines that the term asymptotic dependence is not so useful here, since it offers only partial insight into the structure. In what follows, we thus avoid this term where possible, talking instead about faces of the simplex on which H places mass.

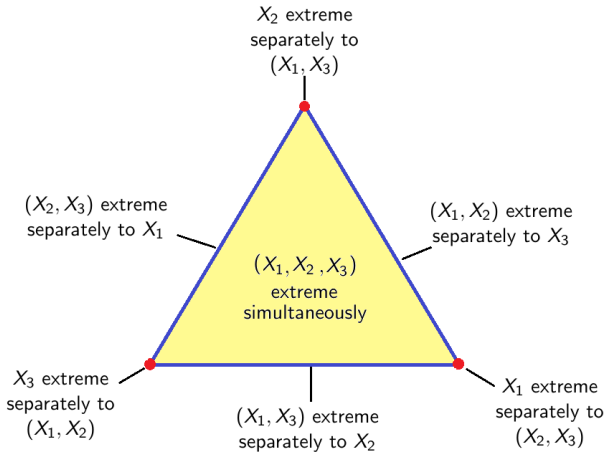

Figure 3.1.1: The simplex $\mathcal{S}_2$. Coordinates are transformed to the equilateral simplex.

In the d-dimensional case, $\mathcal{S}_{d-1}$ can be partitioned into $2^d - 1$ faces, each of which could contain mass. Mass on each of these faces corresponds to a different subset of the variables $(X_1, \ldots, X_d)$ being the only ones taking their largest values concurrently. This is demonstrated in Figure 3.1.1 for $d = 3$. For high, or even moderate, dimensions, there are many faces to consider, and the task of determining which faces truly contain mass, and therefore the extremal dependence structure of the variables, is not straightforward, because for a finite sample with continuous margins, we will never observe points lying exactly on the boundary of the simplex: no W_i equal zero when $R < \infty$.

The multivariate regular variation assumption (3.1.2) can also be phrased in terms of measures on the cone $\mathbb{E} = [0, \infty]^d \setminus \{0\}$, see Section 3.2.1. Each face of $\mathcal{S}_{d-1}$ can

be identified with a sub-cone of $\mathbb{E}$ for which one or more components are identically zero. Intuition and visualization are often simpler with H, but in the sequel we work with $\mathbb{E}$ and the sub-cones corresponding to faces of $\mathcal{S}_{d-1}$. Variants of our methods that directly use the radial-angular framework are presented in Chapter 4.

The problem of determining the extremal dependence structure of variables has been recently studied elsewhere in the literature. Under the assumption that the data are from an elliptical copula, Klüppelberg et al. (2015) use factor analysis on extreme correlations linked to the tail dependence function. Chautru (2015) introduces a non-parametric approach based on statistical learning, combining a principal component analysis algorithm with clustering techniques. A Bayesian clustering method is proposed by Vettori et al. (2018), based on the hierarchical dependence structure of the nested logistic distribution of Tawn (1990). Goix et al. (2016, 2017) propose a non-parametric simplex partitioning approach in which they condition on the radial variable being above some high threshold. They assume that there is mass on a particular face if the number of points in the corresponding region of the simplex is sufficiently large, leading to a sparse representation of the dependence structure. Chiapino and Sabourin (2017) propose an algorithm to group together nearby faces with extremal mass into feature clusters, by exploiting their graphical structure and a measure of extremal dependence. Finally, Chiapino et al. (2019) extend this approach by instead using the coefficient of tail dependence of Ledford and Tawn (1996).

In this chapter we exploit additional, but commonly satisfied, hidden regular variation assumptions on non-standard sub-cones of $\mathbb{E}$, by introducing a new set of parameters that describes the dominant extremal dependence structure. We study properties of these parameters, their link to existing coefficients, and explore their values for a range of examples. Estimation of the parameters provides us with an asymptotically-motivated framework for determining the extremal dependence struc-

ture, as well as allowing us to estimate the proportion of mass associated with each set of variables. We propose two such inferential methods, both with computational complexity $O(dn \log n)$ if $d < n$, for d representing the number of variables and n the number of data points. This is the same complexity as the method of Goix et al. (2017).

3.2 Theoretical motivation

3.2.1 Multivariate regular variation

A function $\lambda : (0, \infty] \to (0, \infty]$ is said to be regularly varying at infinity, with index $\alpha \in \mathbb{R}$, if $\lambda(tx)/\lambda(t) \to x^\alpha$, as $t \to \infty$, for all $x > 0$. For such functions, we write $\lambda \in \mathrm{RV}_\alpha$. We can always express $\lambda(x) = L(x)x^\alpha$, with $L \in \mathrm{RV}_0$ termed a slowly varying function. A cone $G \subset \mathbb{R}^d$ is a set such that for any $\boldsymbol{x} \in G$, $t\boldsymbol{x} \in G$ for all $t > 0$. The assumption of multivariate regular variation on the cone G means that there exists a scaling function $a(t) \to \infty$, and a positive measure μ, such that

$$t \Pr\left(\boldsymbol{X}/a(t) \in \cdot\right) \to \mu(\cdot), \quad t \to \infty, \tag{3.2.1}$$

with vague convergence in the space of non-negative Radon measures on G (Resnick 2007, Chapter 3). If we assume that the margins of $\boldsymbol{X}$ are standard Fréchet or Pareto, we may take $a(t) = t$, and the limit measure μ is homogeneous of order -1. For the remainder of this section, we assume that $\boldsymbol{X}$ has standard Fréchet marginal distributions.

3.2.2 Hidden regular variation

The concept of hidden regular variation was introduced by Resnick (2002), who formalized and extended the ideas of Ledford and Tawn (1996, 1997). Further work has been done by Maulik and Resnick (2004) and Mitra and Resnick (2011), for example, whilst Resnick (2007) provides a textbook treatment. Here, multivariate regular

variation is assumed on some cone in $\mathbb{R}^d$. If there is also regular variation, but with a scaling function of smaller order, on some sub-cone, we have hidden regular variation on that sub-cone.

To our knowledge, the marginal case of this hidden regular variation framework is the only one previously exploited from a statistical perspective; from a theoretical viewpoint, Das et al. (2013) consider hidden regular variation on a series of non-standard cones, although these are mostly different from the ones we will consider. For $\boldsymbol{X}_C = \{X_i : i \in C\}$, $\boldsymbol{x}_C = \{x_i : i \in C\}$, with $C \subseteq D$, Ledford and Tawn (1997) considered multivariate regular variation on the cone $\mathbb{E} = [0, \infty]^d \setminus \{\mathbf{0}\}$ and hidden regular variation on

$$\mathbb{E}_C^* = \left\{ \boldsymbol{x} \in \mathbb{E} : \boldsymbol{x}_C \in (0, \infty]^{|C|},\ \boldsymbol{x}_{D \setminus C} \in [0, \infty]^{|D \setminus C|} \right\}, \tag{3.2.2}$$

with limit measures on $\mathbb{E}_C^*$ homogeneous of order $-1/\eta_C$, and the so-called coefficient of tail dependence η_C taking values in $(0, 1]$. If $\mu(\mathbb{E}_C^*) > 0$, variables $\boldsymbol{X}_C$ can take their largest values simultaneously. If we instead consider sub-cones of $\mathbb{E}$ of the form

$$\mathbb{E}_C = \left\{ \boldsymbol{x} \in \mathbb{E} : \boldsymbol{x}_C \in (0, \infty]^{|C|},\ \boldsymbol{x}_{D \setminus C} \in \{0\}^{|D \setminus C|} \right\}, \tag{3.2.3}$$

where $\{0\}^m$ denotes an m-vector of zeros, then having $\mu(\mathbb{E}_C) > 0$ indicates that variables in $\boldsymbol{X}_C$ can take their largest values simultaneously while variables in $\boldsymbol{X}_{D \setminus C}$ are of smaller order. Our task is to determine the sub-cones $\mathbb{E}_C$ on which μ places mass, equivalent to the problem of detecting where H places mass on $\mathcal{S}_{d-1}$, thus revealing the extremal dependence structure of $\boldsymbol{X}$. For simplicity, we assume that if μ places mass on $\mathbb{E}_C$, then for any measurable $B_C \subset \mathbb{E}_C$, $\mu(B_C) > 0$. More generally, it is only necessary that there exists $B_C \subset \mathbb{E}_C$ such that $\mu(B_C) > 0$, however if the mass lies only in very restricted parts of $\mathbb{E}_C$, then the task of detecting which sub-cones contain mass is naturally more difficult.

For a finite sample, mass will not occur on sub-cones $\mathbb{E}_C$ with $|C| < d$; this is equivalent to all mass for $\boldsymbol{W}$ being placed in the interior of the simplex in Figure 3.1.1. One option is to truncate the variables to zero below some marginal threshold, to ensure mass on at least some of these sub-cones at a finite level. Let us define

$$X^* = \begin{cases} 0, & X \leq -1/\log p, \\ X, & X > -1/\log p, \end{cases} \tag{3.2.4}$$

such that p is the quantile at which we truncate. The variable X^* has the same tail behaviour as X, but in general $\Pr(\boldsymbol{X}^*/t \in B_C) > 0$ for $B_C \subset \mathbb{E}_C$, and in this way we could define a hidden regular variation assumption on $\mathbb{E}_C$. Writing $B_C = \{\boldsymbol{x} \in \mathbb{E} : \boldsymbol{x}_C \in B, \boldsymbol{x}_{D\setminus C} \in \{0\}^{|D\setminus C|}\}$, for some $B \subset (0,\infty]^{|C|}$, then $\Pr(\boldsymbol{X}^*/t \in B_C) = \Pr(\boldsymbol{X}_C/t \in B, \boldsymbol{X}_{D\setminus C} \in [0, -1/\log p]^{|D\setminus C|})$, such that we consider the behaviour when the variables $\boldsymbol{X}_C$ are growing at a common rate, but variables $\boldsymbol{X}_{D\setminus C}$ have a fixed upper bound. However, the latter condition does not capture all possible behaviour that leads to variables $\boldsymbol{X}_{D\setminus C}$ being of smaller order than $\boldsymbol{X}_C$, and in general a more elaborate assumption is needed. We consider how we can allow $\boldsymbol{X}_{D\setminus C}$ to be upper bounded by a function that is growing, but at a potentially slower rate than t.

Define the set $(y,\infty]^C \times [0,z]^{D\setminus C} = \{\boldsymbol{x} \in \mathbb{E} : x_i > y, i \in C; x_j \leq z, j \in D \setminus C\}$. Then under the regular variation assumption (3.2.1),

$$t\Pr\left\{\boldsymbol{X}/t \in (y,\infty]^C \times [0,z]^{D\setminus C}\right\} \to \mu\left((y,\infty]^C \times [0,z]^{D\setminus C}\right) \geq \mu\left((y,\infty]^C \times \{0\}^{D\setminus C}\right). \tag{3.2.5}$$

Therefore, if $\mu\left((y,\infty]^C \times \{0\}^{D\setminus C}\right) > 0$, and hence $\mu\left(\mathbb{E}_C\right) > 0$, this indicates that in (3.2.5) we may be able to consider $z = z_t \to 0$ at a suitable rate in t and still observe

$$\lim_{t\to\infty} t\Pr\left\{\boldsymbol{X}/t \in (y,\infty]^C \times [0,z_t]^{D\setminus C}\right\} > 0. \tag{3.2.6}$$

A consequence of a positive limit in (3.2.6) is that $\Pr\{\boldsymbol{X}/t \in (y,\infty]^C \times [0,z_t]^{D\setminus C}\} \in \mathrm{RV}_{-1}$. When the limit in (3.2.6) is zero, then either $z_t \to 0$ too quickly — consider for

example the case $z_t \equiv 0$ — or μ places no mass on $\mathbb{E}_C$. In these cases we focus on the rate of convergence to zero in (3.2.6). Taking $z_t = zt^{\delta-1}$ for $\delta \in [0,1]$, and rephrasing in terms of min and max projections, our main assumption is as follows.

Assumption 1. *Suppose we have regular variation on the cone* $\mathbb{E} = [0,\infty]^d \setminus \{\mathbf{0}\}$*, so that equation* (3.2.1) *is satisfied with* μ *homogeneous of order* -1*. For* $C \subseteq D$*, let* $X_\wedge^C = \min_{i\in C} X_i$ *and* $X_\vee^{D\setminus C} = \max_{i\in D\setminus C} X_i$*. We assume that*

$$Pr\left\{\left(X_\wedge^C/t, X_\vee^{D\setminus C}/t^\delta\right) \in (y,\infty]\times[0,z]\right\} \in \mathrm{RV}_{-1/\tau_C(\delta)}, \quad t\to\infty, \quad 0<y,z<\infty, \tag{3.2.7}$$

for $\delta \in [0,1]$*, and that there exists* $\delta^* < 1$ *such that* $\tau_C(\delta^*) = 1$ *for all* C *such that* $\mu(\mathbb{E}_C) > 0$*, and* $\tau_C(\delta^*) < 1$ *for all* C *such that* $\mu(\mathbb{E}_C) = 0$*.*

We note that the probability in (3.2.7), and hence $\tau_C(\delta)$, is non-decreasing in δ. The case $\delta = 0$ and $z = -1/\log p$ is identical to a regular variation assumption on the truncated variables $\boldsymbol{X}^*$; allowing $\delta > 0$ produces a more diverse range of possibilities.

Through the final line of Assumption 1, the indices $\tau_C(\delta)$ contain information on the limiting extremal dependence structure; the challenge is to find a suitable δ^*, noting that if δ is too small we could have $\tau_C(\delta) < 1$ even when $\mu(\mathbb{E}_C) > 0$, but if δ is too large, some $\tau_C(\delta)$ could be close to one even when $\mu(\mathbb{E}_C) = 0$, making the detection problem difficult in light of statistical uncertainty. These issues are discussed further below and in Section 3.3.

Overall, examining the regular variation properties in Assumption 1 leads to understanding of the sub-asymptotic behaviour of μ in relation to which of the sub-cones $\mathbb{E}_C$ are charged with mass. This is analogous to determining the support of H in (3.1.2). In the remainder of this section, we illustrate the utility and validity of our hidden regular variation assumption via examples, and discuss properties of the indices $\tau_C(\delta)$. Theorems 1 and 2 clarify some links between $\tau_C(\delta)$ and η_C.

Theorem 1. *Assume (hidden) regular variation on $\mathbb{E}_C^*$ defined in* (3.2.2)*, such that $Pr\left(X_\wedge^C > t\right) \in \mathrm{RV}_{-1/\eta_C}$. Suppose further that for all $\overline{C} \subseteq D$ such that $\overline{C} \supseteq C$, Assumption 1 is satisfied for $\delta = 1$, and $Pr\left(X_\wedge^{\overline{C}} > t, X_\vee^{D\setminus\overline{C}} \leq t\right) \in \mathrm{RV}_{-1/\tau_{\overline{C}}(1)}$. Then $\eta_C = \max_{\overline{C}:C\subseteq\overline{C}} \tau_{\overline{C}}(1)$.*

Proof. We have

$$\Pr\left(X_\wedge^C > t\right) = \Pr\left(\boldsymbol{X}/t \in (1,\infty]^C \times [0,\infty]^{D\setminus C}\right),$$

$$\Pr\left(X_\wedge^{\overline{C}} > t, X_\vee^{D\setminus\overline{C}} \leq t\right) = \Pr\left(\boldsymbol{X}/t \in (1,\infty]^{\overline{C}} \times [0,1]^{D\setminus\overline{C}}\right).$$

By the partition

$$(1,\infty]^C \times [0,\infty]^{D\setminus C} = \bigcup_{\overline{C}:C\subseteq\overline{C}} (1,\infty]^{\overline{C}} \times [0,1]^{D\setminus\overline{C}},$$

$$\Pr\left(\boldsymbol{X}/t \in (1,\infty]^C \times [0,\infty]^{D\setminus C}\right) = \sum_{\overline{C}:C\subseteq\overline{C}} \Pr\left(\boldsymbol{X}/t \in (1,\infty]^{\overline{C}} \times [0,1]^{D\setminus\overline{C}}\right),$$

from which the result follows. □

We note $\tau_D(\delta)$ does not depend on δ; we therefore denote it by τ_D.

Theorem 2. *For all $C \in 2^D \setminus \emptyset$ with $|C| \geq 2$, assume (hidden) regular variation on $\mathbb{E}_C^*$ with coefficient of tail dependence η_C, and suppose Assumption 1 holds for all $\delta \in [0,1]$. For any set C with $|C| \geq 2$, if $\eta_{\overline{C}} < \eta_C$ for all $\overline{C} \supset C$, then $\tau_C(1) = \eta_C$, and $\tau_C(\delta) \leq \eta_C$ for all $\delta \in [0,1]$.*

Proof. Since $\mathbb{E}_D = \mathbb{E}_D^*$, we have $\eta_D = \tau_D$, and by Theorem 1, for any set $C_{d-1} \subset D$ with $|C_{d-1}| = d-1$,

$$\eta_{C_{d-1}} = \max\left\{\tau_{C_{d-1}}(1), \tau_D\right\} = \max\left\{\tau_{C_{d-1}}(1), \eta_D\right\}.$$

Since, by assumption, $\eta_{C_{d-1}} > \eta_D$, we have $\eta_{C_{d-1}} = \tau_{C_{d-1}}(1)$. Similarly, for any set $C_{d-2} \subset C_{d-1}$ with $|C_{d-2}| = d-2$,

$$\eta_{C_{d-2}} = \max\left\{\tau_{C_{d-2}}(1), \tau_{C_{d-1}}(1), \tau_D\right\} = \max\left\{\tau_{C_{d-2}}(1), \eta_{C_{d-1}}, \eta_D\right\}.$$

Again, since $\eta_{C_{d-2}} > \eta_{C_{d-1}} > \eta_D$ for all $C_{d-2} \subset C_{d-1}$, then $\eta_{C_{d-2}} = \tau_{C_{d-2}}(1)$. The result $\tau_C(1) = \eta_C$ follows by iteration for any set C with $|C| \geq 2$. Since $\tau_C(\delta)$ is non-decreasing in δ, $\tau_C(1) = \max_{\delta \in [0,1]} \tau_C(\delta)$, so $\tau_C(\delta) \leq \eta_C$ for all $\delta \in [0,1]$. □

In Appendix A.1 and Appendix A.2, respectively, we calculate the value of $\tau_C(\delta)$, with $C \in 2^D \setminus \emptyset$, for a range of bivariate and multivariate copulas. For the bivariate case, we restrict our investigation to a subclass of bivariate extreme value distributions (Tawn, 1988) that covers all possible structures for sub-cones charged with mass, focusing on the case where the spectral density is regularly varying at 0 and 1. For multivariate cases there are many more possibilities, so we study certain trivariate extreme value distributions (Tawn, 1990), which have $\chi_C > 0$ for at least one set $|C| \geq 2$, and two classes of copula having $\chi_C = 0$ for all $|C| \geq 2$. The results are summarized here.

The bivariate extreme value distribution in standard Fréchet margins has distribution function of the form $F(x,y) = \exp\{-V(x,y)\}$ for some exponent measure

$$V(x,y) = 2\int_0^1 \max\{w/x, (1-w)/y\}\, dH(w), \quad x,y > 0, \tag{3.2.8}$$

where H denotes the spectral measure defined in equation (3.1.2) on the unit simplex $[0,1]$. In the bivariate case, there are three sub-cones: $\mathbb{E}_1$, $\mathbb{E}_2$ and $\mathbb{E}_{1,2}$. If $H(\{0\}) = \theta_2 \in [0, 1/2]$ and $H(\{1\}) = \theta_1 \in [0, 1/2]$, the distribution places mass θ_2, θ_1, $\theta_{1,2} = 1 - (\theta_1 + \theta_2)$ in the sub-cones. If $\theta_1 + \theta_2 = 1$, the variables are independent, and $\mu(\mathbb{E}_{1,2}) = 0$. In this case, all the limiting mass is placed on $\mathbb{E}_1$ and $\mathbb{E}_2$. Here, Assumption 1 holds for $C = \{1\}, \{2\}$ and $\{1,2\}$ with $\tau_1(\delta) = \tau_2(\delta) = 1$ for all $\delta \in [0,1]$, and $\tau_{1,2} = \eta_{1,2} = 1/2$.

When $\theta_1 + \theta_2 < 1$, $\mu(\mathbb{E}_{1,2}) > 0$ and $\tau_{1,2} = \eta_{1,2} = 1$, i.e., both variables can be simultaneously large. If $\theta_i > 0$, it follows that $\tau_i(\delta) = 1$ for $\delta \in [0,1]$ and $i = 1, 2$, and there is mass on the corresponding sub-cone $\mathbb{E}_i$. However, when $\theta_1 = \theta_2 = 0$, there is

no mass on either of these sub-cones, and additional conditions are required for (3.2.7) to hold. We suppose that H is absolutely continuous on $(0,1)$ with Lebesgue density $h(w) = dH(w)/dw$ satisfying $h(w) \sim c_1(1-w)^{s_1}$ as $w \to 1$, and $h(w) \sim c_2 w^{s_2}$ as $w \to 0$, for $s_1, s_2 > -1$ and $c_1, c_2 > 0$. In Appendix A.1, we show that for $i = 1, 2$, $\tau_i(\delta) = \{(s_i + 2) - \delta(s_i + 1)\}^{-1}$. To illustrate this final case, consider the bivariate extreme value distribution with the logistic dependence structure (Tawn, 1988), with $V(x,y) = \left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}$ and

$$h(w) = \frac{1}{2}\left(\alpha^{-1} - 1\right)\left\{w^{-1/\alpha} + (1-w)^{-1/\alpha}\right\}^{\alpha-2}\left\{w(1-w)\right\}^{-1-1/\alpha}, \tag{3.2.9}$$

for $0 < w < 1$ and $\alpha \in (0,1)$. For this model $s_1 = s_2 = -2 + 1/\alpha$, and so $\tau_1(\delta) = \tau_2(\delta) = \alpha/(1 + \alpha\delta - \delta)$ which increases from $\tau_i(\delta) = \alpha < 1$ at $\delta = 0$ to $\tau_i(\delta) = 1$ at $\delta = 1$.

	Copula	$\lvert C\rvert = 1$	$\lvert C\rvert = 2$	$\lvert C\rvert = 3$
(i)	Independence	$\tau_1(\delta) = \tau_2(\delta) = \tau_3(\delta) = 1$	$\tau_{1,2}(\delta) = \tau_{1,3}(\delta) = \tau_{2,3}(\delta) = 1/2$	$\tau_{1,2,3} = 1/3$
(ii)	Independence and bivariate logistic	$\tau_1(\delta) = \tau_2(\delta) = \frac{\alpha}{1+\alpha\delta-\delta}, \tau_3(\delta) = 1$	$\tau_{1,2}(\delta) = 1, \tau_{1,3}(\delta) = \tau_{2,3}(\delta) = \frac{\alpha}{\alpha\delta+1+\alpha-\delta}$	$\tau_{1,2,3} = 1/2$
(iii)	Trivariate logistic	$\tau_1(\delta) = \tau_2(\delta) = \tau_3(\delta) = \frac{\alpha}{1+\alpha\delta-\delta}$	$\tau_{1,2}(\delta) = \tau_{1,3}(\delta) = \tau_{2,3}(\delta) = \frac{\alpha}{2+\alpha\delta-2\delta}$	$\tau_{1,2,3} = 1$
(iv)	Trivariate inverted logistic	$\tau_1(\delta) = \tau_2(\delta) = \tau_3(\delta) = 1$	$\tau_{1,2}(\delta) = \tau_{1,3}(\delta) = \tau_{2,3}(\delta) = 2^{-\alpha}$	$\tau_{1,2,3} = 3^{-\alpha}$

Table 3.2.1: Values of $\tau_C(\delta)$ for some trivariate copula examples. For all logistic models the dependence parameter α satisfies $0 < \alpha < 1$, with larger α values corresponding to weaker dependence.

When $d = 3$, there are many more possibilities for combinations of sub-cones with mass. Table 3.2.1 gives $\tau_C(\delta)$ for four examples, in each case identifying $\tau_C(\delta)$ on sub-cones of dimension one, two and three. Cases (i)-(iii) in Table 3.2.1 are all special cases of the trivariate extreme value copula. Case (i) is the independence copula, which has limit mass on $\mathbb{E}_1$, $\mathbb{E}_2$ and $\mathbb{E}_3$. For the d-dimensional independence copula, $\tau_C(\delta) = m^{-1}$ for a sub-cone of dimension $m \leq d$, and does not depend on the value of δ. Case (ii) is the copula corresponding to variables (X_1, X_2) following a bivariate extreme value logistic distribution (3.2.9), independent of X_3. Here all the limit mass

is placed on $\mathbb{E}_{1,2}$ and $\mathbb{E}_3$. Again, $\tau_C(\delta)$ differs between sub-cones with C of different dimension. The trivariate extreme value logistic model, case (iii), places all extremal mass on $\mathbb{E}_{1,2,3}$, so that $\tau_{1,2,3} = 1$, and $\tau_C(\delta) < 1$ when $\delta < 1$, for all C with $|C| < 3$. Since this is a symmetric model, $\tau_C(\delta)$ is the same on all three sub-cones with $|C| = 1$, and is also equal for each sub-cone with $|C| = 2$.

Copula (iv) is the inverted extreme value copula (Ledford and Tawn, 1997), with a symmetric logistic dependence model. It places all limiting mass on sub-cones with $|C| = 1$, but unlike the independence copula, has sub-asymptotic dependence, reflected by the values of $\tau_C(\delta)$ for sub-cones with $|C| = 2, 3$, which are closer to one than in the independence case. The values of $\tau_C(\delta)$ do not depend on δ for this case.

The Gaussian copula with covariance matrix Σ also exhibits asymptotic independence, with all limit mass on $\mathbb{E}_C$ with $|C| = 1$. We study the values of $\tau_C(\delta)$ for the trivariate case in the Appendix A.2.6. For sets $C = \{i\}$, $i = 1, 2, 3$, $\tau_C(\delta) = 1$ only if $\delta \geq \max\left(\rho_{ij}^2, \rho_{ik}^2\right)$, where $\rho_{ij} > 0$ is the Gaussian correlation parameter for variables i and j; otherwise, $\tau_C(\delta) < 1$. We also know that $\tau_{1,2,3} = \eta_{1,2,3} = \left(\mathbf{1}_3^T \Sigma^{-1} \mathbf{1}_3\right)^{-1}$, with $\mathbf{1}_d \in \mathbb{R}^d$ denoting a vector of 1s. For $C = \{i, j\}$, $i < j$, under Assumption 1, Theorem 2 leads to $\tau_C(1) = \eta_C = \left(\mathbf{1}_2^T \Sigma_{i,j}^{-1} \mathbf{1}_2\right)^{-1}$, with $\Sigma_{i,j}$ denoting the submatrix of Σ corresponding to variables i and j, provided the correlations satisfy $1 + \rho_C \neq \sum_{C' : |C'| = 2, C' \neq C} \rho_{C'}$; for $\delta < 1$, $\tau_C(\delta) \leq \eta_C$.

3.3 Methodology

3.3.1 Introduction to methodology

The coefficient $\tau_C(\delta)$ defined in Assumption 1 reveals whether the measure μ places mass on the sub-cone $\mathbb{E}_C$. For $\mu(\mathbb{E}_C) > 0$, we assume there exists $\delta^* < 1$ such that $\tau_C(\delta^*) = 1$, but we could still have $\tau_C(\delta) < 1$ for values of $\delta < \delta^*$. For sub-cones

with $\mu(\mathbb{E}_C) = 0$, the detection problem becomes easier the further $\tau_C(\delta)$ is from 1, and since $\tau_C(\delta)$ is non-decreasing in δ it is ideal to take δ as small as possible. We therefore have a trade-off between choosing δ large enough that $\tau_C(\delta) = 1$ on sub-cones with extremal mass, but small enough that $\tau_C(\delta)$ is not close to 1 on sub-cones without extremal mass. For the examples in Table 3.2.1, we could take $\delta = 0$, since the sub-cones with $\mu(\mathbb{E}_C) > 0$ have $\tau_C(\delta) = 1$ for all $\delta \in [0, 1]$. However, the Gaussian case reveals that although $\mu(\mathbb{E}_i) > 0$, for $i = 1, \ldots, d$, we can have $\tau_i(0) < 1$, so it is necessary to take $\delta > 0$ for correct identification of sub-cones with extremal mass.

We therefore introduce two approaches for determining the extremal dependence structure of a set of variables. In the first method we set $\delta = 0$, and apply a truncation to the variables $\boldsymbol{X}$ by setting any values below some marginal threshold equal to zero. This transformation is analogous to the approach of Goix et al. (2017), who partition the non-negative orthant in a similar way, but we additionally exploit Assumption 1. In our second method, we consider $\delta > 0$ when exploiting the regular variation assumption. As well as aiming to determine the extremal dependence structure, both methods estimate the proportion of extremal mass associated with each sub-cone $\mathbb{E}_C$.

3.3.2 Method 1: $\delta = 0$

We apply Assumption 1 with δ equal to zero by applying truncation (3.2.4) to variables $\boldsymbol{X}$ for some choice of p. Recall that the cone $\mathbb{E}$ equals $\bigcup_{C \in 2^D \setminus \emptyset} \mathbb{E}_C$, with the components of the union disjoint and defined as in (3.2.3). We wish to partition $\mathbb{E}$ with approximations to $\mathbb{E}_C$, by creating regions where components indexed by C are large and those not in C are small. This is achieved via regions of the form

$$E_C = \left\{ \boldsymbol{x}^* \in \mathbb{E} : \boldsymbol{x}^*_C \in (-1/\log p, \infty]^{|C|},\ \boldsymbol{x}^*_{D \setminus C} \in \{0\}^{|D \setminus C|} \right\}. \tag{3.3.1}$$

Define the variable $Q = \min(X_i^* : X_i^* > 0, i = 1, \ldots, d)$, and recall that we denote $X_\wedge^C = \min_{i \in C} X_i$ and $X_\vee^{D \setminus C} = \max_{i \in D \setminus C} X_i$. Under Assumption 1, as $q \to \infty$,

$$\Pr(Q > q \mid \boldsymbol{X}^* \in E_C) \propto \Pr\left(X_\wedge^C > q, X_\vee^{D \setminus C} < -1/\log p\right) \in \mathrm{RV}_{-1/\tau_C(0)},$$

so that $\Pr(Q > q \mid \boldsymbol{X}^* \in E_C) = L_C(q) q^{-1/\tau_C(0)}$ for some slowly varying function L_C. We now let $\tau_C = \tau_C(0)$, and assume that the model

$$\Pr(Q > q \mid \boldsymbol{X}^* \in E_C) = K_C q^{-1/\tau_C}, \quad q > u_C, \tag{3.3.2}$$

holds for a high threshold u_C, with $\tau_C \in (0, 1]$ and $K_C > 0$ for all $C \in 2^D \setminus \emptyset$. Here, the slowly varying function L_C is replaced by the constant K_C as a modelling assumption, removing the possibility of having $L_C(q) \to 0$ as $q \to \infty$.

Model (3.3.2) may be fitted using a censored likelihood approach. Suppose that we observe n_C values $q_1, \ldots q_{n_C}$ of Q in E_C. The censored likelihood associated with E_C is

$$L_C(K_C, \tau_C) = \prod_{j=1}^{n_C} \left(1 - K_C u_C^{-1/\tau_C}\right)^{\mathbb{1}_{\{q_j \le u_C\}}} \left(\frac{K_C}{\tau_C} q_j^{-1-1/\tau_C}\right)^{\mathbb{1}_{\{q_j > u_C\}}}, \tag{3.3.3}$$

with u_C a high threshold. Analytical maximization of (3.3.3) leads to closed-form estimates of (K_C, τ_C), with the latter corresponding to the Hill estimate (Hill, 1975). In particular,

$$\hat{\tau}_C = \left(\sum_{j=1}^{n_C} \mathbb{1}_{\{q_j > u_C\}}\right)^{-1} \sum_{j=1}^{n_C} \mathbb{1}_{\{q_j > u_C\}} \log\left(\frac{q_j}{u_C}\right), \quad \hat{K}_C = \left(\frac{\sum_{j=1}^{n_C} \mathbb{1}_{\{q_j > u_C\}}}{n_C}\right) u_C^{1/\hat{\tau}_C}.$$

This estimate of τ_C can exceed 1, so we prefer to use $\min(\hat{\tau}_C, 1)$, with an appropriate change to $\hat{K}_C$. The Hill estimator for τ_C is consistent if $u_C \to \infty$, $\sum_j \mathbb{1}_{\{q_j > u_C\}} \to \infty$ and $\sum_j \mathbb{1}_{\{q_j > u_C\}} / n_C \to 0$; the assumption of $L_C(q) \sim K_C > 0$ is not required for this. The second condition ensures that the number of points in E_C with $Q > u_C$ goes to infinity, and since the expected number $n_C \Pr(Q > u_C \mid \boldsymbol{X}^* \in E_C) \sim n_C K_C u_C^{-1/\tau_C}$, this entails $u_C = o(n_C^{\tau_C})$.

The method of Goix et al. (2017) produces empirical estimates of $\Pr(\boldsymbol{X} \in E_C \mid R > r_0)$, for $R = \|\boldsymbol{X}\|_\infty$, and some value of r_0 within the range of observed values. These estimates are then assumed to hold for all $R > r_0$ and are used to approximate the limit. If the conditional probability $\Pr(\boldsymbol{X} \in E_C \mid R > r)$ changes with $r > r_0$, Goix et al. estimate this as a positive constant or as zero. In contrast, our semiparametric method allows us to estimate $\Pr(\boldsymbol{X}^* \in E_C \mid Q > q)$, for all q above a high threshold, via

$$\Pr(\boldsymbol{X}^* \in E_C \mid Q > q) = \frac{\Pr(Q > q \mid \boldsymbol{X}^* \in E_C)\Pr(\boldsymbol{X}^* \in E_C)}{\sum_{C' \in 2^D \setminus \emptyset} \Pr(Q > q \mid \boldsymbol{X}^* \in E_{C'})\Pr(\boldsymbol{X}^* \in E_{C'})}, \quad C \in 2^D \setminus \emptyset, \tag{3.3.4}$$

with E_C as in (3.3.1). Our estimate of probability (3.3.4) varies continuously with q, with this variation being determined by the estimated values $\hat{\tau}_C$, for $C \in 2^D \setminus \emptyset$. In situations where sub-asymptotic dependence leads to many points in a region E_C, but $\mu(\mathbb{E}_C) = 0$ and $\hat{\tau}_C < 1$, this extrapolation can be helpful in obtaining a better approximation to the limit. The relative merits of these differences to the approach of Goix et al., which are common to Methods 1 and 2, are illustrated in Sections 3.4 and 3.5.

The right-hand side of equation (3.3.4) consists of two types of component. We estimate terms of the form $\Pr(\boldsymbol{X}^* \in E_C)$ empirically, and we estimate those of the form $\Pr(Q > q \mid \boldsymbol{X}^* \in E_C)$ as in (3.3.2) by replacing K_C and τ_C by their estimates, and evaluating for some large choice of q, discussed in Section 3.5. This approach yields an estimate for the proportion of mass in each region. We denote the estimated vector of these proportions by $\hat{\boldsymbol{p}} = (\hat{p}_C : C \in 2^D \setminus \emptyset)$. In order to obtain a sparse representation of the mass on the simplex, we follow Goix et al. (2016, 2017) and ignore any mass that has been detected which is considered to be negligible; see item 4 below. Our method can be summarized as follows.

1. Transform the data to standard Fréchet margins, and for a choice of the tuning parameter p, apply transformation (3.2.4).

2. Assign each transformed observation to a region E_C as in (3.3.1), removing any all-zero points.

3. For each region E_C containing more than m points, fit model (3.3.2) for a choice of threshold u_C, and estimate $\Pr(\boldsymbol{X}^* \in E_C \mid Q > q)$ for a large value of q by equation (3.3.4). Set $\Pr(\boldsymbol{X}^* \in E_C \mid Q > q) = 0$ in the remaining regions. Denote the resulting estimate by $\hat{p}_C$.

4. If $\hat{p}_C < \pi$, for a choice of the tuning parameter π, set $\hat{p}_C$ to zero, renormalizing the resulting vector.

The parameter m in step 3 ensures there are enough points to estimate the parameters on each sub-cone. In simulations, where our aim was to estimate the proportion of mass associated with each sub-cone, it was found not to have a significant effect on results; we therefore take $m = 1$. If the parameters τ_C are directly of interest, a higher value of m would be necessary for reliable estimation.

3.3.3 Method 2: $\delta > 0$

An alternative to setting $\delta = 0$ and partitioning the positive orthant using regions E_C, is to consider $\delta > 0$ in the application of Assumption 1, specifically $\Pr\left(X_\wedge^C > t, X_\vee^{D\setminus C} \leq t^\delta\right) \in \mathrm{RV}_{-1/\tau_C(\delta)}$. However, unlike with $\delta = 0$, this does not lead directly to a univariate structure variable with tail index $1/\tau_C(\delta)$. We instead consider $\Pr\left\{X_\wedge^C > t, X_\vee^{D\setminus C} \leq \left(X_\wedge^C\right)^\delta\right\} = \Pr\left(X_\wedge^C > t, \boldsymbol{X} \in \tilde{E}_C\right)$, with $\tilde{E}_C$ defined as

$$\tilde{E}_C = \left\{\boldsymbol{x} \in \mathbb{E} : x_\vee^{D\setminus C} \leq \left(x_\wedge^C\right)^\delta\right\},\ |C| < d; \qquad \tilde{E}_D = \mathbb{E} \setminus \bigcup_{C \in 2^D \setminus \emptyset : |C| < d} \tilde{E}_C,$$

for each $C \in 2^D \setminus \emptyset$. We denote the corresponding tail index as $1/\tilde{\tau}_C(\delta)$, and assume

$$\Pr\left(X_\wedge^C > q, \boldsymbol{X} \in \tilde{E}_C\right) \in \mathrm{RV}_{-1/\tilde{\tau}_C(\delta)}, \qquad q \to \infty.$$

Analogously to equation (3.3.2) of Method 1, for each region $\tilde{E}_C$, we assume the model

$$\Pr(X_\wedge^C > q \mid \boldsymbol{X} \in \tilde{E}_C) = K_C q^{\,-1/\tilde{\tau}_C(\delta)}, \quad q > u_C, \tag{3.3.5}$$

for some large threshold u_C, where estimates of K_C and $\tilde{\tau}_C(\delta)$ are again obtained by maximizing a censored likelihood. In Appendix A.3.1, we examine estimates of $\tilde{\tau}_C(\delta)$, which we find to reasonably approximate the true values of $\tau_C(\delta)$. This indicates that the indices $\tilde{\tau}_C(\delta)$ provide useful information about $\tau_C(\delta)$. We note that the regions $\tilde{E}_C$ are not disjoint. Supposing we have observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, we obtain an empirical estimate of $\Pr(\boldsymbol{X} \in \tilde{E}_C)$ using

$$\frac{1}{n}\sum_{j=1}^{n} \frac{\mathbb{1}_{\{\boldsymbol{x}_j \in \tilde{E}_C\}}}{\sum_{C \in 2^D \setminus \emptyset} \mathbb{1}_{\{\boldsymbol{x}_j \in \tilde{E}_C\}}}, \tag{3.3.6}$$

so that the contribution of each observation sums to one. Combining equations (3.3.5) and (3.3.6), we estimate

$$\Pr\left(X_\wedge^C > q, \boldsymbol{X} \in \tilde{E}_C\right) = \Pr(X_\wedge^C > q \mid \boldsymbol{X} \in \tilde{E}_C)\Pr(\boldsymbol{X} \in \tilde{E}_C), \qquad C \in 2^D \setminus \emptyset, \tag{3.3.7}$$

for some large q. To estimate the proportion of extremal mass associated with each sub-cone, we consider probability (3.3.7) for a given $\tilde{E}_C$ divided by the sum over all such probabilities, corresponding to $\tilde{E}_{C'}$, $C' \in 2^D \setminus \emptyset$. The result is evaluated at a high threshold q, similarly to step 3 of Method 1, and negligible mass is removed as in step 4 of Method 1.

3.4 Simulation study

3.4.1 Overview and metrics

We present simulations to demonstrate Methods 1 and 2, and compare them with the approach of Goix et al. (2017). Here, we consider a max-mixture distribution involving Gaussian and extreme value logistic distributions, described in equation (3.4.2). In Appendix A.3.3, we present results for a special case of this, the asymmetric logistic distribution (Tawn, 1990), that is used by Goix et al. (2017) to assess the performance of their methods. The key difference between these two distributions

is that the Gaussian components in the max-mixture model lead to sub-asymptotic dependence, in contrast to independence, on certain sub-cones. Hence, distinguishing between sub-cones with and without limiting mass is a more difficult task for our max-mixture distribution. We test the efficacy of the methods as classifiers using receiver operating characteristic curves. For the classes of model we consider, it is possible to calculate the proportion of extremal mass on the various sub-cones analytically, allowing us to compare our estimates to the true distribution of mass using the Hellinger distance.

When incorporating a cut-off for sparse representation of the measure μ, as mentioned in Section 3.3.2, the methods can be viewed as classification techniques. Plotting receiver operating characteristic curves is a common method for testing the efficacy of classifiers (Hastie et al., 2009). To obtain such curves, the false positive rate of a method is plotted against the true positive rate, as some parameter of the method varies. In our case, the false positive rate is the proportion of sub-cones incorrectly detected as having mass, while the true positive rate is the proportion of correctly detected sub-cones. To obtain our curves, we vary the threshold, π, above which estimated mass is considered non-negligible. For $\pi = 0$, all sub-cones will be included in the estimated dependence structure, leading to the true and false positive rates both being 1, while $\pi = 1$ includes none of the sub-cones, so both equal 0. A perfect result for a given data set and method would be a false positive rate of 0 and true positive rate of 1: the closer the curve is to the point $(0, 1)$, the better the method. This is often quantified using the area under the curve, with values closer to 1 corresponding to better methods.

Such curves only assess whether the sub-cones determined as having mass are correct. We may also wish to take into account sub-cones detected as having mass that are close to the truth. We therefore propose an extension of the receiver operating

characteristic curve, which we term the neighboured receiver operating characteristic curve, that takes into account mass on adjacent sub-cones. For the sub-cone corresponding to the set C, we define adjacent sub-cones as those corresponding to any sets $\underline{C}$ and $\overline{C}$ with $\underline{C} \subset C \subset \overline{C}$ such that $|\underline{C}| = |C| - 1$ and $|\overline{C}| = |C| + 1$. When constructing neighboured receiver operating characteristic curves, in place of the true positive rate, we consider the proportion of sub-cones that should be detected which are either detected or have an adjacent sub-cone that is detected. The false positive rate is replaced by the proportion of sub-cones that do not truly have mass and are not adjacent to any of the true sub-cones with mass, that are detected as having mass. Interpretation of these curves is analogous to that of standard receiver operating characteristic curves, and one may also consider the area under the curve as a measure for the success of the classifier.

Let $\boldsymbol{p} = \left(p_C; C \in 2^D \setminus \emptyset\right)$ denote the true proportion of mass on each sub-cone, and denote its estimate by $\hat{\boldsymbol{p}}$. The Hellinger distance between $\boldsymbol{p}$ and $\hat{\boldsymbol{p}}$,

$$\mathrm{HD}(\boldsymbol{p}, \hat{\boldsymbol{p}}) = \frac{1}{\sqrt{2}} \left\{ \sum_{C \in 2^D \setminus \emptyset} \left(p_C^{1/2} - \hat{p}_C^{1/2} \right)^2 \right\}^{1/2}, \tag{3.4.1}$$

is used to determine the precision of the estimated proportions. In particular, $\mathrm{HD}(\boldsymbol{p}, \hat{\boldsymbol{p}}) \in [0, 1]$, and equals 0 if and only if $\boldsymbol{p} = \hat{\boldsymbol{p}}$. The closer $\mathrm{HD}(\boldsymbol{p}, \hat{\boldsymbol{p}})$ is to 0, the better $\boldsymbol{p}$ is estimated by $\hat{\boldsymbol{p}}$. Errors on small proportions are penalized more heavily than errors on large proportions. A small positive mass on a region, estimated as zero, will incur a relatively heavy penalty.

3.4.2 Max-mixture distribution

Segers (2012) shows how to construct distributions that place extremal mass on different combinations of sub-cones. Here, we take a different approach by considering max-mixture models with asymptotic and sub-asymptotic dependence in different sub-cones. This can be achieved by using a mixture of extreme value logistic and

multivariate Gaussian copulas, a particular example of which we consider here.

Let $\boldsymbol{Z}_C = \left(Z_{i,C} : i \in C\right)$ be a $|C|$-dimensional random vector with standard Fréchet marginal distributions, and $\{\boldsymbol{Z}_C : C \in 2^D \setminus \emptyset\}$ be independent random vectors. Define the vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ with components

$$X_i = \max_{C \in 2^D \setminus \emptyset : i \in C} \left(\theta_{i,C} Z_{i,C}\right), \quad \theta_{i,C} \in [0,1], \qquad \sum_{C \in 2^D \setminus \emptyset : i \in C} \theta_{i,C} = 1, \tag{3.4.2}$$

for $i = 1, \ldots, d$. The constraints on $\theta_{i,C}$ ensure that $\boldsymbol{X}$ also has standard Fréchet margins. The random vector $\boldsymbol{Z}_C$ may exhibit asymptotic dependence, in which case mass will be placed on the sub-cone $\mathbb{E}_C$, or it may exhibit asymptotic independence, in which case mass will be placed on the sub-cones $\mathbb{E}_i$, $i \in C$.

Here, we consider one particular five-dimensional example. We define $\boldsymbol{Z}_{1,2}$ and $\boldsymbol{Z}_{4,5}$ to have bivariate Gaussian copulas with correlation parameter ρ, and $\boldsymbol{Z}_{1,2,3}$, $\boldsymbol{Z}_{3,4,5}$ and $\boldsymbol{Z}_{1,2,3,4,5}$ to have three-dimensional and five-dimensional extreme value logistic copulas with dependence parameter α. The bivariate Gaussian distribution is asymptotically independent with sub-asymptotic dependence, while the logistic distribution is asymptotically dependent for $\alpha \in (0,1)$. As such, the sub-cones with mass resulting from this construction are $\mathbb{E}_1, \mathbb{E}_2, \mathbb{E}_4, \mathbb{E}_5, \mathbb{E}_{1,2,3}, \mathbb{E}_{3,4,5}$ and $\mathbb{E}_{1,2,3,4,5}$. The Gaussian components mean that sub-cones $\mathbb{E}_{1,2}$ and $\mathbb{E}_{4,5}$ have no mass asymptotically, but the parameter ρ controls the decay rate of the mass. We assign equal mass to each of the seven charged sub-cones by setting

$$\theta_{1,2} = (5,5)/7, \quad \theta_{4,5} = (5,5)/7,$$

$$\theta_{1,2,3} = (1,1,3)/7, \quad \theta_{3,4,5} = (3,1,1)/7, \quad \theta_{1,2,3,4,5} = (1,1,1,1,1)/7.$$

In this model, the sub-cones with mass are fixed, in contrast to the asymmetric logistic examples in Appendix A.3.3, where following Goix et al. (2017), they are chosen at random over different simulation runs. Setting $\rho = 0$ in this max-mixture distribution

gives an asymmetric logistic model.

Each setting in the simulation study is repeated 100 times, taking samples of size 10,000. In Method 1, we set $p = 0.5$, u_C to be the 0.75 quantile of observed Q values in region E_C for each $C \in 2^D \setminus \emptyset$, and the value of q for which we estimate $\Pr(\boldsymbol{X}^* \in E_C \mid Q > q)$ to be the 0.9999 quantile of all observed Q values. In Method 2, we set $\delta = 0.5$, each threshold u_C to be the 0.85 quantile of observed $X_\wedge^C$ values in region $\tilde{E}_C$, and the extrapolation level q to be the 0.9999 quantile of observed values of $\boldsymbol{X}$. The parameters in the method of Goix et al. (2017) are chosen to be $(\epsilon, k) = (0.1, n^{1/2})$, using notation from that paper. When calculating the Hellinger distances, we used $\pi = 0.001$ as the value above which estimated mass is considered significant in all three methods. The tuning parameters are not optimized for individual data sets, but fixed at values that we have found to work well across a range of settings. In Section 3.4.3, we discuss stability plots, which could be used as guide as to which tuning parameter values may be sensible for a given set of data. In Section 3.5, we present tables showing how the estimated extremal dependence structure changes as the tuning parameters vary for a particular data set, allowing us to further examine this mass stability and choose a reasonable value of p in Method 1 or δ in Method 2.

In Figure 3.4.1, we show the mean Hellinger distance achieved by each method for $\rho \in \{0, 0.25, 0.5, 0.75\}$ and $\alpha \in [0.1, 0.9]$. Results for the area under the receiver operating characteristic curves are provided in Table 3.4.1. The performance of all three methods deteriorates as the value of the correlation parameter ρ, or the dependence parameter α, increases. In the former case this is due to the stronger sub-asymptotic dependence on sub-cones without extremal mass; in the latter case, larger values of α in logistic component $\boldsymbol{Z}_C$ lead to larger values of $\tau_{\underline{C}}(\delta)$ for $\underline{C} \subset C$, so it is harder to determine which sub-cones truly contain extremal mass. In terms of the Hellinger distance, Method 1 is the most successful for $\rho = 0, 0.25$, although its performance

deteriorates when there is stronger correlation in the Gaussian components. Method 2 yields the best results for $\rho = 0.5, 0.75$. In terms of estimating the proportion of extremal mass associated with each sub-cone, at least one of our proposed methods is always more successful than Goix et al. for this max-mixture model. The results in Table 3.4.1 reveal that all three methods are successful classifiers for low values of ρ and α. For $\alpha = 0.75$ and $\rho = 0, 0.25$, Method 1 and the approach of Goix et al. demonstrate similarly strong performance, while for $\rho = 0.5, 0.75$, Method 2 again provides the best results.

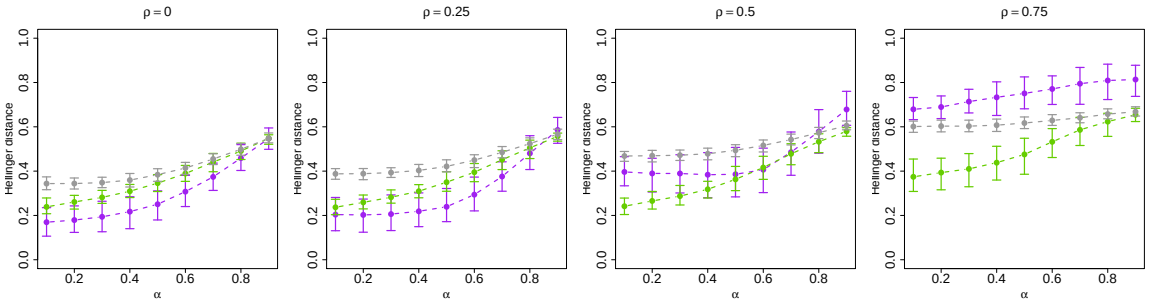

Figure 3.4.1: Mean Hellinger distance, 0.05 and 0.95 quantiles over 100 simulations. Method 1: purple; Method 2: green; Goix et al.: grey.

	$\rho = 0$			$\rho = 0.25$			$\rho = 0.5$			$\rho = 0.75$		
α	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
Goix et al.	100 (0.0)	100 (0.0)	98.0 (1.1)	99.7 (0.4)	99.8 (0.4)	96.3 (1.4)	92.3 (0.6)	91.9 (0.5)	90.1 (1.2)	91.0 (1.0)	90.1 (1.7)	87.6 (1.2)
Method 1	100 (0.0)	100 (0.1)	97.7 (1.4)	100 (0.1)	99.9 (0.3)	96.7 (1.2)	97.3 (1.6)	96.3 (1.9)	91.5 (1.9)	92.9 (1.0)	90.0 (0.9)	87.5 (0.2)
Method 2	100 (0.0)	99.2 (0.7)	96.0 (1.6)	100 (0.1)	98.9 (0.8)	94.6 (1.8)	99.5 (0.6)	97.5 (1.1)	92.7 (1.7)	94.4 (1.9)	92.9 (1.9)	89.1 (2.0)

Table 3.4.1: Average area under the receiver operating characteristic curves, given as percentages, for 100 samples from a five-dimensional mixture of bivariate Gaussian and extreme value logistic distributions; the standard deviation of each result is given in brackets.

Table 3.4.2 provides results for the average area under the neighboured receiver operating characteristic curves obtained in our simulations. In terms of this criterion,

	$\rho = 0$			$\rho = 0.25$			$\rho = 0.5$			$\rho = 0.75$		
α	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
Goix et al.	100 (0.0)	100 (0.0)	96.0 (1.1)	100 (0.0)	100 (0.0)	96.8 (1.5)	100 (0.0)	100 (0.0)	98.5 (0.9)	100 (0.0)	100 (0.0)	98.6 (0.5)
Method 1	100 (0.0)	100 (0.0)	95.9 (1.2)	100 (0.0)	100 (0.0)	94.8 (1.5)	100 (0.0)	100 (0.0)	92.0 (3.1)	100 (0.0)	100 (0.0)	96.2 (3.4)
Method 2	100 (0.0)	98.0 (1.7)	95.2 (0.2)	100 (0.0)	97.7 (1.7)	95.2 (0.2)	100 (0.0)	97.5 (1.3)	94.4 (1.4)	100 (0.0)	98.3 (0.7)	89.2 (2.5)

Table 3.4.2: Average area under the neighboured receiver operating characteristic curves, given as percentages, for 100 samples from a five-dimensional mixture of bivariate Gaussian and extreme value logistic distributions; the standard deviation of each result is given in brackets.

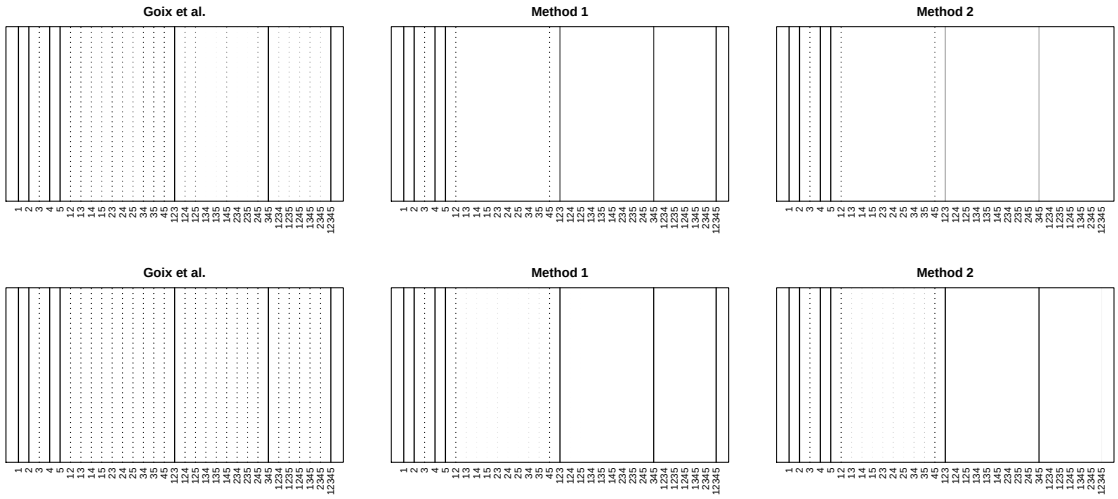

Figure 3.4.2: Plots to show the number of times each sub-cone is assigned mass greater than $\pi = 0.01$ (top) and $\pi = 0.001$ (bottom), for $(\alpha, \rho) = (0.75, 0.5)$. Darker lines correspond to higher detection rates over 100 simulations. True sub-cones with mass: solid lines; sub-cones without mass; dashed lines.

the method of Goix et al. (2017) is generally the most successful, with Method 2 generating the worst results. We investigate why this might be the case in Figure 3.4.2, where we demonstrate how often each sub-cone is detected as having mass above $\pi = 0.01, 0.001$ for the $(\alpha, \rho) = (0.75, 0.5)$ case. For $\pi = 0.001$, the approach of Goix et al. places mass on around three times as many sub-cones as Methods 1 and 2, and over twice as many for the $\pi = 0.01$ case, so our methods provide sparser representations

of the extremal mass that are both much closer to the truth. The reason for this difference is explained by the method of Goix et al. assuming there is extremal mass on a sub-cone $\mathbb{E}_C$ if $\Pr(\boldsymbol{X} \in E_C \mid R > r_0) > \pi$, whereas we recognize that when $\hat{\tau}_C < 1$ or $\hat{\tau}_C(\delta) < 1$, non-limit mass can be on a sub-cone at a finite threshold, but may progressively decrease to zero as the level of extremity of the vector variable is increased to infinity. When $\hat{\tau}_C = 1$ or $\hat{\tau}_C(\delta) = 1$, we estimate mass on sub-cone $\mathbb{E}_C$ similarly to Goix et al. As a consequence, our approach integrates information over the entire tail to estimate which sub-cones have limit mass, as opposed to Goix et al., who use information only at a single quantile. We also observe from Figure 3.4.2 that Method 2 often fails to detect the sub-cone corresponding to all five variables being large simultaneously, and places more mass on lower-dimensional sub-cones, arising from the estimated values of $\tau_C(\delta)$. Method 1 also places mass on these lower-dimensional sub-cones, but more often detects the true higher-dimensional sub-cones with mass.

3.4.3 Stability plots

One way to decide on reasonable tuning parameter values for a given set of data is via a parameter stability plot. Here, we outline how to construct such a plot for an example using the max-mixture distribution of Section 3.4.2 with Method 2, where our aim is to obtain a sensible range of values for the tuning parameter δ, by considering the region of δ where the number of sub-cones determined as having mass is stable.

For $\delta \in \{0.05, 0.075, \ldots, 0.95\}$, we use Method 2 to estimate the proportion of extremal mass on each sub-cone, and find the number of sub-cones whose estimated mass is greater than $\pi = 0.001$ in each case. The remaining parameters are fixed as in Section 3.4.2. Figure 3.4.3 shows the estimates of the number of sub-cones, with a 95% confidence interval constructed from 250 bootstrapped samples: these constitute our stability plot. Analogous plots can be created to choose p in Method 1, or in

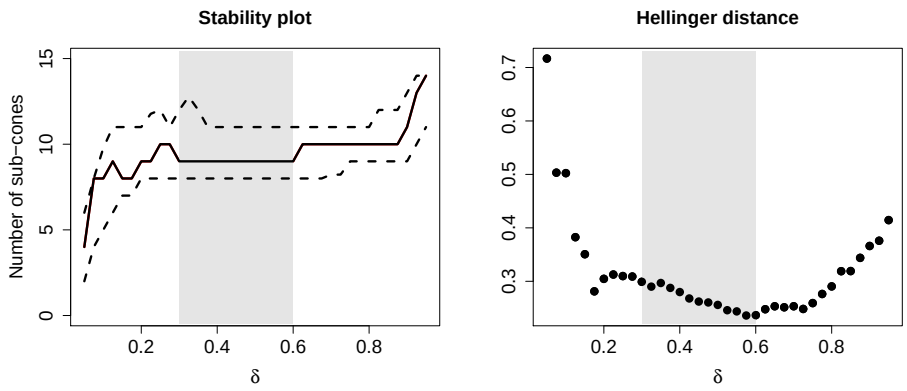

Figure 3.4.3: Stability plot (left) for Method 2, with dashed lines showing a 95% bootstrapped confidence interval for the number of sub-cones with mass, and a plot of the Hellinger distance (right) for each value of δ. The shaded regions correspond to the stable range of tuning parameter values. Data were simulated from the max-mixture distribution of Section 3.4.2 with $n = 10,000$, $\alpha = 0.25$ and $\rho = 0.25$.

each case to choose π. In practice, the choice of threshold π should depend on the dimension of the data; this is not explored here.

The number of sub-cones detected as having mass is most stable for values of δ between 0.3 and 0.6, indicated by the shaded regions in Figure 3.4.3, suggesting values of δ in this range may be appropriate for this sample. The right-hand panel of Figure 3.4.3 shows the Hellinger distance corresponding to the set of estimated proportions obtained for each value of δ. For this particular sample, although values of δ within the stable range slightly overestimate the number of sub-cones with mass, the smallest Hellinger distance occurs for a value of δ within the stable range, and the Hellinger distance is reasonably consistent across these tuning parameter values. In practice, the true proportions on each sub-cone are unknown, so Hellinger plots cannot be constructed; the plot here supports the idea of using stability plots in choosing suitable tuning parameter values. There is no guarantee that stability plots will find the optimal tuning parameter values, but they do offer some insight into

tuning parameter optimization. Consideration of the context of the problem may be useful in determining whether it is reasonable for extremal mass to be placed on particular combinations of sub-cones, and such insight could facilitate the choice of different p or δ values for different sub-cones.

3.5 River flow data

We apply Methods 1 and 2 to daily mean river flow readings from 1980 to 2013, measured in cubic metres per second, at five gauging stations in the North West of England. These data are available from the Centre for Ecology and Hydrology at `nrfa.ceh.ac.uk` (Morris and Flavin, 1990, 1994). Estimates of the extremal dependence structure of the flows could be used to aid model selection, or one could carry out density estimation on each sub-cone to give an overall model.

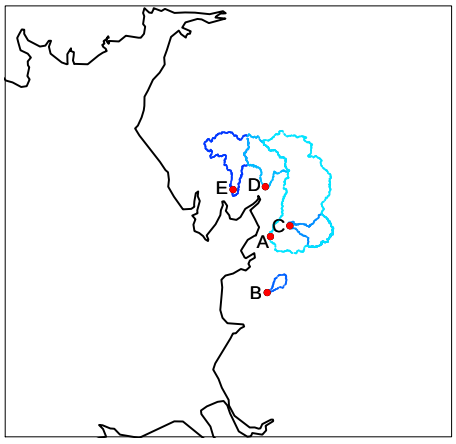

Figure 3.5.1: Locations of the river flow gauges (labelled A to E) and corresponding catchment boundaries.

The locations of the five gauges are shown in Figure 3.5.1; the labels assigned to each location will be used to describe the dependence structures estimated in this section. Figure 3.5.1 also illustrates the boundaries of the catchments associated with

each gauge. These catchments demonstrate the areas from which surface water, usually as a result of precipitation, will drain to each gauge. The spatial dependence of river flow is studied by Keef et al. (2013b) and Asadi et al. (2015). As high river flow is mainly caused by heavy rainfall, we may observe extreme river flow readings at several locations simultaneously if they are affected by the same extreme weather event. Gauges with adjacent or overlapping catchments are expected to take their largest values simultaneously, with stronger dependence between gauges that are closer together.

Table 3.5.1 shows the percentage of the extremal mass assigned to each sub-cone for tuning parameter values $p \in \{0.7, 0.725, \ldots, 0.975\}$ and $\delta \in \{0.2, 0.25, \ldots, 0.75\}$. We set $\pi = 0.01$ to be the threshold below which the proportion of mass is deemed negligible, and the extrapolation levels q to be the 0.999 quantile of the observed Q and $\boldsymbol{X}$ values in Methods 1 and 2, respectively. Remaining parameters are fixed as in Section 3.4.2. By observing how the estimated dependence structure changes over a range of tuning parameter values, we aim to find a 'stable region' in which the results are most reliable. A further consideration is whether the tuning parameters give a feasible estimate of the extremal dependence structure. In particular, each variable should be represented on at least one sub-cone, and moment constraint (3.1.3) should be taken into account. For Method 2, Table 3.5.1 indicates that for $\delta \geq 0.45$, the sub-cone corresponding to location E is assigned more than 20% of the extremal mass, which is not possible due to the moment constraint. Feasible stable regions are demonstrated by the shaded regions in Table 3.5.1. For Method 2, one could also look for a value of δ that give estimates of $\tau_C(\delta)$ satisfying $\max_{C:C\supseteq i} \hat{\tau}_C(\delta) = 1$, subject to estimation uncertainty, for every $i = 1, \ldots, d$.

Focusing on tuning parameter values within each of the stable regions in Table 3.5.1, Method 1 suggests the dependence structure to be {B, E, ABCD, ABCDE}, while Method 2 suggests {B, E, ABC, ABCD, ABCDE}. All the sub-cones detected

p	B	C	D	E	AC	AD	BC	ABC	ACD	ADE	ABCD	ACDE	ABCDE
0.700	2										3		95
0.725	2										4		93
0.750	1										4		95
0.775	1										11		88
0.800	7										11		82
0.825	10			3							16		71
0.850	13			7							17		63
0.875	12			12							19	1	55
0.900	8			31							15		46
0.925	6			57				2	2		15		19
0.950	5			52				4	2		16		21
0.975	14	1		73		2	1					2	6

δ	B	C	D	E	AC	AD	BC	ABC	ACD	ADE	ABCD	ACDE	ABCDE
0.20											1		99
0.25	4										15		81
0.30	3		1					7			24		65
0.35	8			1				9			33		49
0.40	14			3				2			49		30
0.45	27		1	26			1	4			29		12
0.50	22		1	49			1	2			15	2	7
0.55	16			52		1	1	3			18	4	6
0.60	13			55				2	2		19	5	4
0.65	16		3	48		2	2	3			17	6	3
0.70	16	2	4	45		2	3	3	4		13	6	3
0.75	19	3	4	40	1	2	3	5	4	1	11	4	2

Table 3.5.1: The percentage of mass assigned to each sub-cone for varying values of the tuning parameters in Method 1 (left) and Method 2 (right). The grey regions demonstrate the feasible stable ranges.

by Method 2 are either also detected by Method 1, or are neighbours of sub-cones detected by Method 1, showing there is some agreement between the methods. If we had used a higher threshold for the negligible mass, say $\pi = 0.1$, for tuning parameter values in the stable region, both methods would have detected mass on sub-cones {B, ABCD, ABCDE}. We also investigated the behaviour of the methods using the 0.99 and 0.9999 quantiles for extrapolation level q. For both methods, the set of sub-cones estimated as having mass was stable, but for the lower quantile Method 2 placed less mass on ABCDE, and there was more mass assigned to this sub-cone at the higher quantile.

The subsets of locations detected as having simultaneously high river flows seem feasible when considering the geographic positions of the gauging stations. For instance, both methods suggest mass on ABCD; as station E lies towards the edge of the region under consideration, it is possible for weather events to affect only the other four locations. Both methods also suggest that locations B and E can experience high river flows individually; this seems reasonable as they lie at the edge of region we consider. The catchment of gauge C lies entirely within the catchment of gauge A. We observe that location C occurs with location A in the subsets of sites determined

to take their largest values simultaneously, which may be a consequence of this nested structure.

p	A	B	C	D	AC	AD	BC	ABC	ACD	ABCD
0.700										100
0.725										100
0.750										100
0.775										100
0.800										100
0.825		3								97
0.850		3								97
0.875		4								96
0.900		2								98
0.925		5						1		94
0.950		7				2		5	2	84
0.975		35	3			1	2		7	52

δ	A	B	C	D	AC	AD	BC	ABC	ACD	ABCD
0.20										100
0.25										100
0.30		2								98
0.35		4								96
0.40		9								91
0.45		23		1				2	1	72
0.50		28		2		1	1	1	6	60
0.55		19	1	4		2	2	3	14	55
0.60		22	1	4		5	2	4	17	45
0.65		26	2	5		6	3	5	18	36
0.70		26	3	6		5	4	6	20	30
0.75	1	28	4	7	1	6	6	9	17	20

Table 3.5.2: Estimated percentage of extremal mass on to each sub-cone when considering locations A-D, for varying values of the tuning parameters in Method 1 (left) and Method 2 (right).

To assess whether our methods are self-consistent across different dimensions, Table 3.5.2 shows similar results for locations A, B, C and D. We would expect the subsets of locations deemed to be simultaneously large to be the same as in Table 3.5.1 if we ignore location E. Considering the same tuning parameter values as for Table 3.5.1, we see that the extremal dependence structures are estimated to be {B, ABCD} for both methods. For Method 1, this is the set of sub-cones we would expect based on the five-dimensional results. For Method 2, we would also expect to detect the sub-cone labelled ABC, although this was only assigned a relatively small proportion of the mass in the five-dimensional case.

Tables 3.5.1 and 3.5.2 demonstrate the importance of tuning parameter selection in Methods 1 and 2. As p or δ increase, we are more likely to detect mass on the one-dimensional sub-cones, or sub-cones corresponding to subsets of the variables with low cardinality. Likewise, for low values of p or δ, we assign more extremal mass to

the sub-cone representing all variables being simultaneously extreme. In practice, we should consider the feasibility of the detected dependence structures, as well as the stability of the regions determined to have extremal mass as p or δ vary. Our methods could be used to impose structure in more complete models for multivariate extremes. Even if a handful of different options look plausible with some variation in p or δ, this is still a huge reduction over the full set of possibilities.

Chapter 4

Radial-Angular Approaches to Determining Extremal Dependence Structures

4.1 Introduction

In Chapter 3, we discussed methods for determining the dependence structure of multivariate extremes motivated by regular variation assumptions, and presented two inferential approaches based on sub-cones of $\mathbb{E} = [0, \infty]^d \setminus \{\mathbf{0}\}$. As mentioned in Section 3.1, an alternative viewpoint is to consider variables in terms of their radial and angular components, $(R, \boldsymbol{W})$, as in the approach of Goix et al. (2016); we consider three inferential methods for this setting in this chapter.

As in Chapter 3, we consider the random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$, with all variables following a common heavy-tailed distribution. The radial and angular components of $\boldsymbol{X}$ are defined as $R = \|\boldsymbol{X}\|$ and $\boldsymbol{W} = \boldsymbol{X}/\|\boldsymbol{X}\|$. We take the norms to be L_1 and the common marginal distributions to be standard Fréchet. These pseudo-polar coordinates have $R > 0$ and $\boldsymbol{W} \in \mathcal{S}_{d-1} = \{(w_1, \ldots, w_d) \in [0, 1]^d : \sum_{i=1}^{d} w_i = 1\}$,

with the unit simplex $\mathcal{S}_{d-1}$ being the disjoint union of $2^d - 1$ regions of the form

$$\mathbb{B}_C = \left\{ \boldsymbol{w} \in \mathcal{S}_{d-1} : \boldsymbol{w}_C \in (0,1]^{|C|},\ \boldsymbol{w}_{D\setminus C} \in \{0\}^{|D\setminus C|} \right\}, \tag{4.1.1}$$

for $C \in 2^D \setminus \emptyset$ and $\boldsymbol{w}_C = \{w_i : i \in C\}$. We recall the assumption of multivariate regular variation (Resnick, 2007), which states that

$$\lim_{t\to\infty} \Pr(R > tr, \boldsymbol{W} \in B \mid R > t) = H(B) r^{-1}, \quad r \geq 1,$$

for B a measurable subset of $\mathcal{S}_{d-1}$. The measure H placing mass on the region $\mathbb{B}_C$ is equivalent to the measure μ placing mass on $\mathbb{E}_C$ in Section 3.2.2, and having $H(\mathbb{B}_C) > 0$ means that variables in the set $\{X_i : i \in C\}$ can be simultaneously large while the remaining variables in $\boldsymbol{X}$ are of smaller order. Hence, analogously to Chapter 3, our aim is to estimate $\Pr(\boldsymbol{W} \in \mathbb{B}_C \mid R > r)$ as $r \to \infty$, for each face $\mathbb{B}_C$.

When considering the sub-cones $\mathbb{E}_C$ in Chapter 3, we noted that at finite levels, mass can only occur on $\mathbb{E}_D$. We encounter a similar issue in the radial-angular setting, in that no component W_i is exactly zero when $R < \infty$, and all mass occurs in $\mathbb{B}_D$. One option here is to consider $\boldsymbol{X}^*$ constructed via truncation (3.2.4), and define $R^* = \|\boldsymbol{X}^*\|$ and $\boldsymbol{W}^* = \{\boldsymbol{X}^*/R^* : R^* > 0\}$. We then let

$$B_C^* = \left\{ \boldsymbol{w}^* \in \mathcal{S}_{d-1} : \boldsymbol{w}_C^* \in (0,1]^{|C|},\ \boldsymbol{w}_{D\setminus C}^* \in \{0\}^{|D\setminus C|} \right\}. \tag{4.1.2}$$

Considering the angular component $\boldsymbol{W}^*$, mass will now occur on at least some regions B_C^* with $|C| < d$. In this setting, for $C = \{i\}$, we have

$$\Pr\left(R^* > r, \boldsymbol{W}^* \in B_i^*\right) = \Pr\left(R^* > r, W_i^* = 1\right) = \Pr\left(X_i > r; X_j < -1/\log p, j \neq i\right),$$

so that in region B_i^*, the coefficient of regular variation is $\tau_i(0)$, defined as in Assumption 1 of Chapter 3. This suggests that an analogous regular variation assumption may be useful in this radial-angular setting.

For each $C \in 2^D \setminus \emptyset$, we look at approximations B_C to $\mathbb{B}_C$, for which we assume the model

$$\Pr(R > r \mid \boldsymbol{W} \in B_C) = L_C(r) r^{-1/\kappa_C}, \tag{4.1.3}$$

as $r \to \infty$, where L_C is a slowly varying function and $\kappa_C \in (0, 1]$. We assume that regions with $\kappa_C = 1$ and $L_C(r) \not\to 0$, as $r \to \infty$, are those containing extremal mass, and correspond to subsets of variables that take their largest values simultaneously while the remaining variables are of smaller order.

In this chapter, we present three new approaches for estimating extremal dependence structures in the radial-angular setting. We introduce Method 3 in Section 4.2.1, where we use a partitioning approach to approximate the faces $\mathbb{B}_C$. In Methods 4 and 5, introduced in Sections 4.2.3 and 4.2.4, respectively, we instead propose the assignment of weights to points in the simplex based on their proximity to the various faces. This effectively yields a soft-thresholding between the simplex regions, in contrast to the hard-thresholding in the previously discussed methods, and allows each point to provide information to more than one face. In Method 4, we apply this idea to extend the approach of Goix et al. (2016), where the proportion of extremal mass associated with each face is estimated empirically above a high threshold, while Method 5 is a weighted version of our regular variation approach, based on equation (4.1.3). We present simulation results in Section 4.3, and a discussion comparing all our proposed methods is provided in Section 4.4.

4.2 Methodology

4.2.1 Method 3: simplex partitioning

As an alternative to the truncation used to construct regions B_C^* in (4.1.2), we propose approximating each face $\mathbb{B}_C$ using a region of the form

$$B_C = \left\{ \boldsymbol{w} \in \mathcal{S}_{d-1} : \boldsymbol{w}_C \in [\epsilon, 1]^{|C|},\ \boldsymbol{w}_{D\setminus C} \in [0, \epsilon)^{|D\setminus C|} \right\}, \tag{4.2.1}$$

for some small $\epsilon \in (0, 1/d]$. Figure 4.2.1 gives an example of this partition in the trivariate case with $\epsilon = 0.1$. Each region B_C is an ϵ-thickened version of $\mathbb{B}_C$. Figure 4.2.2 provides a comparison of the regions B_C^*, formed via a truncation of the variables $\boldsymbol{X}$, and B_C for the bivariate case. While these two partitions are not identical, they should both provide reasonable approximations of the faces $\mathbb{B}_C$.

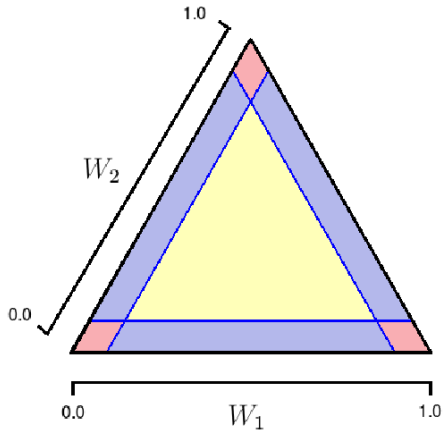

Figure 4.2.1: A partition of $\mathcal{S}_2$, with $\epsilon = 0.1$; the coordinates are transformed to the equilateral simplex for visual purposes.

We now consider probabilities of the form

$$\Pr(\boldsymbol{W} \in B_C \mid R > r) = \frac{\Pr(R > r \mid \boldsymbol{W} \in B_C)\Pr(\boldsymbol{W} \in B_C)}{\sum_{C' \in 2^D \setminus \emptyset} \Pr(R > r \mid \boldsymbol{W} \in B_{C'})\Pr(\boldsymbol{W} \in B_{C'})}, \tag{4.2.2}$$

for $C \in 2^D \setminus \emptyset$ and B_C as in (4.2.1). Analogously to Methods 1 and 2 in Chapter 3, we estimate terms of the form $\Pr(\boldsymbol{W} \in B_C)$ empirically, and for the remaining terms

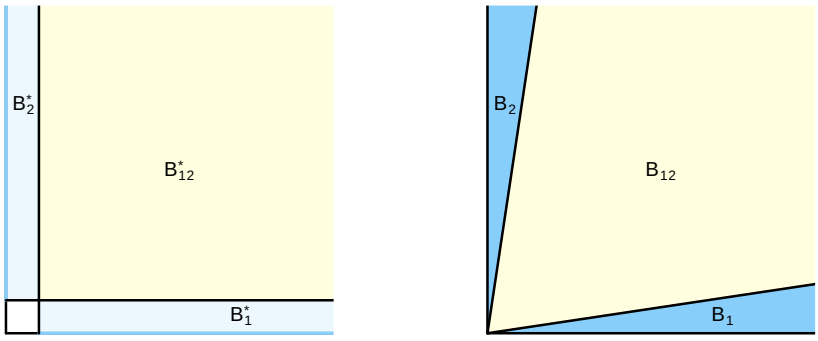

Figure 4.2.2: Comparison of regions B_C^* (left) and B_C (right) in Cartesian coordinates; the paler blue regions are truncated to the axes.

use regular variation assumption (4.1.3), replacing the slowly varying function with a constant J_C. That is, we suppose

$$\Pr(R > r \mid \boldsymbol{W} \in B_C) = J_C r^{-1/\kappa_C}, \quad \text{for } r > u_C, \tag{4.2.3}$$

and some high threshold u_C. Estimates of (J_C, κ_C) can again be obtained via a censored likelihood approach. In particular, suppose we have n_C radial observations $r_1, \ldots, r_{n_C}$ in region B_C. Then we obtain the censored likelihood

$$\prod_{j=1}^{n_C} \left(1 - J_C u_C^{-1/\kappa_C}\right)^{\mathbb{1}_{\{r_j \le u_C\}}} \left(\frac{J_C}{\kappa_C} r_j^{-1-1/\kappa_C}\right)^{\mathbb{1}_{\{r_j > u_C\}}}, \tag{4.2.4}$$

which is maximized by the estimators

$$\hat{\kappa}_C = \left(\sum_{j=1}^{n_C} \mathbb{1}_{\{r_j > u_C\}}\right)^{-1} \sum_{j=1}^{n_C} \mathbb{1}_{\{r_j > u_C\}} \log\left(\frac{r_j}{u_C}\right) \quad \text{and} \quad \hat{J}_C = \left(\frac{\sum_{j=1}^{n_C} \mathbb{1}_{\{r_j > u_C\}}}{n_C}\right) u_C^{1/\hat{\kappa}_C},$$

analogously to the models in Chapter 3. In practice, we use the estimator $\tilde{\kappa}_C = \min(\hat{\kappa}_C, 1)$ with an appropriate change to the estimate of J_C. Method 3 is summarized as follows.

1. Transform data to standard Fréchet margins, and calculate the radial-angular components $(R, \boldsymbol{W})$.

2. Assign each transformed observation to a region B_C as in (4.2.1).

3. For each region B_C containing more than m points, fit model (4.2.3) using censored likelihood (4.2.4) for a choice of threshold u_C, and estimate $\Pr(\boldsymbol{W} \in B_C \mid R > r)$ for a large value of r by equation (4.2.2). Set $\Pr(\boldsymbol{W} \in B_C \mid R > r) = 0$ in the remaining regions. Denote the resulting estimate by $\hat{p}_C$.

4. If $\hat{p}_C < \pi$, for a choice of the tuning parameter π, set $\hat{p}_C$ to zero, renormalizing the resulting vector.

As in Chapter 3, the choice of parameter m was found not have a significant effect on results, and we take the default choice $m = 1$.

From Section 4.1, we know that the coefficient of regular variation is $\tau_1(0)$ in B_1^*. In Figure 4.2.2, we see that for large radial values, region B_1 contains parts of B_{12}^*, as well as B_1^*, so that $\kappa_1 \geq \tau_1(0)$ in general. Similar issues arise when using the ϵ-thickening approach in higher than two dimensions. Although the set of values $\{\kappa_C;\ C \in 2^D \setminus \emptyset\}$ may not be fully informative about the extremal dependence structure, estimates of it should provide additional information over the empirical approach of Goix et al. (2016). We investigate the relative performance of Method 3 in Section 4.3.

4.2.2 Incorporating weights

As discussed in Section 4.1, an alternative to the partitioning approach of Method 3 or Goix et al. (2016), is to assign weights to observations so they can provide information to more than one face. Suppose we have n observations of the radial-angular components, denoted $(r_1, \boldsymbol{w}_1)$, $\ldots$, $(r_n, \boldsymbol{w}_n)$. Let $\delta_{C,j}$ be the weight contributed to the face $\mathbb{B}_C$ by angular observation $\boldsymbol{w}_j$, for $C \in 2^D \setminus \emptyset$ and $j = 1, \ldots, n$. Natural conditions on these weights are

1. $\delta_{C,j} \in [0, 1],$ for each $C \in 2^D \setminus \emptyset$ and $j = 1, \ldots, n$;

2. $\sum_{C \in 2^D \setminus \emptyset} \delta_{C,j} = 1, \text{ for each } j = 1, \ldots, n;$

ensuring that they are non-negative and the overall contribution of each data point is unchanged. We explain how to incorporate these weights into the approach of Goix et al. (2016) and our Method 3 in Sections 4.2.3 and 4.2.4, respectively. There are infinitely many ways to define the $\delta_{C,j}$ values; we propose one such possibility in Section 4.2.5.

4.2.3 Method 4: weighted extension of the approach of Goix et al. (2016)

In the method of Goix et al. (2016), the faces of the angular simplex are approximated via a partition of $\mathcal{S}_{d-1}$ similar to (4.2.1), where we denote by B_C the region of the partition that approximates $\mathbb{B}_C$. In the approach of Goix et al., empirical estimation of $\Pr(\boldsymbol{W} \in B_C \mid R > r_0)$, for some high radial threshold r_0, is used to approximate the proportion of extremal mass on each face.

By instead allowing each point to contribute mass to different faces via the weights $\delta_{C,j}$, this method can be adapted so that the proportion of extremal mass on face $\mathbb{B}_C$ is estimated by

$$\Pr(\boldsymbol{W} \in \mathbb{B}_C \mid R > r_0) = \frac{\sum_{j=1}^{n} \delta_{C,j} \mathbb{1}_{\{r_j > r_0\}}}{\sum_{j=1}^{n} \mathbb{1}_{\{r_j > r_0\}}}, \tag{4.2.5}$$

with r_0 again denoting a high radial threshold. The overall method is summarized in the following steps.

1. Transform data to standard Fréchet margins, and calculate the radial-angular components $(R, \boldsymbol{W})$.

2. Calculate $\delta_{C,j}$ for each $j = 1, \ldots, n$ and $C \in 2^D \setminus \emptyset$, for some definition of the weights $\delta_{C,j}$.

3. Evaluate (4.2.5) for each face, for some choice of threshold r_0.

As in Methods 1, 2 and 3, negligible mass can be removed by setting to zero any estimates below some value π, and renormalizing the resulting vector of proportions.

4.2.4 Method 5: weighted extension of the regular variation approach

We adapt Method 3 to include the weights $\delta_{C,j}$ by changing the estimation of the components in the right hand side of equation (4.2.2), for each $C \in 2^D \setminus \emptyset$. Firstly, the probability $\Pr(\boldsymbol{W} \in B_C)$ is estimated by $\frac{1}{n}\sum_{j=1}^{n} \delta_{C,j}$, i.e., the average weight associated with face $\mathbb{B}_C$. Recall that under assumption (4.2.3), replacing the slowly varying function by a constant, for each face we have a model of the form

$$\Pr(R > r \mid \boldsymbol{W} \in B_C) = J_C r^{-1/\kappa_C},$$

for r greater than some high threshold u_C, $\kappa_C \in (0, 1]$ and $J_C > 0$. Instead of using the partitioning approach of Method 3, this model can be fitted by adapting the censored likelihood approach. In particular, for each $C \in 2^D \setminus \emptyset$, we replace $\mathbb{1}_{\{r_j > u_C\}}$ by $\delta_{C,j}\mathbb{1}_{\{r_j > u_C\}}$, and $\mathbb{1}_{\{r_j \le u_C\}}$ by $\delta_{C,j}\mathbb{1}_{\{r_j \le u_C\}}$, to yield the function

$$\prod_{j=1}^{n} \left(1 - J_C u_C^{-1/\kappa_C}\right)^{\delta_{C,j}\mathbb{1}_{\{r_j \le u_C\}}} \left(\frac{J_C}{\kappa_C} r_j^{-1-1/\kappa_C}\right)^{\delta_{C,j}\mathbb{1}_{\{r_j > u_C\}}},$$

which should be maximized. This is no longer a true likelihood in the sense of corresponding to an underlying statistical model, yet the weights have clear roles in modifying the estimators. The estimators that maximize this modified objective function are

$$\hat{\kappa}_C = \left(\sum_{j=1}^{n} \delta_{C,j}\mathbb{1}_{\{r_j > u_C\}}\right)^{-1} \sum_{j=1}^{n} \delta_{C,j}\mathbb{1}_{\{r_j > u_C\}} \log\left(\frac{r_j}{u_C}\right)$$

and

$$\hat{J}_C = \left(\frac{\sum_{j=1}^{n} \delta_{C,j}\mathbb{1}_{\{r_j > u_C\}}}{\sum_{j=1}^{n} \delta_{C,j}}\right) u_C^{1/\hat{\kappa}_C}.$$

In practice, we take the estimate of κ_C to be $\min(\hat{\kappa}_C, 1)$, with a suitable update to $\hat{J}_C$, as previously. Having determined which weights to use, and computed $\hat{\kappa}_C$ and $\hat{J}_C$, the remainder of the approach is analogous to Method 3. Moreover, Method 3 is a special case of the weighted approach, with $\delta_{C,j} = 1$ if $w_j \in B_C$, the region of the partition that approximates $\mathbb{B}_C$, and $\delta_{C,j} = 0$ otherwise, just as the approach of Goix et al. (2016) is a special case of Method 4. In Method 3, we only carry out the estimation procedure on a face if there are at least m angular observations in the corresponding region of the partitioned simplex. The natural extension of this condition to the weighted approach is to have $\sum_{j=1}^{n} \delta_{C,j}$ being at least m; we again take the default choice $m = 1$.

4.2.5 A proposed weighting

We propose a weighting which takes into account the proximity of a point $\boldsymbol{w} = (w_1, \ldots, w_d)$ to various faces of the unit simplex. We begin by considering the vertex to which the point $\boldsymbol{w}$ is closest. This can be determined by finding $S^{(1)} = \{i : w_i = \max(w_1, \ldots, w_d)\}$. Then, for some $k > 0$, the weight assigned to the region $\mathbb{B}_{S^{(1)}}$, corresponding to the face where only variable $\{X_i : i \in S^{(1)}\}$ is large, is

$$\delta_{S^{(1)}} = \left(\sum_{i \in S^{(1)}} w_i \right)^k = w_{S^{(1)}}^k.$$

For $s = 2, \ldots, d$, we then find the set $S^{(s)}$ with $|S^{(s)}| = s$ such that $\boldsymbol{w}$ lies closer to the region $\mathbb{B}_{S^{(s)}}$ than any other set $\mathbb{B}_{S^{(s)'}}$ with $|S^{(s)'}| = s$. For this set, the associated weight is defined as

$$\delta_{S^{(s)}} = \left(\sum_{i \in S^{(s)}} w_i \right)^k - \left(\sum_{i \in S^{(s-1)}} w_i \right)^k.$$

For all remaining sets of size s, the weight is set to zero. Note that for $s = d$ the only choice is $S^{(d)} = D = \{1, \ldots, d\}$. In Appendix B.1, we verify that this choice of weights satisfies properties 1 and 2 from Section 4.2.2.

In practice, the data are transformed to Fréchet margins using the rank transform; this can lead to components of $\boldsymbol{w}$ being equal. We break such ties randomly, so that each point assigns non-zero weight to exactly d faces, one of each dimension. An alternative would be to divide the weight across faces of the same dimension which are equidistant from the point $\boldsymbol{w}$. The latter approach is more computationally complex, and is unlikely to significantly improve results.

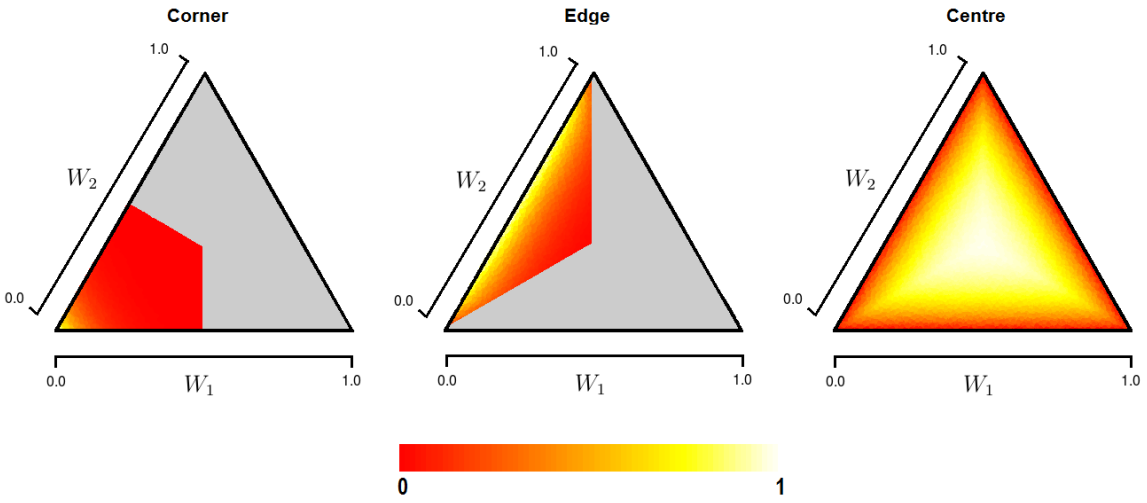

Figure 4.2.3: An example of our proposed weighting in the trivariate case, with $k = 10$. The grey regions in the first two plots show where the weights are exactly 0.

Figure 4.2.3 shows an example of this weighting in the trivariate case. The first diagram illustrates the weight contributed by each point to the face where only variable X_3 is large, based on its position in the simplex. The central diagram shows the weight assigned to the face with variables (X_2, X_3) being large simultaneously. In both these cases, illustrations of the weights assigned to the other faces of the same dimension are obtained by rotation. The final diagram shows the weights associated with the centre of the simplex, where all three variables are simultaneously large. Increasing the value of k assigns more weight to the interior of the simplex, while smaller values of k assign greater weight to the vertices.

4.3 Simulation study

4.3.1 Max-mixture distribution

In this section, we assess the efficacy of Methods 3, 4 and 5 by considering the results of simulations based on the five-dimensional max-mixture distribution introduced in Section 3.4.2 of Chapter 3. We consider the asymmetric logistic model in Section 4.3.2. Once again, our metrics are the Hellinger distance; the area under the receiver operating characteristic curve (now denoted AUC); and the area under the neighboured receiver operating characteristic curve (now denoted AUC*). This final metric was introduced in Chapter 3, and allows us to assess whether an estimated extremal dependence structure is *close* to the truth by taking into account whether or not faces adjacent to those truly having mass are detected.

The max-mixture model consists of two bivariate Gaussian copulas with correlation parameter ρ, as well as two trivariate and one five-dimensional logistic copula with dependence parameter α. In Method 3, we set $\epsilon = 0.05$, each threshold u_C to be the 0.975 quantile of observed radial values over all sets $C \in 2^D \setminus \emptyset$ and r in $\Pr(\boldsymbol{W} \in B_C \mid R > r)$ to be the largest observed radial value for that particular sample. In Methods 4 and 5, we take the parameter controlling the weighting to be $k = 25$. For Method 4, the radial threshold r_0 is set to the 0.995 quantile of the observed radial values. For Method 5, the threshold u_C in each region is taken to be the 0.975 quantile over all the observed radial values, and the value of r for which (4.2.2) is evaluated is taken to be the maximum of the observed radial values. For all three methods, to obtain the Hellinger distances, we take the threshold π below which mass on each face is deemed negligible to be 0.001, as in the simulations of Chapter 3; this parameter is varied to obtain the AUC and AUC* values. Again, these tuning parameter values are not optimized for individual experiments, but parameter stability plots as introduced in Section 3.4.3 could be used to choose the tuning parameter,

and one should ensure that a feasible combination of faces has been determined as having mass. We take samples of size $n = 10,000$ in each simulation, and repeat each setting 100 times.

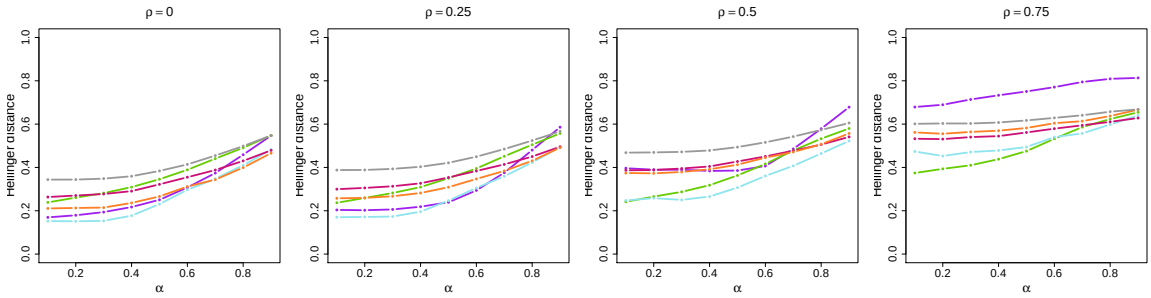

Figure 4.3.1: Mean Hellinger distance over 100 simulations for the max-mixture distribution. Method 1: purple; Method 2: green; Method 3: blue; Method 4: pink; Method 5: orange; Goix et al.: grey.

In Figure 4.3.1, we present results on the average Hellinger distance achieved by Methods 3, 4 and 5, comparing them to the previous results obtained for the approach of Goix et al. and Methods 1 and 2 in Section 3.4.2; the 0.05 and 0.95 quantiles are now omitted to improve readability. These results cover a range of model parameters; $\rho \in \{0, 0.25, 0.5, 0.75\}$ and $\alpha \in \{0.1, 0.2, \ldots, 0.9\}$. Method 3, based on the regular variation assumption and a partition of the angular simplex, is generally the most successful for $\rho \in \{0, 0.25, 0.5\}$, but is outperformed by Method 2 for $\rho = 0.75$ and $\alpha \leq 0.6$. The weighted approach used in Methods 4 and 5 is consistently more successful than the approach of Goix et al., and these methods also provide competitive results compared to Method 1, particularly in the case where $\rho = 0.75$, corresponding to stronger dependence in the asymptotically independent Gaussian components, or higher values of α, which lead to weaker asymptotic dependence in the logistic parts of the model.

	$\rho = 0$			$\rho = 0.25$			$\rho = 0.5$			$\rho = 0.75$		
α	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
Method 3	**100 (0.2)**	99.7 (0.9)	96.3 (2.8)	99.9 (0.3)	99.6 (0.9)	96.0 (3.0)	99.3 (1.1)	**98.7 (1.6)**	**94.4 (3.4)**	93.9 (1.7)	92.5 (2.2)	87.0 (3.6)
Method 4	**100 (0.0)**	**100 (0.1)**	**99.1 (1.4)**	99.9 (0.2)	**99.9 (0.4)**	**98.7 (1.6)**	97.3 (1.8)	96.8 (1.9)	**94.6 (2.7)**	92.1 (1.4)	91.0 (2.0)	88.6 (2.6)
Method 5	**100 (0.2)**	99.8 (0.7)	97.9 (2.3)	99.8 (0.4)	99.7 (0.8)	**97.6 (2.5)**	96.0 (2.0)	95.6 (2.1)	**93.1 (3.2)**	92.0 (1.0)	91.4 (1.4)	87.7 (2.4)
	$\rho = 0$			$\rho = 0.25$			$\rho = 0.5$			$\rho = 0.75$		
α	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
Method 3	**100 (0.0)**	99.8 (0.7)	**96.5 (2.1)**	**100 (0.2)**	99.8 (0.7)	96.3 (2.1)	**100 (0.0)**	99.7 (0.8)	96.3 (2.4)	**100 (0.2)**	99.7 (0.8)	97.6 (2.2)
Method 4	**100 (0.0)**	**100 (0.2)**	**98.6 (1.6)**	**100 (0.0)**	**100 (0.2)**	**98.6 (1.9)**	**100 (0.0)**	**100 (0.2)**	**99.1 (1.9)**	**100 (0.0)**	**100 (0.0)**	**99.7 (0.7)**
Method 5	**100 (0.0)**	**100 (0.2)**	**98.5 (1.4)**	**100 (0.0)**	**100 (0.2)**	**98.4 (1.4)**	**100 (0.2)**	**100 (0.3)**	**98.9 (1.1)**	**100 (0.3)**	99.9 (0.3)	**98.7 (1.3)**

Table 4.3.1: Average AUC (top) and AUC* (bottom) values, given as percentages, for 100 samples from a five-dimensional max-mixture model. Results in bold show where the methods are at least as successful as any of those studied in Chapter 3.

In Table 4.3.1, we present the mean AUC and AUC* values achieved by Methods 3, 4 and 5 for this max-mixture distribution, with $\rho \in \{0, 0.25, 0.5, 0.75\}$ and $\alpha \in \{0.25, 0.5, 0.75\}$; boxplots of these results are provided in Appendix B.2. As was the case when comparing Hellinger distances, the results are reasonably similar for Methods 4 and 5, although the former is often a slightly more successful classifier, and is at least as successful in terms of the average AUC as all methods we consider in six out of the twelve cases. Notably, for many cases where $\alpha = 0.75$, with weaker asymptotic dependence in the logistic components, the weighted approaches provide the greatest improvement over the hard-thresholding methods, although Method 2 remains the most successful classifier for $\rho = 0.75$.

Method 4, the weighted version of Goix et al. (2016) performs at least as well as all other methods in terms of the average AUC* values, with Method 5, our other weighted approach, giving only slightly worse results. This suggests that using weights to allow observations to provide information to more than one face is a successful strategy, allowing us to estimate extremal dependence structures that are close to the truth. For this max-mixture distribution, we note that a large proportion of the faces truly have extremal mass or have a neighbouring face that does, and the AUC* metric is perhaps more informative for the asymmetric logistic distribution studied in

Section 4.3.2.

4.3.2 Asymmetric logistic distribution

We now test Methods 3, 4 and 5 on data simulated from asymmetric logistic distributions (Tawn, 1990), as in Appendix A.3.3 for the methods of Chapter 3. We consider data from models of two different dimensions. For dimension $d = 5$, we sample $n = 10,000$ points and consider $f = 5$, 10 or 15 faces with extremal mass. For $d = 10$, we have $n = 100,000$ and $f \in \{10, 50, 100\}$. Each simulation is repeated 100 times, and we compare methods using Hellinger distances, AUC and AUC*. To obtain the results for Methods 3, 4 and 5, the tuning parameters used are the same as for the max-mixture examples in Section 4.3.1.

We present results on the average Hellinger distance in Figure 4.3.2, for values of the dependence parameter $\alpha \in \{0.1, 0.2, \ldots, 0.9\}$. Results are shown for Methods 3, 4 and 5, alongside results for methods from Chapter 3. Comparing the weighted approaches, Figure 4.3.2 shows that Method 5 performs better than Method 4 in the majority of cases, and both these methods show improvement over those studied in Chapter 3 for larger values of α, corresponding to weaker asymptotic dependence. For low-to-moderate values of α, Method 1 is still generally the most successful in terms of the Hellinger distance, particularly in the two most sparse cases, $(d, f) = (5, 5)$ and $(10, 10)$. While Method 3, based on the partition of the angular simplex, is rarely the most successful in terms of the Hellinger distance, it does provide reasonably competitive results in the five-dimension case.

Table 4.3.2 provides results on the average AUC values obtained for a subset of the cases considered in Figure 4.3.2. In particular, we consider $\alpha \in \{0.25, 0.5, 0.75\}$ to cover a range of strengths of extremal dependence. Boxplots of these results are provided in Appendix B.3. The results for dimension $d = 5$ show that the inclusion of

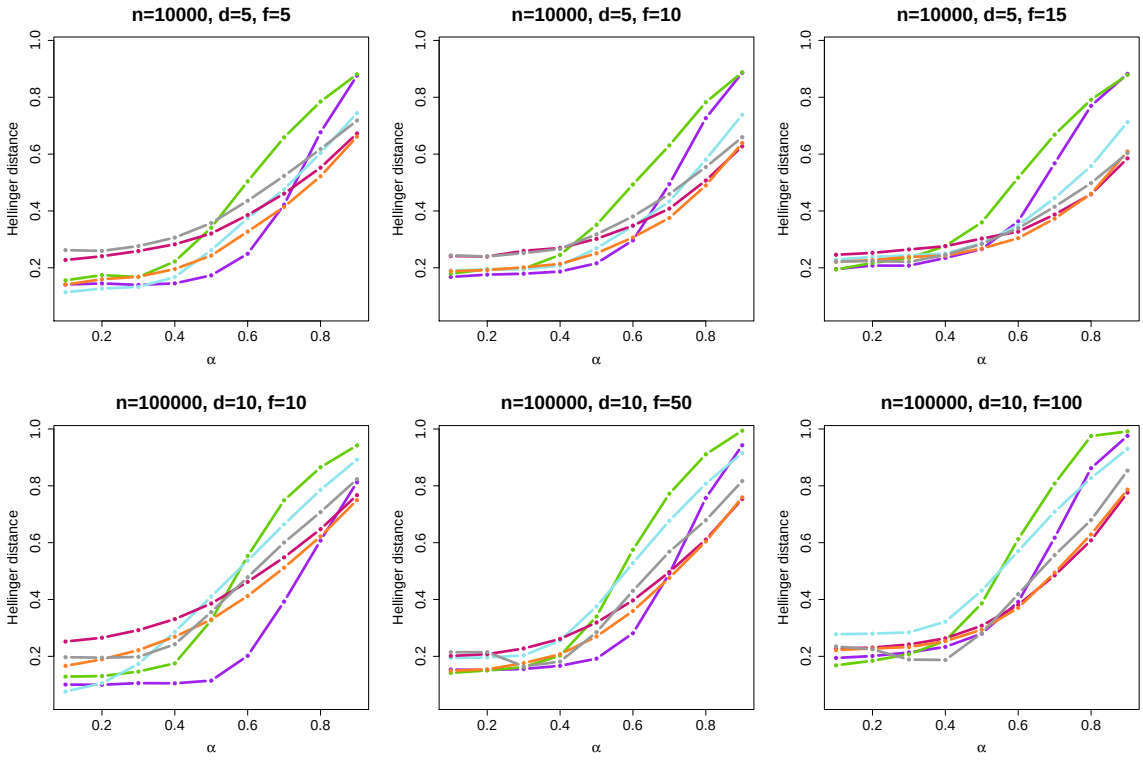

Figure 4.3.2: Mean Hellinger distance over 100 simulations from asymmetric logistic distributions with $d = 5$ (top row) and $d = 10$ (bottom row). Method 1: purple; Method 2: green; Method 3: blue; Method 4: pink; Method 5: orange; Goix et al. grey.

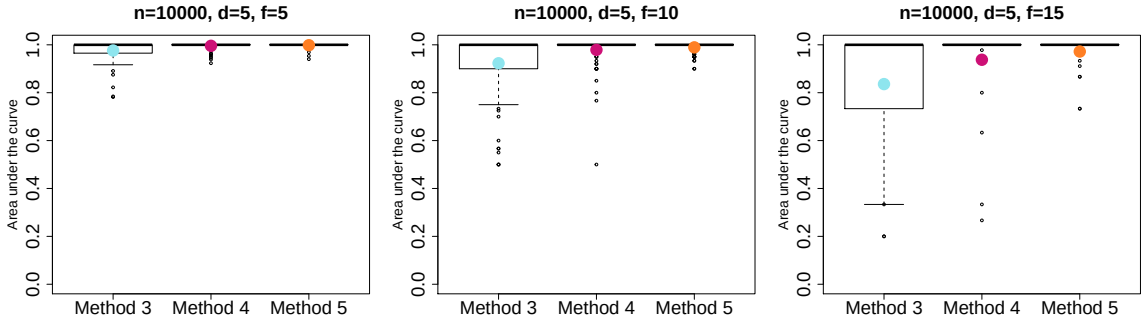

Figure 4.3.3: Boxplots of AUC* values for Methods 3, 4 and 5 with $d = 5$; $f = 5, 10, 15$; $n = 10,000$ and $\alpha = 0.75$. Mean AUC* values are shown by the circles in each plot.

(α, f)	$(0.25, 5)$	$(0.25, 10)$	$(0.25, 15)$	$(0.5, 5)$	$(0.5, 10)$	$(0.5, 15)$	$(0.75, 5)$	$(0.75, 10)$	$(0.75, 15)$
Method 3	99.9 (0.3)	99.7 (0.7)	99.5 (1.0)	99.5 (1.3)	98.3 (2.0)	97.0 (2.2)	94.3 (5.0)	86.2 (6.5)	81.2 (7.9)
Method 4	99.4 (2.4)	97.4 (4.0)	93.4 (5.0)	99.4 (1.9)	97.2 (3.4)	94.0 (5.0)	98.5 (2.2)	**93.3 (4.7)**	**87.6 (6.5)**
Method 5	99.8 (1.1)	99.2 (1.6)	98.2 (2.0)	99.7 (1.3)	98.3 (2.3)	96.8 (2.7)	**98.8 (2.1)**	**93.5 (4.6)**	**87.7 (6.3)**
(α, f)	$(0.25, 10)$	$(0.25, 50)$	$(0.25, 100)$	$(0.5, 10)$	$(0.5, 50)$	$(0.5, 100)$	$(0.75, 10)$	$(0.75, 50)$	$(0.75, 100)$
Method 3	**100 (0.0)**	**100 (0.0)**	99.9 (0.2)	**100 (0.1)**	99.5 (0.3)	98.1 (0.4)	98.5 (1.4)	93.1 (1.4)	87.8 (1.9)
Method 4	**100 (0.0)**	99.8 (0.4)	98.4 (1.0)	**100 (0.0)**	99.9 (0.3)	98.5 (0.8)	**100 (0.0)**	**98.8 (0.4)**	95.9 (1.1)
Method 5	**100 (0.0)**	**100 (0.1)**	99.6 (0.2)	**100 (0.0)**	99.9 (0.2)	99.0 (0.4)	**100 (0.0)**	98.5 (0.5)	95.3 (1.0)

Table 4.3.2: Average AUC values, given as percentages, for Methods 3 and 4 over 100 samples from five-dimensional (top) and ten-dimensional (bottom) asymmetric logistic distributions, with dependence parameter α. Results in bold show where the methods are at least as successful as any of those studied in Chapter 3.

weights gives most improvement when the strength of dependence in the asymmetric logistic model is weaker, with generally better results for $\alpha = 0.75$ for Methods 4 and 5 than all three methods discussed in Chapter 3. However, Method 1 remains the most successful classifier for $\alpha = 0.25$ and $\alpha = 0.5$ for $d = 5$ across all three values of f we consider. For the $d = 10$ cases, all six approaches perform well for $\alpha = 0.25$ and $\alpha = 0.5$. When $\alpha = 0.75$, Methods 4 and 5 perform well in the sparse case of $f = 10$; otherwise, Method 2 from Chapter 3 is the most successful.

Finally, in Figure 4.3.3, we present boxplots of AUC* values for $d = 5$, $\alpha = 0.75$ and $f \in \{5, 10, 15\}$. Methods 4 and 5 are much more successful than all the previous approaches in terms of this criterion. As in the max-mixture case, this suggests that when these weighted methods are incorrect in their assessment of the extremal dependence structure, they are generally closer to the truth than the non-weighted approaches, and may therefore provide more reasonable approximations. Method 5 is slightly more successful than Method 4 according to Figure 4.3.3, although the difference does not appear significant.

4.4 Discussion

Each of the methods we have discussed have been shown to work best in different settings. In general, it appears that the weighted approaches in Methods 4 and 5 are most successful for weak asymptotic dependence on faces with mass, with particularly strong results in terms of the AUC and AUC* values. Methods 1 and 3 often performed better for the max-mixture distribution than the other approaches in terms of the Hellinger distance, particularly for low values of the correlation ρ; for stronger correlation in this model, Method 2 generally remains the most successful. In terms of the asymmetric logistic model, Method 1 generally performs best when the extremal mass is sparse, i.e., the mass is concentrated on a small number of faces, while the approach of Goix et al. and our weighted methods are better in less sparse cases.

In practice, one could apply several methods for determining extremal dependence structures, and compare the results; it may be that the outputs of some methods are more sensible than others in the context of the problem. Alternatively, if the aim is to choose or construct models exhibiting a given extremal dependence structure, one could fit several of these models based on the different methods for assessing the dependence structure of the data, and use model selection criteria to choose the most appropriate.

Chapter 5

An Investigation into the Tail Properties of Vine Copulas

5.1 Introduction

Vine copulas have recently been growing in popularity as a tool for multivariate dependence modelling, due to their flexibility and relatively straightforward extension to moderate or high dimensions. They have the potential to be used in extremal dependence modelling, but questions remain about how well they can capture dependence and independence in multivariate extremes. In this chapter, we aim to investigate some of the tail properties of vine copulas by calculating the tail dependence coefficient of Ledford and Tawn (1996) for various models belonging to this class.

Suppose we are interested in modelling the variables $\boldsymbol{X} = \{X_1, \ldots, X_d\}$, which in this chapter we assume have standard exponential margins, i.e., $\Pr(X_i < x) = 1 - e^{-x}$, $x > 0$, for $i = 1, \ldots, d$. For any subset of these variables, $\boldsymbol{X}_C = \{X_i : i \in C\}$, with $C \subseteq D = \{1, \ldots, d\}$ and $|C| \geq 2$, the coefficient of tail dependence η_C (Ledford and Tawn, 1996) is defined via the relation

$$\Pr\left(X_i > x : i \in C\right) \sim L_C(e^x)e^{-x/\eta_C}, \tag{5.1.1}$$

as $x \to \infty$, for some slowly varying function L_C and $\eta_C \in (0, 1]$. If $\eta_C = 1$ and $L_C(x) \not\to 0$, as $x \to \infty$, the variables in $\boldsymbol{X}_C$ are asymptotically dependent, i.e., take their largest values simultaneously. For $\eta_C < 1$, these variables are asymptotically independent, and the value of the coefficient itself provides a measure of the residual dependence between the variables, that is, the rate at which they tend towards independence, asymptotically.

In this chapter, we focus on models for continuous variables, although vine copulas can also be used in the discrete case. Our main focus is on the three dimensional case, but the techniques we use to calculate η_C are also applicable in higher dimensions. A trivariate vine copula is defined in terms of the joint density

$$\begin{aligned} f(x_1, x_2, x_3) =& f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \\ & \cdot c_{12}\left\{F_1(x_1), F_2(x_2)\right\} \cdot c_{23}\left\{F_2(x_2), F_3(x_3)\right\} \\ & \cdot c_{13|2}\left\{F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)\right\}, \end{aligned} \tag{5.1.2}$$

for marginal densities f_i, $i = 1, 2, 3$; marginal distribution functions F_i, $i = 1, 2, 3$; copula densities c_{12}, c_{23} and $c_{13|2}$; and conditional distribution functions $F_{1|2}$ and $F_{3|2}$. Such a copula can be represented graphically, as in Figure 5.1.1. Each of the edges in the graph corresponds to one of the bivariate copulas that contributes to density (5.1.2). If each of $c_{12}, c_{23}, c_{13|2}$ is a Gaussian copula density, the overall trivariate joint distribution is also Gaussian (Joe, 1996 ; Joe et al., 2010). Since the tail dependence properties of the Gaussian model are well-studied in the literature, we focus on cases where the pair copulas are from inverted extreme value or extreme value classes of distributions.

We aim to investigate the values of the coefficients η_C for $C \in 2^D$, the power set of $D = \{1, \ldots, d\}$, with $|C| \geq 2$, in order to provide a summary of the tail dependence properties of vine copulas. However, the coefficient η_C in (5.1.1) is defined in

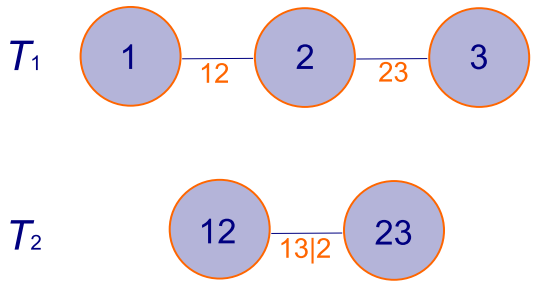

Figure 5.1.1: Graphical representation of a trivariate vine copula, with tree labels T_1, T_2 as introduced in Section 2.5.3.

terms of a survivor function, which can often not be obtained analytically for vine copulas. We therefore turn to the approach of Nolde (2014), who gives a geometric interpretation of the coefficient of tail dependence that enables its calculation from a density; we discuss this approach in Section 5.2.1. One drawback of this method is that it is only applicable when the joint density of $\boldsymbol{X}_C$ can be obtained analytically. In our vine copula examples, we have the form of the overall joint density of $\boldsymbol{X}$, but not necessarily for $\boldsymbol{X}_C$ with $C \subset D = \{1, \ldots, d\}$. In particular, for the trivariate case represented by Figure 5.1.1, for pairs directly linked in the upper tree, i.e., $C = \{1, 2\}$ and $C = \{2, 3\}$, the density can be obtained from the corresponding pair copula, i.e., c_{12} and c_{23}, but for $C = \{1, 3\}$, this is not the case. In Section 5.2.3, we extend the technique of Nolde (2014) using ideas related to the Laplace approximation, in order to find η_C when the density can be obtained for $\boldsymbol{X}$ but not $\boldsymbol{X}_C$.

In Section 5.3, our focus is on situations where inverted extreme value copulas are used to construct the vine. We first apply the techniques of Section 5.2 to calculate η_C for trivariate vine copulas, focussing on the non-trivial cases of $\eta_{\{1,2,3\}}$ and $\eta_{\{1,3\}}$, before demonstrating the approach for higher dimensional examples, where the underlying graphical structure of the vine copula is a further consideration. In all of these cases, there is overall asymptotic independence between the variables, i.e., no combination of variables can take their largest values simultaneously. In Section 5.4, we turn

our attention back to the trivariate case, but consider vine copulas constructed from extreme value and inverted extreme value pair copulas, where different subsets of the variables can exhibit asymptotic independence or asymptotic dependence. A summary of our findings, and possible avenues for future work, is provided in Section 5.5.

5.2 Density-based calculation of η_C

5.2.1 Geometric interpretation of η_C (Nolde, 2014)

Nolde (2014) presents a strategy for determining the coefficient of tail dependence, η_C, for variables $\boldsymbol{X}_C = \{X_i : i \in C\}$, based on the geometry of scaled random samples from the joint distribution of $\boldsymbol{X}_C$. Assuming standard exponential margins, and joint density $f_C(\boldsymbol{x}_C)$, the idea is to study the gauge function $g_C(\boldsymbol{x}_C)$ such that

$$-\log f_C(t\boldsymbol{x}_C) \sim t g_C(\boldsymbol{x}_C), \tag{5.2.1}$$

as $t \to \infty$, with $g_C(\boldsymbol{x}_C)$ being homogeneous of order 1. The boundary of the scaled random sample $(\boldsymbol{X}_1/\log n, \ldots, \boldsymbol{X}_n/\log n)$, for large n, is described by the set $G_C = \{\boldsymbol{x}_C \in \mathbb{R}^{|C|} : g_C(\boldsymbol{x}_C) = 1\}$, with the scaling function $\log n$ chosen due to the exponential margins. The coefficient of tail dependence η_C then corresponds to the smallest value of r such that $G_C \cap [r, \infty)^{|C|} = \emptyset$. In the remainder of this chapter, we drop the subscript C from the set G_C, density f_C and gauge function g_C when discussing the overall vector of variables $\boldsymbol{X}$, i.e., when $C = D$.

As an example, consider a bivariate inverted extreme value copula with exponent measure V, defined in equation (2.3.2) of Chapter 2, and variables (X, Y) having standard exponential margins. This model has distribution function

$$F(x, y) = 1 - e^{-x} - e^{-y} + \exp\left\{-V\left(x^{-1}, y^{-1}\right)\right\},$$

for $x, y > 0$, so that by first differentiating $F(x, y)$ with respect to both components to obtain the density $f(x, y)$, with V_1, V_2 and V_{12} denoting the derivatives of the exponent

measure with respect to the first, second and both components, respectively, we have

$$\begin{aligned}
-\log f(tx,ty) =& 2\log(tx) + 2\log(ty) + V\left\{(tx)^{-1},(ty)^{-1}\right\} \\
&- \log\left[V_1\left\{(tx)^{-1},(ty)^{-1}\right\} V_2\left\{(tx)^{-1},(ty)^{-1}\right\} - V_{12}\left\{(tx)^{-1},(ty)^{-1}\right\}\right] \\
=& 2\log(tx) + 2\log(ty) + tV\left(x^{-1},y^{-1}\right) \\
&- \log\left\{t^4 V_1\left(x^{-1},y^{-1}\right) V_2\left(x^{-1},y^{-1}\right) - t^3 V_{12}\left(x^{-1},y^{-1}\right)\right\} \\
=& tV\left(x^{-1},y^{-1}\right) + O(\log t), \qquad (5.2.2)
\end{aligned}$$

as $t \to \infty$, by exploiting the homogeneity of the exponent measure. That is, the gauge function is given by $g(x,y) = V\left(x^{-1},y^{-1}\right)$. As an example of this class of models, consider the inverted logistic copula with exponent measure $V(x,y) = \left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}$, and therefore gauge function $g(x,y) = \left(x^{1/\alpha} + y^{1/\alpha}\right)^{\alpha}$. The set G corresponding to this gauge function is shown by the red line in the left panel of Figure 5.2.1 for the case where $\alpha = 0.5$, and the grey points show a suitably-normalized sample from the joint distribution. The blue line shows the boundary of the set $[r,\infty)^2$, with the smallest value of r such that the two sets do not intersect being $1/\sqrt{2} = 2^{-\alpha}$, occurring when $x = y$. This corresponds to the known value of $\eta_{\{1,2\}}$ for this copula.

A similar approach shows that the bivariate extreme value logistic copula with dependence parameter $\alpha \in (0,1)$ has gauge function

$$g(x,y) = \frac{1}{\alpha}\max\{x,y\} + \left(1 - \frac{1}{\alpha}\right)\min\{x,y\}.$$

In this case, the point $(x,y) = (1,1)$ satisfies $g(x,y) = 1$, and since both variables are at most 1 in the set G, we must have $\eta_{\{1,2\}} = 1$. This is demonstrated in the right panel of Figure 5.2.1 for $\alpha = 0.5$.

5.2.2 Numerical approximation of η_C

Analytically calculating the largest value of r such that G_C and $[r,\infty)^{|C|}$ intersect may not be possible for certain gauge functions. We therefore propose a strategy to

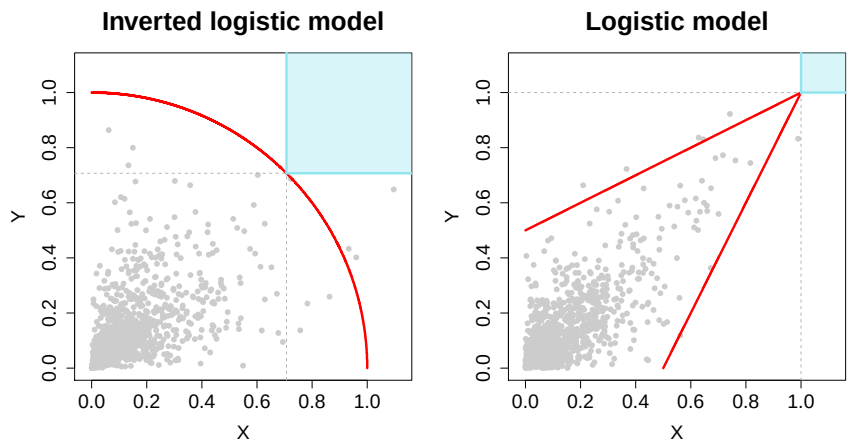

Figure 5.2.1: A scaled sample from a bivariate inverted logistic copula (left) and bivariate logistic copula (right) with $\alpha = 0.5$ (grey); the sets $G = \{(x, y) \in \mathbb{R}^2 : g(x, y) = 1\}$ (red); and the sets $[\eta_{\{1,2\}}, \infty)^2$ (blue).

calculate this value numerically, using the following steps:

1. Simulate n points, $\boldsymbol{x}_{C,j}$, $j = 1, \ldots, n$, in the set G_C by:
   (a) simulating n points, $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$, from a $|C|$-dimensional Dirichlet distribution with all concentration parameters set to 0.5;
   (b) setting $\boldsymbol{x}_{C,j} = \{x_{C,j,i} : i \in C\} = \boldsymbol{w}_j / g_C(\boldsymbol{w}_j)$, for $j = 1, \ldots, n$ (note that $g_C(\boldsymbol{x}_{C,j}) = 1$ by homogeneity of the gauge function).
2. Find $r_i = \max\{x_{C,j,i} : j \in \{1, \ldots, n\}, x_{C,j,k} \geq x_{C,j,i}\ \forall k \neq i\}$, for each $i \in C$.
3. An approximation of η_C is given by $\max_{i \in C} r_i$.

5.2.3 Extension of Nolde (2014) when joint density not available

Applying the definition of the gauge function in (5.2.1) relies on us having the density $f_C(\boldsymbol{x}_C)$ analytically, which may not always be possible. For instance, for vine copula

models, we can construct the density $f(\boldsymbol{x})$ of $\boldsymbol{X} = \{X_1, \ldots, X_d\}$, but obtaining the density for subsets of these variables involves integration that may not result in a closed form expression, particularly if the variables are not directly linked in the base tree of the corresponding graphical model. We therefore propose an extension of the method of Nolde (2014) to deal with such cases.

In the original approach of Nolde, we have $-\log f(t\boldsymbol{x}) \sim tg(\boldsymbol{x})$, as $t \to \infty$. This implies that for large t, $f(t\boldsymbol{x}) \approx \exp\{-tg(\boldsymbol{x})\}$. By ideas related to the Laplace approximation (de Bruijn, 1958), we have

$$\begin{aligned} f_C(t\boldsymbol{x}_C) &= \int_{[0,\infty)^{d-|C|}} f(t\boldsymbol{x}) \prod_{i \notin C} dtx_i \\ &\approx \int_{[0,\infty)^{d-|C|}} \exp\{-tg(\boldsymbol{x})\} \prod_{i \notin C} dtx_i \propto \exp\left\{-t \min_{x_i : i \notin C} g(\boldsymbol{x})\right\}, \end{aligned}$$

as $t \to \infty$. As such, we assume that

$$-\log f_C\left(t\boldsymbol{x}_C\right) \sim t \min_{x_i : i \notin C} g(\boldsymbol{x}),$$

as $t \to \infty$, so that an approximation of the gauge function $g_C(\boldsymbol{x}_C)$ for the vector $\boldsymbol{X}_C$ can be obtained from the gauge function $g(\boldsymbol{x})$ for the vector $\boldsymbol{X}$, for any set $C \in \{1, \ldots, d\}$ with $|C| \in [2, d-1]$.

If analytical calculation of the gauge function $g_C(\boldsymbol{x}_C)$ is not possible, one could use a numerical minimization technique to find points $\boldsymbol{x}_{C,j}$ with $g_C(\boldsymbol{x}_{C,j}) \in [1-\epsilon, 1+\epsilon]$, for some small $\epsilon > 0$, in order to approximate the limit set $G_C = \{\boldsymbol{x}_C \in \mathbb{R}^{|C|} : g_C(\boldsymbol{x}_C) = 1\}$. Then steps 2 and 3 from the numerical method proposed in Section 5.2.2 could be used to approximate η_C.

5.3 Vine copulas with inverted extreme value pair copula components

5.3.1 Overview

In this section, we apply the techniques from Section 5.2 to calculate η_C for vine copulas constructed from inverted extreme value pair copulas. In Section 5.3.2, we focus on the trivariate case, obtaining the corresponding gauge function, and subsequently calculating η_C for $C = \{1, 3\}$ and $C = \{1, 2, 3\}$. We demonstrate the results by focussing on inverted logistic examples. In Section 5.3.3, we obtain results for higher dimensional vine copulas with inverted extreme value components. In this latter case, the underlying graphical structure of the vine copula can take different forms: we focus on D-vines and C-vines here, the concept of which will also be introduced in Section 5.3.3.

5.3.2 Trivariate case

A bivariate inverted extreme value copula with exponent measure V has the form

$$C(u, v) = u + v - 1 + \exp\left[-V\left\{\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)}\right\}\right].$$

Let V_1, V_2 and V_{12} denote the derivative of the exponent measure with respect to the first, second, and both components, respectively. Differentiating $C(u, v)$ with respect to the second component gives the conditional distribution function

$$F(u \mid v) = 1 + \left(\frac{1}{1-v}\right)\{-\log(1-v)\}^{-2} V_2\left\{\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)}\right\} \cdot \exp\left[-V\left\{\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)}\right\}\right], \tag{5.3.1}$$

and subsequently differentiating with respect to the first component gives the copula density

$$c(u, v) = \left(\frac{1}{1-u}\right)\left(\frac{1}{1-v}\right)\{-\log(1-u)\}^{-2}\{-\log(1-v)\}^{-2}$$

$$\cdot \exp\left[-V\left\{\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)}\right\}\right]$$
$$\cdot \left[V_1\left\{\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)}\right\} V_2\left\{\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)}\right\}\right.$$
$$\left. - V_{12}\left\{\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)}\right\}\right]. \tag{5.3.2}$$

In calculating values of η for a trivariate vine, we are interested in the behaviour, as $t \to \infty$, of

$$\begin{aligned} -\log f(t\boldsymbol{x}) = &-\log f_1(tx_1) - \log f_2(tx_2) - \log f_3(tx_3) \\ &- \log c_{12}\left\{F_1(tx_1), F_2(tx_2)\right\} - \log c_{23}\left\{F_2(tx_2), F_3(tx_3)\right\} \\ &- \log c_{13|2}\left\{F_{1|2}(tx_1|tx_2), F_{3|2}(tx_3|tx_2)\right\}. \end{aligned} \tag{5.3.3}$$

Consider a trivariate vine copula with three inverted extreme value components whose exponent measures are denoted $V^{\{12\}}$, $V^{\{23\}}$ and $V^{\{13|2\}}$, corresponding to the links in the vine. Working in standard exponential margins, we have $-\log f_i(tx_i) = -\log\left(e^{-tx_i}\right) = tx_i$, for $i = 1, 2, 3$. We also have marginal distribution functions $F_i(tx_i) = 1 - e^{-tx_i}$, for $i = 1, 2, 3$. Noting that $-\log\left\{1 - F_i(tx_i)\right\} = tx_i$; the exponent measure $V^{\{12\}}$ is homogeneous of order -1; $V_1^{\{12\}}$ and $V_2^{\{12\}}$ are homogeneous of order -2; and $V_{12}^{\{12\}}$ is homogeneous of order -3, we have

$$\begin{aligned} &-\log c_{12}\left\{F_1(tx_1), F_2(tx_2)\right\} = \\ &\quad - tx_1 - tx_2 + 2\log(tx_1) + 2\log(tx_2) + V^{\{12\}}\left\{(tx_1)^{-1}, (tx_2)^{-1}\right\} \\ &\quad - \log\left[V_1^{\{12\}}\left\{(tx_1)^{-1}, (tx_2)^{-1}\right\} V_2^{\{12\}}\left\{(tx_1)^{-1}, (tx_2)^{-1}\right\} - V_{12}^{\{12\}}\left\{(tx_1)^{-1}, (tx_2)^{-1}\right\}\right] \\ &= t\left\{V^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right) - x_1 - x_2\right\} + 2\left\{\log(tx_1) + \log(tx_2)\right\} \\ &\quad - \log\left\{t^4 V_1^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right) V_2^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right) - t^3 V_{12}^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right)\right\} \\ &= t\left\{V^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right) - x_1 - x_2\right\} + O(\log t), \end{aligned} \tag{5.3.4}$$

and by similar calculations,

$$-\log c_{23}\left\{F_2(tx_2), F_3(tx_3)\right\} = t\left\{V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_2 - x_3\right\} + O(\log t). \tag{5.3.5}$$

Substituting the marginal distribution functions $F_i(tx_i)$, $i = 1, 2$, into the corresponding conditional copulas, we have

$$\begin{aligned} F_{1|2}(tx_1 \mid tx_2) &= 1 + e^{tx_2}(tx_2)^{-2} V_2^{\{12\}} \left\{(tx_1)^{-1}, (tx_2)^{-1}\right\} \exp\left[-V^{\{12\}}\left\{(tx_1)^{-1}, (tx_2)^{-1}\right\}\right] \\ &= 1 + x_2^{-2} V_2^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right) \exp\left[t\left\{x_2 - V^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right)\right\}\right], \end{aligned}$$

and similarly,

$$F_{3|2}(tx_3 \mid tx_2) = 1 + x_2^{-2} V_1^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) \exp\left[t\left\{x_2 - V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right)\right\}\right].$$

Setting $h_{1|2} = \log\left\{-x_2^{-2} V_2^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right)\right\}$ and defining $h_{3|2}$ analogously,

$$\begin{aligned} \log\left\{1 - F_{1|2}(tx_1 \mid tx_2)\right\} &= h_{1|2} + t\left\{x_2 - V^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right)\right\}; \\ \log\left\{1 - F_{3|2}(tx_3 \mid tx_2)\right\} &= h_{3|2} + t\left\{x_2 - V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right)\right\}, \end{aligned}$$

with $h_{1|2}$ and $h_{3|2}$ not depending on t. This implies that

$$\begin{aligned} &-\log c_{13|2}\left\{F_{1|2}(tx_1 \mid tx_2), F_{3|2}(tx_3 \mid tx_2)\right\} \\ &\quad = \log\left\{1 - F_{1|2}(tx_1 \mid tx_2)\right\} + \log\left\{1 - F_{3|2}(tx_3 \mid tx_2)\right\} \\ &\qquad + 2\log\left[-\log\left\{1 - F_{1|2}(tx_1 \mid tx_2)\right\}\right] + 2\log\left[-\log\left\{1 - F_{3|2}(tx_3 \mid tx_2)\right\}\right] \\ &\qquad + V^{\{13|2\}}\left[\frac{-1}{\log\left\{1 - F_{1|2}(tx_1 \mid tx_2)\right\}}, \frac{-1}{\log\left\{1 - F_{3|2}(tx_3 \mid tx_2)\right\}}\right] \\ &\qquad - \log\left(V_1^{\{13|2\}}\left[\frac{-1}{\log\left\{1 - F_{1|2}(tx_1 \mid tx_2)\right\}}, \frac{-1}{\log\left\{1 - F_{3|2}(tx_3 \mid tx_2)\right\}}\right]\right. \\ &\qquad\qquad \cdot V_2^{\{13|2\}}\left[\frac{-1}{\log\left\{1 - F_{1|2}(tx_1 \mid tx_2)\right\}}, \frac{-1}{\log\left\{1 - F_{3|2}(tx_3 \mid tx_2)\right\}}\right] \\ &\qquad\qquad \left. - V_{12}^{\{13|2\}}\left[\frac{-1}{\log\left\{1 - F_{1|2}(tx_1 \mid tx_2)\right\}}, \frac{-1}{\log\left\{1 - F_{3|2}(tx_3 \mid tx_2)\right\}}\right]\right) \\ &\quad = h_{1|2} + h_{3|2} \\ &\qquad + t\left\{2x_2 - V^{\{12\}}(x_1^{-1}, x_2^{-1}) - V^{\{23\}}(x_2^{-1}, x_3^{-1})\right\} \\ &\qquad + 2\log\left[t\left\{V^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right) - x_2\right\} - h_{1|2}\right] \\ &\qquad + 2\log\left[t\left\{V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_2\right\} - h_{3|2}\right] \end{aligned}$$

$$
\begin{aligned}
&+ tV^{\{13|2\}}\left\{\frac{1}{V^{\{12\}}(x_1^{-1},x_2^{-1})-x_2-\frac{1}{t}h_{1|2}},\frac{1}{V^{\{23\}}(x_2^{-1},x_3^{-1})-x_2-\frac{1}{t}h_{3|2}}\right\}\\
&-\log\Bigg[t^4V_1^{\{13|2\}}\left\{\frac{1}{V^{\{12\}}(x_1^{-1},x_2^{-1})-x_2-\frac{1}{t}h_{1|2}},\frac{1}{V^{\{23\}}(x_2^{-1},x_3^{-1})-x_2-\frac{1}{t}h_{3|2}}\right\}\\
&\qquad\cdot V_2^{\{13|2\}}\left\{\frac{1}{V^{\{12\}}(x_1^{-1},x_2^{-1})-x_2-\frac{1}{t}h_{1|2}},\frac{1}{V^{\{23\}}(x_2^{-1},x_3^{-1})-x_2-\frac{1}{t}h_{3|2}}\right\}\\
&\qquad-t^3V_{12}^{\{13|2\}}\left\{\frac{1}{V^{\{12\}}(x_1^{-1},x_2^{-1})-x_2-\frac{1}{t}h_{1|2}},\frac{1}{V^{\{23\}}(x_2^{-1},x_3^{-1})-x_2-\frac{1}{t}h_{3|2}}\right\}\Bigg]\\
&=t\left\{2x_2-V^{\{12\}}\left(x_1^{-1},x_2^{-1}\right)-V^{\{23\}}\left(x_2^{-1},x_3^{-1}\right)\right\}\\
&+tV^{\{13|2\}}\left[\left\{V^{\{12\}}\left(x_1^{-1},x_2^{-1}\right)-x_2\right\}^{-1},\left\{V^{\{23\}}\left(x_2^{-1},x_3^{-1}\right)-x_2\right\}^{-1}\right]\\
&+O(\log t).
\end{aligned}
\tag{5.3.6}
$$

Substituting results (5.3.4), (5.3.5) and (5.3.6) into (5.3.3) yields

$$
\begin{aligned}
-\log f(t\boldsymbol{x}) &= t(x_1+x_2+x_3)\\
&\quad+t\left\{V^{\{12\}}\left(x_1^{-1},x_2^{-1}\right)-x_1-x_2\right\}\\
&\quad+t\left\{V^{\{23\}}\left(x_2^{-1},x_3^{-1}\right)-x_2-x_3\right\}\\
&\quad+t\left\{2x_2-V^{\{12\}}\left(x_1^{-1},x_2^{-1}\right)-V^{\{23\}}\left(x_2^{-1},x_3^{-1}\right)\right\}\\
&\quad+tV^{\{13|2\}}\left[\left\{V^{\{12\}}\left(x_1^{-1},x_2^{-1}\right)-x_2\right\}^{-1},\left\{V^{\{23\}}\left(x_2^{-1},x_3^{-1}\right)-x_2\right\}^{-1}\right]\\
&\quad+O(\log t)\\
&=t\left(x_2+V^{\{13|2\}}\left[\left\{V^{\{12\}}\left(x_1^{-1},x_2^{-1}\right)-x_2\right\}^{-1},\left\{V^{\{23\}}\left(x_2^{-1},x_3^{-1}\right)-x_2\right\}^{-1}\right]\right)\\
&\quad+O(\log t),
\end{aligned}
\tag{5.3.7}
$$

so that the gauge function of a trivariate vine copula with inverted extreme value components is

$$
g(\boldsymbol{x})=x_2+V^{\{13|2\}}\left[\left\{V^{\{12\}}\left(x_1^{-1},x_2^{-1}\right)-x_2\right\}^{-1},\left\{V^{\{23\}}\left(x_2^{-1},x_3^{-1}\right)-x_2\right\}^{-1}\right].
$$

Inverted logistic example

An example of an inverted extreme value copula is the inverted logistic copula, where the exponent measure is of the form $V(x,y)=\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha}$. Suppose all three

pair copulas that make up the trivariate vine have this form, with $V^{\{12\}}$, $V^{\{23\}}$ and $V^{\{13|2\}}$ having dependence parameters α, β and γ, respectively. Then the corresponding gauge function is

$$g(\boldsymbol{x}) = x_2 + \left[\left\{ \left(x_1^{1/\alpha} + x_2^{1/\alpha} \right)^{\alpha} - x_2 \right\}^{1/\gamma} + \left\{ \left(x_2^{1/\beta} + x_3^{1/\beta} \right)^{\beta} - x_2 \right\}^{1/\gamma} \right]^{\gamma}. \quad (5.3.8)$$

Figure 5.3.1 demonstrates the sets $G = \{\boldsymbol{x} \in \mathbb{R}^3 : g(\boldsymbol{x}) = 1\}$ and $[\eta_{\{1,2,3\}}, \infty)^3$ for this gauge function, with $\alpha \in \{0.25, 0.5, 0.75\}$, $\beta = 0.25$ and $\gamma = 0.5$.

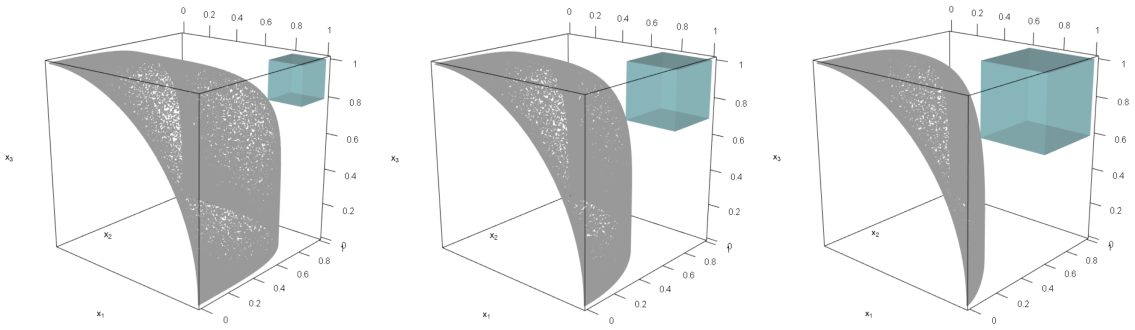

Figure 5.3.1: Points in the set $G = \{\boldsymbol{x} \in \mathbb{R}^3 : g(\boldsymbol{x}) = 1\}$ for a trivariate vine with three inverted logistic pair copula components (grey) and the set $[\eta_{\{1,2,3\}}, \infty)^3$ (blue): $\alpha = 0.25$ (left), $\alpha = 0.5$ (centre), $\alpha = 0.75$ (right); $\beta = 0.25$ and $\gamma = 0.5$.

For gauge functions of the form (5.3.8), numerical investigations suggest that the intersection of G and $[\eta_{\{1,2,3\}}, \infty)^3$ occurs on the diagonal, i.e., when $x_1 = x_2 = x_3$, and this is supported by the plots in Figure 5.3.1. By homogeneity of the gauge function, and since $\eta_{\{1,2,3\}}$ solves the equation $g\left(\eta_{\{1,2,3\}}, \eta_{\{1,2,3\}}, \eta_{\{1,2,3\}}\right) = 1$, we have

$$\eta_{\{1,2,3\}} = \frac{1}{g(1,1,1)} = \frac{1}{1 + \left\{ (2^{\alpha} - 1)^{1/\gamma} + (2^{\beta} - 1)^{1/\gamma} \right\}^{\gamma}}. \quad (5.3.9)$$

For $\alpha = \beta = \gamma = 1$, the model has $\eta_{\{1,2,3\}} = 1/3$, corresponding to complete independence, and as $\alpha, \beta, \gamma \to 0$, $\eta_{\{1,2,3\}} \to 1$, corresponding to asymptotic dependence between all three variables. It will not always be the case that the intersection of G

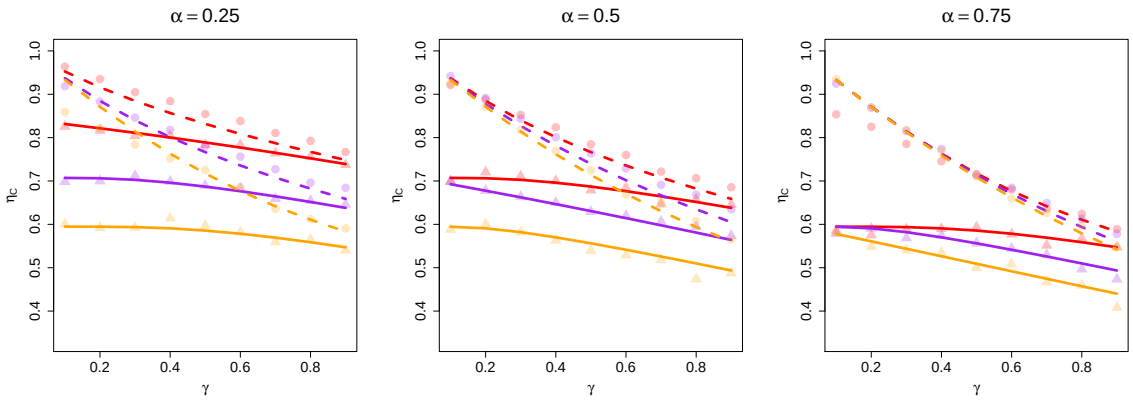

Figure 5.3.2: Values of $\eta_{\{1,3\}}$ (dashed) and $\eta_{\{1,2,3\}}$ (solid) for a trivariate vine copula with inverted logistic components, with $\alpha = 0.25$ (left), $\alpha = 0.5$ (centre), $\alpha = 0.75$ (right); $\beta = 0.25$ (red), $\beta = 0.5$ (purple) and $\beta = 0.75$ (orange); and $\gamma \in (0.1, 0.9)$. Average Hill estimates of $\eta_{\{1,3\}}$ (circles) and $\eta_{\{1,2,3\}}$ (triangles) are provided in each case.

and $[\eta_{\{1,2,3\}}, \infty)^3$ occurs on the diagonal, but we note that the value of $1/g(1,1,1)$ will at least provide a lower bound for the value of $\eta_{\{1,2,3\}}$ in general.

In Figure 5.3.2, we demonstrate the behaviour of $\eta_{\{1,2,3\}}$ in (5.3.9) across a range of parameter values. We also present estimated values of $\eta_{\{1,2,3\}}$ obtained by applying the Hill estimate (Hill, 1975) to samples of size 100,000 from the corresponding vine copula, and taking the average over 100 replications. The threshold in the Hill estimation is taken to be the 0.99 quantile of observed minima in each case, and we present results for $\alpha, \beta \in \{0.25, 0.5, 0.75\}$ and $\gamma \in \{0.1, 0.2, \ldots, 0.9\}$. The average Hill estimates lie close to the values obtained using equation (5.3.9), and therefore support our theoretical results.

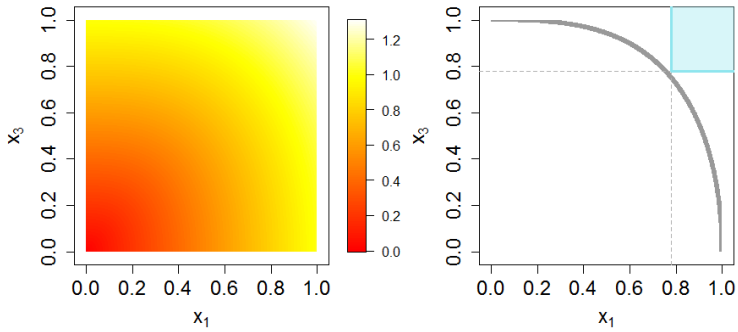

Figure 5.3.3: Left: $g_{\{1,3\}}(x_1, x_3)$ for $(\alpha, \beta, \gamma) = (0.5, 0.25, 0.5)$. Right: an approximation of the corresponding sets $G_{\{1,3\}}$ (grey) and $[\eta_{\{1,3\}}, \infty)^2$ (blue).

$\eta_{\{1,3\}}$ for the inverted logistic example

The joint density of (X_1, X_3) cannot be found analytically for a trivariate vine with inverted logistic pair copula components; we therefore use the method proposed in Section 5.2.3, with the gauge function for this pair of variables being $g_{\{1,3\}}(x_1, x_3) = \min_{x_2} g(\boldsymbol{x})$, for $g(\boldsymbol{x})$ in (5.3.8). We investigate the behaviour of this function by carrying out the required minimization numerically. The left panel of Figure 5.3.3 shows the value of $g_{\{1,3\}}(x_1, x_3)$ in the region $[0, 1]^2$, for $(\alpha, \beta, \gamma) = (0.5, 0.25, 0.5)$, chosen to match the parameter values for the central panel of Figure 5.3.1. In the right panel of Figure 5.3.3, we approximate the set $G_{\{1,3\}}$ by considering values of $g_{\{1,3\}}(x_1, x_3) \in (0.99, 1.01)$. As for the $\eta_{\{1,2,3\}}$ case, we see that the intersection of $G_{\{1,3\}}$ and $[\eta_{\{1,3\}}, \infty)^2$ occurs on the diagonal, i.e., when $x_1 = x_3$. Further investigations suggest this is generally the case for vine copulas constructed from inverted logistic pair copulas.

By focussing on the case where $x_1 = x_3$, to calculate $\eta_{\{1,3\}}$, we need to first minimize $g(x_1, y, x_1)$ with respect to y to obtain the gauge function for (X_1, X_3). Since all gauge functions are homogeneous of order 1, we set $y = vx_1$, so that $g_{\{1,3\}}(x_1, x_1) =$

$\min_v g(x_1, vx_1, x_1)$. Differentiating $g(x_1, vx_1, x_1)$ with respect to v, we have

$$\frac{dg(x_1, vx_1, x_1)}{dv} = x_1 \Bigg\{ 1 + \left[\left\{ (1+v^{1/\alpha})^\alpha - v \right\}^{1/\gamma} + \left\{ (1+v^{1/\beta})^\beta - v \right\}^{1/\gamma} \right]^{\gamma-1} \cdot \Bigg[\left\{ v^{-1+1/\alpha}(1+v^{1/\alpha})^{\alpha-1} - 1 \right\} \left\{ (1+v^{1/\alpha})^\alpha - v \right\}^{-1+1/\gamma} + \left\{ v^{-1+1/\beta}(1+v^{1/\beta})^{\beta-1} - 1 \right\} \left\{ (1+v^{1/\beta})^\beta - v \right\}^{-1+1/\gamma} \Bigg] \Bigg\}.$$

From Figure 5.3.3, we deduce that there is only one value of x_1 satisfying $g_{\{1,3\}}(x_1, x_1) = 1$. This corresponds to the value of $\eta_{\{1,3\}}$, so that $\eta_{\{1,3\}} = 1/g_{\{1,3\}}(1,1) = 1/\min_v g(1, v, 1)$, by homogeneity of the gauge functions. Combining this information, we have

$$\eta_{\{1,3\}} = \left(v + \left[\left\{ \left(1+v^{1/\alpha}\right)^\alpha - v \right\}^{1/\gamma} + \left\{ \left(1+v^{1/\beta}\right)^\beta - v \right\}^{1/\gamma} \right]^\gamma \right)^{-1},$$

with v satisfying $dg(1, v, 1)/dv = 0$, i.e.,

$$1 + \left[\left\{ \left(1+v^{1/\alpha}\right)^\alpha - v \right\}^{1/\gamma} + \left\{ \left(1+v^{1/\beta}\right)^\beta - v \right\}^{1/\gamma} \right]^{\gamma-1} \cdot \Bigg[\left\{ (1+v^{-1/\alpha})^{\alpha-1} - 1 \right\} \left\{ (1+v^{1/\alpha})^\alpha - v \right\}^{-1+1/\gamma} + \left\{ (1+v^{-1/\beta})^{\beta-1} - 1 \right\} \left\{ (1+v^{1/\beta})^\beta - v \right\}^{-1+1/\gamma} \Bigg] = 0. \quad (5.3.10)$$

In Figure 5.3.2, we calculate $\eta_{\{1,3\}}$ for a range of parameter values by solving equation (5.3.10) numerically, and compare to the value of $\eta_{\{1,2,3\}}$ in each case. We also present average Hill estimates of $\eta_{\{1,3\}}$ obtained using a similar approach as for the estimates of $\eta_{\{1,2,3\}}$, which again appear to support our theoretical results. From Figure 5.3.2, we see that a range of tail independence properties can be achieved for the pair (X_1, X_3) using this model, from complete independence ($\eta_{\{1,3\}} = 1/2$) to asymptotic independence with strong residual dependence ($\eta_{\{1,3\}} \to 1$), which is analogous to our observations for $\eta_{\{1,2,3\}}$ for this model.

5.3.3 Higher dimensions and different vine structures

We now extend the results of Section 5.3.2 by considering vine copulas with dimension $d > 3$ constructed from inverted extreme value pair copulas, with the aim being to find

the gauge function and value of η_D in each case. When working with higher than three dimensions, the graphical structure of the vine becomes an important consideration. In particular, the trees used in the graphical representations can take different forms. We focus on two such cases: gauge functions for the class of vine copulas known as D-vines, where all trees in the vine are paths; and gauge functions for C-vines, which have exactly one node that is connected to all other nodes in each tree. These two classes are demonstrated in Figure 5.3.4 for the case $d = 4$. In the final part of this section, we demonstrate the values of η_D calculated using these gauge functions for both classes of model.

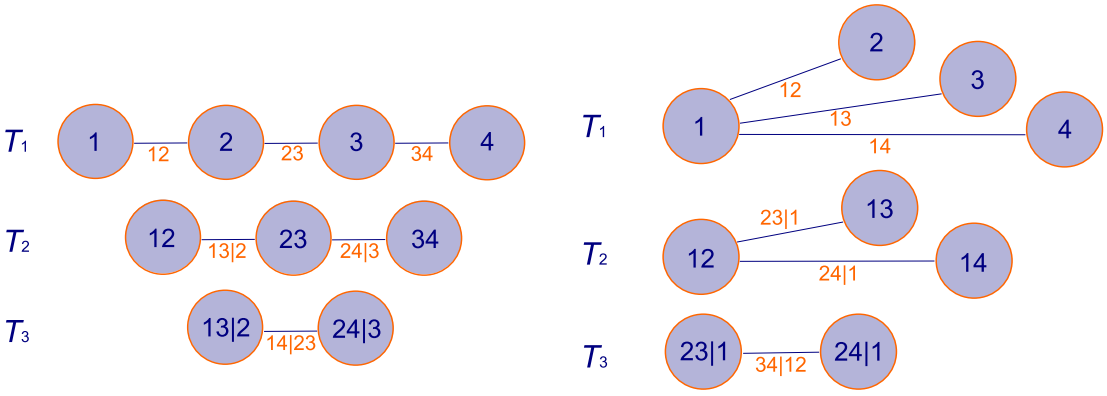

Figure 5.3.4: Graphical representations of four dimensional vine copula models; D-vine (left) and C-vine (right).

Gauge functions for D-vines

A d-dimensional D-vine is made up of $(d-1)$ trees, labelled $T_1, \ldots, T_{d-1}$, and a total of $(d-1)d/2$ edges. We suppose that the pair copula represented by each edge is an inverted extreme value copula, with the superscript on the exponent measure corresponding to the edge-label, as in the trivariate case. For the four-dimensional example in Figure 5.3.4, we have

$$-\log f(t\boldsymbol{x}) = -\log f_1(tx_1) - \log f_2(tx_2) - \log f_3(tx_3) - \log f_4(tx_4)$$

$$
\begin{aligned}
&- \log c_{12}\left\{F_1(tx_1), F_2(tx_2)\right\} - \log c_{23}\left\{F_2(tx_2), F_3(tx_3)\right\} - \log c_{34}\left\{F_3(tx_3), F_4(tx_4)\right\} \\
&- \log c_{13|2}\left\{F_{1|2}(tx_1|tx_2), F_{3|2}(tx_3|tx_2)\right\} - \log c_{24|3}\left\{F_{2|3}(tx_2|tx_3), F_{4|3}(tx_4|tx_3)\right\} \\
&- \log c_{14|23}\left\{F_{1|23}(tx_1|tx_2, tx_3), F_{4|23}(tx_4|tx_2, tx_3)\right\}. \qquad (5.3.11)
\end{aligned}
$$

We note that several of these terms can be thought of in terms of lower-dimensional vine copulas that are subsets of the four-dimensional vine. In particular, all terms in the trivariate formula (5.3.3) for the set of variables (X_1, X_2, X_3) appear in (5.3.11). Let f_{123} denote the joint density corresponding to this trivariate case. The density f_{234} corresponding to variables (X_2, X_3, X_4) also comes from a trivariate vine copula equivalent to f_{123} up to a labelling of the variables. The sections of the four-dimensional vine corresponding to these two trivariate subsets are highlighted in Figure 5.3.5, and can be thought of as sub-vines of the overall vine copula. We note that these two sub-vines overlap in the centre, as they share the variables (X_2, X_3). This suggests that if we try to represent $-\log f$ for the overall model in terms of $-\log f_{123}$ and $-\log f_{234}$, we will count the section corresponding to $-\log f_{23}$ twice, with f_{23} denoting the joint density of (X_2, X_3). Taking this inclusion-exclusion into account, equation (5.3.11) can be simplified to

$$
\begin{aligned}
-\log f(t\boldsymbol{x}) = &- \log f_{123}(tx_1, tx_2, tx_3) - \log f_{234}(tx_2, tx_3, tx_4) + \log f_{23}(tx_2, tx_3) \\
&- \log c_{14|23}\left\{F_{1|23}(tx_1|tx_2, tx_3), F_{4|23}(tx_4|tx_2, tx_3)\right\}. \qquad (5.3.12)
\end{aligned}
$$

In Section 5.3.2, we studied gauge functions of trivariate vine copulas with inverted extreme value components. From equation (5.3.7), we have

$$
\begin{aligned}
&-\log f_{123}(tx_1, tx_2, tx_3) \\
&\quad = t\left(x_2 + V^{\{13|2\}}\left[\left\{V^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right) - x_2\right\}^{-1}, \left\{V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_2\right\}^{-1}\right]\right) \\
&\qquad + O(\log t), \qquad (5.3.13)
\end{aligned}
$$

and considering f_{234} as equivalent to f_{123} up to labelling of the variables, we have

$$
-\log f_{234}(tx_2, tx_3, tx_4)
$$

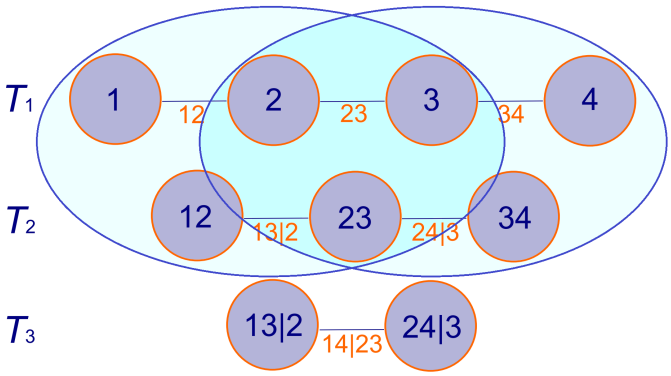

Figure 5.3.5: Trivariate subsets of the four-dimensional D-vine.

$$= t\left(x_3 + V^{\{24|3\}}\left[\left\{V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_3\right\}^{-1}, \left\{V^{\{34\}}\left(x_3^{-1}, x_4^{-1}\right) - x_3\right\}^{-1}\right]\right) + O(\log t). \tag{5.3.14}$$

Moreover, from the four-dimensional vine, the pair (X_2, X_3) is modelled by an inverted extreme value copula with exponent measure $V^{\{23\}}$, so by equation (5.2.2) we have

$$\log f_{23}(tx_2, tx_3) = -tV^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) + O(\log t). \tag{5.3.15}$$

As such, $-\log c_{14|23}\left\{F_{1|23}(tx_1|tx_2, tx_3), F_{4|23}(tx_4|tx_2, tx_3)\right\}$ is the only term of (5.3.11) left for us to study. Joe (1996) shows that

$$F_{1|23}(x_1|x_2, x_3) = \frac{\partial C_{13|2}\left\{F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)\right\}}{\partial F_{3|2}(x_3|x_2)},$$

and

$$F_{4|23}(x_4|x_2, x_3) = \frac{\partial C_{24|3}\left\{F_{2|3}(x_2|x_3), F_{4|3}(x_4|x_3)\right\}}{\partial F_{2|3}(x_2|x_3)}.$$

Since we focus on inverted extreme value copulas, the derivatives have the form (5.3.1), where the arguments u and v are replaced by the appropriate bivariate conditional distributions. We find that for some functions $h_{1|23}, h_{4|23}$ not depending on t,

$$1 - F_{1|23}(tx_1 \mid tx_2, tx_3) \sim \exp(h_{1|23}) \cdot \exp\left\{-t\left(x_2 - V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right)\right.\right.$$

$$+ V^{\{13|2\}}\left[\left\{V^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right) - x_2\right\}^{-1}, \left\{V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_2\right\}^{-1}\right]\Bigg)\Bigg\} \tag{5.3.16}$$

and

$$\begin{aligned}1 - F_{4|23}(tx_4 \mid tx_2, tx_3) \sim \\ \exp(h_{4|23}) \cdot \exp\Bigg\{ - t\Bigg(x_3 - V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) \\ + V^{\{24|3\}}\left[\left\{V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_3\right\}^{-1}, \left\{V^{\{34\}}\left(x_3^{-1}, x_4^{-1}\right) - x_3\right\}^{-1}\right]\Bigg)\Bigg\}.\end{aligned} \tag{5.3.17}$$

This is similar to the results for the trivariate vine, and applying an argument analogous to (5.3.6), we see that

$$\begin{aligned}&- \log c_{14|23}\left\{F_{1|23}(tx_1|tx_2, tx_3), F_{4|23}(tx_4|tx_2, tx_3)\right\} \sim \\ &- x_2 + V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - V^{\{13|2\}}\left[\left\{V^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right) - x_2\right\}^{-1}, \left\{V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_2\right\}^{-1}\right] \\ &- x_3 + V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - V^{\{24|3\}}\left[\left\{V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_3\right\}^{-1}, \left\{V^{\{34\}}\left(x_3^{-1}, x_4^{-1}\right) - x_3\right\}^{-1}\right] \\ &+ V^{\{14|23\}}\Bigg\{\Bigg(x_2 - V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) \\ &\qquad + V^{\{13|2\}}\left[\left\{V^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right) - x_2\right\}^{-1}, \left\{V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_2\right\}^{-1}\right]\Bigg)^{-1}, \\ &\quad \Bigg(x_3 - V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) \\ &\qquad + V^{\{24|3\}}\left[\left\{V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_3\right\}^{-1}, \left\{V^{\{34\}}\left(x_3^{-1}, x_4^{-1}\right) - x_3\right\}^{-1}\right]\Bigg)^{-1}\Bigg\}.\end{aligned} \tag{5.3.18}$$

We can now substitute the results (5.3.13), (5.3.14), (5.3.15) and (5.3.18) into equation (5.3.12) to obtain the gauge function for this four-dimensional vine copula as

$$\begin{aligned}g(\boldsymbol{x}) =& V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) \\ &+ V^{\{14|23\}}\Bigg\{\Bigg(x_2 - V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right)\end{aligned}$$

$$+ V^{\{13|2\}}\left[\left\{V^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right) - x_2\right\}^{-1}, \left\{V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_2\right\}^{-1}\right]\Bigg)^{-1},$$

$$\Bigg(x_3 - V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right)$$

$$+ V^{\{24|3\}}\left[\left\{V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_3\right\}^{-1}, \left\{V^{\{34\}}\left(x_3^{-1}, x_4^{-1}\right) - x_3\right\}^{-1}\right]\Bigg)^{-1}\Bigg\}.$$

This gauge function can be written more simply in terms of the gauge functions of the three sub-vines highlighted in Figure 5.3.5 and the exponent measure corresponding to the pair copula in tree T_3. That is,

$$g(\boldsymbol{x}) = g_{\{2,3\}}(x_2, x_3) + V^{\{14|23\}}\left\{\frac{1}{g_{\{1,2,3\}}(x_1, x_2, x_3) - g_{\{2,3\}}(x_2, x_3)}, \frac{1}{g_{\{2,3,4\}}(x_2, x_3, x_4) - g_{\{2,3\}}(x_2, x_3)}\right\}. \tag{5.3.19}$$

For D-vine copulas, this same structure can be extended to higher dimensions, creating an iterative formula for calculating the gauge function; this is stated in Theorem 3.

Theorem 3. *The gauge function for a d-dimensional D-vine with inverted extreme value pair copula components is given by*

$$g(\boldsymbol{x}) = g_{D\setminus\{1,d\}}(\boldsymbol{x}_{-\{1,d\}}) + V^{\{1,d|D\setminus\{1,d\}\}}\left\{\frac{1}{g_{D\setminus\{d\}}(\boldsymbol{x}_{-\{d\}}) - g_{D\setminus\{1,d\}}(\boldsymbol{x}_{-\{1,d\}})}, \frac{1}{g_{D\setminus\{1\}}(\boldsymbol{x}_{-\{1\}}) - g_{D\setminus\{1,d\}}(\boldsymbol{x}_{-\{1,d\}})}\right\}.$$

Theorem 3 is proved in Appendix C.1. We discuss how to obtain η_D from this gauge function at the end of this section.

Gauge functions for C-vines

Using similar arguments as for the D-vines in the previous section, we can construct an iterative formula for the gauge functions of d-dimensional C-vines. We now consider the sub-vines as corresponding to the sets of variables $\boldsymbol{X}_{-d}$ and $\boldsymbol{X}_{-(d-1)}$, which overlap at $\boldsymbol{X}_{-\{(d-1),d\}}$. This is demonstrated in Figure 5.3.6 for the four-dimensional

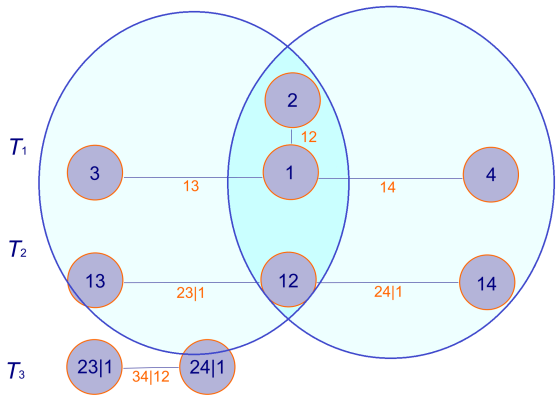

Figure 5.3.6: Trivariate subsets of the four-dimensional C-vine.

case.

Following the same steps as in the previous section, we obtain the gauge function

$$g(\boldsymbol{x}) = g_{D\setminus\{(d-1,d\}}(\boldsymbol{x}_{-\{d-1,d\}}) + V^{\{d-1,d|D\setminus\{d-1,d\}\}}\left\{\frac{1}{g_{D\setminus\{d\}}(\boldsymbol{x}_{-\{d\}}) - g_{D\setminus\{d-1,d\}}(\boldsymbol{x}_{-\{d-1,d\}})}, \frac{1}{g_{D\setminus\{d-1\}}(\boldsymbol{x}_{-\{d-1\}}) - g_{D\setminus\{d-1,d\}}(\boldsymbol{x}_{-\{d-1,d\}})}\right\},$$

and discuss using this to calculate η_D in the remainder of this section.

Calculating η_D for d-dimensional D-vines and C-vines with inverted logistic components

As for the trivariate vine copula examples with inverted logistic pair copula components, numerical results suggest that the intersection of the set $G = \{\boldsymbol{x} \in \mathbb{R}^d : g(\boldsymbol{x}) = 1\}$ and $[\eta_D, \infty)^d$ for these D-vines and C-vines occurs when $x_1 = x_2 \cdots = x_d$. As before, we can exploit the homogeneity of the gauge function, and conclude that $\eta_D = g(1, \ldots, 1)^{-1}$ in this case.

Due to the nested structure of the gauge functions, the value of η_D can be written

in terms of the values of η_C for various sub-vines of the copula, and the exponent measure corresponding to tree T_{d-1} of the vine. In particular, for D-vines, we have

$$\eta_D = \left\{ \eta_{D\setminus\{1,d\}}^{-1} + V^{\{1,d|D\setminus\{1,d\}\}} \left(\frac{1}{\eta_{D\setminus\{d\}}^{-1} - \eta_{D\setminus\{1,d\}}^{-1}}, \frac{1}{\eta_{D\setminus\{1\}}^{-1} - \eta_{D\setminus\{1,d\}}^{-1}} \right) \right\}^{-1} \tag{5.3.20}$$

and for C-vines,

$$\eta_D = \left\{ \eta_{D\setminus\{d-1,d\}}^{-1} + V^{\{d-1,d|D\setminus\{d-1,d\}\}} \left(\frac{1}{\eta_{D\setminus\{d\}}^{-1} - \eta_{D\setminus\{d-1,d\}}^{-1}}, \frac{1}{\eta_{D\setminus\{(d-1)\}}^{-1} - \eta_{D\setminus\{d-1,d\}}^{-1}} \right) \right\}^{-1}.$$

Setting $\eta_C = 1$ for $|C| = 1$, we now have an iterative method for calculating the values of η_D for these classes of model for $d \geq 3$ dimensions.

As an example, we consider the case where all the pair copulas of the vine are inverted logistic with the same dependence parameter $\alpha \in (0, 1)$. In this case, the known value of η_D for the bivariate copula is $2^{-\alpha}$. We can therefore use our iterative formulas to calculate η_D for higher dimensional vine copulas. Since the exponent is homogeneous of order -1, the expression for η_D in (5.3.20) in this case simplifies to

$$\eta_D = \left\{ \eta_{D\setminus\{1,d\}}^{-1} + 2^{\alpha} \left(\eta_{D\setminus\{d\}}^{-1} - \eta_{D\setminus\{1,d\}}^{-1} \right) \right\}^{-1}, \tag{5.3.21}$$

and we can use the iterative method to derive the exact value of η_D for any d-dimensional D-vine copula. For a trivariate vine copula we have,

$$\eta_D = \left\{ 1 + 2^{\alpha} \left(2^{\alpha} - 1 \right) \right\}^{-1}, \tag{5.3.22}$$

and for a four-dimensional vine copula,

$$\eta_D = \left[2^{\alpha} + 2^{\alpha} \left\{ 1 + 2^{\alpha} \left(2^{\alpha} - 1 \right) - 2^{\alpha} \right\} \right]^{-1} = \left\{ 2^{\alpha} + 2^{\alpha} \left(2^{\alpha} - 1 \right)^2 \right\}^{-1}. \tag{5.3.23}$$

For this particular example, we can extend the results to higher d-dimensional vine copulas, yielding, for $d \geq 3$,

$$\eta_D = \begin{cases} \left\{ 1 + 2^{\alpha} \sum_{k=1}^{(d-1)/2} \left(2^{\alpha} - 1 \right)^{2(k-1)+1} \right\}^{-1}, & \text{for } d \text{ odd}, \\ \left\{ 2^{\alpha} \sum_{k=1}^{d/2} \left(2^{\alpha} - 1 \right)^{2(k-1)} \right\}^{-1}, & \text{for } d \text{ even}. \end{cases} \tag{5.3.24}$$

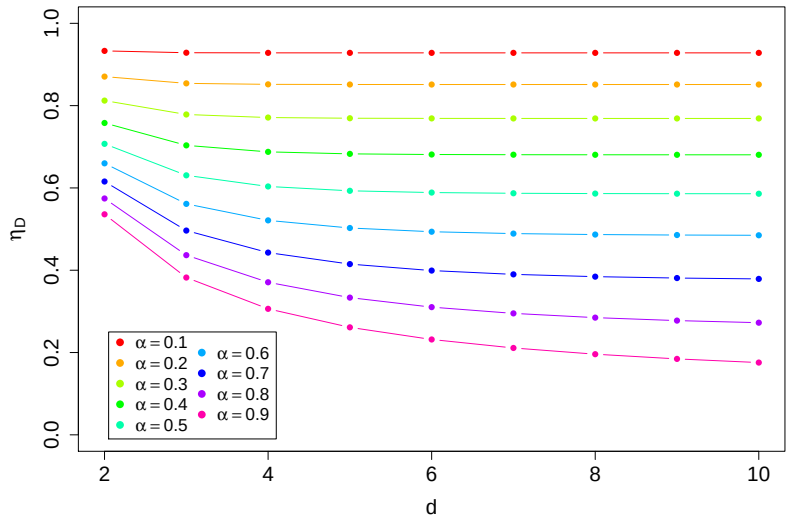

Figure 5.3.7: Values of η_D for $d \in \{2, \ldots, d\}$ for a d-dimensional D-vine or C-vine with inverted logistic pair copulas with dependence parameters $\alpha \in \{0.1, 0.2, \ldots, 0.9\}$.

We prove result (5.3.24) by induction in Appendix C.2. We note that since the pair copulas, and therefore the corresponding exponent measures, are all taken to be identical, the value of η_D is the same for the D-vines and C-vines of the same dimension. These values are demonstrated in Figure 5.3.7 for $\alpha \in \{0.1, 0.2, \ldots, 0.9\}$ and $d \in \{2, \ldots, 10\}$, where we have $\eta_D < 1$ in all cases, corresponding to asymptotic independence. Complete independence in the d-dimensional vine copula corresponds to $\eta_D = 1/d$. We see from Figure 5.3.7 that for $\alpha = 0.9$, we approach this case, while for $\alpha = 0.1$, the values of η_D are close to 1, corresponding to strong residual dependence. These models are therefore able to capture a range of sub-asymptotic dependence strengths in the asymptotic independence case.

5.4 Trivariate vine copulas with extreme value and inverted extreme value pair copula components

5.4.1 Overview

We have so far focussed on the tail dependence properties of vine copulas with inverted extreme value pair copula components. In this section, we investigate these same properties for trivariate vine copulas where the components are either extreme value or inverted extreme value copulas. We consider five such cases, which along with the results in Section 5.3.2 cover the range of possible scenarios: the two copulas in tree T_1 being inverted extreme value with an extreme value copula in tree T_2; tree T_1 having one extreme value and one inverted extreme value copula with the copula in tree T_2 being either extreme value or inverted extreme value; and both copulas in tree T_1 being from the extreme value family with the copula in tree T_2 being either extreme value or inverted extreme value. In Section 5.4.2, we demonstrate how to calculate the gauge function for each of these cases, for the extreme value components satisfying certain conditions, and then investigate their corresponding values of $\eta_{\{1,2,3\}}$ and $\eta_{\{1,3\}}$ for (inverted) logistic examples in Section 5.4.3.

5.4.2 Gauge functions for trivariate vines with extreme value and inverted extreme value components

For any trivariate vine copula, $-\log f(t\boldsymbol{x})$ has the form (5.3.3). In our gauge function calculations, we work with variables having exponential margins, so that the first three terms of this expression always satisfy

$$-\log f_1(tx_1) - \log f_2(tx_2) - \log f_3(tx_3) = t(x_1 + x_2 + x_3), \tag{5.4.1}$$

and we have $F_i(tx_i) = 1 - e^{-tx_i}$, for $i = 1, 2, 3$. For inverted extreme value copulas, we have shown in equations (5.3.4) and (5.3.5) that

$$-\log c_{12}\left\{F_1(tx_1), F_2(tx_2)\right\} = t\left\{V^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right) - x_1 - x_2\right\} + O(\log t) \tag{5.4.2}$$

and

$$-\log c_{23}\left\{F_2(tx_2), F_3(tx_3)\right\} = t\left\{V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_2 - x_3\right\} + O(\log t). \tag{5.4.3}$$

In order to investigate the behaviour of the extreme value pair copula components, we impose the condition that the corresponding spectral densities place no mass on $\{0\}$ or $\{1\}$ and have regularly varying tails. Let $h^{\{12\}}(w)$, $h^{\{23\}}(w)$, $h^{\{13|2\}}(w)$ denote the spectral density for each pair copula component. We assume that each of these densities has $h^{\{\cdot\}}(w) \sim c_1^{\{\cdot\}}(1-w)^{s_1^{\{\cdot\}}}$ as $w \to 1$ and $h^{\{\cdot\}}(w) \sim c_2^{\{\cdot\}} w^{s_2^{\{\cdot\}}}$ as $w \to 0$, for some $c_1^{\{\cdot\}}, c_2^{\{\cdot\}} \in \mathbb{R}$ and $s_1^{\{\cdot\}}, s_2^{\{\cdot\}} > -1$. In Appendix C.3.2, we show that

$$\begin{aligned}-\log c_{12}\{F_1(tx_1), F_2(tx_2)\} \sim & t\Big\{\left(1 + s_1^{\{12\}}\mathbb{1}_{\{x_1 \geq x_2\}} + s_2^{\{12\}}\mathbb{1}_{\{x_1 < x_2\}}\right)\max(x_1, x_2) \\ & - \left(2 + s_1^{\{12\}}\mathbb{1}_{\{x_1 \geq x_2\}} + s_2^{\{12\}}\mathbb{1}_{\{x_1 < x_2\}}\right)\min(x_1, x_2)\Big\},\end{aligned} \tag{5.4.4}$$

and

$$\begin{aligned}-\log c_{23}\{F_2(tx_2), F_3(tx_3)\} \sim & t\Big\{\left(1 + s_1^{\{23\}}\mathbb{1}_{\{x_2 \geq x_3\}} + s_2^{\{23\}}\mathbb{1}_{\{x_2 < x_3\}}\right)\max(x_2, x_3) \\ & - \left(2 + s_1^{\{23\}}\mathbb{1}_{\{x_2 \geq x_3\}} + s_2^{\{23\}}\mathbb{1}_{\{x_2 < x_3\}}\right)\min(x_2, x_3)\Big\},\end{aligned} \tag{5.4.5}$$

so that the final term of (5.3.3), i.e., $-\log c_{13|2}\left\{F_{1|2}(tx_1|tx_2), F_{3|2}(tx_3|tx_2)\right\}$, is the only one left for us to consider.

In Section 5.3.2, we showed that for the components in tree T_1 being inverted extreme value copulas, we have

$$\begin{aligned}F_{1|2}(tx_1 \mid tx_2) &= 1 + x_2^{-2}V_2^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right)\exp\left[t\left\{x_2 - V^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right)\right\}\right] \\ &= 1 - a_{1|2}\exp\left[-t\left\{V^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right) - x_2\right\}\right];\end{aligned} \tag{5.4.6}$$

$$F_{3|2}(tx_3 \mid tx_2) = 1 + x_2^{-2} V_1^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) \exp\left[t\left\{x_2 - V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right)\right\}\right]$$
$$= 1 - a_{3|2} \exp\left[-t\left\{V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_2\right\}\right], \tag{5.4.7}$$

for $a_{1|2} = -x_2^{-2} V_2^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right)$ and $a_{3|2} = -x_2^{-2} V_1^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right)$. We consider the extreme value case in Appendix C.3.3. Placing the same conditions on the spectral density as for the $-\log c_{12}$ and $-\log c_{23}$ calculations, we obtain

$$F_{1|2}(tx_1 \mid tx_2) = \begin{cases} \frac{2c_2^{\{12\}}}{(s_2^{\{12\}}+1)} \exp\left\{t(x_1 - x_2)(s_2^{\{12\}} + 1)\right\}\{1 + o(1)\}, & x_1 < x_2, \\ -V_2^{\{12\}}(1,1)\{1 + o(1)\}, & x_1 = x_2, \\ 1 - \frac{2c_1^{\{12\}}}{(s_1^{\{12\}}+2)} \exp\left\{-t(x_1 - x_2)(s_1^{\{12\}} + 2)\right\}\{1 + o(1)\}, & x_1 > x_2, \end{cases} \tag{5.4.8}$$

and

$$F_{3|2}(tx_3 \mid tx_2) = \begin{cases} \frac{2c_1^{\{23\}}}{(s_1^{\{23\}}+1)} \exp\left\{t(x_3 - x_2)(s_1^{\{23\}} + 1)\right\}\{1 + o(1)\}, & x_3 < x_2, \\ -V_1^{\{23\}}(1,1)\{1 + o(1)\}, & x_3 = x_2, \\ 1 - \frac{2c_2^{\{23\}}}{(s_2^{\{23\}}+2)} \exp\left\{-t(x_3 - x_2)(s_2^{\{23\}} + 2)\right\}\{1 + o(1)\}, & x_3 > x_2. \end{cases} \tag{5.4.9}$$

So, as $t \to \infty$, the conditional distributions of the extreme value copula in equation (5.4.8) tends towards either 0, $-V_2^{\{12\}}(1,1)$ or 1, and expression (5.4.9) tends towards 0, $-V_1^{\{23\}}(1,1)$ or 1. We must therefore consider three different cases in our gauge function calculations for any extreme value copula in tree T_1. Based on the asymptotic forms of $F_{1|2}(tx_1 \mid tx_2)$ and $F_{3|2}(tx_3 \mid tx_2)$ for (inverted) extreme value copulas, we focus on investigating $-\log c_{13|2}(u, v)$ for u and v of the form

$$a\{1 + o(1)\} \quad ; \quad b_1 \exp(-b_2 t)\{1 + o(1)\} \quad ; \quad 1 - c_1 \exp(-c_2 t)\{1 + o(1)\},$$

for $b_2, c_2 > 0$. We provide these results in Appendix C.4 for all nine combinations of the asymptotic forms of u and v, and for $c_{13|2}(u, v)$ being either an extreme value or inverted extreme value copula density.

Together, these results provide all the necessary information to calculate the gauge functions. We first demonstrate how to combine all these results to obtain the gauge function of a vine copula with two inverted extreme value copulas in tree T_1, and an extreme value copula in tree T_2, and subsequently find the gauge functions for the remaining cases.

Inverted extreme value copulas in T_1; extreme value copula in T_2

For this case, we use results (5.4.1), (5.4.2) and (5.4.3) in equation (5.3.3), which gives the form of $-\log f(t\boldsymbol{x})$ for a trivariate vine copula. We have

$$\begin{aligned}
-\log f(t\boldsymbol{x}) &= t(x_1+x_2+x_3) + t\left\{V^{\{12\}}\left(x_1^{-1},x_2^{-1}\right) - x_1 - x_2\right\} \\
&\quad + t\left\{V^{\{23\}}\left(x_2^{-1},x_3^{-1}\right) - x_2 - x_3\right\} - \log c_{13|2}\left\{F_{1|2}(tx_1|tx_2), F_{3|2}(tx_3|tx_2)\right\} + O(\log t) \\
&= t\left\{V^{\{12\}}\left(x_1^{-1},x_2^{-1}\right) + V^{\{23\}}\left(x_2^{-1},x_3^{-1}\right) - x_2\right\} - \log c_{13|2}\left\{F_{1|2}(tx_1|tx_2), F_{3|2}(tx_3|tx_2)\right\} \\
&\quad + O(\log t). \qquad (5.4.10)
\end{aligned}$$

From equations (5.4.6) and (5.4.7), we see that $F_{1|2}(tx_1 \mid tx_2) = 1 - a_{1|2}\exp\{-b_{1|2}t\}$ and $F_{3|2}(tx_3 \mid tx_2) = 1 - a_{3|2}\exp\{-b_{3|2}t\}$, for $b_{1|2} = V^{\{12\}}(x_1^{-1}, x_2^{-1}) - x_2$ and $b_{3|2} = V^{\{23\}}(x_2^{-1}, x_3^{-1}) - x_2$. Using results from case 9 of Appendix C.4, we deduce that

$$\begin{aligned}
&-\log c_{13|2}\left\{F_{1|2}(tx_1|tx_2), F_{3|2}(tx_3|tx_2)\right\} \sim \\
&\quad t\Big\{\left(1 + s_1^{\{13|2\}}\mathbb{1}_{\{b_{1|2}\geq b_{3|2}\}} + s_2^{\{13|2\}}\mathbb{1}_{\{b_{1|2}<b_{3|2}\}}\right)\max(b_{1|2}, b_{3|2}) \\
&\qquad - \left(2 + s_1^{\{13|2\}}\mathbb{1}_{\{b_{1|2}\geq b_{3|2}\}} + s_2^{\{13|2\}}\mathbb{1}_{\{b_{1|2}<b_{3|2}\}}\right)\min(b_{1|2}, b_{3|2})\Big\}.
\end{aligned}$$

Combining this with result (5.4.10), we find that the required gauge function has the form

$$\begin{aligned}
g(\boldsymbol{x}) &= x_2 + b_{1|2} + b_{3|2} + \left(1 + s_{\mathrm{m}}^{\{13|2\}}\right)\max(b_{1|2}, b_{3|2}) - \left(2 + s_{\mathrm{m}}^{\{13|2\}}\right)\min(b_{1|2}, b_{3|2}) \\
&= x_2 + \left(2 + s_{\mathrm{m}}^{\{13|2\}}\right)\max(b_{1|2}, b_{3|2}) - \left(1 + s_{\mathrm{m}}^{\{13|2\}}\right)\min(b_{1|2}, b_{3|2}) \\
&= \left(2 + s_{\mathrm{m}}^{\{13|2\}}\right)\max(b_{1|2} - x_2, b_{3|2} - x_2) - \left(1 + s_{\mathrm{m}}^{\{13|2\}}\right)\min(b_{1|2} - x_2, b_{3|2} - x_2),
\end{aligned}$$

i.e.,

$$g(\boldsymbol{x}) = \left(2 + s_{\mathrm{m}}^{\{13|2\}}\right) \max\left\{V^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right), V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right)\right\} - \left(1 + s_{\mathrm{m}}^{\{13|2\}}\right) \min\left\{V^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right), V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right)\right\}, \tag{5.4.11}$$

with

$$s_{\mathrm{m}}^{\{13|2\}} = s_1^{\{13|2\}} \mathbb{1}_{\left\{V^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right) \geq V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right)\right\}} + s_2^{\{13|2\}} \mathbb{1}_{\left\{V^{\{12\}}\left(x_1^{-1}, x_2^{-1}\right) < V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right)\right\}}.$$

In Section 5.4.3, we study this gauge function in more detail for the (inverted) logistic case, and use it to find the corresponding values of $\eta_{\{1,2,3\}}$ and $\eta_{\{1,3\}}$.

Extreme value and inverted extreme value copulas in T_1; inverted extreme value copula in T_2

To calculate the gauge function for this model, we use results (5.4.1), (5.4.3) and (5.4.4) to give the asymptotic form of the first five terms of equation (5.3.3). For the final term, equations (5.4.7) and (5.4.8) give the required form of the conditional distributions, and we apply the inverted extreme value results from cases 3, 6 and 9 of Appendix C.4 to yield the gauge function

$$g(\boldsymbol{x}) = \begin{cases} \left(2 + s_1^{\{13|2\}}\right)\left(1 + s_2^{\{12\}}\right)(x_2 - x_1) + V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right), & x_1 \leq x_2, \\ x_2 + V^{\{13|2\}}\left[\left\{(x_1 - x_2)\left(2 + s_1^{\{12\}}\right)\right\}^{-1}, \left\{V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_2\right\}^{-1}\right], & x_1 > x_2. \end{cases} \tag{5.4.12}$$

Extreme value and inverted extreme value copulas in T_1; extreme value copula in T_2

Since the pair copulas in tree T_1 of this model are the same as in the previous example, the first five terms of equation (5.3.3) will also be the same. To study the final term, we apply the extreme value results from cases 3, 6 and 9 of Appendix C.4, and find

that the gauge function is

$$g(\boldsymbol{x}) = \begin{cases} x_2 + \left(1 + s_2^{\{12\}}\right)(x_2 - x_1) + \left(2 + s_2^{\{13|2\}}\right)\left\{V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_2\right\}, & x_1 \leq x_2, \\ x_2 + \left(2 + s_{\mathrm{m}}^{\{13|2\}}\right)\max\left\{\left(2 + s_1^{\{12\}}\right)(x_1 - x_2), V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_2\right\} \\ \quad - \left(1 + s_{\mathrm{m}}^{\{13|2\}}\right)\min\left\{\left(2 + s_1^{\{12\}}\right)(x_1 - x_2), V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_2\right\}, & x_1 > x_2, \end{cases} \tag{5.4.13}$$

with

$$\begin{aligned} s_{\mathrm{m}}^{\{13|2\}} = {} & s_1^{\{13|2\}} \mathbb{1}_{\left\{\left(2+s_1^{\{12\}}\right)(x_1-x_2) \geq V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_2\right\}} \\ & + s_2^{\{13|2\}} \mathbb{1}_{\left\{\left(2+s_1^{\{12\}}\right)(x_1-x_2) < V^{\{23\}}\left(x_2^{-1}, x_3^{-1}\right) - x_2\right\}}. \end{aligned}$$

Extreme value copulas in T_1; inverted extreme value copula in T_2

For this model, the first five terms of equation (5.3.3) are given by results (5.4.1), (5.4.4) and (5.4.5). Since both of the extreme value copulas in tree T_1 can have three different asymptotic behaviours, we require all nine inverted extreme value cases from Appendix C.4, and the conditional distributions (5.4.8) and (5.4.9) to study the final term of (5.3.3). The gauge function in this case is

$$g(\boldsymbol{x}) = \begin{cases} x_2 + \left(2 + s_{\mathrm{m}}^{\{13|2\}}\right)\max\left\{\left(1 + s_2^{\{12\}}\right)(x_2 - x_1), \left(1 + s_1^{\{23\}}\right)(x_2 - x_3)\right\} \\ \quad - \left(1 + s_{\mathrm{m}}^{\{13|2\}}\right)\min\left\{\left(1 + s_2^{\{12\}}\right)(x_2 - x_1), \left(1 + s_1^{\{23\}}\right)(x_2 - x_3)\right\}, \\ \qquad\qquad x_1 < x_2, x_3 < x_2, \\ x_2 + \left(2 + s_1^{\{13|2\}}\right)\left(1 + s_2^{\{12\}}\right)(x_2 - x_1) + \left(2 + s_2^{\{23\}}\right)(x_3 - x_2), \\ \qquad\qquad x_1 < x_2, x_3 \geq x_2, \\ x_2 + \left(2 + s_2^{\{13|2\}}\right)\left(1 + s_1^{\{23\}}\right)(x_2 - x_3) + \left(2 + s_1^{\{12\}}\right)(x_1 - x_2), \\ \qquad\qquad x_1 \geq x_2, x_3 < x_2, \\ x_2 + V^{\{13|2\}}\left[\left\{\left(2 + s_1^{\{12\}}\right)(x_1 - x_2)\right\}^{-1}, \left\{\left(2 + s_2^{\{23\}}\right)(x_3 - x_2)\right\}^{-1}\right], \\ \qquad\qquad x_1 \geq x_2, x_3 \geq x_2, \end{cases} \tag{5.4.14}$$

with

$$s_{\mathrm{m}}^{\{13|2\}} = s_1^{\{13|2\}} \mathbb{1}_{\left\{\left(1+s_2^{\{12\}}\right)(x_2-x_1) \geq \left(1+s_1^{\{23\}}\right)(x_2-x_3)\right\}} + s_2^{\{13|2\}} \mathbb{1}_{\left\{\left(1+s_2^{\{12\}}\right)(x_2-x_1) < \left(1+s_1^{\{23\}}\right)(x_2-x_3)\right\}}.$$

Extreme value copulas in T_1; extreme value copula in T_2

Finally, for a vine copula where all three components are extreme value copulas, the first five terms of (5.3.3) are the same as in the previous section, and we require all nine extreme value cases from Appendix C.4 to obtain the gauge function

$$g(\boldsymbol{x}) = \begin{cases} x_2 + V^{\{13|2\}}\left[\left\{\left(1+s_2^{\{12\}}\right)(x_2-x_1)\right\}^{-1}, \left\{\left(1+s_1^{\{23\}}\right)(x_2-x_3)\right\}^{-1}\right], \\ \qquad\qquad x_1 \leq x_2, x_3 \leq x_2, \\ x_2 + \left(2+s_2^{\{13|2\}}\right)\left(2+s_2^{\{23\}}\right)(x_3-x_2) + \left(1+s_2^{\{12\}}\right)(x_2-x_1), \\ \qquad\qquad x_1 \leq x_2, x_3 > x_2, \\ x_2 + \left(2+s_1^{\{13|2\}}\right)\left(2+s_1^{\{12\}}\right)(x_1-x_2) + \left(1+s_1^{\{23\}}\right)(x_2-x_3), \\ \qquad\qquad x_1 > x_2, x_3 \leq x_2, \\ x_2 + \left(2+s_{\mathrm{m}}^{\{13|2\}}\right)\max\left\{\left(2+s_1^{\{12\}}\right)(x_1-x_2), \left(2+s_2^{\{23\}}\right)(x_3-x_2)\right\} \\ \quad - \left(1+s_{\mathrm{m}}^{\{13|2\}}\right)\min\left\{\left(2+s_1^{\{12\}}\right)(x_1-x_2), \left(2+s_2^{\{23\}}\right)(x_3-x_2)\right\}, \\ \qquad\qquad x_1 > x_2, x_3 > x_2, \end{cases} \tag{5.4.15}$$

with

$$s_{\mathrm{m}}^{\{13|2\}} = s_1^{\{13|2\}} \mathbb{1}_{\left\{\left(2+s_1^{\{12\}}\right)(x_1-x_2) \geq \left(2+s_2^{\{23\}}\right)(x_3-x_2)\right\}} + s_2^{\{13|2\}} \mathbb{1}_{\left\{\left(2+s_1^{\{12\}}\right)(x_1-x_2) < \left(2+s_2^{\{23\}}\right)(x_3-x_2)\right\}}.$$

We study (inverted) logistic examples of this gauge function and the others discussed in this section in Section 5.4.3.

5.4.3 $\eta_{\{1,2,3\}}$ and $\eta_{\{1,3\}}$ for vines with logistic and inverted logistic components

In Section 5.4.2, we obtained the gauge functions of trivariate vine copulas constructed from extreme value and inverted extreme value pair copula components. We now focus on the logistic case, and aim to calculate the values of $\eta_{\{1,2,3\}}$ and $\eta_{\{1,3\}}$ for each possible construction. The logistic model has exponent measure of the form

$$V(x,y) = \left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha},$$

and we focus on the dependence parameter α taking values in $(0,1)$. For the logistic model, the corresponding spectral density has $h(w) \sim c_1(1-w)^{s_1}$ as $w \to 1$, and $h(w) \sim c_2 w^{s_2}$ as $w \to 0$, with $s_1 = s_2 = 1/\alpha - 2$. In the remainder of this section, we denote the parameters associated with copulas c_{12}, c_{23} and $c_{13|2}$ by $\alpha, \beta, \gamma \in (0,1)$, respectively, whether the copula is logistic or inverted logistic, and note that in the logistic case, we have $s_1^{\{12\}} = s_2^{\{12\}} = 1/\alpha - 2$; $s_1^{\{23\}} = s_2^{\{23\}} = 1/\beta - 2$ and $s_1^{\{13|2\}} = s_2^{\{13|2\}} = 1/\gamma - 2$. We now take each of the constructions discussed in Section 5.4.2 in turn, and consider the gauge function, as well as the values of $\eta_{\{1,2,3\}}$ and $\eta_{\{1,3\}}$ in each case.

Inverted logistic copulas in T_1; logistic copula in T_2

For this vine copula, using the form of the gauge function obtained in Section 5.4.2, we have

$$g(\boldsymbol{x}) = (1/\gamma)\max\left\{\left(x_1^{1/\alpha} + x_2^{1/\alpha}\right)^{\alpha}, \left(x_2^{1/\beta} + x_3^{1/\beta}\right)^{\beta}\right\} + (1-1/\gamma)\min\left\{\left(x_1^{1/\alpha} + x_2^{1/\alpha}\right)^{\alpha}, \left(x_2^{1/\beta} + x_3^{1/\beta}\right)^{\beta}\right\}.$$

In Figure 5.4.1, we demonstrate the sets $G = \{\boldsymbol{x} \in \mathbb{R}^3 : g(\boldsymbol{x}) = 1\}$ for this gauge function, with $\alpha \in \{0.25, 0.5, 0.75\}$, $\beta = 0.25$ and $\gamma = 0.5$. These plots, alongside further numerical studies, suggest that the intersection of G and $[\eta_{\{1,2,3\}}, \infty)^3$ occurs when $x_1 = x_2 \leq x_3$ if $\alpha \geq \beta$, and $x_1 > x_2 = x_3$ if $\alpha < \beta$. We first consider

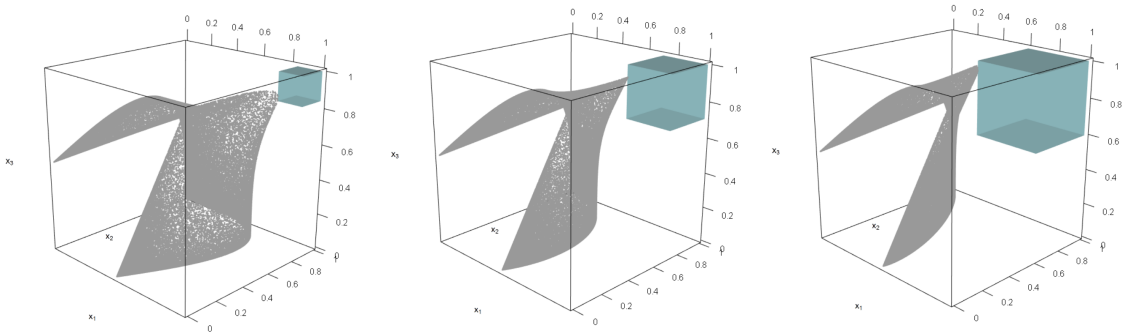

Figure 5.4.1: Points in the set $G = \{\boldsymbol{x} \in \mathbb{R}^3 : g(\boldsymbol{x}) = 1\}$ for a trivariate vine with inverted logistic copulas in T_1 and a logistic copula in T_2 (grey) and the set $[\eta_{\{1,2,3\}}, \infty)^3$ (blue): $\alpha = 0.25$ (left), $\alpha = 0.5$ (centre), $\alpha = 0.75$ (right); $\beta = 0.25$ and $\gamma = 0.5$.

the former case, with $x_1 = x_2 \leq x_3$, where we have two possibilities to study. If $\left(x_1^{1/\alpha} + x_2^{1/\alpha}\right)^\alpha \geq \left(x_2^{1/\beta} + x_3^{1/\beta}\right)^\beta$, i.e., $2^\alpha x_1 \geq \left(x_1^{1/\beta} + x_3^{1/\beta}\right)^\beta$, then the value of x_3 must satisfy

$$x_1 \leq x_3 \leq \left(2^{\alpha/\beta} - 1\right)^\beta x_1,$$

or, equivalently, we can set $x_3 = rx_1$, with $1 \leq r \leq \left(2^{\alpha/\beta} - 1\right)^\beta$. Our aim is to find the maximum value of x_1 such that $g(x_1, x_1, rx_1) = 1$ under this constraint. We have

$$\begin{aligned} g(x_1, x_1, rx_1) = 1 &\Rightarrow \frac{2^\alpha x_1}{\gamma} + \left(1 - \frac{1}{\gamma}\right)\left\{x_1^{1/\beta} + (rx_1)^{1/\beta}\right\}^\beta = 1 \\ &\Rightarrow x_1 = \frac{\gamma}{2^\alpha + (\gamma - 1)\left(1 + r^{1/\beta}\right)^\beta}. \end{aligned}$$

We note that x_1 is increasing with r, and therefore takes its maximum value when $r = \left(2^{\alpha/\beta} - 1\right)^\beta$, which corresponds to $x_1 = 1/2^\alpha$. This is our first candidate for $\eta_{\{1,2,3\}}$. The other possibility is that $\left(x_1^{1/\alpha} + x_2^{1/\alpha}\right)^\alpha < \left(x_2^{1/\beta} + x_3^{1/\beta}\right)^\beta$, i.e., $2^\alpha x_1 < \left(x_1^{1/\beta} + x_3^{1/\beta}\right)^\beta$, so that the value of x_3 must satisfy

$$\left(2^{\alpha/\beta} - 1\right)^\beta x_1 < x_3 \leq 1.$$

If we set $x_3 = rx_1$, this is equivalent to having $\left(2^{\alpha/\beta} - 1\right)^\beta < r \leq 1/x_1$. We have

$$g(x_1, x_1, rx_1) = 1 \Rightarrow \frac{1}{\gamma}\left\{x_1^{1/\beta} + (rx_1)^{1/\beta}\right\}^\beta + 2^\alpha\left(1 - \frac{1}{\gamma}\right)x_1 = 1$$

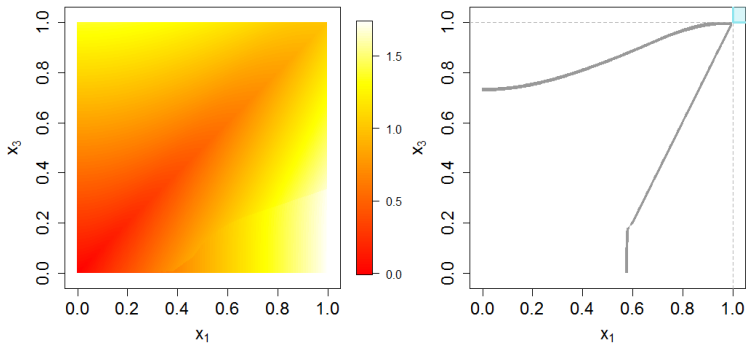

Figure 5.4.2: Left: $g_{\{1,3\}}(x_1, x_3)$ with $(\alpha, \beta, \gamma) = (0.5, 0.25, 0.5)$. Right: an approximation of the corresponding sets $G_{\{1,3\}}$ (grey) and $[\eta_{\{1,3\}}, \infty)^2$ (blue).

$$\Rightarrow x_1 = \frac{\gamma}{(1 + r^{1/\beta})^{\beta} + 2^{\alpha}(\gamma - 1)}.$$

The value of x_1 is now decreasing with r, so again takes its maximum value with $r = \left(2^{\alpha/\beta} - 1\right)^{\beta}$, giving $x_1 = 1/2^{\alpha}$. As such, we find that for this model, if $\alpha \geq \beta$ then $\eta_{\{1,2,3\}} = 1/2^{\alpha}$. By a symmetric argument, if $\alpha < \beta$, $\eta_{\{1,2,3\}} = 1/2^{\beta}$. So the value of this coefficient of tail dependence can be summarized as $\eta_{\{1,2,3\}} = \min\left\{1/2^{\alpha}, 1/2^{\beta}\right\}$.

To find the value of $\eta_{\{1,3\}}$, as in Section 5.3.2, numerical results suggest that we consider the $x_1 = x_3$ case, i.e., $g_{\{1,3\}}(x_1, x_1) = \min_v g(x_1, vx_1, x_1)$. We find that

$$\eta_{\{1,3\}} = \left[\frac{1}{\gamma} \max\left\{\left(1 + v^{1/\alpha}\right)^{\alpha}, \left(1 + v^{1/\beta}\right)^{\beta}\right\} + \left(1 - \frac{1}{\gamma}\right) \min\left\{\left(1 + v^{1/\alpha}\right)^{\alpha}, \left(1 + v^{1/\beta}\right)^{\beta}\right\}\right]^{-1},$$

with v satisfying $dg(1, v, 1)/dv = 0$. We find that $v = 0$ solves this equation, so that $\eta_{\{1,3\}} = 1$, i.e., there X_1 and X_3 can take their largest values simultaneously without X_2. This is supported by Figure 5.4.2; in the left panel we show the value of $g_{\{1,3\}}(x_1, x_3)$ for $(x_1, x_3) \in [0, 1]^2$ and $(\alpha, \beta, \gamma) = (0.5, 0.25, 0.5)$, and in the right panel we approximate the set $G_{\{1,3\}}$ by considering values of $g_{\{1,3\}}(x_1, x_3) \in (0.99, 1.01)$.

Logistic and inverted logistic copulas in T_1; inverted logistic copula in T_2

Using the form of the gauge function in equation (5.4.12) we have

$$g(\boldsymbol{x}) = \begin{cases} (1/\gamma)\,(1/\alpha - 1)\,(x_2 - x_1) + \left(x_2^{1/\beta} + x_3^{1/\beta}\right)^{\beta}, & x_1 \leq x_2, \\ x_2 + \left[\left\{(x_1 - x_2)/\alpha\right\}^{1/\gamma} + \left\{\left(x_2^{1/\beta} + x_3^{1/\beta}\right)^{\beta} - x_2\right\}^{1/\gamma}\right]^{\gamma}, & x_1 > x_2. \end{cases}$$

For this gauge function, we find that the intersection of the set $G = \{\boldsymbol{x} \in \mathbb{R}^3 : g(\boldsymbol{x}) = 1\}$ and $[\eta_{\{1,2,3\}}, \infty)^3$ occurs when $x_1 \geq x_2 = x_3$, which reduces the part of the gauge function that we need to investigate. By following a similar approach to the one used in the previous example, we find that $\eta_{\{1,2,3\}} = 1/2^{\beta}$ in this case. Moreover, following the approach of Section 5.3.2 for calculating the bivariate coefficient of tail dependence between X_1 and X_3, we find that

$$\eta_{\{1,3\}} = \left(v + \left[\left\{(1-v)/\alpha\right\}^{1/\gamma} + \left\{\left(1 + v^{1/\beta}\right)^{\beta} - v\right\}^{1/\gamma}\right]^{\gamma}\right)^{-1}$$

with v such that

$$1 + \left[\left\{(1-v)/\alpha\right\}^{1/\gamma} + \left\{\left(1 + v^{1/\beta}\right)^{\beta} - v\right\}^{1/\gamma}\right]^{\gamma-1} \cdot \left[-\frac{1}{\alpha}\left\{(1-v)/\alpha\right\}^{-1+1/\gamma} + \left\{\left(1 + v^{1/\beta}\right)^{\beta} - v\right\}^{-1+1/\gamma}\left\{\left(1 + v^{-1/\beta}\right)^{\beta-1} - 1\right\}\right] = 0.$$

Logistic and inverted logistic copulas in T_1; logistic copula in T_2

From equation (5.4.13), we have

$$g(\boldsymbol{x}) = \begin{cases} (1/\alpha)\,x_2 + (1 - 1/\alpha)\,x_1 + (1/\gamma)\left\{\left(x_2^{1/\beta} + x_3^{1/\beta}\right)^{\beta} - x_2\right\}, & x_1 \leq x_2, \\ x_2 + (1/\gamma)\max\left\{(x_1 - x_2)/\alpha, \left(x_2^{1/\beta} + x_3^{1/\beta}\right)^{\beta} - x_2\right\} & \\ \qquad + (1 - 1/\gamma)\min\left\{(x_1 - x_2)/\alpha, \left(x_2^{1/\beta} + x_3^{1/\beta}\right)^{\beta} - x_2\right\}, & x_1 > x_2. \end{cases}$$

As in the previous example, we find that the intersection of G and $[\eta_{\{1,2,3\}}, \infty)^3$ occurs when $x_1 \geq x_2 = x_3$, and once again $\eta_{\{1,2,3\}} = 1/2^{\beta}$, suggesting that the inverted logistic copula in tree T_1 particularly controls the level of asymptotic independence in the overall model. We also find that

$$\eta_{\{1,3\}} = \left(1 + v^{1/\beta}\right)^{-\beta}, \quad \text{with } v \text{ such that } (1 + v^{1/\beta})^{\beta} - (1-v)/\alpha - v = 0.$$

Logistic copulas in T_1; inverted logistic copula in T_2

Using the result from equation (5.4.14), we have

$$g(\boldsymbol{x}) = \begin{cases} x_2 + (1/\gamma)\max\left\{(1/\alpha - 1)(x_2 - x_1), (1/\beta - 1)(x_2 - x_3)\right\} \\ \qquad + (1 - 1/\gamma)\min\left\{(1/\alpha - 1)(x_2 - x_1), (1/\beta - 1)(x_2 - x_3)\right\}, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad x_1 < x_2, x_3 < x_2, \\ x_2 + (1/\gamma)(1/\alpha - 1)(x_2 - x_1) + (1/\beta)(x_3 - x_2), \qquad x_1 < x_2, x_3 \geq x_2, \\ x_2 + (1/\gamma)(1/\beta - 1)(x_2 - x_3) + (1/\alpha)(x_1 - x_2), \qquad x_1 \geq x_2, x_3 < x_2, \\ x_2 + \left[\left\{(x_1 - x_2)/\alpha\right\}^{1/\gamma} + \left\{(x_3 - x_2)/\beta\right\}^{1/\gamma}\right]^{\gamma}, \qquad x_1 \geq x_2, x_3 \geq x_2. \end{cases}$$

For this gauge function, we observe that $g(1,1,1) = 1$. Since the set $G = \{\boldsymbol{x} \in \mathbb{R}^3 : g(\boldsymbol{x}) = 1\}$ is bounded by the set $[0,1]^3$, the smallest value of r such that $G \cap [r,\infty)^3 = \emptyset$ must be 1. As such, we have $\eta_{\{1,2,3\}} = 1$. Moreover, if $\eta_D = 1$, then $\eta_C = 1$ for any set $C \subset D$ with $|C| \geq 2$. As such, we also find that $\eta_{\{1,3\}} = 1$ in this case. This result agrees with the findings of Joe et al. (2010), who show that a vine copula will have overall upper tail dependence if each of the copulas in tree T_1 also have this property and the copula in tree T_2 has support on $(0,1)$, as is the case here.

Logistic copulas in T_1; logistic copula in T_2

Finally, using the gauge function obtained in equation (5.4.15), we have

$$g(\boldsymbol{x}) = \begin{cases} x_2 + \left[\left\{(1/\alpha - 1)(x_2 - x_1)\right\}^{1/\gamma} + \left\{(1/\beta - 1)(x_2 - x_3)\right\}^{1/\gamma}\right]^{\gamma}, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad x_1 \leq x_2, x_3 \leq x_2, \\ x_2 + (1/\gamma)(1/\beta)(x_3 - x_2) + (1/\alpha - 1)(x_2 - x_1), \qquad x_1 \leq x_2, x_3 > x_2, \\ x_2 + (1/\gamma)(1/\alpha)(x_1 - x_2) + (1/\beta - 1)(x_2 - x_3), \qquad x_1 > x_2, x_3 \leq x_2, \\ x_2 + (1/\gamma)\max\left\{(x_1 - x_2)/\alpha, (x_3 - x_2)/\beta\right\} \\ \qquad + (1 - 1/\gamma)\min\left\{(x_1 - x_2)/\alpha, (x_3 - x_2)/\beta\right\}, \qquad x_1 > x_2, x_3 > x_2. \end{cases}$$

As for the previous example, in this case, we note that $g(1,1,1) = 1$, which suggests that $\eta_{\{1,2,3\}} = \eta_{\{1,3\}} = 1$.

5.5 Discussion

The aim of this chapter was to investigate some of the tail dependence properties of vine copulas, via the coefficient of tail dependence η_C. We demonstrated how to apply the geometric approach of Nolde (2014) to calculate these values from a density, and extended the approach for cases where the joint density of $\{X_i : i \in C\}$ cannot be obtained analytically, but the joint density of $\{X_i : i \in C'\}$ with $C' \supset C$ is known.

We focussed on trivariate vine copulas constructed from extreme value and inverted extreme value pair copulas, and higher dimensional D-vine and C-vine copulas constructed only from inverted extreme value pair copulas. In the latter case, there is overall asymptotic independence between the variables. In the former case, the copulas in tree T_1 particularly influence the overall tail dependence properties of the vine. If there are two asymptotically dependent extreme value copulas in tree T_1, there is overall asymptotic dependence in the vine, as found by Joe et al. (2010), otherwise, all three variables cannot be large together, although other subsets of the variables could take their largest values simultaneously while the others are of smaller order.

As discussed in Chapters 3 and 4, the values of η_C, for $C \in \{1,2,3\}$ and $|C| \geq 2$, do not necessarily tell us the full story about the extremal dependence structure of the variables. One possible avenue for future work is to calculate the coefficients τ_C for $C \in \{1,2,3\}$, introduced in Chapter 3, which provide additional information about the tail dependence properties of the variables. As an initial investigation into the types of extremal dependence structure that might be possible using vine copulas, we consider some trivariate vine copula examples in the radial-angular setting of Chap-

ter 4. For each possible construction of a trivariate vine copula from logistic and inverted logistic components, with dependence parameters $\alpha = 0.25$ and $\alpha = 0.75$, respectively, we simulate a sample of size 1,000,000 and plot the angular components whose corresponding radial values are above the observed 0.9999 quantile; the results are shown in Figure 5.5.1.

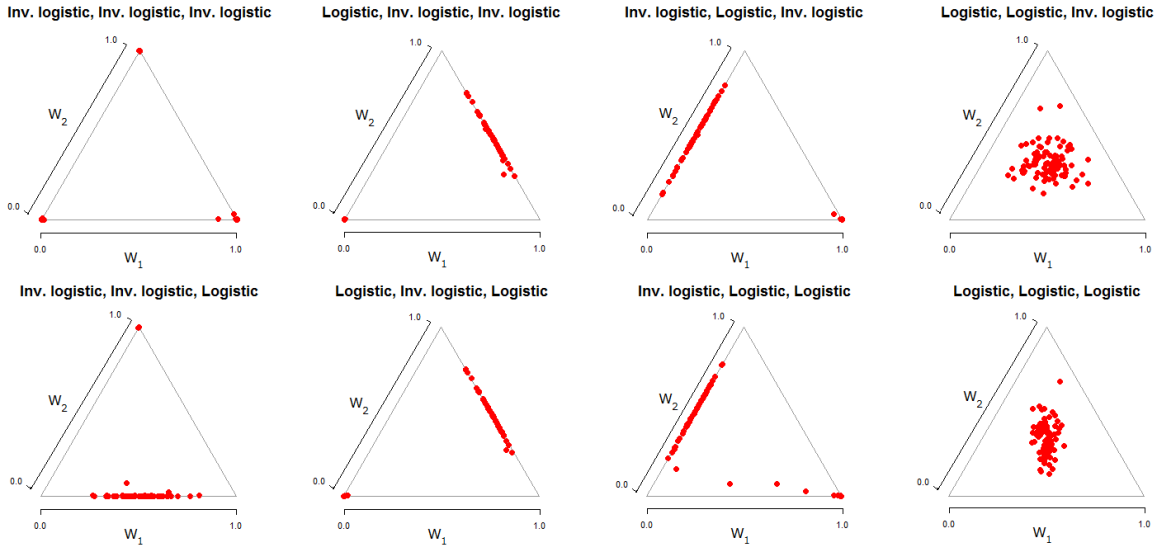

Figure 5.5.1: Possible extremal dependence structures for trivariate vine copulas constructed from logistic and inverted logistic components.

From these plots, it appears that the possible extremal dependence structures are $\{\{X_1\},\{X_2\},\{X_3\}\}$, $\{\{X_1\},\{X_2,X_3\}\}$, $\{\{X_2\},\{X_1,X_3\}\}$, $\{\{X_3\},\{X_1,X_2\}\}$ and $\{\{X_1,X_2,X_3\}\}$, so that each variable only has extreme values on one face of the angular simplex. We also expect this to be the case for other vine copulas constructed from bivariate copulas whose extremal dependence structures are $\{\{X_1\},\{X_2\}\}$, as for the asymptotically independent inverted logistic copula, or $\{\{X_1,X_2\}\}$, as for the logistic copula. To achieve more complicated extremal dependence structures in the vine copula, the pair copulas must exhibit other extremal dependence structures. In the example from Figure 5.5.1, this could be achieved, for instance, by replacing the

bivariate logistic copulas with the asymmetric logistic copula of Tawn (1990), which can exhibit asymptotic dependence but with the spectral density placing mass on $\{0\}$, $\{1\}$, or both of these endpoints; having asymptotically dependent extreme value pair copulas with such mass on the boundaries is not currently covered by our work. Although this may increase the number of possible extremal dependence structures, the underlying vine structure may still mean that some cases are not possible, and investigating this also presents a possible avenue for further work. The inverted asymmetric logistic copula, as studied by Papastathopoulos and Tawn (2016), exhibits asymptotic independence, so replacing the inverted logistic components of Figure 5.5.1 by this copula will result in the same extremal dependence structures.

Chapter 6

A Bayesian Spatio-Temporal Model for Precipitation Extremes

6.1 Introduction

Recently, there have been numerous examples of devastating rainfall events - these include Storm Desmond, which hit northern England and Scotland, and Hurricane Harvey which affected the southern United States. In both cases, a large amount of damage and disruption was caused by severe flooding. By better understanding the probability of extreme rainfall events occurring, we can prepare more suitably for these potential flood events by adapting infrastructure appropriately.

This chapter outlines the STOR-i team's approach to the EVA2017 challenge. The challenge data set is comprised of precipitation readings for multiple weather stations in the Netherlands; the training set consists of data collected between 1972 and 1995 whilst the validation set was collected from 1996 to 2016, with different numbers of observations for each site. A detailed description of the data is provided in Wintenberger (2018). The aim of the competition was to predict extreme quantiles for the years 1996 to 2016 and predictions were assessed via a predefined error metric; see

Wintenberger (2018).

There exists a rich literature within the extreme value theory framework for modelling precipitation extremes. A classical approach is to utilize block maxima. Suppose that we have independent and identically distributed (i.i.d.) random variables $X_1, \ldots, X_n$, with $M_n = \max\{X_1, \ldots, X_n\}$. When normalized appropriately, and as $n \to \infty$, M_n follows a generalized extreme value (GEV) distribution (von Mises, 1936; Jenkinson, 1955), which has distribution function

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-\frac{1}{\xi}}\right\}, \tag{6.1.1}$$

where $\{z\}_+ = \max\{0, z\}$, and has parameters $(\mu, \sigma, \xi) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$, corresponding to location, scale and shape parameters respectively.

An alternative technique is to follow Pickands (1975) and use exceedances of a threshold u. For some suitably large u, the conditional distribution function of $(X_i - u) \mid X_i > u$ is approximately given by the generalized Pareto distribution (GPD), which has the form

$$H(x) = 1 - \left(1 + \frac{\xi x}{\psi}\right)_+^{-\frac{1}{\xi}}, \qquad x > 0, \tag{6.1.2}$$

where $(\psi, \xi) \in \mathbb{R}_+ \times \mathbb{R}$ are the scale and shape parameters respectively. In the context of the challenge at hand, both the GEV and GPD may be fitted separately at each site to give a model fit whereby any dependence is ignored.

By considering the physical process of rainfall, one can expect that nearby locations will exhibit similar behaviour, which invites improved inference by sharing information across sites. One popular method for the modelling of spatial extremes is to use max-stable processes (Brown and Resnick, 1977; Smith, 1990; Schlather, 2002). These arise as the limiting process from replications of spatial processes which have been suitably normalized (de Haan, 1984) and have been used to analyse rainfall data

previously; see, for example, Davison et al. (2012) and Reich and Shaby (2012). However, such processes assume dependence of the extremes across sites; an investigation of pairwise dependence using scatter plots showed no clear evidence for this behaviour across the spatial grid. Moreover, max-stable models are difficult to fit and this would have been further impeded by the lack of data available at some sites.

Another approach is to impose spatial structure on the model parameters via a Bayesian hierarchical model; this is closer in nature to the method we propose. Spatial hierarchical models have been used previously to model spatial count data (Diggle et al., 1998) and, more recently, have been utilized in extreme value analysis. Cooley et al. (2007) describe a model, applied to rainfall data, whereby a GPD is fitted at the sampling locations, and allow the model parameters to vary according to a spatial process structure - in particular the authors use a Gaussian process for this. A spatio-temporal hierarchical modelling method for extreme events is given by Sang and Gelfand (2009), who apply their methods to precipitation data.

We define a Bayesian hierarchical model which accounts for the spatial and seasonal variation in the data. Our approach captures the frequency of non-zero events of precipitation and introduces an extremal mixture model, combining Gamma and generalized Pareto distributions, for positive amounts of rainfall. Spatio-temporal structure in the parameters for the extremal mixture model is imposed via a separate autoregressive prior for each of them, which takes the form of a Gaussian Markov random field. Model estimates are then obtained using spatial interpolation and Markov chain Monte Carlo (MCMC) techniques. Cooley et al. (2007) defines a similar approach for continuous space, whereas we consider a finite number of sites and additionally incorporate seasonality.

6.2 Methodology

6.2.1 Likelihood

Interest lies in modelling the daily rainfall amounts for each site and month. Due to seasonality in the rainfall data, the weak extremal dependence of the daily amount of rainfall across sites and the nature of the challenge, we model each month and site individually. Specifically, daily rainfall events within a month at a site are assumed to be i.i.d. Our model is motivated by an analysis of the sites for which data have been recorded for at least five years.

Let $R_{j,m}$ denote the random variable corresponding to the daily rainfall amount at site j for a day in month $m = 1, \ldots, 12$. We consider the transformed random variable

$$\tilde{R}_{j,m} = \log\left(1 + R_{j,m}\right). \tag{6.2.1}$$

Wadsworth et al. (2010) show that such a transformation may increase the rate of convergence of the distribution tails to an extreme value form, in particular for distributions which appear as heavy-tailed as our rainfall data. Predictions on the extreme quantiles of $R_{j,m}$ are later obtained in Section 6.3 by reversing this transformation. We note that the transformed observations are non-negative and an observation of $R_{j,m} = 0$ remains unchanged.

We infer on the distribution of $\tilde{R}_{j,m}$ by defining a hierarchical model. The first model component considers occurrences of non-zero amounts of rainfall on a day, $\tilde{R}_{j,m} > 0$, and we denote their probability by $p_{j,m}$. A temporal trend in $p_{j,m}$ was investigated, but we did not find evidence of this for any site. Next, we consider the distribution $\tilde{R}_{j,m} \mid (\tilde{R}_{j,m} > 0)$. There exists a rich literature on modelling positive rainfall amounts, such as Wilks (2006), So et al. (2015) and Yunus et al. (2017). By investigating QQ plots, we find that an estimated Gamma distribution works quite well

for non-extreme amounts of precipitation. However, most of the observed monthly extremes are not captured well.

To improve the model fit, we define an extremal mixture model (Frigessi et al., 2002; Behrens et al., 2004; MacDonald et al., 2011) which combines the Gamma distribution with a GPD as defined in (6.1.2). Given a threshold $u_{j,m}$, $\tilde{R}_{j,m} \mid (\tilde{R}_{j,m} \leq u_{j,m})$ follows a truncated Gamma distribution, while $\tilde{R}_{j,m} \mid (\tilde{R}_{j,m} > u_{j,m})$ is generalized Pareto distributed. Formally, let $G_{j,m} \sim \text{Gamma}(\alpha_{j,m}, \beta_{j,m})$ with shape $\alpha_{j,m}$ and rate $\beta_{j,m}$, and $H_{j,m} \sim \text{GPD}(\psi_{j,m}, \xi_{j,m})$ with scale $\psi_{j,m} = \tilde{\psi}_{j,m} - \xi u_{j,m}$ and shape $\xi_{j,m}$. The reparametrisation of the scale parameter in $H_{j,m}$ removes the effect of the threshold on inference and has been used in previous studies (Fawcett and Walshaw, 2006). Then, the cumulative distribution function of $\tilde{R}_{j,m} \mid (\tilde{R}_{j,m} > 0)$ is given by

$$\Pr\left(\tilde{R}_{j,m} > r \mid \tilde{R}_{j,m} > 0\right) = \begin{cases} \Pr\left(G_{j,m} > r\right) & r \leq u_{j,m}, \\ \Pr\left(G_{j,m} > u_{j,m}\right) \Pr\left(H_{j,m} > r - u_{j,m}\right) & r > u_{j,m}. \end{cases} \tag{6.2.2}$$

Combining the model components defined above, the event $\tilde{R}_{j,m} > r$, for $r > u_{j,m}$, occurs with probability

$$\Pr\left(\tilde{R}_{j,m} > r\right) = p_{j,m} \Pr\left(G_{j,m} > u_{j,m}\right) \Pr\left(H_{j,m} > r - u_{j,m}\right).$$

Due to the empirical mean of $R_{j,m} \mid (R_{j,m} > 0)$ being similar for all j, we fix $\alpha_{j,m},\ m = 1, \ldots 12$, in the Gamma distribution to be constant across sites and, thus, refer to this parameter as α_m in the rest of this chapter.

6.2.2 Prior model

Prior selection is critical in this analysis due to the varying degrees of data availability at each site; inference at sites where data are lacking or unavailable will be dominated by the prior distribution. We considered uninformative, improper Uniform priors on $\log \alpha_m$, $\log \beta_{j,m}$, $\log \tilde{\psi}_{j,m}$ and $\xi_{j,m}$. However, these produced unrealistic estimates of

$\xi_{j,m}$, mostly due to the difficulty in estimating $\xi_{j,m}$ given short data records. Studies on extreme rainfall often feature the prior used in Martins and Stedinger (2000) which constrains the shape parameter to be in a sensible interval.

For each of these parameters, we instead introduce a prior aimed at exploiting spatial and seasonal structure, assuming that parameters for neighbouring sites and adjacent months are likely to be similar. In particular, let $\phi_{j,m}$ denote an arbitrary parameter at site j and month m. We propose that

$$\phi_{j,m} \sim \mathcal{N}\left(\frac{\phi_{j,m-1} + \phi_{j,m+1} + \sum_{j' \neq j} \phi_{j',m} d_{j,j'}}{2 + \sum_{j' \neq j} d_{j,j'}}, \frac{1}{(2 + \sum_{j' \neq j} d_{j,j'})\tau_\phi}\right), \tag{6.2.3}$$

where $\tau_\phi > 0$ denotes the precision that is common to all sites and months. The constant $d_{j,j'} \geq 0$ describes our prior belief concerning the degree of similarity of $\phi_{j,m}$ and $\phi_{j',m}$. This prior is a variant of the intrinsic autoregressive (IAR) prior as described in Banerjee et al. (2004) and allows us to pool information across neighbouring sites and months, which helps to produce more stable parameter estimates and to reduce uncertainty in these estimates. The cyclical nature of the sequence of months means that values 0 and 13 for $m-1$ and $m+1$ should be replaced by the values 12 and 1 respectively in order to ensure that December and January are correctly identified as being adjacent months. We define a flat, conjugate Gamma$(1, 0.001)$ prior for τ_ϕ.

6.2.3 Threshold selection and estimation

We detail our approach to estimate the model defined in Sections 6.2.1 and 6.2.2 in the following. First, we consider $p_{j,m}$, which can be estimated independently from the remaining parameters due to the hierarchical model structure. Next, the selection of the thresholds $u_{j,m}$ is described. Finally, we infer on the remaining model parameters via an MCMC algorithm which is outlined at the end of this subsection.

For sites with more than five years of data, we estimate $p_{j,m}$ empirically due to

the high number of observations available. We infer on the remaining sites via spatial interpolation. Let $\mathcal{J}$ denote the indices of the sites with at least five years of data. We further define a pairwise weighting between arbitrary sites j and j' by introducing the weight

$$d_{j,j'} = \exp\left(-\|\boldsymbol{x}_j - \boldsymbol{x}_{j'}\|\right), \tag{6.2.4}$$

where $\boldsymbol{x}_j$ denotes the longitude and latitude coordinates of site j and $\|\cdot\|$ corresponds to the Euclidean distance. As the study region is small, the curvature of the earth is negligible and the Euclidean distance in the two-dimensional space is close to the true distance between the sites. Then for a site $j \notin \mathcal{J}$, the estimate $\hat{p}_{j,m}$ for $p_{j,m}$ is derived as

$$\hat{p}_{j,m} = \sum_{j' \in \mathcal{J}} d_{j,j'} \hat{p}_{j',m}. \tag{6.2.5}$$

The weights $d_{j,j'}$ defined in (6.2.4) are identical to the ones which we set in the prior density (6.2.3). As the weighting function (6.2.4) produces larger values for locations close together, a higher weight is given to neighbouring sites.

We now consider how to select the thresholds, $u_{j,m}$, of our model (6.2.2). These thresholds must be large enough for the asymptotic argument of Pickands (1975) to approximately hold whilst also low enough so that we have a sufficient number of observations for reliable model fitting. We use the classical fixed threshold approach as described in Coles (2001) for the sites in $\mathcal{J}$. Specifically, by inspection of threshold stability plots, we find the smallest threshold above which the GPD is an appropriate model for the exceedances. For the other sites, we estimate these thresholds in an equivalent manner to (6.2.5). Other threshold selection methods are outlined by Scarrott and MacDonald (2012).

The parameters of our Gamma-GPD mixture model are estimated using MCMC methods. We sample from the posterior distribution using a Metropolis-within-Gibbs scheme. In particular, proposal values of each parameter are generated sequentially

from a Gaussian distribution and accepted with a probability defined as the posterior ratio of the proposed state relative to the current state of the Markov chain. The hyperparameter τ_ϕ in (6.2.3) is updated by sampling from the full conditional Gamma posterior as described by Knorr-Held (2003). We tune the parameters of the MCMC algorithm to ensure an acceptance rate of 20-25% in accordance with the optimality criterion of Roberts et al. (1997).

6.3 Results and discussion

We begin this section by considering the results of the MCMC implementation. We run our MCMC chains for 20000 iterations, and discard the first 5000 iterations as burn-in to aid convergence. Examples of the chains produced are provided in Figure 6.3.1 for scale and shape parameters $\psi_{10,6}$ and $\xi_{10,6}$. Estimates of these parameters were obtained using the posterior means of their respective MCMC chains. These plots demonstrate that good mixing has been achieved for this case; similar results were obtained across other stations and months.

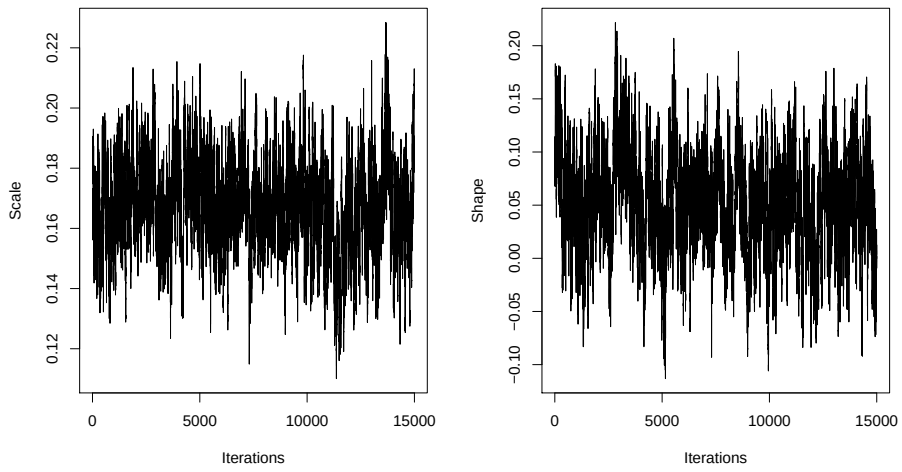

Figure 6.3.1: MCMC chains for the scale and shape parameters for station 10 in June.

We now explore the monthly variation in the estimated model parameters by focussing on results at four nearby stations. The locations of these stations are shown in the top left panel of Figure 6.3.2. The data set contains over 8000 observations for stations 2 and 5, and no observations for stations 7 and 10. The top right and bottom left panels of Figure 6.3.2 show our estimates of the scale and shape parameters, respectively, at these four locations. These plots demonstrate the seasonality in the parameter estimates, with higher values of both the scale and shape generally corresponding to summer and autumn months. This effect is maintained in the predicted 0.998 quantiles, shown in the bottom right panel of Figure 6.3.2, which are typically highest between June and October. A similar trend was observed at other sites, particularly those with limited data where estimates are more heavily influenced by information from other locations, due to the spatial smoothing imposed by the model.

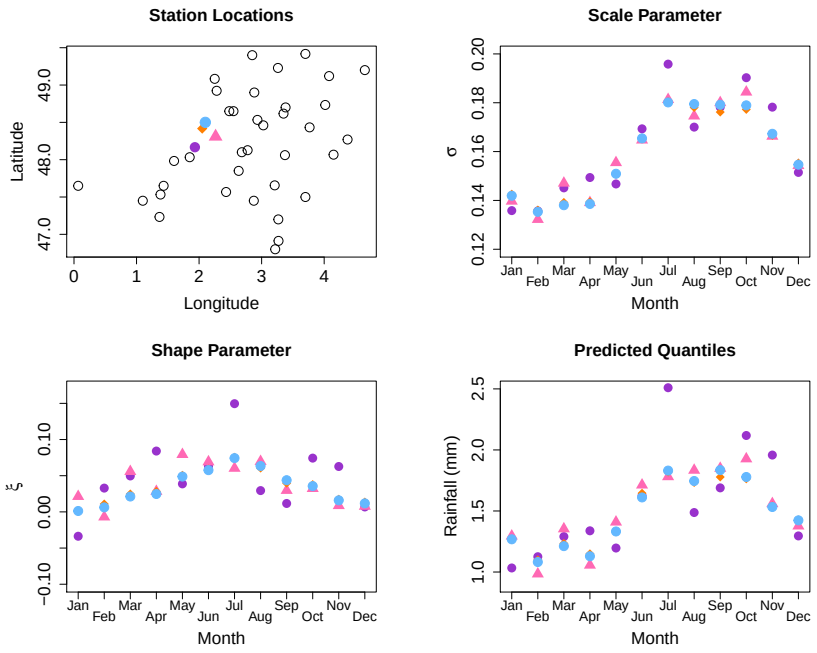

Figure 6.3.2: Location of stations 2 (purple), 5 (pink), 7 (orange) and 10 (blue), as well as estimates of the corresponding scale and shape parameters and predicted 0.998 quantiles.

We now consider our estimates in the context of the competition, which used the quantile loss function by Koenker (2005). In particular, as in the challenge, we consider the percentage improvement provided by our method over benchmark predictions. The competition was split into two challenges: Challenge 1 involved only sites where observations were available, with the benchmark quantile estimates being given by the monthly maxima at each station; Challenge 2 included predictions for all sites, with the benchmark for those sites with no data being taken as the average of the quantiles predicted in Challenge 1 for each month. Our method gave a 59.9% improvement over the benchmark for Challenge 1, and a 57.7% improvement for Challenge 2. Table 6.3.1 shows the performance of our approach using this same metric, but with the results separated by month.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Challenge 1	57.7	71.1	60.0	65.0	43.7	62.8	65.9	77.0	38.7	38.4	52.2	33.4
Challenge 2	54.4	69.3	57.4	61.9	43.1	60.7	64.2	75.4	37.9	36.4	49.3	31.3

Table 6.3.1: Percentage improvement over the benchmark for Challenges 1 and 2 across each month.

As is to be expected, our method performed better in Challenge 1, where only predictions for sites with observations were considered, across all months. Looking at these results separately for each month allows us to identify possible areas for improvement. In particular, the scores for September, October and December are lower than for other months, suggesting that the method could be improved by focussing on the modelling of autumn and winter months.

Chapter 7

Discussion

7.1 Thesis summary

The aim of this thesis was to present novel theoretical and methodological results for multivariate extremes, with particular emphasis on extremal dependence between multivariate random variables. In this chapter, we first summarize the contributions of this work, before presenting some ideas for further work in Section 7.2.

In Chapter 3, we proposed approaches for identifying subsets of variables that can take their largest values simultaneously, while the others are of smaller order. This involved the introduction of a novel set of indices, based on a regular variation assumption, that describe the extremal dependence structure between variables. We proposed two inferential methods to estimate these parameters, as well as the proportion of extremal mass associated with our $2^d - 1$ sub-cones, each chosen to represent a different subset of variables taking its largest values simultaneously. This methodology could be applied to aid model selection, or in the construction of mixture models that exhibit the required extremal dependence structures. We demonstrated these methods through a simulation study and an application to river flow data.

In Chapter 4, we discussed extensions of the results of Chapter 3, by considering variables in terms of their radial-angular components. The first of these radial-angular methods approximates the various faces of the angular simplex using a partitioning approach similar to the methods in Chapter 3. The remaining proposed methods use a soft-thresholding technique that takes into account the distance of points from different faces of the simplex; we presented a version of this for our hidden regular variation assumption, as well as weighted variant of the method proposed by Goix et al. (2016). All of these methods were compared in a simulation study, where different methods were shown to perform best in different cases.

In Chapter 5, we carried out a theoretical investigation into the extremal dependence properties of vine copulas. In particular, we studied the coefficient of tail dependence η_C (Ledford and Tawn, 1996) for certain vine copula examples. This involved applying the geometric approach of Nolde (2014) for calculating η_C when the joint density is known, and extending the approach for cases where only higher order joint densities can be obtained analytically. We focussed on trivariate vine copulas constructed from extreme value and inverted extreme value components, and obtained results for higher dimensional D-vines and C-vines formed from inverted extreme value pair copulas. We demonstrated our results through a series of (inverted) logistic examples.

Finally, in Chapter 6 we presented the results of a team competition for the EVA 2017 conference. The challenge involved predicting extreme precipitation quantiles for several stations in the Netherlands. Our team proposed a Bayesian approach with a hierarchical structure that modelled spatio-temporal dependence through the model parameters; our estimation procedure involved MCMC techniques and spatial interpolation. The approach performed well in terms of the quantile loss metric proposed by the challenge organizer.

7.2 Further work

7.2.1 Overview

In this section, we discuss some possible ways to improve the methods presented in the previous chapters, as well as other potential avenues for future work. In Section 7.2.2, we discuss additional constraints on the parameters of the models in Chapter 3 that may improve model fitting, while in Section 7.2.3 we propose an alternative approach for redistributing negligible mass in these methods that takes into account the proximity of sub-cones with mass. In Section 7.2.4, we present an extension of the conditional approach of Heffernan and Tawn (2004) involving mixture distributions, which could be used to capture combinations of different extremal dependence structures.

7.2.2 Parameter estimation in Chapter 3

Under Assumption 1, each sub-cone $\mathbb{E}_C$ containing extremal mass should have $\tau_C(\delta^*) = 1$ for some $\delta^* < 1$, and a consequence of moment constraint (3.1.3) is that each variable should be represented on at least one sub-cone with extremal mass. Combining these two statements gives a constraint on the set of $\tau_C(\delta)$ values. In particular, we have that $\max_{C:C\supset i} \tau_C(\delta^*) = 1$, for $i = 1, \ldots, d$ and some $\delta^* < 1$. This result is not currently guaranteed by our methods, and presents a potential area for improvement. For some models, such as those presented in Table 3.2.1, the constraint is satisfied for $\delta^* = 0$. We impose this as an additional assumption in fitting Method 1; recalling that we set $\tau_C = \tau_C(0)$, our constraint becomes $\max_{C:C\supset i} \tau_C = 1$, for $i = 1, \ldots, d$. We now suggest estimating all the (τ_C, K_C) parameters simultaneously, rather than maximizing the likelihood separately on each sub-cone. This is achieved by including a penalty term that encourages the condition on the τ_C values to be satisfied. We propose maximizing a function of the form

$$\sum_{C\in 2^D\setminus\emptyset} \log\{L_C(K_C, \tau_C)\} - \lambda \left| \sum_{i=1}^{d} \max_{C:C\supset i} \tau_C - d \right|, \tag{7.2.1}$$

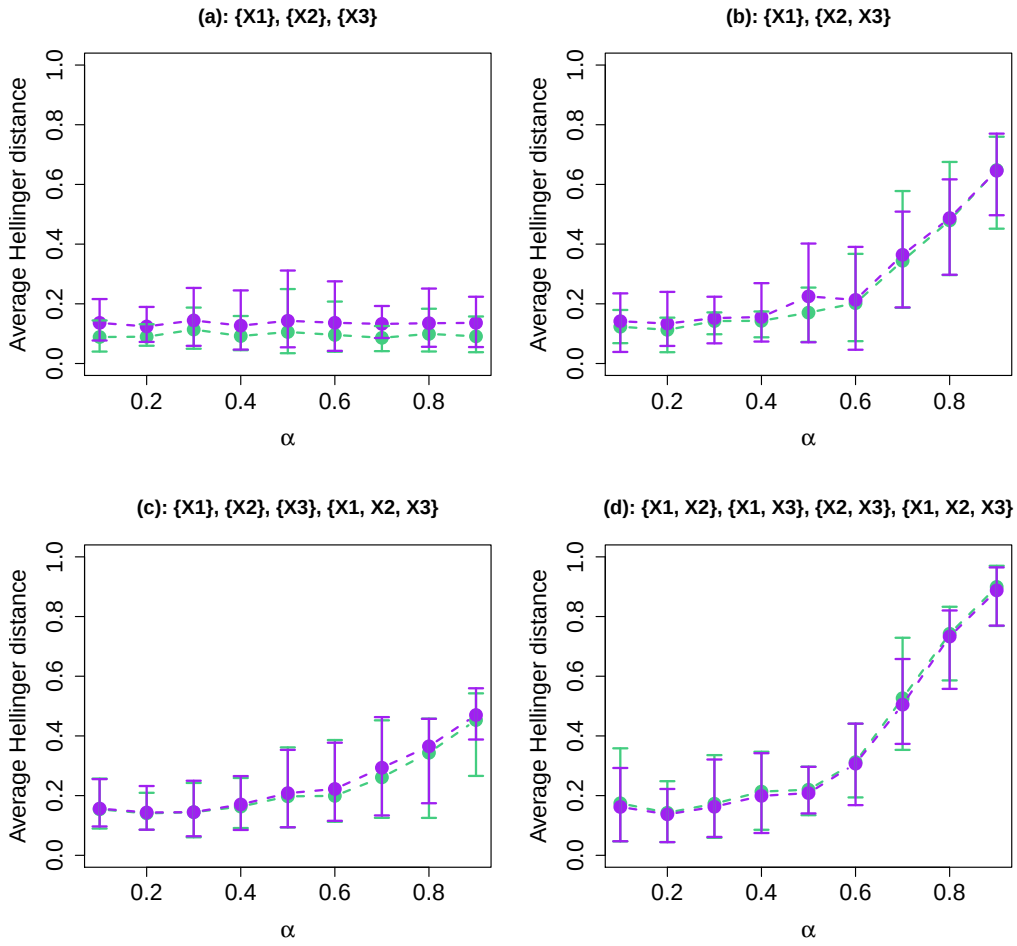

Figure 7.2.1: Average Hellinger distance, 0.05 and 0.95 quantiles over 25 simulations. Method 1: purple; penalized censored log-likelihood: green.

for the censored likelihood L_C associated with each sub-cone defined as in (3.3.3), and some $\lambda > 0$, imposing that $\tau_C \in (0, 1]$ for all $C \in 2^D \setminus \emptyset$. Parameter estimates that maximize (7.2.1) do not have a closed form, so the function must be optimized numerically.

We demonstrate the use of the penalized censored log-likelihood (7.2.1) in the trivariate case, comparing to results obtained using our original Method 1. Figure 7.2.1 shows the average Hellinger distance achieved using both these approaches

for data sampled from asymmetric logistic distributions with four different underlying extremal dependence structures: case (a) corresponds to mass on $\mathbb{E}_1$, $\mathbb{E}_2$ and $\mathbb{E}_3$; case (b) places extremal mass on $\mathbb{E}_1$ and $\mathbb{E}_{2,3}$; case (c) has mass on $\mathbb{E}_1$, $\mathbb{E}_2$, $\mathbb{E}_3$ and $\mathbb{E}_{1,2,3}$; and in case (d), extremal mass is placed on the four sub-cones $\mathbb{E}_{1,2}$, $\mathbb{E}_{1,3}$, $\mathbb{E}_{2,3}$ and $\mathbb{E}_{1,2,3}$. In this study, the dependence parameter α takes values in $\{0.1, 0.2, \ldots, 0.9\}$, and the sample size in each case is $n = 1000$. The tuning parameters are chosen to be the same as in the simulation study of Chapter 3 for Method 1. For the proposed penalized approach, function (7.2.1) is optimized numerically using `optim` in `R`. We set the additional tuning parameter to be $\lambda = 5$, which we found to work well in practice across the simulations. In most situations, the two methods give similar results in terms of the Hellinger distance, with the penalized approach giving most improvement in case (a), where the extreme values occur separately in each of the three variables.

We may expect the penalized approach to be most successful in sparse cases, where the extremal mass is placed on fewer sub-cones, as it is more likely that the correct τ_C estimates will be increased to 1. This is demonstrated in Figure 7.2.2, where we provide boxplots of the estimated τ_C values using Method 1 and the penalized approach. Cases (a) and (b) are the most sparse examples here, with each variable being extreme on exactly one sub-cone, and the penalized approach is reasonably successful at estimating the correct τ_C values as 1 for these examples. It is also interesting to note that the estimates of the τ_C values which should be less than 1 are similar using both methods. For cases (c) and (d), there is generally less improvement in the τ_C estimates using the penalized approach. For case (d) with $\alpha = 0.8$, the values of τ_1, τ_2 and τ_3 are sometimes incorrectly increased using the penalized approach, although this may be expected with such weak asymptotic dependence and no true extremal mass on $\mathbb{E}_1$, $\mathbb{E}_2$ and $\mathbb{E}_3$. As the results are usually no worse for the penalized approach than Method 1, but there is sometimes evidence of improvement, it may be worth considering this alternative estimation approach in future.

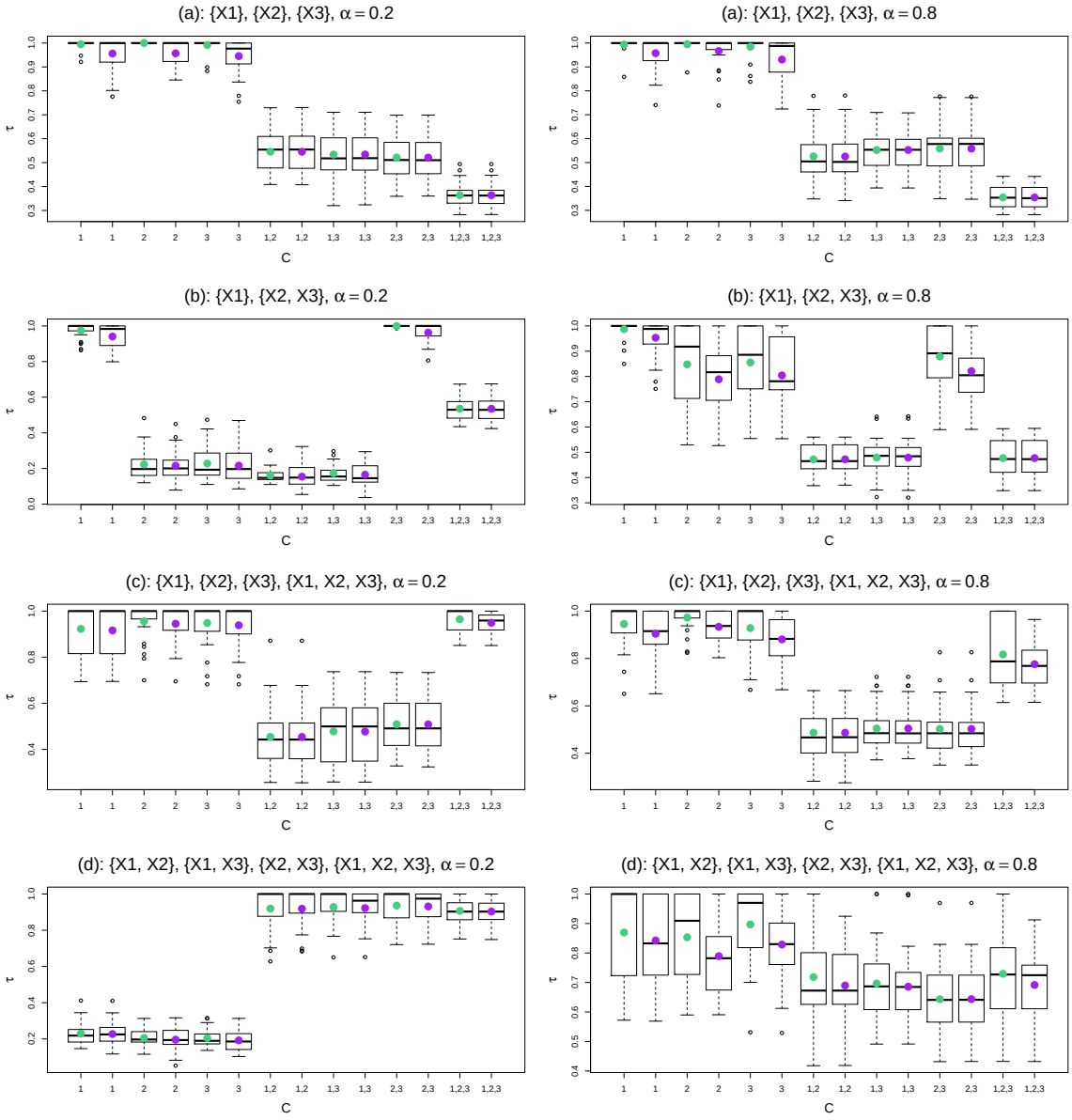

Figure 7.2.2: Comparison of estimated τ_C values for cases (a)-(d) with $\alpha = 0.2$ and $\alpha = 0.8$, over 25 simulations. The points show the average estimated τ_C value in each case. Method 1: purple; penalized censored log-likelihood: green.

The moment constraint (3.1.3) can also be used to infer conditions on the proportion of extremal mass assigned to each sub-cone. With $(\hat{\tau}_C, \hat{K}_C)$ denoting our estimates of (τ_C, K_C), let $C_{\max} = \left\{ C \in 2^D \setminus \emptyset : \hat{\tau}_C = \max_{C'} \hat{\tau}_{C'} \right\}$. Recall that $\boldsymbol{X}^*$ is defined

by truncating $\boldsymbol{X}$ according to (3.2.4), and we set $Q = \min(X_i^* : X_i^* > 0, i = 1, \ldots, d)$. We consider $\Pr(\boldsymbol{X}^* \in E_C \mid Q > q)$ to estimate the proportion of extremal mass assigned to each sub-cone, for E_C defined in (3.3.1). As $q \to \infty$, non-zero mass ψ_C will be estimated only on sub-cones corresponding to $C \in C_{\max}$, with

$$\psi_C = \frac{\hat{K}_C \widehat{\Pr}(\boldsymbol{X}^* \in E_C)}{\sum_{C' \in C_{\max}} \hat{K}_{C'} \widehat{\Pr}(\boldsymbol{X}^* \in E_{C'})},$$

for $\widehat{\Pr}(\boldsymbol{X}^* \in E_C)$ denoting the empirical estimate of $\Pr(\boldsymbol{X}^* \in E_C)$. The values of ψ_C naturally satisfy $\sum_{C \in C_{\max}} \psi_C = 1$, but there are additional consequences of (3.1.3) that could be considered. In particular, for all $i = 1, \ldots, d$, we should have

$$\sum_{C : C \supset i} \psi_C \geq \frac{1}{d}, \tag{7.2.2}$$

otherwise the integral in (3.1.3) would be less than $1/d$. Note that the situation where mass is only placed on sub-cones $\mathbb{E}_i$, $i \in \{1, \ldots, d\}$, corresponds to the boundary case where $C_{\max} = \{C : C \in 2^D \setminus \emptyset, |C| = 1\}$, and we must have $\psi_C = 1/d$ for all C with $|C| = 1$. Further, an upper bound on the proportion of mass on each sub-cone with $C \in C_{\max}$ is given by

$$\psi_C \leq \frac{|C|}{d}, \tag{7.2.3}$$

otherwise the integral in the moment constraint would be greater than $1/d$ for some $i \in \{1, \ldots, d\}$. Full asymptotic dependence, where all extremal mass is assigned to $\mathbb{E}_D$, is a boundary case here, with $\psi_D = |D|/d = 1$. Conditions (7.2.2) and (7.2.3) on the location of extremal mass on the simplex could also be incorporated into the estimation procedure to potentially improve results.

7.2.3 Redistribution of mass in Chapters 3 and 4

Another area of potential improvement in the methods to determine extremal dependence structures is in the redistribution of negligible mass. We currently propose setting to zero any estimated mass $\hat{p}_C$ below some small value π, and to subsequently

renormalize the vector of estimated proportions. This results in the redistribution of mass across all sub-cones corresponding to sets C' with $\hat{p}_{C'} > \pi$, proportionally to $\hat{p}_{C'}$. A better alternative may be to reassign mass to sub-cones that are close in some way to where the mass was originally assigned. To do this, we need some measure of similarity between the sets of variables that are considered extreme on the various sub-cones.

One possibility is to consider the disjunctive union of pairs of sets,

$$C \triangle C' = (C \cup C') \setminus (C \cap C'),$$

i.e., the union without the intersection, and to use the cardinality of this set as a measure of the difference between C and C'; small values of $|C \triangle C'|$ correspond to sets C and C' being similar, while larger values suggest greater differences between the components of C and C'. In particular, suppose we determine that the sub-cone corresponding to set C contains negligible mass, i.e., $\hat{p}_C < \pi$. We could reassign the mass equally to all sets in

$$\underset{C':\hat{p}_{C'}>\pi}{\arg\min} |C \triangle C'|,$$

to redistribute only to the closest sub-cones with extremal mass. This echoes our definition of neighboured receiver operating characteristic curves, where we considered adjacent sub-cones as those with $|C \triangle C'| = 1$.

In terms of classification, this new redistribution technique will make no difference to the performance of our methods, since we will still end up estimating extremal mass on the same sub-cones. However, it may improve results based on the average Hellinger distance, which measures how close the estimated proportions are to the truth. This may only result in small improvements since the value of the tuning parameter π is usually taken to be close to zero. However, if a more sparse representation of the extremal dependence structure is required, and π is taken to be slightly larger than in our current examples, this may present a more successful procedure for redistributing mass.

7.2.4 Conditional mixture model

The conditional approach to modelling multivariate extremes, introduced by Heffernan and Tawn (2004), and discussed in Section 2.3.11, is a flexible model that is able to capture a range of tail dependence features, including asymptotic independence and asymptotic dependence. In some cases, data may exhibit a mixture of two or more such dependence behaviours, linked to the position of extremal mass in the angular simplex discussed in Chapters 3 and 4. In order to capture these characteristics, one possibility is to exploit the conditional model of Heffernan and Tawn by introducing mixture components. We consider the bivariate case here, but later discuss possible extensions to higher dimensions.

Assuming that the variables X_1 and X_2 have standard Gumbel margins, which can be achieved via a transformation, and conditioning on $X_1 = x$, for x above some large threshold u, we propose the model

$$X_2 \mid (X_1 = x) = \begin{cases} \alpha_1 x + x^{\beta_1} Z_1, & \text{with probability } p, \\ \alpha_2 x + x^{\beta_2} Z_2, & \text{with probability } 1 - p, \end{cases}$$

with $p \in [0, 1]$; $\alpha_1, \alpha_2 \in [0, 1]$; $\beta_1, \beta_2 < 1$; and residuals $Z_1 \sim G_1$, $Z_2 \sim G_2$ following non-degenerate distributions. For inference, following Heffernan and Tawn (2004), we suggest the working assumptions that $Z_1 \sim N(\mu_1, \sigma_1^2)$ and $Z_2 \sim N(\mu_2, \sigma_2^2)$, for $\mu_1, \mu_2 \in \mathbb{R}$ and $\sigma_1, \sigma_2 > 0$.

If we assume a mixture of asymptotic independence and asymptotic dependence in our data, we can impose that $\alpha_2 = 1$ and $\beta_2 = 0$ to ensure the second mixture component corresponds to asymptotic dependence, and allow the usual constraints on (α_1, β_1) so that the first mixture component corresponds to asymptotic independence. That is, for observations of X_2 corresponding to extreme values of $X_1 = x$, we would

fit the mixture model

$$X_2 \mid (X_1 = x) \sim \begin{cases} N\left(\alpha_1 x + \mu_1 x^{\beta_1},\ x^{2\beta_1}\sigma_1^2\right), & \text{with probability } p, \\ N\left(x + \mu_2,\ \sigma_2^2\right), & \text{with probability } 1-p. \end{cases} \tag{7.2.4}$$

In our investigations, we found that maximum likelihood techniques were not successful in providing reasonable estimates of the parameters in this model, which is most likely due to the dependence between some of the parameters; we therefore turn to our attention to Bayesian inference via Markov chain Monte Carlo (MCMC) techniques. We found that applying the Gibbs sampler and Metropolis-Hastings algorithm generally resulted in unsatisfactory performance, with poor mixing in the MCMC chains. An alternative to this is to use Hamiltonian Monte Carlo (Duane et al., 1987; Neal, 1994), which is designed to overcome some of the problems associated with other MCMC methods. We use the software Stan (Stan Development Team, 2018), which provides a relatively straightforward framework for implementing this method, with parameter estimates obtained from the mean of the corresponding posterior densities.

In Figure 7.2.3 we demonstrate the performance of Hamiltonian Monte Carlo for bivariate data exhibiting a mixture of asymptotic dependence and asymptotic independence, simulated from model (7.2.4). The chains were run for 10000 iterations, with the first half discarded as burn-in. The MCMC chains for all seven parameters show reasonable mixing, and true parameter values all lie within the 90% credible intervals obtained from the posterior distributions, demonstrating that this is a promising approach for fitting mixture models for conditional extremes.

Having obtained parameter estimates for the proposed mixture model, it is possible to cluster the observed values of $X_2 \mid (X_1 > u)$ into two groups by evaluating the likelihood associated with each mixture component, and assigning each observation to the cluster corresponding to the highest likelihood. Let C_1 and C_2 denote the

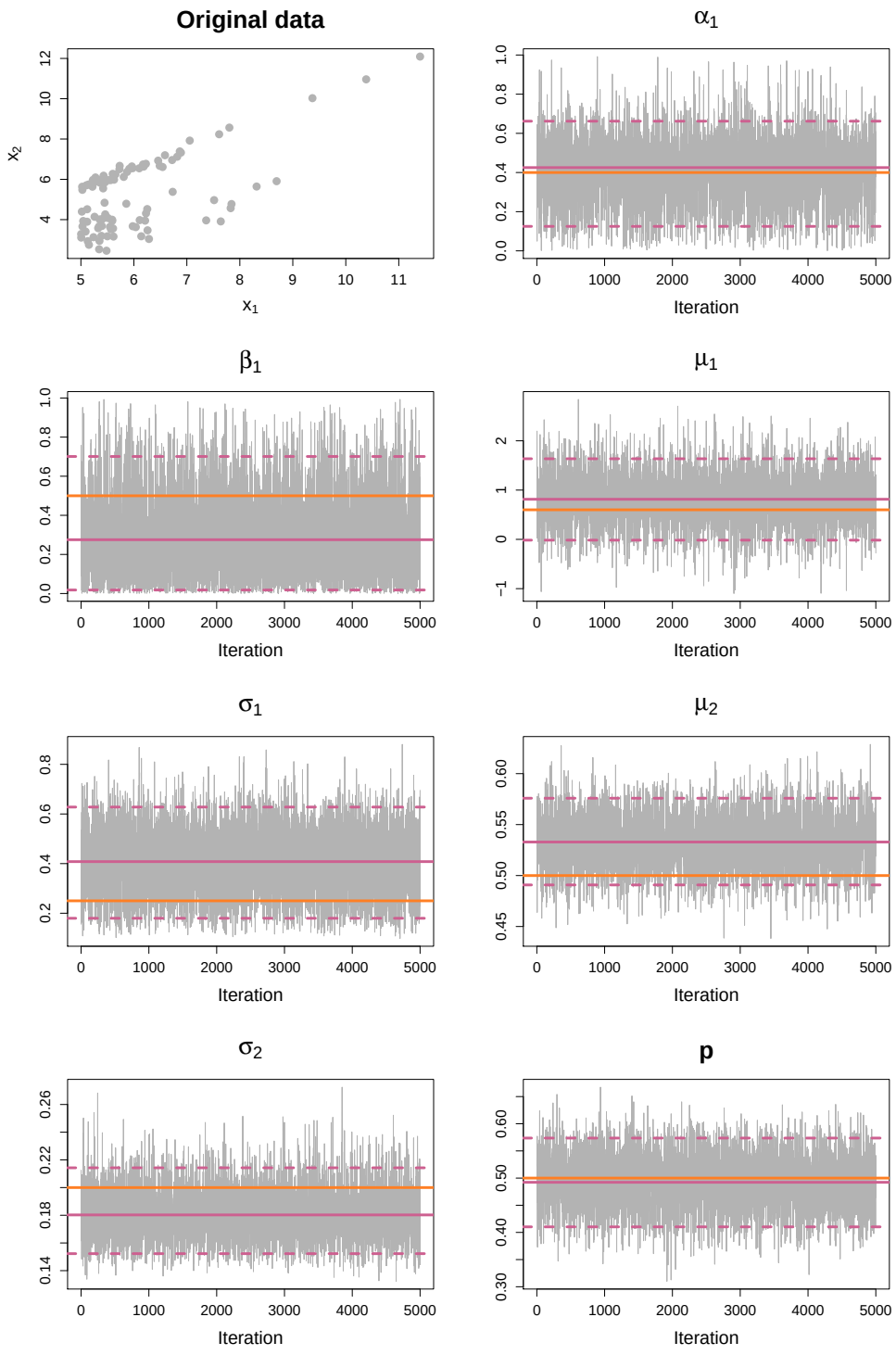

Figure 7.2.3: A sample of bivariate data exhibiting a mixture of asymptotic dependence and asymptotic independence, and results from applying Hamiltonian Monte Carlo with 5000 iterations. True parameter values: orange; mean of the posterior: solid pink; 90% credible interval; dashed pink.

clusters associated with the first and second mixture components, respectively. As in Section 2.3.11, we suppose we have n observations, $(x_{1,1}, x_{2,1}), \ldots, (x_{1,n}, x_{2,n})$, and let $i_1, \ldots, i_{n_u}$ be the indices i in the set $\{1, \ldots, n\}$ corresponding to $x_{1,i} > u$. The residual associated with the i_lth observation is now given by $z_l = (x_{2,i_l} - \hat{\alpha}_1 x_{1,i_l})/x_{1,i_l}^{\hat{\beta}_1}$ for observations corresponding to cluster C_1, and $z = (x_{2,i_l} - x_{1,i_l})$ for cluster C_2. From these residuals, simulation can be carried out using an analogous approach to the one in Section 2.3.11, allowing for extrapolation beyond the observed conditional extreme values, although we must also take into account the cluster label associated with each sampled residual in step 3. That is, to simulate values $(x^*_{1,1}, x^*_{2,1}), \ldots, (x^*_{1,m}, x^*_{2,m})$, for $j = 1, \ldots, m$, we have the algorithm

1. set $x^*_{1,j} = u + e_j$, where $e_j \sim \text{Exp}(1)$,
2. sample z^*_j with replacement from the set of observed residuals $\{z_1, \ldots, z_{n_u}\}$, independently of $x^*_{1,j}$,
3. set $x^*_{2,j} = \begin{cases} \hat{\alpha}_1 x^*_{1,j} + z^*_j (x^*_{1,j})^{\hat{\beta}_1}, & \text{if } z^*_j \in C_1, \\ x^*_{1,j} + z^*_j, & \text{if } z^*_j \in C_2. \end{cases}$

We compare simulated values from this algorithm to the standard (one component) Heffernan and Tawn approach in Figure 7.2.4. It is clear from the second plot that using two components in the mixture model allows for extrapolation in two distinct directions capturing the behaviour of both the cluster corresponding to asymptotic dependence and the one exhibiting asymptotic independence. In the one component model the directions of extrapolation are averaged across the two clusters, which does not properly capture either component. This may, for example, lead to the two component model giving better estimation of return levels for data that exhibit mixtures of extremal dependence types.

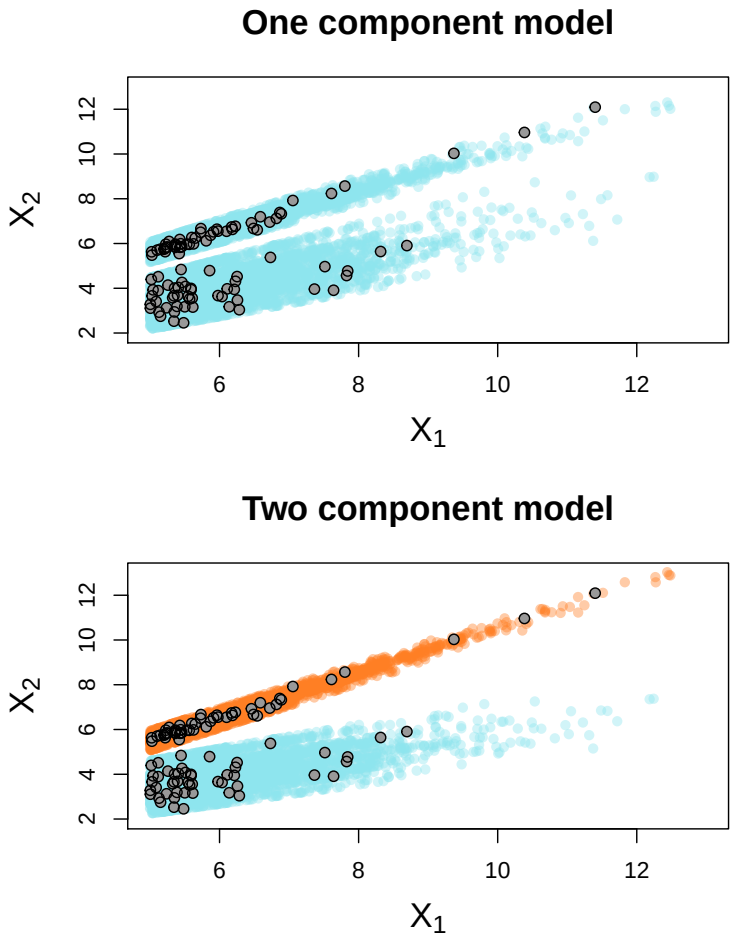

Figure 7.2.4: Extrapolation using the original approach of Heffernan and Tawn (top) and our proposed two component mixture model (bottom); the orange and blue regions show the two clusters. The original data are shown in grey.

Moving into higher than two dimensions will present challenges due to the number of parameters involved in the model. We could first use dimension reduction techniques, such as those introduced in Chapters 3 and 4, to identify the number of mixture components necessary in the model. This may also inform us about the tail dependence properties that each mixture component should capture, and provide some constraints on the parameter values, which may aid model fitting.

Appendix A

Supplementary Material for Chapter 3

A.1 Calculation of $\tau_C(\delta)$ for a bivariate extreme value distribution

We determine the value of $\tau_C(\delta)$, defined in (3.2.7), by establishing the index of regular variation of

$$\Pr\left(X_i > t, i \in C; X_j < t^{\delta}, j \in D \setminus C\right).$$

Here, we calculate $\tau_1(\delta)$, $\tau_2(\delta)$ and $\tau_{1,2}$ for a bivariate extreme value distribution, with distribution function given in (3.2.8). The exponent measure V can be written as

$$V(x,y) = \frac{2}{y}\int_0^1 (1-w)dH(w) - \frac{2}{y}\int_{\frac{x}{x+y}}^1 (1-w)h(w)dw + \frac{2}{x}\int_{\frac{x}{x+y}}^1 wh(w)dw + \frac{2\theta_1}{x}.$$

To study $\tau_1(\delta)$, suppose that $h(w) \sim c_1(1-w)^{s_1}$ as $w \to 1$, for $s_1 > -1$. For $x \to \infty$ and $y = o(x)$, applying Karamata's theorem (Resnick, 2007, Theorem 2.1), we have

$$V(x,y) = \frac{1}{y} - \frac{2c_1}{y(s_1+2)}\left(\frac{y}{x+y}\right)^{s_1+2}\{1+o(1)\} + \frac{2c_1}{x(s_1+1)}\left(\frac{y}{x+y}\right)^{s_1+1}\{1+o(1)\} + \frac{2\theta_1}{x}$$

$$= \frac{1}{y} + 2c_1 \left(\frac{y}{x+y}\right)^{s_1+1} \left\{\frac{1}{x(s_1+1)} - \frac{1}{(s_1+2)(x+y)}\right\} \{1+o(1)\} + \frac{2\theta_1}{x}$$
$$= \frac{1}{y} + \frac{2c_1 y^{s_1+1} x^{-(s_1+2)}}{(s_1+1)(s_1+2)} \{1+o(1)\} + \frac{2\theta_1}{x}.$$

By this result,

$$\begin{aligned}
&\Pr\left(X_1 > t, X_2 < t^\delta\right) = \Pr\left(X_2 < t^\delta\right) - \Pr\left(X_1 < t, X_2 < t^\delta\right) \\
&= \exp\left(-t^{-\delta}\right) - \exp\left\{-V\left(t, t^\delta\right)\right\} \\
&= \exp\left(-t^{-\delta}\right) - \exp\left\{-\frac{1}{t^\delta} - \frac{2c_1 t^{\delta(s_1+1)} t^{-(s_1+2)}}{(s_1+1)(s_1+2)} \{1+o(1)\} - \frac{2\theta_1}{t}\right\} \\
&= \left\{1 - t^{-\delta} + o\left(t^{-\delta}\right)\right\} \\
&\qquad \cdot \left(1 - \left[1 - \frac{2c_1 t^{\delta(s_1+1)} t^{-(s_1+2)}}{(s_1+1)(s_1+2)} + o\left\{t^{\delta(s_1+1)-(s_1+2)}\right\}\right] \left\{1 - 2\theta_1 t^{-1} + o\left(t^{-1}\right)\right\}\right) \\
&= \left\{2\theta_1 t^{-1} + \frac{2c_1 t^{\delta(s_1+1)-(s_1+2)}}{(s_1+1)(s_1+2)}\right\} \{1+o(1)\}.
\end{aligned}$$

If $\theta_1 > 0$, i.e., the spectral measure places mass on $\{1\}$, we see that $\Pr\left(X_1 > t, X_2 < t^\delta\right) \sim 2\theta_1 t^{-1}$ as $t \to \infty$, hence $\tau_1(\delta) = 1$ for all $\delta \in [0,1]$. If $\theta_1 = 0$, we have $\tau_1(\delta) = \{(s_1+2) - \delta(s_1+1)\}^{-1}$, which increases from $(s_1+2)^{-1}$ at $\delta = 0$ to 1 at $\delta = 1$. By similar calculations, if $h(w) \sim c_2 w^{s_2}$ as $w \to 0$ for $s_2 > -1$, we have $\tau_2(\delta) = 1$ if $\theta_2 > 0$, and $\tau_2(\delta) = \{(s_2+2) - \delta(s_2+1)\}^{-1}$ otherwise. Since $\tau_{1,2} = \eta_{1,2}$, we have $\tau_{1,2} = 1$ if $\theta_1 + \theta_2 < 1$, and $\tau_{1,2} = 1/2$ if $\theta_1 + \theta_2 = 1$.

A.2 Calculation of $\tau_C(\delta)$

A.2.1 Overview

In Appendix A.1, we derived $\tau_1(\delta)$, $\tau_2(\delta)$ and $\tau_{1,2}$ for a particular subclass of bivariate extreme value distribution. Here, we present further calculations of $\tau_C(\delta)$ for several trivariate copula models.

In general, there are two cases to consider: $\delta = 0$ and $\delta > 0$. For many models, the asymptotic relations we study will differ by a constant in these two cases, while

the tail index remains the same. For this reason, we focus on $\delta > 0$, and present $\delta = 0$ calculations separately only when the slowly varying function is no longer a constant, but instead varies with t.

A.2.2 Independence copula

We begin by considering the case where all three variables X_1, X_2, X_3 are independent. To calculate the value of $\tau_C(\delta)$, we need to determine the index of regular variation of

$$\Pr\left(X_i > t, i \in C; X_j < t^\delta, j \in D \setminus C\right).$$

In the independence case, this is equivalent to

$$\Pr\left(X_i > t\right)^{|C|} \Pr\left(X_i < t^\delta\right)^{|D\setminus C|} = \left(1 - e^{-1/t}\right)^{|C|} \left(e^{-1/t^\delta}\right)^{|D\setminus C|} \sim t^{-|C|},$$

so that $\tau_C(\delta) = 1/|C|$, which does not depend on the value of δ. That is, $\tau_1(\delta) = \tau_2(\delta) = \tau_3(\delta) = 1$, $\tau_{1,2}(\delta) = \tau_{1,3}(\delta) = \tau_{2,3}(\delta) = 1/2$, and $\tau_{1,2,3} = 1/3$.

A.2.3 Trivariate logistic distribution

The trivariate extreme value logistic distribution belongs to the class of trivariate extreme value distributions. The exponent measure of the logistic distribution has the form

$$V(x, y, z) = \left(x^{-1/\alpha} + y^{-1/\alpha} + z^{-1/\alpha}\right)^\alpha, \tag{A.2.1}$$

for $\alpha \in (0, 1]$. Since $\alpha = 1$ corresponds to the independence case, we restrict our calculations to $\alpha \in (0, 1)$. This distribution exhibits asymptotic dependence, with all limiting mass on $\mathbb{E}_{1,2,3}$. Since $\tau_{1,2,3} = \eta_{1,2,3} = 1$, our interest lies with the values of $\tau_C(\delta)$ for $|C| = 1$ and $|C| = 2$, and we consider each of these in turn.

$|C| = 1$: $\tau_1(\delta)$, $\tau_2(\delta)$, $\tau_3(\delta)$. In a similar approach to the bivariate case, we calculate $\tau_1(\delta)$ by considering

$$\Pr\left(X_1 > t, X_2 < t^\delta, X_3 < t^\delta\right) = \Pr\left(X_2 < t^\delta, X_3 < t^\delta\right) - \Pr\left(X_1 < t, X_2 < t^\delta, X_3 < t^\delta\right)$$

$$
\begin{aligned}
&= \exp\left(-2^{\alpha}t^{-\delta}\right) - \exp\left[-2^{\alpha}t^{-\delta}\left\{1+2^{-1}t^{(\delta-1)/\alpha}\right\}^{\alpha}\right] \\
&= 1 - 2^{\alpha}t^{-\delta} + O(t^{-2\delta}) - \exp\left(-2^{\alpha}t^{-\delta}\left[1+2^{-1}\alpha t^{(\delta-1)/\alpha} + O\left\{t^{2(\delta-1)/\alpha}\right\}\right]\right) \\
&= 2^{\alpha-1}\alpha t^{(\delta-1-\alpha\delta)/\alpha} + O(t^{-2\delta}) + O\left\{t^{(2\delta-2-\alpha\delta)/\alpha}\right\} \sim 2^{\alpha-1}\alpha t^{(\delta-1-\alpha\delta)/\alpha},
\end{aligned}
$$

yielding $\tau_1(\delta) = \alpha/(1+\alpha\delta-\delta)$. By similar calculations, we have $\tau_2(\delta) = \tau_3(\delta) = \alpha/(1+\alpha\delta-\delta)$, which increase from $\alpha < 1$ at $\delta = 0$ to 1 at $\delta = 1$.

$|C| = 2$: $\tau_{1,2}(\delta)$, $\tau_{1,3}(\delta)$, $\tau_{2,3}(\delta)$. We carry out a similar calculation to find the value of $\tau_{1,2}(\delta)$. Here, we have

$$
\begin{aligned}
&\Pr\left(X_1 > t, X_2 > t, X_3 < t^{\delta}\right) \\
&\quad = \Pr\left(X_3 < t^{\delta}\right) - \Pr\left(X_1 < t, X_3 < t^{\delta}\right) \\
&\qquad\qquad - \Pr\left(X_2 < t, X_3 < t^{\delta}\right) + \Pr\left(X_1 < t, X_2 < t, X_3 < t^{\delta}\right) \\
&\quad = \exp\left(-t^{-\delta}\right) - 2\exp\left\{-\left(t^{-1/\alpha}+t^{-\delta/\alpha}\right)^{\alpha}\right\} + \exp\left\{-\left(2t^{-1/\alpha}+t^{-\delta/\alpha}\right)^{\alpha}\right\} \\
&\quad = \exp\left(-t^{-\delta}\right)\left\{1 - 2\exp\left(-t^{-\delta}\left[\alpha t^{(\delta-1)/\alpha} + \frac{1}{2}\alpha(\alpha-1)t^{2(\delta-1)/\alpha} + O\left\{t^{3(\delta-1)/\alpha}\right\}\right]\right)\right. \\
&\qquad\qquad \left. + \exp\left(-t^{-\delta}\left[2\alpha t^{(\delta-1)/\alpha} + 2\alpha(\alpha-1)t^{2(\delta-1)/\alpha} + O\left\{t^{3(\delta-1)/\alpha}\right\}\right]\right)\right\} \\
&\quad = \left\{1 - t^{-\delta} + O\left(t^{-2\delta}\right)\right\}\left[\alpha(1-\alpha)t^{(2\delta-2-\alpha\delta)\alpha} + O\left\{t^{(3\delta-3-\alpha\delta)/\alpha}\right\}\right] \\
&\quad \sim \alpha\left(1-\alpha\right)t^{(2\delta-2-\alpha\delta)/\alpha}.
\end{aligned}
$$

This implies that $\tau_{1,2}(\delta) = \alpha/\left(2+\alpha\delta-2\delta\right)$, which varies from $\alpha/2$ at $\delta = 0$ to 1 at $\delta = 1$. Similarly, we have $\tau_{1,3}(\delta) = \tau_{2,3}(\delta) = \alpha/\left(2+\alpha\delta-2\delta\right)$.

These calculations reveal different indices of regular variation on cones with $|C| = 1, 2$ in the trivariate logistic case.

A.2.4 Trivariate distribution with extremal mass on one vertex and one edge

Now we consider a trivariate example where the extremal mass is placed on one sub-cone $\mathbb{E}_C$ with $|C| = 1$, and another with $|C| = 2$. This can be achieved by taking (X_1, X_2) to have a bivariate extreme value logistic distribution, and X_3 to be a standard Fréchet random variable independent of (X_1, X_2). The exponent measure in this case has the form

$$V(x, y, z) = \left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha} + z^{-1}, \quad \alpha \in (0, 1).$$

$|C| = 1$: $\tau_1(\delta)$, $\tau_2(\delta)$, $\tau_3(\delta)$. We first consider the index of regular variation on the cone corresponding to only X_1 being large. Following a similar procedure to previously, and exploiting the independence of (X_1, X_2) and X_3, we have

$$\begin{aligned}
\Pr\left(X_1 > t, X_2 < t^{\delta}, X_3 < t^{\delta}\right) &= \Pr\left(X_3 < t^{\delta}\right)\left\{\Pr\left(X_2 < t^{\delta}\right) - \Pr\left(X_1 < t, X_2 < t^{\delta}\right)\right\} \\
&= \exp\left(-t^{-\delta}\right)\left(\exp\left(-t^{-\delta}\right) - \exp\left[-t^{-\delta}\left\{1 + t^{(\delta-1)/\alpha}\right\}^{\alpha}\right]\right) \\
&= \exp\left(-2t^{-\delta}\right)\left\{1 - \exp\left(-t^{-\delta}\left[\alpha t^{(\delta-1)/\alpha} + O\left\{t^{2(\delta-1)/\alpha}\right\}\right]\right)\right\} \\
&= \left\{1 - 2t^{-\delta} + O\left(t^{-2\delta}\right)\right\}\left[\alpha t^{(\delta-1-\alpha\delta)/\alpha} + O\left\{t^{(2\delta-2-\alpha\delta)/\alpha}\right\}\right] \\
&\sim \alpha t^{(\delta-1-\alpha\delta)/\alpha},
\end{aligned}$$

revealing that $\tau_1(\delta) = \alpha/(1 + \alpha\delta - \delta)$. By similar calculations, $\tau_2(\delta) = \alpha/(1 + \alpha\delta - \delta)$. For the sub-cone corresponding to only variable X_3 being large, we have $\tau_3(\delta) = 1$, since

$$\begin{aligned}
\Pr\left(X_1 < t^{\delta}, X_2 < t^{\delta}, X_3 > t\right) &= \Pr\left(X_3 > t\right)\Pr\left(X_1 < t^{\delta}, X_2 < t^{\delta}\right) \\
&= \left(1 - e^{-1/t}\right)\exp\left\{-\left(t^{-\delta/\alpha} + t^{-\delta/\alpha}\right)^{\alpha}\right\} \\
&= \left(1 - e^{-1/t}\right)\exp\left(-2^{\alpha}t^{-\delta}\right) \\
&= \left\{t^{-1} + O\left(t^{-2}\right)\right\}\left\{1 - 2^{\alpha}t^{-\delta} + O\left(t^{-2\delta}\right)\right\} \sim t^{-1}.
\end{aligned}$$

$|C| = 2$: $\tau_{1,2}(\delta)$, $\tau_{1,3}(\delta)$, $\tau_{2,3}(\delta)$. We begin by showing that $\tau_{1,2}(\delta) = 1$. We have

$$\begin{aligned}\Pr\left(X_1 > t, X_2 > t, X_3 < t^\delta\right) &= \Pr\left(X_3 < t^\delta\right)\left\{1 - \Pr\left(X_1 < t\right)\right. \\ &\qquad \left. - \Pr\left(X_2 < t\right) + \Pr\left(X_1 < t, X_2 < t\right)\right\} \\ &= \exp\left(-t^{-\delta}\right)\left\{1 - 2\exp\left(-1/t\right) + \exp\left(-2^\alpha t^{-1}\right)\right\} \\ &= \left\{1 - t^{-\delta} + O\left(t^{-2\delta}\right)\right\}\left\{\left(2 - 2^\alpha\right)t^{-1} + O(t^{-2})\right\} \\ &\sim \left(2 - 2^\alpha\right)t^{-1}.\end{aligned}$$

Next, we consider the sub-cone $\mathbb{E}_{1,3}$. In this case, we have

$$\begin{aligned}\Pr\left(X_1 > t, X_2 < t^\delta, X_3 > t\right) &= \Pr\left(X_3 > t\right)\left\{\Pr\left(X_2 < t^\delta\right) - \Pr\left(X_1 < t, X_2 < t^\delta\right)\right\} \\ &= \left\{1 - \exp\left(-1/t\right)\right\}\left[\exp\left(-t^{-\delta}\right) - \exp\left\{-\left(t^{-1/\alpha} + t^{-\delta/\alpha}\right)^\alpha\right\}\right] \\ &= \left\{t^{-1} + O\left(t^{-2}\right)\right\}\exp\left(-t^{-\delta}\right)\left(1 - \exp\left[-\alpha t^{(\delta-1-\alpha\delta)/\alpha} + O\left\{t^{(2\delta-2-\alpha\delta)/\alpha}\right\}\right]\right) \\ &= \left\{t^{-1} + O\left(t^{-2}\right)\right\}\left\{1 - t^{-\delta} + O\left(t^{-2\delta}\right)\right\}\left[\alpha t^{(\delta-1-\alpha\delta)/\alpha} + O\left\{t^{(2\delta-2-\alpha\delta)/\alpha}\right\}\right] \\ &\sim \alpha t^{(\delta-1-\alpha\delta-\alpha)/\alpha},\end{aligned}$$

so we have $\tau_{1,3}(\delta) = \alpha/(\alpha\delta+1+\alpha-\delta)$. Again by symmetry, $\tau_{2,3}(\delta) = \alpha/(\alpha\delta+1+\alpha-\delta)$. These indices vary from $\alpha/(1+\alpha)$ at $\delta = 0$ to $1/2$ at $\delta = 1$.

$|C| = 3$: $\tau_{1,2,3}$. Finally, we consider sub-cone $\mathbb{E}_{1,2,3}$, where

$$\begin{aligned}&\Pr\left(X_1 > t, X_2 > t, X_3 > t\right) \\ &\qquad = \Pr\left(X_3 > t\right)\left\{1 - \Pr\left(X_1 < t\right) - \Pr\left(X_2 < t\right) + \Pr\left(X_1 < t, X_2 < t\right)\right\} \\ &\qquad = \left\{1 - \exp\left(-1/t\right)\right\}\left\{1 - 2\exp\left(-1/t\right) + \exp\left(-2^\alpha t^{-1}\right)\right\} \\ &\qquad = \left\{t^{-1} + O\left(t^{-2}\right)\right\}\left\{\left(2 - 2^\alpha\right)t^{-1} + O\left(t^{-2}\right)\right\} \sim \left(2 - 2^\alpha\right)t^{-2},\end{aligned}$$

i.e., $\tau_{1,2,3} = 1/2$.

A.2.5 Trivariate inverted logistic distribution

Next, we consider an inverted trivariate extreme value distribution, defined via its distribution function

$$\begin{aligned}\Pr(X_1 < x, X_2 < y, X_3 < z) = 1 &- \{F_{X_1}(x') + F_{X_2}(y') + F_{X_3}(z')\} \\ &+ \{F_{X_1,X_2}(x', y') + F_{X_1,X_3}(x', z') + F_{X_2,X_3}(y', z')\} \\ &- F_{X_1,X_2,X_3}(x', y', z'),\end{aligned}$$

where F_{X_1,X_2,X_3} denotes the corresponding trivariate extreme value distribution function; F_{X_1,X_2}, F_{X_1,X_3} and F_{X_2,X_3} are the corresponding bivariate distribution functions; F_{X_1}, F_{X_2} and F_{X_3} are the marginal distributions of X_1, X_2 and X_3; and noting that $-\log\left(1 - e^{-1/x}\right) = -\log\{x^{-1} + O(x^{-2})\}$,

$$x' = -\frac{1}{\log\left(1 - e^{-1/x}\right)} \sim \frac{1}{\log x},$$

as $x \to \infty$, with y', z' defined analogously. In the case of the trivariate inverted logistic distribution, which we focus on here, $F_{X_1,X_2,X_3}(x, y, z) = \exp\{-V(x, y, z)\}$, for V defined as in (A.2.1). The inverted logistic distribution exhibits asymptotic independence, placing all extremal mass on the cones with $|C| = 1$. We will show that $\tau_1(\delta) = \tau_2(\delta) = \tau_3(\delta) = 1$ for this model, and then calculate $\tau_{1,2}(\delta), \tau_{1,3}(\delta), \tau_{2,3}(\delta)$ and $\tau_{1,2,3}$.

$|C| = 1$: $\tau_1(\delta)$, $\tau_2(\delta)$, $\tau_3(\delta)$. To begin, we focus on calculating $\tau_1(\delta)$ for $\delta > 0$, by considering

$$\begin{aligned}\Pr\left(X_1 > t, X_2 < t^\delta, X_3 < t^\delta\right) &= \Pr\left(X_1 > t\right) - \Pr\left(X_1 > t, X_2 > t^\delta\right) \\ &\quad - \Pr\left(X_1 > t, X_3 > t^\delta\right) + \Pr\left(X_1 > t, X_2 > t^\delta, X_3 > t^\delta\right) \\ &= \left(1 - e^{-1/t}\right) - 2\exp\left(-\left[\left\{\log t + O\left(t^{-1}\right)\right\}^{1/\alpha} + \left\{\delta \log t + O\left(t^{-2\delta}\right)\right\}^{1/\alpha}\right]^\alpha\right) \\ &\quad + \exp\left(-\left[\left\{\log t + O\left(t^{-1}\right)\right\}^{1/\alpha} + 2\left\{\delta \log t + O\left(t^{-2\delta}\right)\right\}^{1/\alpha}\right]^\alpha\right) \\ &= \left\{t^{-1} + O\left(t^{-2}\right)\right\} - 2\exp\left[-\left(1 + \delta^{1/\alpha}\right)^\alpha \log t + O\left\{(t \log t)^{-1}\right\}\right]\end{aligned}$$

$$
\begin{aligned}
&\quad + \exp\left[-\left(1+2\delta^{1/\alpha}\right)^{\alpha}\log t + O\left\{(t\log t)^{-1}\right\}\right] \\
&= \left\{t^{-1} - 2t^{-\left(1+\delta^{1/\alpha}\right)^{\alpha}} + t^{-\left(1+2\delta^{1/\alpha}\right)^{\alpha}}\right\}\{1+o(1)\} \sim t^{-1},
\end{aligned}
$$

so we have $\tau_1(\delta) = 1$ for $\delta > 0$.

For $\delta = 0$, we consider variables X_1^*, X_2^*, X_3^* defined via truncation (3.2.4), and study

$$
\begin{aligned}
&\Pr\left(X_1^* > t, X_2^* = 0, X_3 = 0\right) = \Pr\left(X_1 > t, X_2 < -1/\log p, X_3 < -1/\log p\right) \\
&= \Pr\left(X_1 > t\right) - \Pr\left(X_1 > t, X_2 > -1/\log p\right) - \Pr\left(X_1 > t, X_3 > -1/\log p\right) \\
&\qquad + \Pr\left(X_1 > t, X_2 > -1/\log p, X_3 > -1/\log p\right) \\
&= \left\{t^{-1} + O\left(t^{-2}\right)\right\} - 2\exp\left(-\left[\left\{\log t + O\left(t^{-1}\right)\right\}^{1/\alpha} + \left\{-\log(1-p)\right\}^{1/\alpha}\right]^{\alpha}\right) \\
&\qquad + \exp\left(-\left[\left\{\log t + O\left(t^{-1}\right)\right\}^{1/\alpha} + 2\left\{-\log(1-p)\right\}^{1/\alpha}\right]^{\alpha}\right) \\
&= \left\{t^{-1} + O\left(t^{-2}\right)\right\} - 2\exp\left\{-\left\{\log t + O\left(t^{-1}\right)\right\}\left(1 + \left[\frac{-\log(1-p)}{\left\{\log t + O\left(t^{-1}\right)\right\}}\right]^{1/\alpha}\right)^{\alpha}\right\} \\
&\qquad + \exp\left\{-\left\{\log t + O\left(t^{-1}\right)\right\}\left(1 + 2\left[\frac{-\log(1-p)}{\left\{\log t + O\left(t^{-1}\right)\right\}}\right]^{1/\alpha}\right)^{\alpha}\right\} \\
&= \left\{t^{-1} + O\left(t^{-2}\right)\right\} \\
&\quad - 2\exp\left[-\left\{\log t + O\left(t^{-1}\right)\right\} - \alpha\frac{\left\{-\log(1-p)\right\}^{1/\alpha}}{\left\{\log t + O\left(t^{-1}\right)\right\}^{1/\alpha-1}}\right. \\
&\qquad \left. - \frac{\alpha(\alpha-1)}{2}\frac{\left\{-\log(1-p)\right\}^{2/\alpha}}{\left\{\log t + O\left(t^{-1}\right)\right\}^{2/\alpha-1}} + o\left\{(\log t)^{1-2/\alpha}\right\}\right] \\
&\quad + \exp\left[-\left\{\log t + O\left(t^{-1}\right)\right\} - 2\alpha\frac{\left\{-\log(1-p)\right\}^{1/\alpha}}{\left\{\log t + O\left(t^{-1}\right)\right\}^{1/\alpha-1}}\right. \\
&\qquad \left. - 2\alpha(\alpha-1)\frac{\left\{-\log(1-p)\right\}^{2/\alpha}}{\left\{\log t + O\left(t^{-1}\right)\right\}^{2/\alpha-1}} + o\left\{(\log t)^{1-2/\alpha}\right\}\right] \\
&= \left\{t^{-1} + O\left(t^{-2}\right)\right\} \\
&\quad + t^{-1}\left(-2\exp\left[-\alpha\frac{\left\{-\log(1-p)\right\}^{1/\alpha}}{\left\{\log t + O\left(t^{-1}\right)\right\}^{1/\alpha-1}} + \frac{\alpha(1-\alpha)}{2}\frac{\left\{-\log(1-p)\right\}^{2/\alpha}}{\left\{\log t + O\left(t^{-1}\right)\right\}^{2/\alpha-1}}\right.\right. \\
&\qquad \left.\left. + o\left\{(\log t)^{1-2/\alpha}\right\} + O\left(t^{-1}\right)\right]
\end{aligned}
$$

$$+\exp\Bigg[-2\alpha\frac{\{-\log(1-p)\}^{1/\alpha}}{\{\log t+O\left(t^{-1}\right)\}^{1/\alpha-1}}+2\alpha(1-\alpha)\frac{\{-\log(1-p)\}^{2/\alpha}}{\{\log t+O\left(t^{-1}\right)\}^{2/\alpha-1}} + o\left\{(\log t)^{1-2/\alpha}\right\}+O\left(t^{-1}\right)\Bigg]\Bigg)$$

$$\sim \frac{t^{-1}}{(\log t)^{2/\alpha-1}}\left[\alpha(1-\alpha)\left\{-\log(1-p)\right\}^{2/\alpha}\right],$$

as $t\to\infty$, which is regularly varying of order -1. As such, the index of regular variation is $\tau_1(\delta)=1$ for $\delta=0$. Combining these results, $\tau_1(\delta)=1$ for all $\delta\in[0,1]$. By symmetric arguments, $\tau_2(\delta)=\tau_3(\delta)=1$ for all $\delta\in[0,1]$.

$|C|=2$: $\tau_{1,2}(\delta)$, $\tau_{1,3}(\delta)$, $\tau_{2,3}(\delta)$. We first consider the sub-cone $\mathbb{E}_{1,2}$. For $\delta>0$, we have

$$\begin{aligned}
&\Pr\left(X_1>t,X_2>t,X_3<t^\delta\right)=\Pr\left(X_1>t,X_2>t\right)-\Pr\left(X_1>t,X_2>t,X_3>t^\delta\right)\\
&\quad=\exp\left[-2^\alpha\left\{-\log\left(1-e^{-1/t}\right)\right\}\right]\\
&\quad\quad-\exp\left(-\left[2\left\{-\log\left(1-e^{-1/t}\right)\right\}^{1/\alpha}+\left\{-\log\left(1-e^{-1/t^\delta}\right)\right\}^{1/\alpha}\right]^\alpha\right)\\
&\quad=\left(1-e^{-1/t}\right)^{2^\alpha}-\exp\left(-\left[2\left\{\log t+O\left(t^{-1}\right)\right\}^{1/\alpha}+\left\{\delta\log t+O\left(t^{-2\delta}\right)\right\}^{1/\alpha}\right]^\alpha\right)\\
&\quad=t^{-2^\alpha}+O\left(t^{-1-2^\alpha}\right)-\exp\left[-\left(2+\delta^{1/\alpha}\right)^\alpha\log t+O\left\{(t\log t)^{-1}\right\}\right]\sim t^{-2^\alpha},
\end{aligned}$$

i.e., $\tau_{1,2}(\delta)=2^{-\alpha}$, $\delta>0$. For $\delta=0$, using similar arguments as for the $|C|=1$ case, we have

$$\begin{aligned}
&\Pr\left(X_1^*>t,X_2^*>t,X_3^*=0\right)=\Pr\left(X_1>t,X_2>t,X_3<-1/\log p\right)\\
&\quad=\Pr\left(X_1>t,X_2>t\right)-\Pr\left(X_1>t,X_2>t,X_3>-1/\log p\right)\\
&\quad=\exp\left[-2^\alpha\left\{\log t+O\left(t^{-1}\right)\right\}\right]\\
&\quad\quad-\exp\left(-\left[2\left\{\log t+O\left(t^{-1}\right)\right\}^{1/\alpha}+\left\{-\log(1-p)\right\}^{1/\alpha}\right]^\alpha\right)\\
&\quad=\exp\left[-2^\alpha\left\{\log t+O\left(t^{-1}\right)\right\}\right]\\
&\quad\quad\cdot\left(1-\exp\left[-\alpha2^{\alpha-1}\frac{\{-\log(1-p)\}^{1/\alpha}}{\{\log t+O\left(t^{-1}\right)\}^{-1+1/\alpha}}+o\left\{(\log t)^{1-2/\alpha}\right\}\right]\right)\\
&\quad\sim\frac{t^{-2^\alpha}}{(\log t)^{-1+1/\alpha}}\alpha2^{\alpha-1}\{-\log(1-p)\}^{1/\alpha},
\end{aligned}$$

as $t \to \infty$. As such, the index of regular variation is $\tau_{1,2}(\delta) = 2^{-\alpha}$ for all $\delta \in [0,1]$. By analogous arguments, we also have $\tau_{1,3}(\delta) = \tau_{2,3}(\delta) = 2^{-\alpha}$, $\delta \in [0,1]$.

$|C| = 3$**:** $\tau_{1,2,3}$**.** To calculate the index of regular variation for sub-cone $\mathbb{E}_{1,2,3}$, we consider

$$\begin{aligned}\Pr(X_1 > t, X_2 > t, X_3 > t) &= \exp\left\{\log\left(1 - e^{-1/t}\right) V(1,1,1)\right\} \\ &= \left(1 - e^{-1/t}\right)^{V(1,1,1)} = \left\{1 - 1 + t^{-1} + O(t^{-2})\right\}^{3^\alpha} \sim t^{-3^\alpha},\end{aligned}$$

so $\tau_{1,2,3} = 3^{-\alpha}$. This corresponds to the known value of $\eta_{1,2,3}$ for the trivariate inverted logistic distribution, as $\eta_{1,2,3} = V(1,1,1)^{-1} = 3^{-\alpha}$.

A.2.6 Multivariate Gaussian distribution

The multivariate Gaussian provides a further example of a distribution which asymptotically places all mass on sub-cones $\mathbb{E}_C$ with $|C| = 1$. In the case where $d = 3$, for a multivariate Gaussian distribution with covariance matrix Σ, Nolde (2014), for example, shows that

$$\eta_{1,2,3} = \left(\mathbf{1}_3^T \Sigma^{-1} \mathbf{1}_3\right)^{-1}; \quad \eta_{i,j} = \left(\mathbf{1}_2^T \Sigma_{i,j}^{-1} \mathbf{1}_2\right)^{-1}, \quad i < j \in \{1,2,3\},$$

where $\Sigma_{i,j}$ is the submatrix of Σ corresponding to variables i and j, and $\mathbf{1}_d \in \mathbb{R}^d$ is a vector of 1s.

The covariance matrix Σ may be written as

$$\Sigma = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix} = \begin{bmatrix} \Sigma_{12} & B \\ B^T & 1 \end{bmatrix}, \text{ where } \Sigma_{12} = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} \rho_{13} \\ \rho_{23} \end{bmatrix}.$$

We note that since Σ and Σ_{12} are covariance matrices, they must be positive definite, with $\det(\Sigma) = 1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23} > 0$ and $\det(\Sigma_{12}) = 1 - \rho_{12}^2 > 0$.

The inverse of Σ is given by the block matrix

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{12}^{-1} + \Sigma_{12}^{-1}B(1 - B^T\Sigma_{12}^{-1}B)^{-1}B^T\Sigma_{12}^{-1} & -\Sigma_{12}^{-1}B(1 - B^T\Sigma_{12}^{-1}B)^{-1} \\ -(1 - B^T\Sigma_{12}^{-1}B)^{-1}B^T\Sigma_{12}^{-1} & (1 - B^T\Sigma_{12}^{-1}B)^{-1} \end{bmatrix},$$

so that

$$\begin{aligned}
&\mathbf{1}_3^T\Sigma^{-1}\mathbf{1}_3 \\
&\quad = \mathbf{1}_2^T\Sigma_{12}^{-1}\mathbf{1}_2 + (1 - B^T\Sigma_{12}^{-1}B)^{-1}\left(1 - \mathbf{1}_2^T\Sigma_{12}^{-1}B - B^T\Sigma_{12}^{-1}\mathbf{1}_2 + \mathbf{1}_2^T\Sigma_{12}^{-1}BB^T\Sigma_{12}^{-1}\mathbf{1}_2\right) \\
&\quad = \mathbf{1}_2^T\Sigma_{12}^{-1}\mathbf{1}_2 + \frac{1 - \rho_{12}^2}{1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}}\left(1 - 2\mathbf{1}_2^T\Sigma_{12}^{-1}B + \mathbf{1}_2^T\Sigma_{12}^{-1}BB^T\Sigma_{12}^{-1}\mathbf{1}_2\right) \\
&\quad = \mathbf{1}_2^T\Sigma_{12}^{-1}\mathbf{1}_2 + \frac{\det(\Sigma_{12})}{\det(\Sigma)}\left(1 - \mathbf{1}_2^T\Sigma_{12}^{-1}B\right)^2 \\
&\quad = \mathbf{1}_2^T\Sigma_{12}^{-1}\mathbf{1}_2 + \frac{\det(\Sigma_{12})}{\det(\Sigma)}\left(1 - \frac{\rho_{13} + \rho_{23}}{1 + \rho_{12}}\right)^2 \geq \mathbf{1}_2^T\Sigma_{12}^{-1}\mathbf{1}_2,
\end{aligned}$$

with equality if and only if $1 + \rho_{12} = \rho_{13} + \rho_{23}$. By similar calculations,

$$\mathbf{1}_3^T\Sigma^{-1}\mathbf{1}_3 \geq \mathbf{1}_2^T\Sigma_{i,j}^{-1}\mathbf{1}_2, \quad i < j \in \{1, 2, 3\},$$

with equality if and only if $1 + \rho_{ij} = \rho_{ik} + \rho_{jk}$, in which case $\eta_{1,2,3} = \eta_{i,j}$. Applying Theorem 2, for this trivariate case, if $1 + \rho_C \neq \sum_{C':|C'|=2, C' \neq C} \rho_{C'}$ for all $C \subset \{1, 2, 3\}$ with $|C| = 2$, then $\tau_C(1) = \eta_C$ for any set $C \subseteq \{1, 2, 3\}$ with $|C| \geq 2$. Since $\tau_C(\delta)$ is non-decreasing in δ, $\tau_C(\delta) \leq \eta_C$ for $\delta \in [0, 1)$. We also know $\tau_{1,2,3} = \eta_{1,2,3}$.

In this case, calculation of the explicit formulas for $\delta < 1$, in a manner similar to the inverted logistic case, is complicated by the need to consider asymptotic approximations of Gaussian cumulative distribution functions beyond first order. As such, we do not attempt this here, but note that since $\tau_C(1) \leq \eta_C$, $|C| \geq 2$, we would be likely to estimate $\tau_C(\delta) < 1$ in practice.

To gain some insight into the remaining cases of $C = \{i\}$, $i = 1, 2, 3$, we consider the conditional extreme value model of Heffernan and Tawn (2004). Let $\boldsymbol{Y} = \log \boldsymbol{X}$, so that $\boldsymbol{Y}$ has standard Gumbel marginal distributions, and all correlations be positive.

Then conditioning on Y_i gives

$$\Pr\left(Y_j - \rho_{ij}^2 t \leq t^{1/2} z_j, Y_k - \rho_{ik}^2 t \leq t^{1/2} z_k, Y_i > t\right) \sim N(z_j, z_k)\Pr(Y_i > t), \quad t \to \infty,$$

for $N(z_j, z_k)$ denoting the distribution function of a particular Gaussian distribution (Heffernan and Tawn, 2004, Section 8). For $z_j = z_k = 0$, this equates to

$$\Pr\left(Y_j \leq \rho_{ij}^2 t, Y_k \leq \rho_{ik}^2 t, Y_i > t\right) \sim N(0,0)\Pr(Y_i > t), \quad t \to \infty. \tag{A.2.2}$$

In the trivariate case with Gumbel margins, equation (3.2.7) of Assumption 1 can be written as

$$\Pr\left(Y_j \leq \delta t, Y_k \leq \delta t, Y_i > t\right) \in \mathrm{RV}_{-1/\tau_C(\delta)}, \quad t \to \infty.$$

Considering equation (A.2.2) again, we see that if $\delta \geq \max\left(\rho_{ij}^2, \rho_{ik}^2\right)$, in place of the limiting Gaussian distribution, mass will occur at $(-\infty, -\infty)$ for variables (Y_j, Y_k), which implies that $\tau_i(\delta) = 1$. Alternatively, if $\rho_{ij}^2 \leq \delta < \rho_{ik}^2$ or $\rho_{ik}^2 \leq \delta < \rho_{ij}^2$, we have mass at $(-\infty, \infty)$ or $(\infty, -\infty)$, respectively, and for $\delta < \min\left(\rho_{ij}^2, \rho_{ik}^2\right)$ mass occurs at (∞, ∞). In these cases, the left-hand side of equation (A.2.2) is $o\left\{\Pr\left(Y_i > t\right)\right\}$, which is consistent with $\tau_i(\delta) < 1$.

A.3 Simulation study

A.3.1 Estimation of $\tau_C(\delta)$ in Method 2

In Method 2, introduced in Section 3.3.3, we consider regions of the form

$$\tilde{E}_C = \left\{\boldsymbol{x} \in \mathbb{E} : x_\vee^{D\setminus C} \leq \left(x_\wedge^C\right)^\delta\right\}, \ |C| < d; \quad \tilde{E}_D = \mathbb{E} \setminus \bigcup_{C \in 2^D \setminus \emptyset : |C| < d} \tilde{E}_C,$$

for $C \in 2^D \setminus \emptyset$, and assume that $\Pr\left(X_\wedge^C > q, \boldsymbol{X} \in \tilde{E}_C\right) \in \mathrm{RV}_{-1/\tilde{\tau}_C(\delta)}$, for $X_\wedge^C = \min_{i \in C} X_i$. This is used as an alternative to considering $\Pr\left(X_\wedge^C > t, X_\vee^{D\setminus C} \leq t^\delta\right) \in \mathrm{RV}_{-1/\tau_C(\delta)}$, for which there is no clear structure variable. In Figure A.3.1, we demonstrate how well the parameter $\tau_C(\delta)$ is approximated using Method 2. We consider the trivariate logistic distribution, with theoretical $\tau_C(\delta)$ values given in case (iii)

of Table 3.2.1. For $\alpha = 0.25$ and $\alpha = 0.5$, we take samples of size 100,000 from this distribution, and use Method 2 to estimate $\tilde{\tau}_1(\delta)$, $\tilde{\tau}_{1,2}(\delta)$ and $\tilde{\tau}_{1,2,3}$ for values of $\delta \in \{0.1, \ldots, 0.95\}$. The thresholds used correspond to the 0.985 quantile of observed $X_\wedge^C$ values in each region $\tilde{E}_C$. Each simulation is repeated 100 times, and the true $\tau_C(\delta)$ parameter values are shown in red. The results indicate that the estimator derived from considering $\Pr\left\{X_\wedge^C > t, X_\vee^{D\setminus C} \leq \left(X_\wedge^C\right)^\delta\right\}$ yields $\tau_C(\delta)$ as defined through $\Pr\left(X_\wedge^C > t, X_\vee^{D\setminus C} \leq t^\delta\right)$. We observe increased variability in the estimates of $\tilde{\tau}_1(\delta)$ and $\tilde{\tau}_{1,2}(\delta)$ for small values of δ, and $\tilde{\tau}_{1,2,3}$ for large values of δ, which is likely due to the limited data in the corresponding regions $\tilde{E}_C$ for these cases. There is also some bias present in these results, which is an issue associated with the Hill estimator. While alternative estimators are available, each will have issues with bias or variance (Beirlant et al., 2004), and we continue with the Hill estimator.

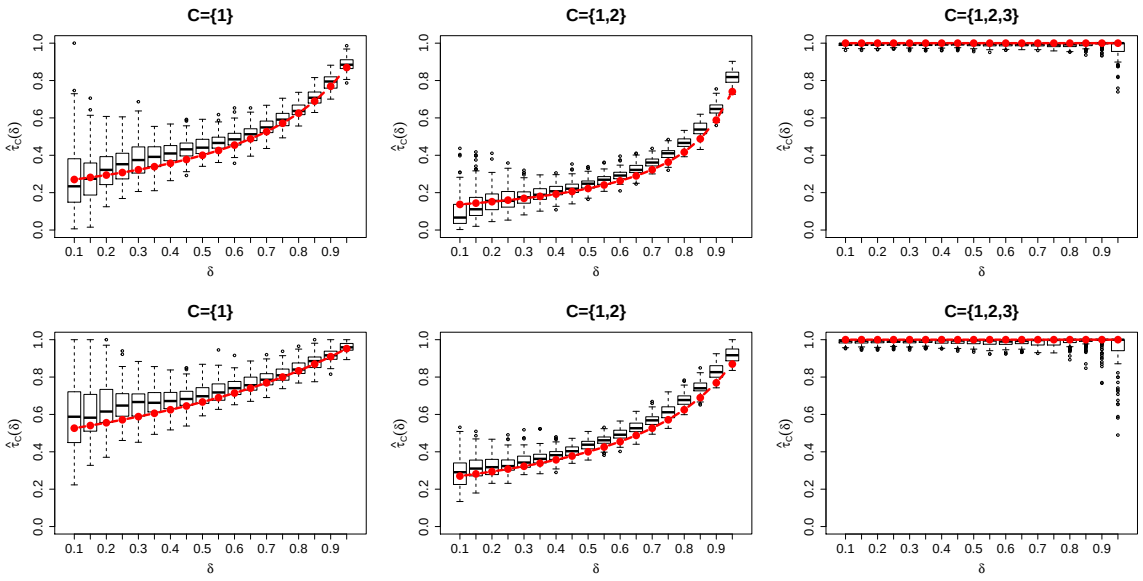

Figure A.3.1: Estimates of $\tau_1(\delta)$, $\tau_{1,2}(\delta)$ and $\tau_{1,2,3}$ for data simulated from trivariate logistic distributions with $\alpha = 0.25$ (top) and $\alpha = 0.5$ (bottom).

A.3.2 Area under the receiver operating characteristic curve results for the max-mixture distribution

In Tables 3.4.1 and 3.4.2 of Chapter 3, we present the average area under the (neighboured) receiver operating characteristic curve for Method 1, Method 2 and the approach of Goix et al. applied to simulations from a particular max-mixture distribution. Figures A.3.2 and A.3.3 provide boxplots of all the values from these simulations.

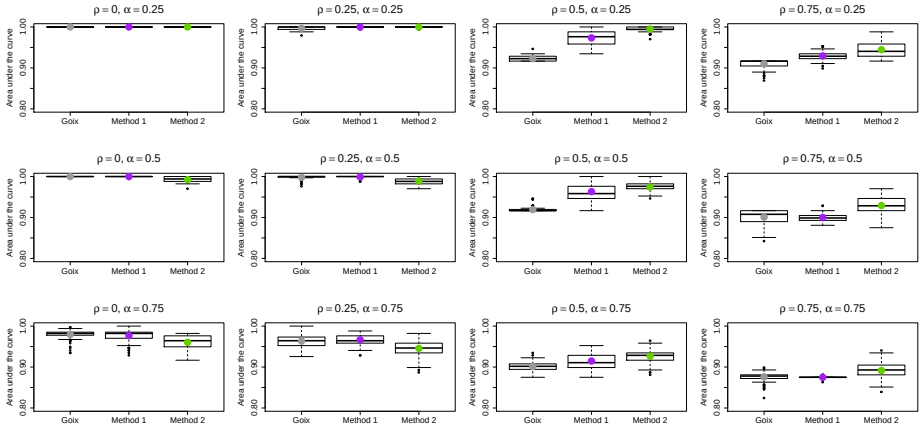

Figure A.3.2: Area under the receiver operating characteristic curve results for 100 simulations from a five-dimensional max-mixture distribution.

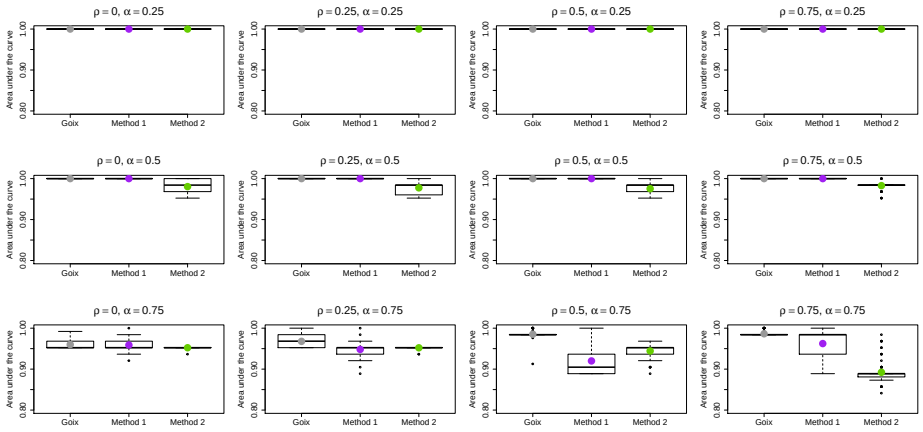

Figure A.3.3: Area under the neighboured receiver operating characteristic curve results for 100 simulations from a five-dimensional max-mixture distribution.

A.3.3 Asymmetric logistic distribution

We now present simulation results for the asymmetric logistic distribution, using the same metrics as for the max-mixture distribution in Section 3.4. This model belongs to the class of multivariate extreme value distributions, and it is possible to calculate the proportion of extremal mass associated with each sub-cone.

In standard Fréchet margins, the multivariate extreme value distribution function is of the form $\exp\{-V(\boldsymbol{x})\}$. Coles and Tawn (1991) show that the spectral density corresponding to sub-cone $\mathbb{E}_C$ is $h_C(\boldsymbol{w}_C) = -V^{\{C\}}(\boldsymbol{w}_C)/d$, where

$$V^{\{C\}}(\boldsymbol{x}_C) = \lim_{x_j \to 0: j \in D \setminus C} \left(\prod_{i \in C} \frac{\partial}{\partial x_i} \right) V(\boldsymbol{x}), \qquad \boldsymbol{x}_C = \{x_i : i \in C\},$$

for $\boldsymbol{w}_C = \boldsymbol{x}_C / r_C$ and $r_C = \sum_{i \in C} x_i$. Hence, for $\mathbb{B}_C = \{\boldsymbol{w} \in \mathcal{S}_{d-1} : w_i \in (0,1], i \in C;\ w_j = 0, j \in D \setminus C\}$, the proportion of mass on corresponding sub-cone $\mathbb{E}_C$ is

$$p_C = -\frac{1}{d} \int_{\mathbb{B}_C} V^{\{C\}}(\boldsymbol{w}_C) \prod_{i \in C} dw_i. \tag{A.3.1}$$

For the asymmetric logistic model (Tawn, 1990), the exponent measure V is defined as

$$V(\boldsymbol{x}) = \sum_{C \in 2^D \setminus \emptyset} \left\{ \sum_{i \in C} (\theta_{i,C}/x_i)^{1/\alpha_C} \right\}^{\alpha_C}, \qquad \theta_{i,C} \in [0,1], \tag{A.3.2}$$

with $\theta_{i,C} = 0$ if $i \notin C$, $\sum_{C \in 2^D \setminus \emptyset} \theta_{i,C} = 1$ for all $i = 1, \ldots, d$ and $C \in 2^D \setminus \emptyset$, and dependence parameters $\alpha_C \in (0,1]$. In Proposition 1 of Section A.4, we show that for the d-dimensional asymmetric logistic model with all $\alpha_C \equiv \alpha$, the proportion of mass on sub-cone $\mathbb{E}_C$ is

$$p_C^{(d)} = \sum_{i \in C} \theta_{i,C}/d, \qquad C \in 2^D \setminus \emptyset.$$

Using this new result, we can compare our estimated proportions to the truth using the Hellinger distance defined in equation (3.4.1) of Chapter 3.

Following Goix et al. (2017), we simulate data from an asymmetric logistic distribution with $\alpha_C \equiv \alpha$, whose extremal mass is concentrated on f randomly chosen sub-cones, ensuring that moment constraint (3.1.3) is satisfied. Suppose the sub-cones chosen correspond to subsets $F_1, \ldots, F_f \in 2^D \setminus \emptyset$. The conditions on the parameters of the asymmetric logistic distribution are satisfied by setting

$$\theta_{i,C} = |\{j : i \in F_j, j \in \{1, \ldots, f\}\}|^{-1}, \qquad C \in \{F_1, \ldots, F_f\}, \tag{A.3.3}$$

and $\theta_{i,C} = 0$ otherwise. We present results for dimensions $d = 5$ and $d = 10$. For $d = 5$, we simulate samples of size $n = 10,000$, and test both our methods when there are truly 5, 10 and 15 sub-cones with extremal mass. For $d = 10$, we have $n = 100,000$ samples, and consider 10, 50 and 100 sub-cones with extremal mass. We set the tuning parameters as in Section A.3.3, and repeat each setting 100 times. In Table A.3.1, we present the average area under the receiver operating characteristic curve for $\alpha \in \{0.25, 0.5, 0.75\}$. Boxplots of the full results obtained are provided in Figures A.3.4 and A.3.5. In the asymmetric logistic model, the closer α_C is to 1, the larger the values of $\tau_{\underline{C}}(\delta)$ for any fixed δ and $\underline{C} \subset C$, as demonstrated by cases (ii) and (iii) in Table 3.2.1. Thus, the larger the value of α in our simulations, the more difficult it is to determine which sub-cones truly contain extremal mass. In Figure A.3.6, we present average values for the area under the neighboured receiver operating characteristic curve, as defined in Section 3.4.1, for the same five-dimensional cases as above, but only consider $\alpha = 0.75$, where classification is most difficult. If every sub-cone without mass has at least one neighbouring sub-cone truly having mass, the false positive rate, and hence the area under the curve, will always be zero; these cases have been removed in Figure A.3.6.

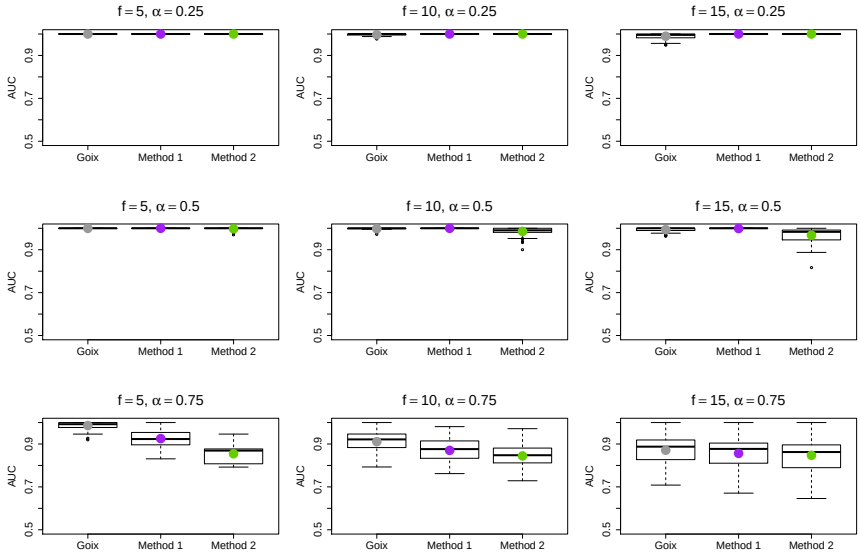

Figure A.3.4: Areas under the receiver operating characteristic curves for 100 simulations from a five-dimensional asymmetric logistic distribution.

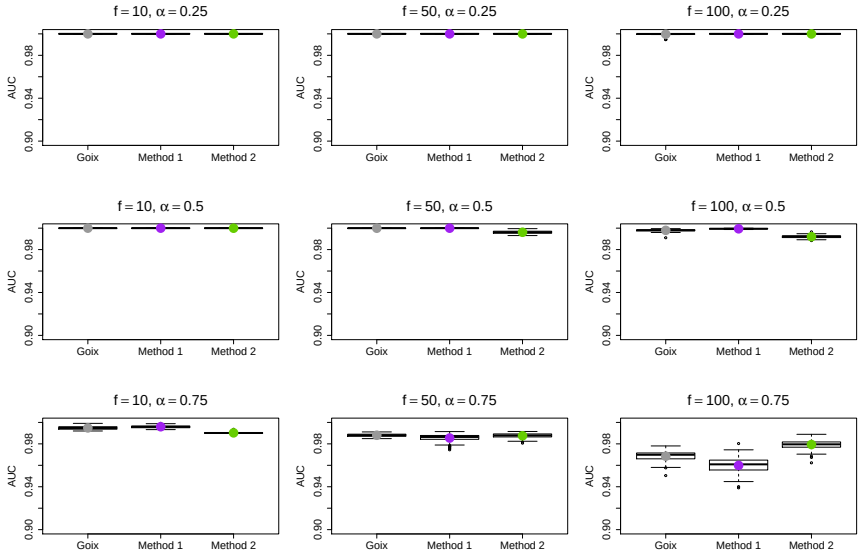

Figure A.3.5: Areas under the receiver operating characteristic curves for 100 simulations from a ten-dimensional asymmetric logistic distribution.

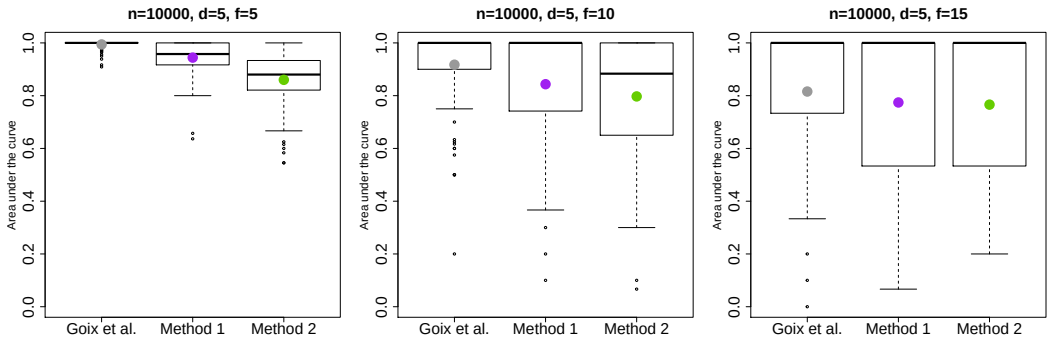

Figure A.3.6: Boxplots of the areas under the neighboured receiver operating characteristic curves for $d = 5$; $f = 5, 10, 15$; $n = 10,000$ and $\alpha = 0.75$. The average values are shown by the circles in each plot.

(α, f)	$(0.25, 5)$	$(0.25, 10)$	$(0.25, 15)$	$(0.5, 5)$	$(0.5, 10)$	$(0.5, 15)$	$(0.75, 5)$	$(0.75, 10)$	$(0.75, 15)$
Goix et al.	100 (0.1)	99.7 (0.5)	99.0 (1.3)	100 (0.2)	99.8 (0.5)	99.4 (0.9)	98.6 (1.7)	91.1 (4.9)	87.1 (7.1)
Method 1	100 (0.0)	100 (0.0)	100 (0.0)	100 (0.0)	100 (0.1)	99.9 (0.3)	92.5 (4.1)	87.0 (5.5)	85.6 (7.3)
Method 2	100 (0.0)	100 (0.0)	100 (0.0)	99.7 (0.6)	98.5 (1.8)	96.8 (3.2)	85.4 (4.3)	84.4 (5.7)	84.7 (7.5)
(α, f)	$(0.25, 10)$	$(0.25, 50)$	$(0.25, 100)$	$(0.5, 10)$	$(0.5, 50)$	$(0.5, 100)$	$(0.75, 10)$	$(0.75, 50)$	$(0.75, 100)$
Goix et al.	100 (0.0)	100 (0.0)	100 (0.1)	100 (0.0)	100 (0.0)	99.8 (0.1)	99.5 (0.2)	98.8 (0.1)	96.9 (0.5)
Method 1	100 (0.0)	100 (0.0)	100 (0.0)	100 (0.0)	100 (0.0)	99.9 (0.0)	99.6 (0.1)	98.5 (0.4)	96.0 (0.8)
Method 2	100 (0.0)	100 (0.0)	100 (0.0)	100 (0.0)	99.6 (0.1)	99.2 (0.1)	99.0 (0.1)	98.8 (0.2)	97.9 (0.4)

Table A.3.1: Average areas under the receiver operating characteristic curves, given as percentages, for 100 samples from five-dimensional (top) and ten-dimensional (bottom) asymmetric logistic distributions, with dependence parameter α and $\theta_{i,C}$ determined via (A.3.3). Standard deviations of these results are given in brackets.

The average areas under the receiver operating characteristic curves in Table A.3.1 show that all three methods perform well when $\alpha = 0.25$ and $\alpha = 0.5$, for both $d = 5$ and $d = 10$, with Method 1 slightly outperforming the other two approaches. The results suggest that the method of Goix et al. (2017) is generally the most successful classifier when $\alpha = 0.75$, followed by Method 1, although the most substantial difference in results occurs for $(d, f, \alpha) = (10, 100, 0.75)$, where Method 2 is most suc-

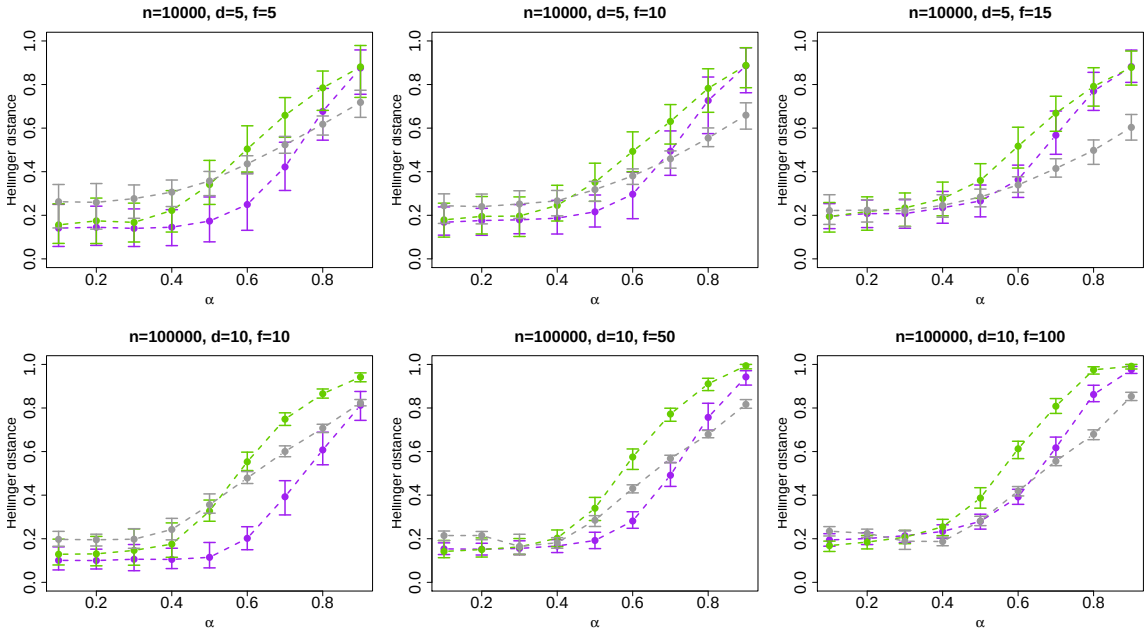

Figure A.3.7: Mean Hellinger distance, 0.05 and 0.95 quantiles over 100 simulations. Method 1: purple; Method 2: green; Goix et al.: grey.

cessful; this is supported by the boxplots of results in Figures A.3.4 and A.3.5. In principle, Method 1 should be better than Method 2 here, so greater assessment of tuning parameters may be required. It is possible that the method of Goix et al. is most successful for larger values of α since it is more difficult for Methods 1 and 2 to distinguish between regions where $\tau_C(\delta)$ does and does not equal 1 in these cases. For the neighboured receiver operating characteristic curve results in Figure A.3.6, Method 1 outperforms Method 2 in all cases, with the method of Goix et al. performing the best under this criterion; this is similar to the results observed for the max-mixture distribution.

Figure A.3.7 shows the average Hellinger distance for $\alpha \in [0.1, 0.9]$ for each of the cases described above. For the most sparse cases, $(d, f) = (5, 5)$ and $(d, f) = (10, 10)$, Method 1 performs significantly better than the other two approaches overall. For the less sparse cases, $(d, f) = (5, 15)$ and $(d, f) = (10, 100)$, the three methods give similar

results in terms of the Hellinger distance for $\alpha \leq 0.5$, but the method of Goix et al. is most successful for larger α values. When the extreme values are concentrated on fewer sub-cones, it may be easier to estimate true values of $\tau_C(\delta) = 1$ using Method 1 than in less sparse examples. For the asymmetric logistic distribution, Method 1 often performs better in terms of estimating the proportion of extremal mass on each sub-cone, while the method of Goix et al. (2017) is better at classification, although this method does tend to place mass on too many sub-cones, as shown in Figure 3.4.2 of Chapter 3.

A.4 Calculating the mass on each sub-cone for an asymmetric logistic model

Proposition 1. *For the d-dimensional asymmetric logistic model with exponent measure* (A.3.2) *and* $\alpha_C \equiv \alpha \in (0,1)$ *for all* $C \in 2^D \setminus \emptyset$,

$$dp_C^{(d)} = \sum_{i \in C} \theta_{i,C},$$

where $p_C^{(d)}$ *denotes the proportion of mass on sub-cone* $\mathbb{E}_C$.

Proof. Consider the exponent measure of the asymmetric logistic model $V(\boldsymbol{x})$ as a sum of functions $V_C(\boldsymbol{x}_C)$, for $C \in 2^D \setminus \emptyset$, i.e.,

$$V(\boldsymbol{x}) = \sum_{C \in 2^D \setminus \emptyset} V_C(\boldsymbol{x}_C), \qquad V_C(\boldsymbol{x}_C) = \left\{ \sum_{i \in C} \left(\frac{\theta_{i,C}}{x_i} \right)^{1/\alpha} \right\}^{\alpha}.$$

Then for any dimension $d \geq |C|$,

$$\begin{aligned} V^{\{C\}}(\boldsymbol{x}_C) &= \left(\prod_{i \in C} \frac{\partial}{\partial x_i} \right) V_C(\boldsymbol{x}_C) \\ &= \left\{ \prod_{i=0}^{|C|-1} -\left(\frac{\alpha - i}{\alpha} \right) \right\} \left(\prod_{i \in C} \frac{\theta_{i,C}^{1/\alpha}}{x_{C,i}^{1+1/\alpha}} \right) \left\{ \sum_{i \in C} \left(\frac{\theta_{i,C}}{x_{C,i}} \right)^{1/\alpha} \right\}^{\alpha - |C|}, \end{aligned}$$

since for $\overline{C} \supset C$, $\lim_{x_i \to 0: i \in \overline{C} \setminus C} \left(\prod_{j \in C} \frac{\partial}{\partial x_j} \right) V_{\overline{C}}(\boldsymbol{x}_{\overline{C}}) = 0$. Hence, by result (A.3.1),

$$dp_C^{(d)} = -\int_{\mathbb{B}_C} V^{\{C\}}(\boldsymbol{w}_C) \prod_{i \in C} dw_i, \tag{A.4.1}$$

which we note does not depend on d. We claim that

$$-\int_{\mathbb{B}_C} V^{\{C\}}(\boldsymbol{w}_C) \prod_{i \in C} dw_i = \sum_{i \in C} \theta_{i,C}. \tag{A.4.2}$$

First consider $|C| = 1$, i.e. $\mathbb{B}_C = \{\boldsymbol{w} : w_i = 1\}$ for $C = \{i\}$. Here,

$$V^{\{i\}}(x_i) = \frac{\partial}{\partial x_i} V_i(x_i) = -\frac{\theta_{i,i}}{x_i^2},$$

so

$$dp_i^{(d)} = \frac{\theta_{i,i}}{w_i^2}\Big|_{w_i=1} = \theta_{i,i}, \qquad i = 1, \ldots, d.$$

Now consider $|C| = 2$. We have

$$V^{\{i,j\}}(x_i, x_j) = \left(\frac{\alpha - 1}{\alpha}\right) \frac{\left\{(1-\theta_{i,i})(1-\theta_{j,j})\right\}^{1/\alpha}}{(x_i x_j)^{1+1/\alpha}} \left\{ \left(\frac{\theta_{i,ij}}{x_i}\right)^{1/\alpha} + \left(\frac{\theta_{j,ij}}{x_j}\right)^{1/\alpha} \right\}^{\alpha-2},$$

so

$$h_{i,j}(w_i) = \left(\frac{1 - \alpha}{\alpha}\right) \frac{\left(\theta_{i,ij}\theta_{j,ij}\right)^{1/\alpha}}{\left\{w_i(1-w_i)\right\}^{1+1/\alpha}} \left\{ \left(\frac{\theta_{i,ij}}{w_i}\right)^{1/\alpha} + \left(\frac{\theta_{j,ij}}{1-w_i}\right)^{1/\alpha} \right\}^{\alpha-2},$$

and

$$dp_{i,j}^{(d)} = \int_0^1 h_{i,j}(w_i) dw_i. \tag{A.4.3}$$

However, taking $d = 2$, we know $2p_{1,2}^{(2)} + 2p_1^{(2)} + 2p_2^{(2)} = 2$, so $dp_{1,2}^{(d)} = \theta_{1,12} + \theta_{2,12}$, and similarly, by (A.4.3), $\int_0^1 h_{i,j}(w_i) dw_i = dp_{i,j}^{(d)} = \theta_{i,ij} + \theta_{j,ij}$. So, (A.4.1) holds for $|C| = 1$ and $|C| = 2$, and we suppose it holds for $|C| \leq k$, i.e. $dp_C^{(d)} = -\int_{\mathbb{B}_C} V^{\{C\}}(\boldsymbol{w}_C) \prod_{i \in C} dw_i = \sum_{i \in C} \theta_{i,C}$.

Since (A.4.1) does not depend on d, take $d = k + 1$. So for all $\underline{C}$ with $|\underline{C}| \leq k$, $(k+1)p_{\underline{C}}^{(k+1)} = \sum_{i \in \underline{C}} \theta_{i,\underline{C}}$. Now take $C = \{1, \ldots, k+1\}$. Then,

$$(k+1)p_C^{(k+1)} = (k+1) - \sum_{\underline{C} \subset C} \sum_{i \in \underline{C}} \theta_{i,\underline{C}} = (k+1) - \sum_{i \in C} \sum_{\underline{C} \subset C} \theta_{i,\underline{C}}$$

$$= \sum_{i \in C} \left(1 - \sum_{\underline{C} \subset C} \theta_{i,\underline{C}} \right)$$
$$= \sum_{i \in C} \theta_{i,C}.$$

As such, (A.4.2) holds by induction. □

Appendix B

Supplementary Material for Chapter 4

B.1 Properties of weights

In this section, we show that the desired properties of weights from Section 4.2.2 are satisfied by the definition in Section 4.2.5. For property 1, first note that the angular components satisfy $\sum_{i \in C} w_{j,i} \in [0,1]$, for all $C \in 2^D \backslash \emptyset$, which implies that

$$\left(\sum_{i \in C} w_{j,i} \right)^k \in [0,1],$$

for all $k > 0$ and $C \in 2^D \backslash \emptyset$. Moreover, for $\underline{C} \subset C$, we have

$$\sum_{i \in \underline{C}} w_{j,i} \leq \sum_{i \in C} w_{j,i} \quad \Rightarrow \quad \left(\sum_{i \in \underline{C}} w_{j,i} \right)^k \leq \left(\sum_{i \in C} w_{j,i} \right)^k,$$

for $k > 0$. Recall that the set $S^{(s)}$ with $|S^{(s)}| = s$ is the set such that $\boldsymbol{w}$ lies closer to the region $\mathbb{B}_{S^{(s)}}$ than any other set $\mathbb{B}_{S^{(s)'}}$ with $|S^{(s)'}| = s$. By the above inequality, we have

$$\delta_{S^{(s)},j} = \left(\sum_{i \in S^{(s)}} w_{j,i} \right)^k - \left(\sum_{i \in S^{(s-1)}} w_{j,i} \right)^k \in [0,1],$$

for all $s = 2, \ldots, d$. As such, $\delta_{C,j} \in [0, 1]$ for all $C \in 2^D \backslash \emptyset$. To show that the proposed weighting satisfies property 2, we have

$$\begin{aligned}\sum_{C \in 2^D \backslash \emptyset} \delta_{C,j} &= \sum_{s=1}^{d} \delta_{S^{(s)},j} \\ &= \Big(\sum_{i \in S^{(1)}} w_{j,i} \Big)^k + \sum_{s=2}^{d} \Big\{ \Big(\sum_{i \in S^{(s)}} w_{j,i} \Big)^k - \Big(\sum_{i \in S^{(s-1)}} w_{j,i} \Big)^k \Big\} \\ &= \Big(\sum_{i \in S^{(d)}} w_{j,i} \Big)^k = 1,\end{aligned}$$

noting that $S^{(d)} = D$.

B.2 Max-mixture AUC and AUC* results

In Table 4.3.1, we presented average area under the (neighboured) receiver operating characteristic curve values for simulations based on a five-dimensional max-mixture distribution. We present boxplots of these results in Figures B.2.1 and B.2.2.

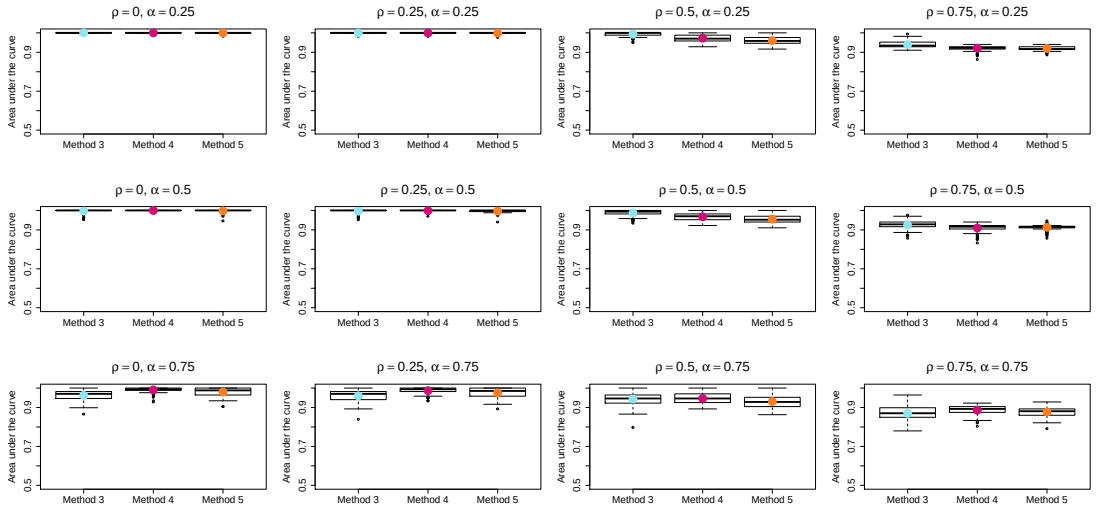

Figure B.2.1: AUC values for Methods 3, 4 and 5 over 100 simulations from a five-dimensional max-mixture distribution.

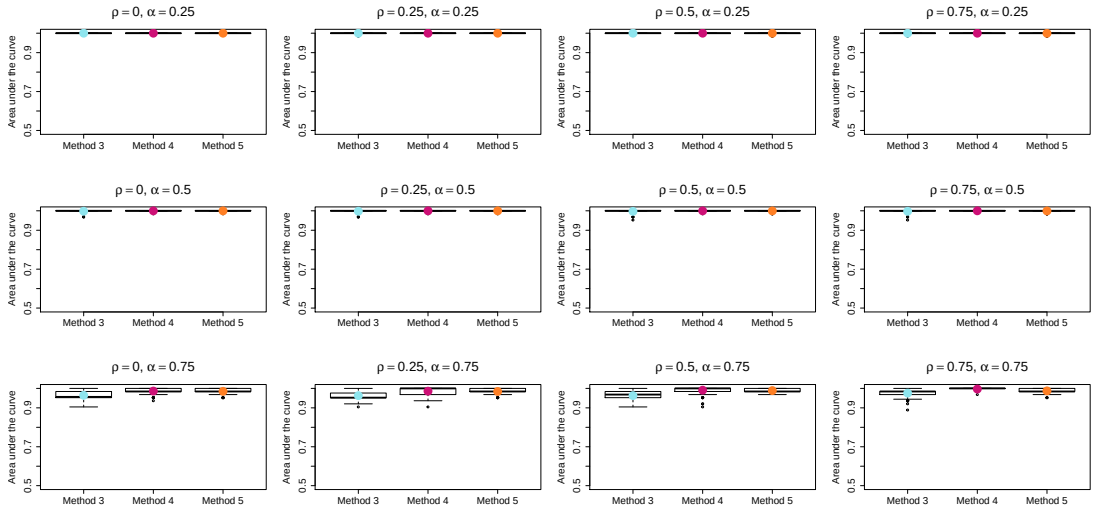

Figure B.2.2: AUC* values for Methods 3, 4 and 5 over 100 simulations from a five-dimensional max-mixture distribution.

B.3 Asymmetric logistic AUC results

In Table 4.3.2, we presented average AUC values for simulations based on asymmetric logistic models of five or ten dimensions. We present boxplots of these results in Figures B.3.1 and B.3.2.

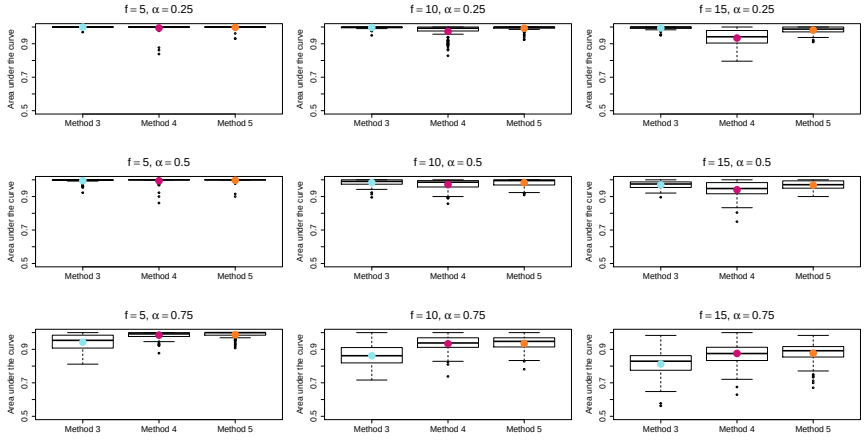

Figure B.3.1: AUC results for Methods 3, 4 and 5 over 100 simulations from a five-dimensional asymmetric logistic distribution.

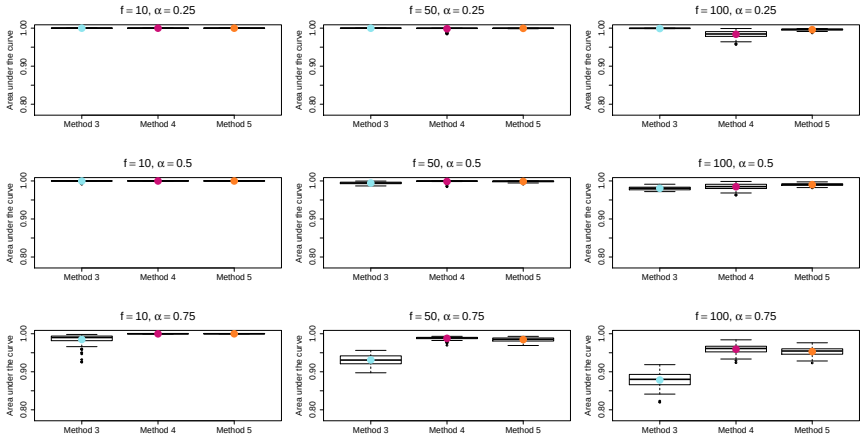

Figure B.3.2: AUC results for Methods 3, 4 and 5 over 100 simulations from a ten-dimensional asymmetric logistic distribution.

Appendix C

Supplementary Material for Chapter 5

C.1 Proof of Theorem 3

C.1.1 Identifying sub-vines of D-vines to construct the gauge function

A D-vine is represented graphically by a series of $d-1$ trees, labelled $T_1, \ldots, T_{d-1}$. Each of these trees is a path, and we suppose that the nodes are labelled in ascending order, as in the left plot of Figure 5.3.4 for the case where $d = 4$. Moving from a D-vine of dimension $d \geq 4$ to one of dimension $d+1$ involves first adding an extra node and edge onto each tree in the graph. In tree T_1, the extra node has label $d+1$, and the extra edge label is $\{d, d+1\}$. In tree T_2 the extra node is labelled $\{d, d+1\}$ and the edge is labelled $\{d-1, d+1\}|d$, and this continues until we reach tree T_{d-1}, where the extra node is labelled $\{3, d+1\}|\{4, \ldots, d\}$ and the corresponding edge is labelled $\{2, d+1\}|\{3, \ldots, d\}$. We finally must also add the tree T_d, with nodes labelled $\{1, d\}|\{2, \ldots, d-1\}$ and $\{2, d+1\}|\{3, \ldots, d\}$, and corresponding edge label $\{1, d+1\}|\{2, \ldots, d\}$. This is demonstrated in Figure C.1.1, for an example where we

move from a D-vine of dimension four to one of dimension five.

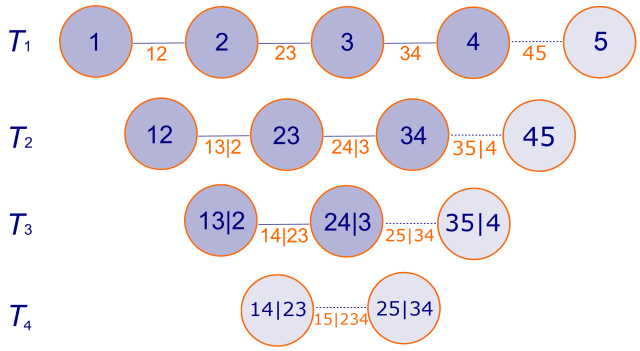

Figure C.1.1: Example of the extending a four-dimensional D-vine to a five-dimensional D-vine.

Due to this iterative construction, we can consider a d-dimensional D-vine in terms of three lower dimensional sub-vines in a similar way to in Figure 5.3.5 for the $d = 4$ case. In particular, in trees $T_1, \ldots, T_{d-2}$, we have two sub-vines of dimension $d-1$; the first corresponds to variables with labels in $\{1, \ldots, d-1\} = D\backslash\{d\}$, and the second to variables with labels in $\{2, \ldots, d\} = D\backslash\{1\}$. In the graph, these two sub-vines will overlap in the region corresponding to a further sub-vine, this time of dimension $d-2$ and corresponding to variables with labels in $\{2, \ldots, d-1\} = D\backslash\{1, d\}$.

In order to calculate the gauge function, we consider the behaviour of $-\log f(t\boldsymbol{x})$, as $t \to \infty$. By considering these three sub-vines, we see that this can be written as

$$\begin{aligned} -\log f(t\boldsymbol{x}) = &-\log f_{D\backslash\{d\}}\left(t\boldsymbol{x}_{-\{d\}}\right) - \log f_{D\backslash\{1\}}\left(t\boldsymbol{x}_{-\{1\}}\right) + \log f_{D\backslash\{1,d\}}\left(t\boldsymbol{x}_{-\{1,d\}}\right) \\ &- \log c_{\{1,d\}|D\backslash\{1,d\}}\left\{F_{1|D\backslash\{1,d\}}\left(tx_1|t\boldsymbol{x}_{-\{1,d\}}\right), F_{d|D\backslash\{1,d\}}\left(tx_d|t\boldsymbol{x}_{-\{1,d\}}\right)\right\}. \end{aligned}$$

Note that this is the form given for the $d = 4$ case in equation (5.3.12). We can therefore infer that the d-dimensional gauge function $g(\boldsymbol{x})$, defined as $-\log f(t\boldsymbol{x}) \sim tg(\boldsymbol{x})$ as $t \to \infty$, satisfies

$$g(\boldsymbol{x}) = g_{D\backslash\{d\}}\left(\boldsymbol{x}_{-\{d\}}\right) + g_{D\backslash\{1\}}\left(\boldsymbol{x}_{-\{1\}}\right) - g_{D\backslash\{1,d\}}\left(\boldsymbol{x}_{-\{1,d\}}\right) + \tilde{g}_D(\boldsymbol{x}), \qquad \text{(C.1.1)}$$

where, as $t \to \infty$,

$$-\log c_{\{1,d\}|D\setminus\{1,d\}}\left\{F_{1|D\setminus\{1,d\}}\left(tx_1|t\boldsymbol{x}_{-\{1,d\}}\right), F_{d|D\setminus\{1,d\}}\left(tx_d|t\boldsymbol{x}_{-\{1,d\}}\right)\right\} \sim t\tilde{g}_D(\boldsymbol{x}). \tag{C.1.2}$$

In Section C.1.2, we present some properties of inverted extreme value copulas that will be used in Section C.1.3 to find $\tilde{g}_D(\boldsymbol{x})$, and hence the form of the gauge function for a d-dimensional D-vine with inverted extreme value components.

C.1.2 Properties of inverted extreme value copulas

Lemma C.1.1. *For the density $c(u,v)$ of an inverted extreme value copula, of the form (5.3.2), if $u = 1 - a_1 e^{-b_1 t}\{1+o(1)\}$ and $v = 1 - a_2 e^{-b_2 t}\{1+o(1)\}$ for some $a_1, a_2, b_1, b_2 > 0$, then as $t \to \infty$,*

$$-\log c(u,v) \sim t\left\{-b_1 - b_2 + V\left(1/b_1, 1/b_2\right)\right\}.$$

Proof.

$$\begin{aligned}
-\log c(u,v) =& \log(1-u) + \log(1-v) + 2\log\{-\log(1-u)\} + 2\log\{-\log(1-v)\} \\
&+ V\left\{\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)}\right\} - \log\left[V_1\left\{\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)}\right\}\right. \\
&\qquad \left. \cdot V_2\left\{\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)}\right\} - V_{12}\left\{\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)}\right\}\right] \\
=& \log a_1 + \log a_2 - (b_1+b_2)t + 2\log\{1+o(1)\} \\
&+ 2\log[-\log a_1 + b_1 t - \log\{1+o(1)\}] + 2\log[-\log a_2 + b_2 t - \log\{1+o(1)\}] \\
&+ tV\left[\frac{1}{-t^{-1}\log a_1 + b_1 - t^{-1}\log\{1+o(1)\}}, \frac{1}{-t^{-1}\log a_2 + b_2 - t^{-1}\log\{1+o(1)\}}\right] \\
&- \log\left(t^4 V_1\left[\frac{1}{-t^{-1}\log a_1 + b_1 - t^{-1}\log\{1+o(1)\}}, \frac{1}{-t^{-1}\log a_2 + b_2 - t^{-1}\log\{1+o(1)\}}\right]\right. \\
&\qquad \cdot V_2\left[\frac{1}{-t^{-1}\log a_1 + b_1 - t^{-1}\log\{1+o(1)\}}, \frac{1}{-t^{-1}\log a_2 + b_2 - t^{-1}\log\{1+o(1)\}}\right] \\
&\left. - t^3 V_{12}\left[\frac{1}{-t^{-1}\log a_1 + b_1 - t^{-1}\log\{1+o(1)\}}, \frac{1}{-t^{-1}\log a_2 + b_2 - t^{-1}\log\{1+o(1)\}}\right]\right) \\
\sim& t\left\{-b_1 - b_2 + V\left(1/b_1, 1/b_2\right)\right\}.
\end{aligned}$$

□

Lemma C.1.2. *For the conditional distribution function $F(u \mid v)$ of an inverted extreme value copula, of the form (5.3.1), if $u = 1 - a_1 e^{-b_1 t}\{1 + o(1)\}$ and $v = 1 - a_2 e^{-b_2 t}\{1 + o(1)\}$ for some $a_1, a_2, b_1, b_2 > 0$, then as $t \to \infty$,*

$$F(u \mid v) = 1 - a \exp\left[-t\left\{V(1/b_1, 1/b_2) - b_2\right\}\right]\{1 + o(1)\},$$

for some $a > 0$.

Proof.

$$\begin{aligned}
F(u \mid v) &= 1 + \left(\frac{1}{1-v}\right)\{-\log(1-v)\}^{-2} V_2\left\{\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)}\right\} \\
&\qquad \cdot \exp\left[-V\left\{\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)}\right\}\right] \\
&= 1 + \frac{e^{b_2 t}}{a_2\{1 + o(1)\}}\left[-\log a_2 + b_2 t - \log\{1 + o(1)\}\right]^{-2} \\
&\quad \cdot V_2\left[\frac{1}{-\log a_1 + b_1 t - \log\{1 + o(1)\}}, \frac{1}{-\log a_2 + b_2 t - \log\{1 + o(1)\}}\right] \\
&\quad \cdot \exp\left(-V\left[\frac{1}{-\log a_1 + b_1 t - \log\{1 + o(1)\}}, \frac{1}{-\log a_2 + b_2 t - \log\{1 + o(1)\}}\right]\right) \\
&= 1 + \frac{e^{b_2 t}}{a_2\{1 + o(1)\}}\left[-t^{-1}\log a_2 + b_2 - t^{-1}\log\{1 + o(1)\}\right]^{-2} \\
&\quad \cdot V_2\left[\frac{1}{-t^{-1}\log a_1 + b_1 - t^{-1}\log\{1 + o(1)\}}, \frac{1}{-t^{-1}\log a_2 + b_2 - t^{-1}\log\{1 + o(1)\}}\right] \\
&\quad \cdot \exp\left(-tV\left[\frac{1}{-t^{-1}\log a_1 + b_1 - t^{-1}\log\{1 + o(1)\}}, \frac{1}{-t^{-1}\log a_2 + b_2 - t^{-1}\log\{1 + o(1)\}}\right]\right) \\
&= 1 - a \exp\left[-t\left\{V(1/b_1, 1/b_2) - b_2\right\}\right]\{1 + o(1)\},
\end{aligned}$$

for $a = -a_2^{-1} V_2(b_2/b_1, 1) \exp\left\{-\log a_1 V_1\left(1, b_1/b_2\right) - \log a_2 V_2\left(b_2/b_1, 1\right)\right\}$. □

C.1.3 Calculation of the gauge function

We claim that the d-dimensional D-vine has a gauge function of the form stated in Theorem 3. We note that from equation (5.3.19), we have already shown this to be the case for $d = 4$. To prove this more generally, we assume that the result holds for

the two $(d-1)$-dimensional sub-vines of the d-dimensional D-vine, i.e.,

$$g_{D\setminus\{d\}} = g_{D\setminus\{1,d-1,d\}} + V^{\{1,d-1|D\setminus\{1,d-1,d\}\}}\left(\frac{1}{g_{D\setminus\{d-1,d\}} - g_{D\setminus\{1,d-1,d\}}}, \frac{1}{g_{D\setminus\{1,d\}} - g_{D\setminus\{1,d-1,d\}}}\right), \tag{C.1.3}$$

and

$$g_{D\setminus\{1\}} = g_{D\setminus\{1,2,d\}} + V^{\{2,d|D\setminus\{1,2,d\}\}}\left(\frac{1}{g_{D\setminus\{1,d\}} - g_{D\setminus\{1,2,d\}}}, \frac{1}{g_{D\setminus\{1,2\}} - g_{D\setminus\{1,2,d\}}}\right),$$

where we have dropped the arguments from the gauge functions to simplify notation. Further, we claim that the conditional distribution functions used in the calculation of (C.1.2) have the form

$$F_{1|D\setminus\{1,d\}}\left(tx_1|t\boldsymbol{x}_{-\{1,d\}}\right) = 1 - k_{1|D\setminus\{1,d\}}\exp\left\{-t\left(g_{D\setminus\{d\}} - g_{D\setminus\{1,d\}}\right)\right\}\{1+o(1)\}, \tag{C.1.4}$$

and

$$F_{d|D\setminus\{1,d\}}\left(tx_d|t\boldsymbol{x}_{-\{1,d\}}\right) = 1 - k_{d|D\setminus\{1,d\}}\exp\left\{-t\left(g_{D\setminus\{1\}} - g_{D\setminus\{1,d\}}\right)\right\}\{1+o(1)\}, \tag{C.1.5}$$

as $t\to\infty$, for some $k_{1|D\setminus\{1,d\}}, k_{d|D\setminus\{1,d\}} > 0$ not depending on t. From results (5.3.16) and (5.3.17), we see that this claim holds for $d=4$. To prove this more generally, we assume that (C.1.4) and (C.1.5) hold in the $(d-1)$-dimension case, so that as $t\to\infty$,

$$F_{1|D\setminus\{1,d-1,d\}}\left(tx_1|t\boldsymbol{x}_{-\{1,d-1,d\}}\right) = 1 - k_{1|D\setminus\{1,d-1,d\}}\exp\left\{-t\left(g_{D\setminus\{d-1,d\}} - g_{D\setminus\{1,d-1,d\}}\right)\right\}\{1+o(1)\},$$

and

$$F_{d-1|D\setminus\{1,d-1,d\}}\left(tx_{d-1}|t\boldsymbol{x}_{-\{1,d-1,d\}}\right) = 1 - k_{d-1|D\setminus\{1,d-1,d\}}\exp\left\{-t\left(g_{D\setminus\{1,d\}} - g_{D\setminus\{1,d-1,d\}}\right)\right\}\{1+o(1)\},$$

for some $k_{1|D\setminus\{1,d-1,d\}}, k_{d-1|D\setminus\{1,d-1,d\}} > 0$. Results from Joe (1996) show that

$$F_{1|D\setminus\{1,d\}}\left(x_1|\boldsymbol{x}_{-\{1,d\}}\right) =$$

$$\frac{\partial C_{1,d-1|D\setminus\{1,d-1,d\}}\left\{F_{1|D\setminus\{1,d-1,d\}}(x_1|\boldsymbol{x}_{-\{1,d-1,d\}}), F_{d-1|D\setminus\{1,d-1,d\}}(x_{d-1}|\boldsymbol{x}_{-\{1,d-1,d\}})\right\}}{\partial F_{d-1|D\setminus\{1,d-1,d\}}(x_{d-1}|\boldsymbol{x}_{-\{1,d-1,d\}})},$$

with result (5.3.1) giving the form of the required derivative of an inverted extreme value copula. Applying Lemma C.1.2, with $b_1 = g_{D\setminus\{d-1,d\}} - g_{D\setminus\{1,d-1,d\}}$ and $b_2 = g_{D\setminus\{1,d\}} - g_{D\setminus\{1,d-1,d\}}$, we see that for some $k_{1|D\setminus\{1,d\}}$, as $t \to \infty$,

$$\begin{aligned}
&F_{1|D\setminus\{1,d\}}\left(tx_1|t\boldsymbol{x}_{-\{1,d\}}\right) = 1 - k_{1|D\setminus\{1,d\}} \\
&\quad\cdot \exp\Bigg(- t\Bigg[V^{\{1,d-1|D\setminus\{1,d-1,d\}\}}\left\{\frac{1}{g_{D\setminus\{d-1,d\}} - g_{D\setminus\{1,d-1,d\}}}, \frac{1}{g_{D\setminus\{1,d\}} - g_{D\setminus\{1,d-1,d\}}}\right\} \\
&\qquad\qquad - g_{D\setminus\{1,d\}} + g_{D\setminus\{1,d-1,d\}}\Bigg]\Bigg)\{1 + o(1)\} \\
&\quad = 1 - k_{1|D\setminus\{1,d\}} \exp\left\{-t\left(g_{D\setminus\{d\}} - g_{D\setminus\{1,d\}}\right)\right\}\{1 + o(1)\} \qquad \text{by assumption (C.1.3).}
\end{aligned}$$

Result (C.1.5) can be proved by a similar argument. From results (C.1.4) and (C.1.5), we see that $F_{1|D\setminus\{1,d\}}\left(tx_1|t\boldsymbol{x}_{-\{1,d\}}\right)$ and $F_{d|D\setminus\{1,d\}}\left(tx_d|t\boldsymbol{x}_{-\{1,d\}}\right)$ can be written in the form required to apply Lemma C.1.1, with $b_1 = g_{D\setminus\{d\}} - g_{D\setminus\{1,d\}}$ and $b_2 = g_{D\setminus\{1\}} - g_{D\setminus\{1,d\}}$. Applying Lemma C.1.1, we have

$$\begin{aligned}
&-\log c_{\{1,d\}|D\setminus\{1,d\}}\left\{F_{1|D\setminus\{1,d\}}\left(tx_1|t\boldsymbol{x}_{-\{1,d\}}\right), F_{d|D\setminus\{1,d\}}\left(tx_d|t\boldsymbol{x}_{-\{1,d\}}\right)\right\} \\
&\quad\sim t\left\{2g_{D\setminus\{1,d\}} - g_{D\setminus\{d\}} - g_{D\setminus\{1\}} + V^{\{1,d|D\setminus\{1,d\}\}}\left(\frac{1}{g_{D\setminus\{d\}} - g_{D\setminus\{1,d\}}}, \frac{1}{g_{D\setminus\{1\}} - g_{D\setminus\{1,d\}}}\right)\right\} \\
&\quad = t\tilde{g}_D(\boldsymbol{x}),
\end{aligned}$$

and combining this with the gauge function result in (C.1.1), we have

$$\begin{aligned}
&g(\boldsymbol{x}) = g_{D\setminus\{1,d\}}(\boldsymbol{x}_{-\{1,d\}}) \\
&\quad + V^{\{1,d|D\setminus\{1,d\}\}}\left\{\frac{1}{g_{D\setminus\{d\}}(\boldsymbol{x}_{-\{d\}}) - g_{D\setminus\{1,d\}}(\boldsymbol{x}_{-\{1,d\}})}, \frac{1}{g_{D\setminus\{1\}}(\boldsymbol{x}_{-\{1\}}) - g_{D\setminus\{1,d\}}(\boldsymbol{x}_{-\{1,d\}})}\right\},
\end{aligned}$$

hence proving Theorem 3 by induction.

C.2 Proof of result (5.3.24)

Proposition 2. *For any even value of $n \geq 4$,*

$$\left\{2^\alpha \sum_{k=1}^{n/2} (2^\alpha - 1)^{2(k-1)}\right\} - 1 - \left\{2^\alpha \sum_{k=1}^{(n-2)/2} (2^\alpha - 1)^{2(k-1)+1}\right\} = (2^\alpha - 1)^{n-1}.$$

Proof. We first show this is true for $n = 4$. We have

$$\begin{aligned}
&\left\{2^\alpha \sum_{k=1}^{2}(2^\alpha-1)^{2(k-1)}\right\} - 1 - \left\{2^\alpha \sum_{k=1}^{1}(2^\alpha-1)^{2(k-1)+1}\right\} \\
&\qquad = 2^\alpha\left\{(2^\alpha-1)^0 + (2^\alpha-1)^2 - (2^\alpha-1)^1\right\} - 1 \\
&\qquad = 2^\alpha - 1 + 2^\alpha(2^\alpha-1)\left\{(2^\alpha-1)-1\right\} \\
&\qquad = (2^\alpha-1)\left\{-2^\alpha + 1 + 2^\alpha(2^\alpha-1)\right\} \\
&\qquad = (2^\alpha-1)^3.
\end{aligned}$$

Now we assume the result holds for some $n = 2m$ with $m \in \mathbb{Z}$ and $m \geq 2$, and show the result also holds for $n = 2(m+1)$. We have

$$\begin{aligned}
&\left\{2^\alpha \sum_{k=1}^{m+1}(2^\alpha-1)^{2(k-1)}\right\} - 1 - \left\{2^\alpha \sum_{k=1}^{m}(2^\alpha-1)^{2(k-1)+1}\right\} \\
&\qquad = \left\{2^\alpha \sum_{k=1}^{m}(2^\alpha-1)^{2(k-1)}\right\} - 1 - \left\{2^\alpha \sum_{k=1}^{m-1}(2^\alpha-1)^{2(k-1)+1}\right\} \\
&\qquad\qquad + 2^\alpha\left\{(2^\alpha-1)^{2m} - (2^\alpha-1)^{2m-1}\right\} \\
&\qquad = (2^\alpha-1)^{2m-1} + 2^\alpha\left\{(2^\alpha-1)^{2m} - (2^\alpha-1)^{2m-1}\right\} \\
&\qquad = (2^\alpha-1)^{2m-1}\left\{1 - 2^\alpha + 2^\alpha(2^\alpha-1)\right\} \\
&\qquad = (2^\alpha-1)^{2m+1}.
\end{aligned}$$

Hence the result is proved by induction. □

To prove result (5.3.24), we first show that it holds for $d = 3$ and $d = 4$. In the former case, we have

$$\eta_D = \left\{1 + 2^\alpha \sum_{k=1}^{1}(2^\alpha-1)^{2(k-1)+1}\right\}^{-1} = \left\{1 + 2^\alpha(2^\alpha-1)\right\}^{-1},$$

as in equation (5.3.22), and in the latter case we have

$$\eta_D = \left\{2^\alpha \sum_{k=1}^{2}(2^\alpha-1)^{2(k-1)}\right\}^{-1} = \left\{2^\alpha + 2^\alpha(2^\alpha-1)^2\right\}^{-1},$$

as in equation (5.3.23). We now assume the result holds for some $d = n-1$ and $d = n$ being odd and even, respectively, and show that it also holds for $d = n+1, n+2$.

To make the notation clearer in the proof, we let $\eta_{D,(n)}$ denote the value of η_D for an n-dimensional D-vine copula with inverted logistic components with parameter α. From equation (5.3.21), we have

$$\eta_{D,(n+1)} = \left\{ \eta_{D,(n-1)}^{-1} + 2^\alpha \left(\eta_{D,(n)}^{-1} - \eta_{D,(n-1)}^{-1} \right) \right\}^{-1};$$

$$\eta_{D,(n+2)} = \left\{ \eta_{D,(n)}^{-1} + 2^\alpha \left(\eta_{D,(n+1)}^{-1} - \eta_{D,(n)}^{-1} \right) \right\}^{-1}.$$

Under the assumption that (5.3.24) holds for $d = n-1, n$, we have

$$\begin{aligned}
\eta_{D,(n+1)} &= \Bigg[1 + 2^\alpha \sum_{k=1}^{(n-2)/2} (2^\alpha - 1)^{2(k-1)+1} \\
&\qquad + 2^\alpha \left\{ 2^\alpha \sum_{k=1}^{n/2} (2^\alpha - 1)^{2(k-1)} - 1 - 2^\alpha \sum_{k=1}^{(n-2)/2} (2^\alpha - 1)^{2(k-1)+1} \right\} \Bigg]^{-1} \\
&= \left\{ 1 + 2^\alpha \sum_{k=1}^{(n-2)/2} (2^\alpha - 1)^{2(k-1)+1} + 2^\alpha (2^\alpha - 1)^{n-1} \right\}^{-1} \qquad \text{by Proposition 2} \\
&= \left\{ 1 + 2^\alpha \sum_{k=1}^{n/2} (2^\alpha - 1)^{2(k-1)+1} \right\}^{-1}.
\end{aligned}$$

Further, we see that

$$\begin{aligned}
\eta_{D,(n+2)} &= \Bigg[2^\alpha \sum_{k=1}^{n/2} (2^\alpha - 1)^{2(k-1)} \\
&\qquad + 2^\alpha \left\{ 1 + 2^\alpha \sum_{k=1}^{n/2} (2^\alpha - 1)^{2(k-1)+1} - 2^\alpha \sum_{k=1}^{n/2} (2^\alpha - 1)^{2(k-1)} \right\} \Bigg]^{-1} \\
&= \left[2^\alpha \sum_{k=1}^{n/2} (2^\alpha - 1)^{2(k-1)} + 2^\alpha \left\{ 1 + 2^\alpha (2^\alpha - 2) \sum_{k=1}^{n/2} (2^\alpha - 1)^{2(k-1)} \right\} \right]^{-1} \\
&= \left(2^\alpha \sum_{k=1}^{n/2} (2^\alpha - 1)^{2(k-1)} + 2^\alpha \left[1 + 2^\alpha (2^\alpha - 2) \sum_{k=0}^{n/2-1} \left\{ (2^\alpha - 1)^2 \right\}^k \right] \right)^{-1} \\
&= \left[2^\alpha \sum_{k=1}^{n/2} (2^\alpha - 1)^{2(k-1)} + 2^\alpha \left\{ 1 + 2^\alpha (2^\alpha - 2) \frac{1 - (2^\alpha - 1)^n}{1 - (2^\alpha - 1)^2} \right\} \right]^{-1} \\
&= \left[2^\alpha \sum_{k=1}^{n/2} (2^\alpha - 1)^{2(k-1)} + 2^\alpha \left\{ 1 + 2^\alpha (2^\alpha - 2) \frac{1 - (2^\alpha - 1)^n}{2^\alpha (2 - 2^\alpha)} \right\} \right]^{-1}
\end{aligned}$$

$$= \left\{ 2^{\alpha} \sum_{k=1}^{n/2} (2^{\alpha}-1)^{2(k-1)} + 2^{\alpha} (2^{\alpha}-1)^{n} \right\}^{-1} = \left\{ 2^{\alpha} \sum_{k=1}^{(n+2)/2} (2^{\alpha}-1)^{2(k-1)} \right\}^{-1}.$$

Hence result (5.3.24) is proved by induction.

C.3 Properties of extreme value copulas

C.3.1 Some properties of the exponent measure

Proposition 3. *If the exponent measure V has corresponding spectral density $h(w)$ placing no mass on $\{0\}$, and with $h(w) \sim c_2 w^{s_2}$ as $w \to 0$, for some $c_2 \in \mathbb{R}$ and $s_2 > -1$, then*

$$V(r,1) = r^{-1} + \frac{2c_2}{(s_2+1)(s_2+2)} r^{s_2+1}\{1+o(1)\};$$

$$-V_2(r,1) \sim \frac{2c_2}{(s_2+1)} r^{s_2+1}, \quad \text{as } r \to 0.$$

Proof. By the definition of the exponent measure,

$$V(x_1,x_2) = \frac{2}{x_1}\int_0^1 w dH(w) - \frac{2}{x_1}\int_0^{\frac{x_1}{x_1+x_2}} w h(w) dw + \frac{2}{x_2}\int_0^{\frac{x_1}{x_1+x_2}} (1-w)h(w)dw.$$

For $x_1 \to 0$ and $x_2 = O(1)$, by Karamata's theorem, we have

$$\begin{aligned} V(x_1,x_2) &= \frac{1}{x_1} - \frac{2c_2}{x_1(s_2+2)} \left(\frac{x_1}{x_1+x_2}\right)^{s_2+2} \{1+o(1)\} \\ &\qquad + \frac{2c_2}{x_2(s_2+1)} \left(\frac{x_1}{x_1+x_2}\right)^{s_2+1} \{1+o(1)\} \\ &= \frac{1}{x_1} + 2c_2 \left(\frac{x_1}{x_1+x_2}\right)^{s_2+1} \left\{ \frac{1}{x_2(s_2+1)} - \frac{1}{(s_2+2)(x_1+x_2)} \right\} \{1+o(1)\} \\ &= \frac{1}{x_1} + \frac{2c_2 x_1^{s_2+1} x_2^{-(s_2+2)}}{(s_2+1)(s_2+2)} \{1+o(1)\}. \end{aligned} \tag{C.3.1}$$

Hence, we have

$$V(r,1) = \frac{1}{r} + \frac{2c_2}{(s_2+1)(s_2+2)} r^{s_2+1}\{1+o(1)\}, \quad \text{as } r \to 0.$$

Moreover, differentiating expression (C.3.1) with respect to x_2 gives

$$V_2(x_1, x_2) = -\frac{2c_2 x_1^{s_2+1} x_2^{-(s_2+3)}}{(s_2+1)}\{1 + o(1)\},$$

and we can therefore infer that

$$-V_2(r, 1) = \frac{2c_2}{(s_2+1)} r^{s_2+1}\{1 + o(1)\}, \quad \text{as } r \to 0.$$

□

Proposition 4. *If the exponent measure V has corresponding spectral density $h(w)$ placing no mass on $\{1\}$, and with $h(w) \sim c_1(1-w)^{s_1}$ as $w \to 1$, for some $c_1 \in \mathbb{R}$ and $s_1 > -1$, then*

$$V(r, 1) = 1 + \frac{2c_1}{(s_1+1)(s_1+2)} r^{-(s_1+2)}\{1 + o(1)\};$$

$$-V_2(r, 1) = 1 - \frac{2c_1}{(s_1+2)} r^{-(s_1+2)}\{1 + o(1)\}, \quad \textit{as } r \to \infty.$$

Proof. By the definition of the exponent measure,

$$V(x_1, x_2) = \frac{2}{x_2}\int_0^1 (1-w)dH(w) - \frac{2}{x_2}\int_{\frac{x_1}{x_1+x_2}}^1 (1-w)h(w)dw + \frac{2}{x_1}\int_{\frac{x_1}{x_1+x_2}}^1 wh(w)dw.$$

For $x_1 \to \infty$ and $x_2 = o(x_1)$, by Karamata's theorem, we have

$$\begin{aligned} V(x_1, x_2) &= \frac{1}{x_2} - \frac{2c_1}{x_2(s_1+2)}\left(\frac{x_2}{x_1+x_2}\right)^{s_1+2}\{1 + o(1)\} \\ &\qquad + \frac{2c_1}{x_1(s_1+1)}\left(\frac{x_2}{x_1+x_2}\right)^{s_1+1}\{1 + o(1)\} \\ &= \frac{1}{x_2} + 2c_1\left(\frac{x_2}{x_1+x_2}\right)^{s_1+1}\left\{\frac{1}{x_1(s_1+1)} - \frac{1}{(s_1+2)(x_1+x_2)}\right\}\{1 + o(1)\} \\ &= \frac{1}{x_2} + \frac{2c_1 x_2^{s_1+1} x_1^{-(s_1+2)}}{(s_1+1)(s_1+2)}\{1 + o(1)\}. \end{aligned} \tag{C.3.2}$$

Hence, we have

$$V(r, 1) = 1 + \frac{2c_1}{(s_1+1)(s_1+2)} r^{-(s_1+2)}\{1 + o(1)\}, \quad \text{as } r \to \infty.$$

Moreover, differentiating expression (C.3.2) with respect to x_2 gives

$$V_2(x_1, x_2) = -\frac{1}{x_2^2} + \frac{2c_1 x_2^{s_1} x_1^{-(s_1+2)}}{(s_1+2)}\{1 + o(1)\},$$

and we can therefore infer that

$$-V_2(r, 1) = 1 - \frac{2c_1}{(s_1+2)} r^{-(s_1+2)}\{1 + o(1)\}, \quad \text{as } r \to \infty.$$

□

C.3.2 Asymptotic behaviour of $-\log c\{F_1(tx_1), F_2(tx_2)\}$

For V representing an exponent measure defined as in equation (2.3.2) of Chapter 2, an extreme value copula has the form

$$C(u, v) = \exp\left\{-V\left(\frac{-1}{\log u}, \frac{-1}{\log v}\right)\right\}.$$

Differentiating $C(u, v)$ with respect to the second argument, we have

$$F(u \mid v) = \frac{\partial C(u, v)}{\partial u} = -\frac{1}{v}(-\log v)^{-2} V_2\left(\frac{-1}{\log u}, \frac{-1}{\log v}\right) \exp\left\{-V\left(\frac{-1}{\log u}, \frac{-1}{\log v}\right)\right\},$$

and differentiating $C(u, v)$ with respect to both arguments yields the copula density

$$\begin{aligned} c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v} =& \frac{1}{uv}(-\log u)^{-2}(-\log v)^{-2} \exp\left\{-V\left(\frac{-1}{\log u}, \frac{-1}{\log v}\right)\right\} \\ & \cdot \left\{V_1\left(\frac{-1}{\log u}, \frac{-1}{\log v}\right) V_2\left(\frac{-1}{\log u}, \frac{-1}{\log v}\right) - V_{12}\left(\frac{-1}{\log u}, \frac{-1}{\log v}\right)\right\}, \end{aligned}$$

with V_1, V_2 and V_{12} denoting the derivatives of the exponent measure with respect to the first, second, and both components, respectively.

For this class of models, setting exponential margins with $F_i(tx_i) = 1 - e^{-tx_i}$, $i = 1, 2$, we have

$$\begin{aligned} -\log c\{F_1(tx_1), F_2(tx_2)\} &= \log\left(1 - e^{-tx_1}\right) + \log\left(1 - e^{-tx_2}\right) \\ &\quad + 2\log\left\{-\log\left(1 - e^{-tx_1}\right)\right\} + 2\log\left\{-\log\left(1 - e^{-tx_2}\right)\right\} \end{aligned}$$

$$
\begin{aligned}
&+ V\left\{\frac{-1}{\log(1-e^{-tx_1})}, \frac{-1}{\log(1-e^{-tx_2})}\right\} \\
&- \log\left[V_1\left\{\frac{-1}{\log(1-e^{-tx_1})}, \frac{-1}{\log(1-e^{-tx_2})}\right\} V_2\left\{\frac{-1}{\log(1-e^{-tx_1})}, \frac{-1}{\log(1-e^{-tx_2})}\right\}\right. \\
&\qquad\qquad \left. - V_{12}\left\{\frac{-1}{\log(1-e^{-tx_1})}, \frac{-1}{\log(1-e^{-tx_2})}\right\}\right]
\end{aligned}
$$

We impose the assumption that the corresponding spectral density $h(w)$ places no mass on $\{0\}$ or $\{1\}$ and has regularly varying tails. In particular, that $h(w) \sim c_1(1-w)^{s_1}$ as $w \to 1$ and $h(w) \sim c_2 w^{s_2}$ as $w \to 0$, for some $c_1, c_2 \in \mathbb{R}$ and $s_1, s_2 > -1$. By a result from Coles and Tawn (1991), we have

$$
-V_{12}(x_1, x_2) = \frac{2}{(x_1+x_2)^3} h\left(\frac{x_1}{x_1+x_2}\right),
$$

and can therefore deduce that

$$
-V_{12}(1, r) \sim 2c_1 r^{s_1}, \quad \text{and} \quad -V_{12}(r, 1) \sim 2c_2 r^{s_2}, \quad \text{as } r \to 0. \tag{C.3.3}
$$

To investigate the behaviour of $-\log c\{F_1(tx_1), F_2(tx_2)\}$ as $t \to \infty$, we consider three cases; $x_1 < x_2$, $x_1 = x_2$, and $x_1 > x_2$.

Case 1: $x_1 < x_2$

$$
\begin{aligned}
&-\log c\{F_1(tx_1), F_2(tx_2)\} = -e^{-tx_1} - e^{-tx_2} \\
&\quad + 2\log(e^{-tx_1}) + e^{-tx_1} + 2\log(e^{-tx_2}) + e^{-tx_2} + O(e^{-2tx_1}) + O(e^{-2tx_2}) \\
&\quad + V\left\{\frac{1}{e^{-tx_1}+O(e^{-2tx_1})}, \frac{1}{e^{-tx_2}+O(e^{-2tx_2})}\right\} \\
&\quad - \log\left[V_1\left\{\frac{1}{e^{-tx_1}+O(e^{-2tx_1})}, \frac{1}{e^{-tx_2}+O(e^{-2tx_2})}\right\}\right. \\
&\qquad\qquad \cdot V_2\left\{\frac{1}{e^{-tx_1}+O(e^{-2tx_1})}, \frac{1}{e^{-tx_2}+O(e^{-2tx_2})}\right\} \\
&\qquad\qquad\qquad \left. - V_{12}\left\{\frac{1}{e^{-tx_1}+O(e^{-2tx_1})}, \frac{1}{e^{-tx_2}+O(e^{-2tx_2})}\right\}\right] \\
&= -2t(x_1+x_2) + O(e^{-2tx_1}) + O(e^{-2tx_2}) + V\left\{\frac{e^{tx_1}}{1+O(e^{-tx_1})}, \frac{e^{tx_2}}{1+O(e^{-tx_2})}\right\} \\
&\quad - \log\left[e^{-4tx_2} V_1\left\{\frac{e^{t(x_1-x_2)}}{1+O(e^{-tx_1})}, \frac{1}{1+O(e^{-tx_2})}\right\} V_2\left\{\frac{e^{t(x_1-x_2)}}{1+O(e^{-tx_1})}, \frac{1}{1+O(e^{-tx_2})}\right\}\right.
\end{aligned}
$$

$$
\begin{aligned}
&\quad - e^{-3tx_2} V_{12}\left\{\frac{e^{t(x_1-x_2)}}{1+O(e^{-tx_1})}, \frac{1}{1+O(e^{-tx_2})}\right\}\Bigg] \\
&\sim -2t(x_1+x_2) - \log\left(e^{-3tx_2}\right) - \log\left\{e^{ts_2(x_1-x_2)}\right\} \quad \text{by result (C.3.3)} \\
&= t\left\{(1+s_2)x_2 - (2+s_2)x_1\right\}
\end{aligned}
$$

Case 2: $x_1 = x_2$

$$
\begin{aligned}
-\log c\{F_1(tx_1), F_2(tx_2)\} &= 2\log(1-e^{-tx_1}) + 4\log\{-\log(1-e^{-tx_1})\} \\
&\quad - \log(1-e^{-tx_1}) V(1,1) \\
&\quad - \log\Big[\left\{-\log(1-e^{-tx_1})\right\}^4 V_1(1,1)V_2(1,1) \\
&\qquad - \left\{-\log(1-e^{-tx_1})\right\}^3 V_{12}(1,1)\Big] \\
&= -4tx_1 - \log\left\{-V_{12}(1,1)e^{-3tx_1}\right\} + O\left(e^{-tx_1}\right) \\
&= -tx_1 + O\left(e^{-tx_1}\right)
\end{aligned}
$$

Case 3: $x_1 > x_2$

$$
\begin{aligned}
&-\log c\{F_1(tx_1), F_2(tx_2)\} = -e^{-tx_1} - e^{-tx_2} \\
&\quad + 2\log(e^{-tx_1}) + e^{-tx_1} + 2\log(e^{-tx_2}) + e^{-tx_2} + O(e^{-2tx_1}) + O(e^{-2tx_2}) \\
&\quad + V\left\{\frac{1}{e^{-tx_1}+O(e^{-2tx_1})}, \frac{1}{e^{-tx_2}+O(e^{-2tx_2})}\right\} \\
&\quad - \log\Bigg[V_1\left\{\frac{1}{e^{-tx_1}+O(e^{-2tx_1})}, \frac{1}{e^{-tx_2}+O(e^{-2tx_2})}\right\} \\
&\qquad \cdot V_2\left\{\frac{1}{e^{-tx_1}+O(e^{-2tx_1})}, \frac{1}{e^{-tx_2}+O(e^{-2tx_2})}\right\} \\
&\qquad - V_{12}\left\{\frac{1}{e^{-tx_1}+O(e^{-2tx_1})}, \frac{1}{e^{-tx_2}+O(e^{-2tx_2})}\right\}\Bigg] \\
&= -2t(x_1+x_2) + O(e^{-2tx_1}) + O(e^{-2tx_2}) + V\left\{\frac{e^{tx_1}}{1+O(e^{-tx_1})}, \frac{e^{tx_2}}{1+O(e^{-tx_2})}\right\} \\
&\quad - \log\Bigg[e^{-4tx_1} V_1\left\{\frac{1}{1+O(e^{-tx_1})}, \frac{e^{t(x_2-x_1)}}{1+O(e^{-tx_2})}\right\} V_2\left\{\frac{1}{1+O(e^{-tx_1})}, \frac{e^{t(x_2-x_1)}}{1+O(e^{-tx_2})}\right\} \\
&\qquad - e^{-3tx_1} V_{12}\left\{\frac{1}{1+O(e^{-tx_1})}, \frac{e^{t(x_2-x_1)}}{1+O(e^{-tx_2})}\right\}\Bigg] \\
&\sim -2t(x_1+x_2) - \log\left(e^{-3tx_1}\right) - \log\left\{e^{ts_1(x_2-x_1)}\right\} \quad \text{by result (C.3.3)}
\end{aligned}
$$

$$= t\left\{(1+s_1)x_1 - (2+s_1)x_2\right\}.$$

These three cases can be combined into a single expression, so that as $t \to \infty$, we have

$$-\log c\{F_1(tx_1), F_2(tx_2)\} \sim t\Big\{ \left(1 + s_1 \mathbb{1}_{\{x_1 \geq x_2\}} + s_2 \mathbb{1}_{\{x_1 < x_2\}}\right) \max(x_1, x_2) - \left(2 + s_1 \mathbb{1}_{\{x_1 \geq x_2\}} + s_2 \mathbb{1}_{\{x_1 < x_2\}}\right) \min(x_1, x_2)\Big\}.$$

C.3.3 Asymptotic behaviour of $F_{1|2}(tx_1 \mid tx_2)$

For a bivariate extreme value copula, the conditional distribution function has the form

$$F_{1|2}(tx_1 \mid tx_2) = -\frac{1}{1-e^{-tx_2}}\{-\log(1-e^{-tx_2})\}^{-2} V_2\left\{\frac{-1}{\log(1-e^{-tx_1})}, \frac{-1}{\log(1-e^{-tx_2})}\right\} \exp\left[-V\left\{\frac{-1}{\log(1-e^{-tx_1})}, \frac{-1}{\log(1-e^{-tx_2})}\right\}\right].$$

To investigate the behaviour of $F_{1|2}(tx_1 \mid tx_2)$ as $t \to \infty$, we again consider three cases.

Case 1: $x_1 < x_2$

$$\begin{aligned}
F_{1|2}(tx_1 \mid tx_2) &= -\frac{1}{1-e^{-tx_2}} V_2\left\{\frac{\log(1-e^{-tx_2})}{\log(1-e^{-tx_1})}, 1\right\} \exp\left[\log\left(1-e^{-tx_2}\right) V\left\{\frac{\log(1-e^{-tx_2})}{\log(1-e^{-tx_1})}, 1\right\}\right] \\
&= -\left[1 + e^{-tx_2}\{1+o(1)\}\right] V_2\left[\exp\{t(x_1-x_2)\}\{1+o(1)\}, 1\right] \\
&\quad \exp\left[\log\left(1-e^{-tx_2}\right) V\left\{\exp\{t(x_1-x_2)\}\{1+o(1)\}, 1\right\}\right] \\
&= \left[1 + e^{-tx_2}\{1+o(1)\}\right]\left[\frac{2c_2}{(s_2+1)} \exp\left\{t(x_1-x_2)(s_2+1)\right\}\{1+o(1)\}\right] \\
&\quad \exp\left[\log\left(1-e^{-tx_2}\right)\left\{\exp\{-t(x_1-x_2)\}\{1+o(1)\}\right]\right. \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(by Proposition 3)} \\
&= \left[1 + e^{-tx_2}\{1+o(1)\}\right]\left[\frac{2c_2}{(s_2+1)} \exp\left\{t(x_1-x_2)(s_2+1)\right\}\{1+o(1)\}\right]
\end{aligned}$$

$$\left[1 - e^{-tx_1}\{1 + o(1)\}\right]$$
$$= \frac{2c_2}{(s_2+1)} \exp\left\{t(x_1 - x_2)(s_2+1)\right\}\{1 + o(1)\}.$$

Case 2: $x_1 = x_2$

$$\begin{aligned} F_{1|2}(tx_1 \mid tx_2) &= -\frac{1}{1 - e^{-tx_1}} V_2(1,1) \exp\left\{\log(1 - e^{-tx_1}) V(1,1)\right\} \\ &= -V_2(1,1)(1 - e^{-tx_1})^{V(1,1)-1} \\ &= -V_2(1,1)\{1 + o(1)\}. \end{aligned}$$

Case 3: $x_1 > x_2$

$$\begin{aligned} F_{1|2}(tx_1 \mid tx_2) &= -\frac{1}{1 - e^{-tx_2}} V_2\left\{\frac{\log(1 - e^{-tx_2})}{\log(1 - e^{-tx_1})}, 1\right\} \\ &\qquad \exp\left[\log\left(1 - e^{-tx_2}\right) V\left\{\frac{\log(1 - e^{-tx_2})}{\log(1 - e^{-tx_1})}, 1\right\}\right] \\ &= -\frac{1}{1 - e^{-tx_2}} V_2\left[\exp\{t(x_1 - x_2)\}\{1 + o(1)\}, 1\right] \\ &\qquad \exp\left[\log\left(1 - e^{-tx_2}\right) V\left\{\exp\{t(x_1 - x_2)\}\{1 + o(1)\}, 1\right\}\right] \\ &= \frac{1}{1 - e^{-tx_2}}\left[1 - \frac{2c_1}{(s_1+2)} \exp\left\{-t(x_1 - x_2)(s_1+2)\right\}\{1 + o(1)\}\right] \\ &\qquad \exp\left(\log\left(1 - e^{-tx_2}\right)\left[1 + \frac{2c_1}{(s_1+1)(s_1+2)} \exp\{-t(s_1+2)(x_1 - x_2)\}\{1 + o(1)\}\right]\right) \end{aligned}$$

(by Proposition 4)

$$\begin{aligned} &= \frac{1 - e^{-tx_2}}{1 - e^{-tx_2}}\left[1 - \frac{2c_1}{(s_1+2)} \exp\left\{-t(x_1 - x_2)(s_1+2)\right\}\{1 + o(1)\}\right] \\ &\qquad \left[1 + \frac{2c_1}{(s_1+1)(s_1+2)} \exp\{-t(s_1+2)(x_1 - x_2) - tx_2\}\{1 + o(1)\}\right] \\ &= 1 - \frac{2c_1}{(s_1+2)} \exp\left\{-t(x_1 - x_2)(s_1+2)\right\}\{1 + o(1)\}. \end{aligned}$$

C.4 $-\log c_{13|2}(u,v)$ for (inverted) extreme value copulas

In Section 5.4, we investigate cases where the copula density in tree T_2 belongs to either the extreme value or inverted extreme value families of distributions. Here, we denote these by $c^{EV}_{13|2}(u,v)$ and $c^{IEV}_{13|2}(u,v)$, respectively. In Section 5.4.2, we find that we should focus on the conditional distributions $F_{1|2}(tx_1 \mid tx_2)$ and $F_{3|2}(tx_3 \mid tx_2)$ having three different asymptotic forms. In this section, we therefore consider the behaviour of $-\log c^{EV}_{13|2}(u,v)$ and $-\log c^{IEV}_{13|2}(u,v)$ for u and v of the form

$$a\{1+o(1)\} \quad ; \quad b_1\exp(-b_2 t)\{1+o(1)\} \quad ; \quad 1-c_1\exp(-c_2 t)\{1+o(1)\},$$

for $b_2, c_2 > 0$, using the results that

$$\begin{aligned}-\log c^{EV}_{13|2}(u,v) =& \log u + \log v + 2\log(-\log u) + 2\log(-\log v) + V^{\{13|2\}}\left(\frac{-1}{\log u}, \frac{-1}{\log v}\right)\\ &- \log\left\{V_1^{\{13|2\}}\left(\frac{-1}{\log u}, \frac{-1}{\log v}\right)V_2^{\{13|2\}}\left(\frac{-1}{\log u}, \frac{-1}{\log v}\right) - V_{12}^{\{13|2\}}\left(\frac{-1}{\log u}, \frac{-1}{\log v}\right)\right\},\end{aligned}$$

and

$$\begin{aligned}-\log c^{IEV}_{13|2}(u,v) =& \log(1-u) + \log(1-v) + 2\log\{-\log(1-u)\} + 2\log\{-\log(1-v)\}\\ &+ V^{\{13|2\}}\left\{\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)}\right\}\\ &- \log\left[V_1^{\{13|2\}}\left\{\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)}\right\}V_2^{\{13|2\}}\left\{\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)}\right\}\right.\\ &\left. - V_{12}^{\{13|2\}}\left\{\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)}\right\}\right].\end{aligned}$$

We also have the assumption that the spectral density $h^{\{13|2\}}(w)$ corresponding to the copula in tree T_2 places no mass on $\{0\}$ or $\{1\}$, and has $h^{\{13|2\}}(w) \sim c_1^{\{13|2\}}(1-w)^{s_1^{\{13|2\}}}$ as $w \to 1$, and $h^{\{13|2\}}(w) \sim c_2^{\{13|2\}} w^{s_2^{\{13|2\}}}$ as $w \to 0$, for some $c_1^{\{13|2\}}, c_2^{\{13|2\}} \in \mathbb{R}$ and $s_1^{\{13|2\}}, s_2^{\{13|2\}} > -1$. In the following nine cases, we provide results for the asymptotic behaviour of $-\log c^{EV}_{13|2}(u,v)$ and $-\log c^{IEV}_{13|2}(u,v)$, as $t \to \infty$, for u and v taking different combinations of the forms stated above.

Case 1: $u = a_u\{1 + o(1)\}$, $v = a_v\{1 + o(1)\}$

$$-\log c^{EV}_{13|2}(u, v) = o(t)$$

$$-\log c^{IEV}_{13|2}(u, v) = o(t)$$

Case 2: $u = a_u\{1 + o(1)\}$, $v = b_{v,1}\exp(-b_{v,2}t)\{1 + o(1)\}$

$$-\log c^{EV}_{13|2}(u, v) = o(t)$$

$$-\log c^{IEV}_{13|2}(u, v) \sim tb_{v,2}\left(1 + s_2^{\{13|2\}}\right)$$

Case 3: $u = a_u\{1 + o(1)\}$, $v = 1 - c_{v,1}\exp(-c_{v,2}t)\{1 + o(1)\}$

$$-\log c^{EV}_{13|2}(u, v) \sim tc_{v,2}\left(1 + s_2^{\{13|2\}}\right)$$

$$-\log c^{IEV}_{13|2}(u, v) = o(t)$$

Case 4: $u = b_{u,1}\exp(-b_{u,2}t)\{1 + o(1)\}$, $v = a_v\{1 + o(1)\}$

$$-\log c^{EV}_{13|2}(u, v) = o(t)$$

$$-\log c^{IEV}_{13|2}(u, v) \sim tb_{u,2}\left(1 + s_1^{\{13|2\}}\right)$$

Case 5: $u = b_{u,1}\exp(-b_{u,2}t)\{1 + o(1)\}$, $v = b_{v,1}\exp(-b_{v,2}t)\{1 + o(1)\}$

$$-\log c^{EV}_{13|2}(u, v) \sim t\left\{-b_{u,2} - b_{v,2} + V^{\{13|2\}}\left(1/b_{u,2}, 1/b_{v,2}\right)\right\}$$

$$\begin{aligned}-\log c^{IEV}_{13|2}(u, v) \sim t\Big\{ &\left(1 + s_1^{\{13|2\}}\mathbb{1}_{\{b_{u,2}\geq b_{v,2}\}} + s_2^{\{13|2\}}\mathbb{1}_{\{b_{u,2}<b_{v,2}\}}\right)\max(b_{u,2}, b_{v,2}) \\ &- \left(2 + s_1^{\{13|2\}}\mathbb{1}_{\{b_{u,2}\geq b_{v,2}\}} + s_2^{\{13|2\}}\mathbb{1}_{\{b_{u,2}<b_{v,2}\}}\right)\min(b_{u,2}, b_{v,2})\Big\}\end{aligned}$$

Case 6: $u = b_{u,1}\exp(-b_{u,2}t)\{1+o(1)\}$**,** $v = 1 - c_{v,1}\exp(-c_{v,2}t)\{1+o(1)\}$

$$-\log c^{EV}_{13|2}(u,v) \sim tc_{v,2}\left(1 + s_2^{\{13|2\}}\right)$$

$$-\log c^{IEV}_{13|2}(u,v) \sim tb_{u,2}\left(1 + s_1^{\{13|2\}}\right)$$

Case 7: $u = 1 - c_{u,1}\exp(-c_{u,2}t)\{1+o(1)\}$**,** $v = a_v\{1+o(1)\}$

$$-\log c^{EV}_{13|2}(u,v) \sim tc_{u,2}\left(1 + s_1^{\{13|2\}}\right)$$

$$-\log c^{IEV}_{13|2}(u,v) = o(t)$$

Case 8: $u = 1 - c_{u,1}\exp(-c_{u,2}t)\{1+o(1)\}$**,** $v = b_{v,1}\exp(-b_{v,2}t)\{1+o(1)\}$

$$-\log c^{EV}_{13|2}(u,v) \sim tc_{u,2}\left(1 + s_1^{\{13|2\}}\right)$$

$$-\log c^{IEV}_{13|2}(u,v) \sim tb_{v,2}\left(1 + s_2^{\{13|2\}}\right)$$

Case 9: $u = 1 - c_{u,1}\exp(-c_{u,2}t)\{1+o(1)\}$**,** $v = 1 - c_{v,1}\exp(-c_{v,2}t)\{1+o(1)\}$

$$\begin{aligned}-\log c^{EV}_{13|2}(u,v) \sim t\Big\{ &\left(1 + s_1^{\{13|2\}}\mathbb{1}_{\{c_{u,2}\geq c_{v,2}\}} + s_2^{\{13|2\}}\mathbb{1}_{\{c_{u,2}<c_{v,2}\}}\right)\max(c_{u,2}, c_{v,2}) \\ &- \left(2 + s_1^{\{13|2\}}\mathbb{1}_{\{c_{u,2}\geq c_{v,2}\}} + s_2^{\{13|2\}}\mathbb{1}_{\{c_{u,2}<c_{v,2}\}}\right)\min(c_{u,2}, c_{v,2})\Big\}\end{aligned}$$

$$-\log c^{IEV}_{13|2}(u,v) \sim t\left\{-c_{u,2} - c_{v,2} + V^{\{13|2\}}\left(1/c_{u,2}, 1/c_{v,2}\right)\right\}$$

Bibliography

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198.

Asadi, P., Davison, A. C., and Engelke, S. (2015). Extremes on river networks. *The Annals of Applied Statistics*, 9(4):2023–2050.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press, Boca Raton.

Barnett, V. (1976). The ordering of multivariate data. *Journal of the Royal Statistical Society. Series A (General)*, 139(3):318–355.

Bedford, T. and Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32(1-4):245–268.

Bedford, T. and Cooke, R. M. (2002). Vines: a new graphical model for dependent random variables. *The Annals of Statistics*, 30(4):1031–1068.

Behrens, C. N., Lopes, H. F., and Gamerman, D. (2004). Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling*, 4(3):227–244.

Beirlant, J., Goegebeur, Y., Teugels, J. L., and Segers, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley Series in Probability and Statistics. Wiley, Chichester.

Beirlant, J. and Vandewalle, B. (2002). Some comments on the estimation of a dependence index in bivariate extreme value statistics. *Statistics & Probability Letters*, 60(3):265–278.

Brown, B. M. and Resnick, S. I. (1977). Extreme values of independent stochastic processes. *Journal of Applied Probability*, 14(4):732–739.

Chautru, E. (2015). Dimension reduction in multivariate extreme value analysis. *Electronic Journal of Statistics*, 9(1):383–418.

Chiapino, M. and Sabourin, A. (2017). Feature clustering for extreme events analysis, with application to extreme stream-flow data. In *New Frontiers in Mining Complex Patterns*, pages 132–147. Springer, Cham.

Chiapino, M., Sabourin, A., and Segers, J. (2019). Identifying groups of variables with the potential of being large simultaneously. *Extremes*, 22(2):193–222.

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

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

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(2):377–392.

Coles, S. G. and Tawn, J. A. (1994). Statistical methods for multivariate extremes: an application to structural design (with discussion). *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43(1):1–48.

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102(479):824–840.

Das, B., Mitra, A., and Resnick, S. I. (2013). Living on the multidimensional edge: seeking hidden risks using regular variation. *Advances in Applied Probability*, 45(1):139–163.

Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, 27(2):161–186.

de Bruijn, N. G. (1958). *Asymptotic Methods in Analysis*. North-Holland Publishing Company, Amsterdam.

de Haan, L. (1984). A spectral representation for max-stable processes. *The Annals of Probability*, 12(4):1194–1204.

de Haan, L. and Zhou, C. (2011). Extreme residual dependence for random vectors and processes. *Advances in Applied Probability*, 43(1):217–242.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 47(3):299–350.

Draisma, G., Drees, H., Ferreira, A., and de Haan, L. (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli*, 10(2):251–280.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222.

Eastoe, E. F. and Tawn, J. A. (2012). Modelling the distribution of the cluster maxima of exceedances of subasymptotic thresholds. *Biometrika*, 99(1):43–55.

Engelke, S. and Hitz, A. S. (2018). Graphical models for extremes. *arXiv:1812.01734*.

Fawcett, L. and Walshaw, D. (2006). A hierarchical model for extreme wind speeds. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 55(5):631–646.

Frigessi, A., Haug, O., and Rue, H. (2002). A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes*, 5(3):219–235.

Galambos, J. (1975). Order statistics of samples from multivariate distributions. *Journal of the American Statistical Association*, 70(351):674–680.

Gissibl, N. and Klüppelberg, C. (2018). Max-linear models on directed acyclic graphs. *Bernoulli*, 24(4A):2693–2720.

Goix, N., Sabourin, A., and Clémençon, S. (2016). Sparse representation of multivariate extremes with applications to anomaly ranking. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51, pages 75–83, Cadiz, Spain. PMLR.

Goix, N., Sabourin, A., and Clémençon, S. (2017). Sparse representation of multivariate extremes with applications to anomaly detection. *Journal of Multivariate Analysis*, 161:12–31.

Gumbel, E. J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publications de l'Institute de Statistique de l'Université de Paris*, 9:171–173.

Hastie, T. J., Tibshirani, R. J., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2nd edition.

Heffernan, J. E. and Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, 17(2):537–571.

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

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174.

Hitz, A. S. and Evans, R. (2016). One-component regular variation and graphical modeling of extremes. *Journal of Applied Probability*, 53(3):733–746.

Hua, L. and Joe, H. (2011). Tail order and intermediate tail dependence of multivariate copulas. *Journal of Multivariate Analysis*, 102(10):1454–1471.

Hüsler, J. and Reiss, R. D. (1989). Maxima of normal random vectors: between independence and complete dependence. *Statistics and Probability Letters*, 7(4):283–286.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81:158–171.

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

Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. *Lecture Notes-Monograph Series*, 28:120–141.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, New York.

Joe, H. (2014). *Dependence Modeling with Copulas*. Chapman & Hall/CRC Press, New York.

Joe, H., Li, H., and Nikoloulopoulos, A. K. (2010). Tail dependence functions and vine copulas. *Journal of Multivariate Analysis*, 101(1):252–270.

Keef, C., Papastathopoulos, I., and Tawn, J. A. (2013a). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: additional constraints for the Heffernan and Tawn model. *Journal of Multivariate Analysis*, 115:396–404.

Keef, C., Tawn, J. A., and Lamb, R. (2013b). Estimating the probability of widespread flood events. *Environmetrics*, 24(1):13–21.

Klüppelberg, C., Haug, S., and Kuhn, G. (2015). Copula structure analysis based on extreme dependence. *Statistics and its Interface*, 8(1):93–107.

Knorr-Held, L. (2003). Some remarks on Gaussian Markov random field models for disease mapping. In Green, P. J., Hjort, N. L., and Richardson, S., editors, *Highly Structured Stochastic Systems*, pages 203–207. Oxford University Press, Oxford.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, Cambridge.

Kurowicka, D. and Cooke, R. M. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley, Chichester.

Kurowicka, D. and Joe, H., editors (2010). *Dependence Modeling: Vine Copula Handbook*. World Scientific, Singapore.

Leadbetter, M. R., Lindgren, G., and Rootzén., H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 59(2):475–499.

MacDonald, A., Scarrott, C. J., Lee, D., Darlow, B., Reale, M., and Russell, G. (2011). A flexible extreme value mixture model. *Computational Statistics & Data Analysis*, 55(6):2137–2157.

Martins, E. S. and Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(3):737–744.

Maulik, K. and Resnick, S. I. (2004). Characterizations and examples of hidden regular variation. *Extremes*, 7(1):31–67.

Mitra, A. and Resnick, S. I. (2011). Hidden regular variation and detection of hidden risks. *Stochastic Models*, 27(4):591–614.

Morris, D. G. and Flavin, R. W. (1990). A digital terrain model for hydrology. In *Proceedings of the 4th International Symposium on Spatial Data Handling*, volume 1, pages 250–262, Zurich.

Morris, D. G. and Flavin, R. W. (1994). Sub-set of UK 50m by 50m hydrological digital terrain model grids. *NERC, Institute of Hydrology, Wallingford*.

Neal, R. M. (1994). An improved acceptance procedure for the hybrid Monte Carlo algorithm. *Journal of Computational Physics*, 111(1):194–203.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer-Verlag, New York, 2nd edition.

Nolde, N. (2014). Geometric interpretation of the residual dependence coefficient. *Journal of Multivariate Analysis*, 123:85–95.

Northrop, P. J., Attalides, N., and Jonathan, P. (2017). Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 66(1):93–120.

Papastathopoulos, I., Strokorb, K., Tawn, J. A., and Butler, A. (2017). Extreme events of Markov chains. *Advances in Applied Probability*, 49(1):134–161.

Papastathopoulos, I. and Tawn, J. A. (2016). Conditioned limit laws for inverted max-stable processes. *Journal of Multivariate Analysis*, 150:214–228.

Peng, L. (1999). Estimation of the coefficient of tail dependence in bivariate extremes. *Statistics & Probability Letters*, 43(4):399–409.

Perfekt, R. (1994). Extremal behaviour of stationary Markov chains with applications. *The Annals of Applied Probability*, 4(2):529–548.

Pickands, J. (1971). The two-dimensional Poisson process and extremal processes. *Journal of Applied Probability*, 8(4):745–756.

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

Pickands, J. (1981). Multivariate extreme value distributions (with a discussion). In *Proceedings of the 43rd session of the International Statistical Institute*, volume 49, pages 859–878, 894–902, Buenos Aires.

Reich, B. J. and Shaby, B. A. (2012). A hierarchical max-stable spatial model for extreme precipitation. *The Annals of Applied Statistics*, 6(4):1430–1451.

Resnick, S. I. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5(4):303–336.

Resnick, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer-Verlag, New York.

Roberts, G. O., Gelman, A., and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120.

Rootzén, H. (1988). Maxima and exceedances of stationary Markov chains. *Advances in Applied Probability*, 20(2):371–390.

Sang, H. and Gelfand, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, 16(3):407–426.

Scarrott, C. J. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT - Statistical Journal*, 10(1):33–60.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1):33–44.

Segers, J. (2012). Max-stable models for multivariate extremes. *REVSTAT - Statistical Journal*, 10(1):61–82.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.

Smith, R. L. (1990). Max-stable processes and spatial extremes. *Unpublished manuscript*.

Smith, R. L. (1992). The extremal index for a Markov chain. *Journal of Applied Probability*, 29(1):37–45.

Smith, R. L., Tawn, J. A., and Coles, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika*, 84(2):249–268.

So, B.-J., Kwon, H.-H., Kim, D., and Lee, S. O. (2015). Modeling of daily rainfall sequence and extremes based on a semiparametric Pareto tail approach at multiple locations. *Journal of Hydrology*, 529(3):1442–1450.

Stan Development Team (2018). *RStan: the R interface to Stan*. R package version 2.17.3.

Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika*, 75(3):397–415.

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

Vettori, S., Huser, R., Segers, J., and Genton, M. G. (2018). Bayesian model averaging over tree-based dependence structures for multivariate extremes. *arXiv:1705.10488*.

von Mises, R. (1936). La distribution de la plus grande de n valeurs. *Rev. Math. Union Interbalcanique*, 1:141–160.

Wadsworth, J. L. (2016). Exploiting structure of maximum likelihood estimators for extreme value threshold selection. *Technometrics*, 58(1):116–126.

Wadsworth, J. L. and Tawn, J. A. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19(5B):2689–2714.

Wadsworth, J. L., Tawn, J. A., and Jonathan, P. (2010). Accounting for choice of measurement scale in extreme value modeling. *The Annals of Applied Statistics*, 4(3):1558–1578.

Wilks, D. S. (2006). *Statistical Methods in the Atmospheric Sciences*. Academic Press, London, 2nd edition.

Wintenberger, O. (2018). Editorial: special issue on the extreme value analysis conference challenge "prediction of extremal precipitation". *Extremes*, 21(3):425–429.

Yunus, R. M., Hasan, M. M., Razak, N. A., Zubairi, Y. Z., and Dunn, P. K. (2017). Modelling daily rainfall with climatological predictors: Poisson-gamma generalized linear modelling approach. *International Journal of Climatology*, 37(3):1391–1399.