A Geometric Interpretation of Multivariate Extreme Value Analysis

Ryan Campbell, BSc, MSc

School of Mathematical Sciences

Lancaster University

A thesis submitted for the degree of

Doctor of Philosophy

September 2025

A Geometric Interpretation of Multivariate Extreme Value Analysis

Ryan Campbell, BSc, MSc.

School of Mathematical Sciences, Lancaster University

A thesis submitted for the degree of *Doctor of Philosophy*. September 2025.

Abstract

In multivariate extreme value analysis, interest lies in characterising the tail behaviour of random vectors. The lack of natural ordering in the multivariate setting, and the many possible combinations of tail behaviour for all the subgroups of components of random vectors make this a difficult task. In this work, the geometry of scaled copies of random vectors is used to characterise the tail of the underlying probability distribution. While this geometry has been shown to provide useful information about the tail behaviour of random vectors, we introduce methodology to estimate it from data in a parametric, Bayesian semiparametric, and piecewise-linear semiparametric manner. The geometry is used to model both the radial and angular components of the pseudo-polar decomposition, a key feature of the geometric framework. Links are made to the classical approach of multivariate extremes by investigating the geometry of generalised Pareto random vectors, an important model used in a variety of practical applications. Both the geometric and the classical approach have their benefits and drawbacks. These will be discussed along with a commentary on future work to be done in the multivariate geometric framework.

Acknowledgements

First and foremost, thank you Jenny Wadsworth for your guidance and patience in the last four years. It has been an honour to be your PhD student, and I will always be grateful for the amazing experience you gave me. Thank you Ioannis Papastathopoulos for always believing in me, and for your constant words of encouragement in difficult times. Thank you to my fellow colleagues at Lancaster – there are too many of you to name. It has been an immense pleasure to be part of such a rich and vibrant extreme value community. Thank you to my examiners Kirstin Strokorb and Emma Eastoe for making the viva experience pleasant and pain-free. Your helpful comments have drastically improved this thesis document. Lastly, thank you to my friends in Lonsdale Fell Runners. I have truly cherished the times we spent up in't fells.

Throughout my PhD, I have been grateful to receive funding from Engineering and Physical Sciences Research Council Doctoral Training Partnership (United Kingdom, award EP/W523811/1), the Natural Sciences and Engineering Research Council of Canada postgraduate scholarship (doctoral program, award 577646), and the Fonds de Recherche du Québec doctoral training scholarship (award B2X-325590).

Against the run-of-the-mill, swimming against the stream.
Life in two dimensions is a mass production scheme.

–Neil Peart

Declaration

I declare that the work presented in this thesis is, to the best of my knowledge and belief, original and my own work, except where stated otherwise. I was part-author on the work presented in Chapters 3 and 4, with my contributions highlighted in disclaimers at the start of each of these chapters, and I was principal author on the work presented in Chapters 5 and 6. The material has not been submitted, either in whole or in part, for a degree at this, or any other university. This thesis does not exceed the maximum permitted word length of 80,000 words including appendices and footnotes, but excluding the bibliography. A rough estimate of the word count is 46,000.

Ryan Campbell

Chapter 3 in this PhD informed the following article:

- J. L. Wadsworth and R. Campbell. Statistical inference for multivariate extremes via a geometric approach. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 86(5):1243–1265, 03 2024.

Chapter 4 in this PhD informed the following article (in revision):

- I. Papastathopoulos, L. de Monte, R. Campbell, and H. Rue. Statistical inference for radial generalized Pareto distributions and return sets in geometric extremes. *arXiv:2310.06130*, 2025.

Chapter 5 in this PhD informed the following article (in revision):

- R. Campbell and J. Wadsworth. Piecewise-linear modeling of multivariate geometric extremes. *arXiv:2412.05195*, 2024.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Multivariate extreme value analysis concerns itself with the characterisation of the tail behaviour of random vectors. Difficulties arise due to the fact that data is often not observed far into these tails, making statistical inference challenging. It could also be difficult to account for the possible combinations of extremal dependencies between components, that is, which components have the possibility to grow large together. Furthermore, the lack of natural ordering of vectors make this characterisation difficult. It's easy to imagine the meaning of a single scalar quantity to be large, but what does it mean for a collection of several scalar values to be large?

Take, for example, daily measurements of carbon monoxide (CO, $\mathrm{mg/m^3}$) and nitrogen dioxide (NO_2, $\mu\mathrm{g/m^3}$) in North Kensington, London. In urban areas in particular, exposure to high air pollution levels can have negative health effects (Holgate, 2017). For these two pollutants, exposure to high levels have been linked to a reduction of oxygen supply to the heart in people with pre-existing conditions (Wu and Wang, 2005) and lung damage in people susceptible to respiratory illnesses (Seals and Krasner, 2020). Understanding their behaviour at extreme levels is therefore of interest to the general population. The daily maximum measurements of CO and NO_2 is shown from October to April 1996–2024 in Figure 1.1. Questions arise naturally when trying to describe how these pollutants behave when their levels

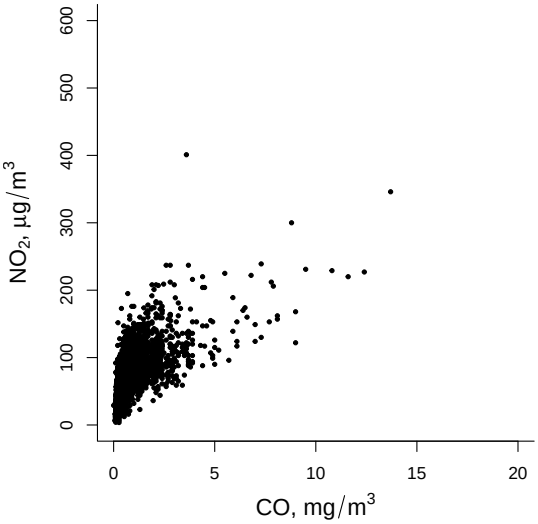

Figure 1.1: Measurements of CO and NO_2 from the North Kensington, London station of the UK-based Automatic Urban and Rural Network (AURN) air quality monitoring network.

become dangerous to human health. For one, do values of CO and NO_2 depend on each other? When one becomes large, does the other? If not, do they have an inverse relationship in the tails? When one is large, does that guarantee the other being large? Furthermore, given the data we have available to us, can we estimate the probability of previously unseen levels of these pollutants? Suppose we wanted to know low likely it is to simultaneously experience CO levels greater that 15 mg/m^3 and NO_2 levels greater than 500 μg/m^3. From Figure 1.1, it is shown that this occurrence is never observed in the available data, so statistical estimation becomes difficult.

It is exactly in these settings that multivariate extreme value analysis originated. In multivariate extremes, assumptions are made on the data's underlying unobserved distribution to be able to make statements in unprecedented areas of a given dataset's domain. Suppose the process we are interested in, such as measurements of CO and NO_2, are thought of as observations from a random vector. More traditional

methods based on the assumptions of multivariate regular variation (Resnick, 1987) or hidden regular variation (Ledford and Tawn, 1996, 1997) assume the components of this random vector grow large together at the same rate in order to estimate the occurrence of rare events. For example, when CO is large, it is assumed that NO_2 is as well. This is often not guaranteed in practice, and if assumed may lead to inaccurate estimates of the occurrence of extreme events. Furthermore, traditional methods may be limited in the locations of the tail in which one could characterise extreme events. While they can be useful to estimate the frequency of all components being large, they often struggle when at least one value is small.

In recent decades, other methods have gained popularity. Notably, rather than modelling the behaviour of the entire random vector, it can be useful in practice to instead condition on one component being large and model the remaining (Heffernan and Tawn, 2004). While often viable for random vectors of length two, statistical inference can be complicated for a practitioner in higher dimensions. For one, a decision must be made as to which component to condition on. Furthermore, in higher dimensions, one might be tempted to fit separate models, each one with a different conditional component. When this is done, there is no theoretically-sound way to link the models and use them to characterise the entire multivariate tail, leaving the practitioner to make compromises and use other methods to join the conditional models. Furthermore, in as low as two dimensions, the dataset may exhibit complex tail behaviour. For example, values of CO and NO_2 can potentially grow together, while values of CO might also have extreme values when NO_2 is low. In this type of simulation the conditional approach will not be suitable. The conditional approach can only model for when either both variables are large or when one variable is large when the other is small, not both.

To remedy the issues of flexibility in multivariate extremes, the work introduced in this thesis will instead use a new approach relying on the geometry of the underlying distribution of the dataset. If you imagine a dataset forming a multidimensional cloud, then the edges of this cloud describe the so-called *limit set*.

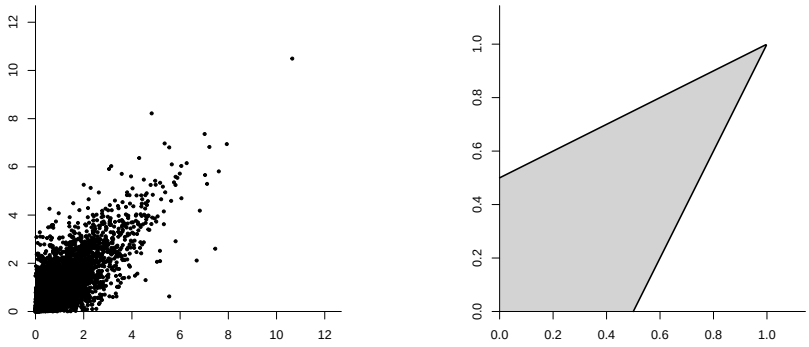

Figure 1.2: Bivariate data (left) and the corresponding limit set in grey (right).

Take, for example, the two-dimensional data shown in Figure 1.2. If points from this sample are scaled appropriately, and if the sample size of this dataset were allowed to grow arbitrarily large, then it has been shown that the points would collect inside the accompanying limit set. Supposing we knew the underlying probability distribution for this data, i.e., how the data was generated, then it is possible to obtain a closed-form expression for the limit set's boundary. It has been shown in Nolde (2014) and Nolde and Wadsworth (2022) that this set can be used to describe the extremal dependence structure of a random vector. That is, the geometry can be used to report if all the components of our random vector experience simultaneous extremes, and which subgroups can be extreme together when the remaining components are at moderate levels. Furthermore, the geometry can be use to describe several key coefficients proposed in recent decades to describe the rate at which extremes can occur in the random vector, linking several key ideas once thought to be disjointed.

Until recently, all work done in the geometric framework has been purely probabilistic, using knowledge of data's underlying distribution. With this knowledge, the limit set has been derived and the extremal dependence structure has been described. What has yet to be done is use an observed, finite-dimensional dataset to estimate the limit set corresponding to the unseen underlying distribution of the data. In this work, we show that this is possible when considering a radial-

angular coordinate system rather than the Cartesian system. When conditioning above a high radial threshold, we derive the radial component's distribution for several parametric examples, showing that it depends on the limit set boundary. Maximizing the corresponding likelihood effectively estimates the limit set boundary and therefore the geometry of the data. This is first done assuming parametric forms on the geometry. We then extend to the semi-parametric setting, were we show in smooth Bayesian and piecewise-linear frequentist approaches the limit set boundary can be obtained without making any parametric assumptions on the data. A candidate model for the angular distribution can also be obtained which directly depends on the dataset's geometry. With these distributions being defined above a high radial threshold, we can draw radial-angular observations that, once converted to Cartesian coordinates, represent an extreme sample lying in previously unprecedented regions of the joint tail. This allows for the accurate estimation of the frequency of potentially catastrophic events. We work in the entire angular domain, making the geometric approach the first to perform accurate inference across the entire tail region of random vectors.

This thesis is organised in the following manner. Chapter 2 covers some of the background material mentioned above in greater detail. In it, the previous developments for extreme value analysis in the univariate and multivariate setting will be described, and the possible situations in which each method can fail will be discussed. These shortcomings motivate the geometric framework, whose recent probabilistic developments will be discussed further. A method to estimate the geometry of data using parametric assumptions will be presented in Chapter 3. A Bayesian semiparametric method is introduced in Chapter 4, along with a piecewise-linear method in Chapter 5. In Chapter 6, an attempt is made to link the geometric framework to the more classical multivariate extremes framework by inspecting the geometry of generalised Pareto random vectors, a key family of distributions in the classical extremes setting. Chapter 6 also aims to highlight the pros, cons, and main differences between classical multivariate extremes and the geometric framework.

Chapter 2

Background Material

2.1 Extreme value analysis in the univariate case

Extreme value analysis was first developed in the univariate setting. Suppose the random variable X describes the behaviour of some process with continuous-valued observations. For example, measurements of CO levels can be thought of as observations from the random variable X. Perhaps the most typical way of estimating the behaviour of the distribution of X at unseen levels is through the maximum of a sequence of n observations $X_1, \ldots, X_n$ so that interest lies in the modelling of $M_n = \max\{X_1, \ldots, X_n\}$. If there exists sequences $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq z\right) \longrightarrow G(z) \quad ; \quad n \to \infty, \tag{2.1}$$

where G is a non-degenerate distribution function, then G is the distribution function corresponding to a generalised extreme value (GEV) distribution, and has the form

$$G(z; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\} \tag{2.2}$$

with $a_+ = \max(a, 0)$, for some location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$, and shape parameter $\xi \in \mathbb{R}$ (Fisher and Tippett (1928); Gnedenko (1943); see also Coles (2001)). The GEV has support $\{z \in \mathbb{R} : 1 + \xi(z-\mu)/\sigma > 0\}$. The shape parameter ξ is of particular interest, and through the Extremal Types Theorem

(Theorem 3.1 of Coles (2001)), the value of ξ can dictate which family G belongs to. When $\xi = 0$, G belongs to the Gumbel family with support $z \in \mathbb{R}$. When $\xi > 0$, G belongs to the Fréchet family with support $z \in [\mu - \sigma/\xi, \infty)$. When $\xi < 0$, G belongs to the Weibull family with support $z \in (-\infty, \mu - \sigma/\xi]$. However, when performing statistical analysis it is more convenient to consider the unifying form presented in equation (2.2).

Given a set of observations $x_1, \ldots, x_n$ of a random variable X, suppose one wishes to make statements about previously unseen extreme observations of X by inferring the parameters (μ, σ, ξ) of the GEV. Given data, it is generally not known what sequences of constants $\{a_n\}$ and $\{b_n\}$ satisfy (2.1) in order to assume the GEV form. However, by rearranging within (2.2), it follows that

$$\Pr(M_n \leq z) \approx G(z; \mu^*, \sigma^*, \xi)$$

for large n, where $\mu^* = a_n\mu + b_n$ and $\sigma^* = a_n\sigma$. As a result, it is not necessary to consider the normalising sequences when fitting the GEV, and a likelihood-based approach can be obtained to estimate unknown location, scale, and shape parameters (μ, σ, ξ). To do this, the n datapoints are split into m partitions (or "blocks"), and the maximum value in each block is taken resulting in m block maxima observations $z_1, \ldots, z_m$. The block maxima are assumed to be iid observations from a GEV distribution, and parameter estimates can be obtained by minimising the negative log-likelihood constructed from the derivative of distribution function (2.2):

$$\begin{aligned}\ell(\mu, \sigma, \xi; z_1, \ldots, z_m) =& m \log \sigma + \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^{m} \log \left\{ \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]_+ \right\} \\ &+ \sum_{i=1}^{m} \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]_+^{-1/\xi}. \end{aligned} \tag{2.3}$$

This is only possible if every block maximum point z_i lies in the domain of the GEV, $z_i \in (\mu - \sigma/\xi, \infty)$ for $i = 1, \ldots, m$. With parameter estimates obtained from standard optimisation tools, one could extrapolate into unobserved regions of the tail of the distribution of maxima. For example, given n observation of CO, one could estimate parameters (μ, σ, ξ), then use the maximum likelihood estimates

(MLEs) $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ to estimate the probability that the maxima of a block of size n/m exceeding $\tilde{z} > z_j$, $j = 1, \ldots, m$, through $\widehat{\Pr}(Z > \tilde{z}) = 1 - G(\tilde{z}; \hat{\mu}, \hat{\sigma}, \hat{\xi})$. Within the block maxima approach, the notion of *return-levels* is a popular and interpretable tool. For a return period $1/p$, $p \in (0, 1)$, and m block maxima observations, a quantity of interest is the level z_p that is expected to be exceeded on average once every $1/p$ blocks. As an example, suppose $m = 365$ days, interest may lie in the level of CO expected once every $1/p$ years (ignoring leap years). Given parameter estimates, one can obtain an estimate $\hat{z}_p$ by solving for z_p in the relation $p = 1 - G(z_p; \hat{\mu}, \hat{\sigma}, \hat{\xi})$. Since the resulting estimate is simply a function of the MLEs, associated uncertainties related to $\hat{z}_p$ can be obtained via the delta method. Return levels of the original random variable X can be similarly obtained by solving for z_p in $p = 1 - G^{1/n}(z_p; \hat{\mu}, \hat{\sigma}, \hat{\xi})$, a result of the property that n iid copies of X satisfy $\Pr(\max(X_1, \ldots, X_n) \leq z) = [\Pr(X \leq z)]^n$.

With this simple inference procedure, the block maxima approach is potentially a powerful tool for practitioners. The GEV distribution within the block maxima framework has been used for modelling tasks in areas like climatology (Carter and Challenor, 1981; Buishand, 1989; Padoan and Rizzelli, 2024) and oceanography (de Haan, 1990; Tawn, 1992; Robinson and Tawn, 1997; Jonathan and Ewans, 2013). Despite its practicality, statistical inference of the GEV model may shed light on some drawbacks. For one, in minimising the objective function in (2.3), it is nearly impossible to estimate $\hat{\xi}$ near 0 when desired. Therefore, if a practitioner notices small estimates of ξ, they could then re-estimate the parameters using the Gumbel distribution function in place of (2.2), but there is no principled way of deciding the threshold around 0 for such a re-fit. Furthermore, asymptotic normality of the maximum likelihood estimates $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ is only available if $\hat{\xi} > -0.5$ (Smith, 1985; Bücher and Segers, 2017). Though values of the shape parameter rarely lie below -0.5 in practice, it still may occur, leaving the practitioner unable to get the usual uncertainty quantities related to maximum likelihood estimates. Another downside of the block maxima approach is that much of the initial data is discarded during

inference. Several of these discarded extremal observations may be otherwise useful in describing the tail behaviour of X.

Given the limitations of the block maxima approach for inference of univariate extremes, it may be beneficial to instead consider the behaviour of X above some high threshold value u. For a random vector X, the Pickands–Balkema–De Haan theorem (Balkema and de Haan, 1974; Pickands, 1975) states that for a large class of distributions, the conditional distribution $X \mid X > u$ converges to a non-degenerate excess distribution as $u \to \infty$, akin to the convergence presented in equation (2.1). For high enough u, a consequence of Theorem 1.5.1 of Ledbetter et al. (1983) is that we can use the approximation

$$\Pr\left(X > x | X > u\right) \approx \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]_+^{-1/\xi}. \tag{2.4}$$

The form (2.4) suggests the upper-tail behaviour follows the generalised Pareto (GP) distribution, with location parameter $u \in \mathbb{R}$, scale parameter $\sigma > 0$, shape parameter $\xi \in \mathbb{R}$. The GP distribution is commonly characterised through its survival function,

$$\Pr\left(X > x\right) \approx \zeta_u \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]_+^{-1/\xi} \tag{2.5}$$

for x in the domain $\{x \in \mathbb{R} : 1 + \xi(x-u)/\sigma > 0,\, x > u\}$, where $\zeta_u = \Pr\left(X > u\right)$ can be estimated empirically as u is often chosen to be a high quantile yet still in the range of observed data (Pickands (1975); Davison and Smith (1990); see also Coles (2001)).

Use of the GP distribution to model threshold exceedances is referred to as the *peaks-over-threshold* framework. Given a set of observations $x_1, \ldots, x_n$, the first step to statistical inference is choosing a high threshold u. Once the threshold u is chosen, maximum likelihood estimates of the remaining GP parameters (σ, ξ) can be obtained by minimising the negative log-likelihood constructed from the derivative of the left-tail probability associated with (2.4):

$$\ell(\sigma, \xi; x_1, \ldots, x_n) = -\sum_{i:x_i > u} \log\left[\sigma^{-1}\left[1 + \xi\left(\frac{x_i - u}{\sigma}\right)\right]_+^{-1/\xi - 1}\right]$$

With MLEs $(\hat{\zeta}_u, \hat{\sigma}, \hat{\xi})$, one can extrapolate far into the tail of X by directly computing the probability of exceeding x^* via (2.4), where $x^* > x_i$, $i = 1, \ldots, n$. As in the block maxima setting, the notion of return-levels is a popular tool in the peaks-over-threshold framework. Commonly formulated as the level that is exceeded on average once every m observation, the return level x_m is the solution to $1/m = \Pr(X > x_m)$. Given MLEs $(\hat{\sigma}, \hat{\xi})$, an estimate $\hat{x}_m$ could be obtained through inverting the GP upper-tail probability in the approximation (2.5), with the associated uncertainty obtained via the delta method, as in the block maxima setting.

Due to its intuitive nature and easy applicability, the GP distribution within the peaks-over-threshold framework has been influential, having been applied to fields such as climatology (Grady, 1992; Tabari, 2021; Pacifici et al., 2025), extreme wind events (Walshaw, 1994; Simiu and Heckert, 1996; Outten and Sobolowski, 2021), and hydrology (Fitzgerald, 1989; Acero et al., 2011; Agilan et al., 2021). While the peaks-over-threshold method provides the practitioner with a better use of data to study the tail behaviour of a random process, the numerical issues when $\xi \approx 0$ that were present in the model fitting stage of the GEV are present here as well. In addition, a major drawback of the peaks-over-threshold approach is in the selection of the threshold u. Davison (1984), Smith (1984), and Davison and Smith (1990) all advocate for the use of the *mean residual life plot* in selecting u. The idea is that when modelling the excesses of n iid copies $X_1, \ldots, X_n$ of X above a high threshold u, a consequence of the GP distribution is that the expected value $\mathbb{E}[X - u \mid x > u]$ should be linear in u. Therefore, in plotting an empirical estimate of this expectation against u, select the lowest value of u above which this plot seems linear. In addition to this being a visual tool and thus prone to user error, there may be several values of u above which the plot appears linear. There may also be no such points. Also outlined in Coles (2001), an alternative method is by fitting the GP distribution at a variety of thresholds u and selecting the lowest value such that threshold stability holds, a key property of the GP distribution. In this method, at a proposed threshold u, a GP is fitted at u and at a higher threshold, and the parameter estimates for

the σ and ξ should be the same at both thresholds. This may not be ideal, as high variability in MLEs may be introduced if fitting above a threshold that is too high. Since the peaks-over-threshold framework has been introduced, a number of alternative methods for selecting u have been introduced, with more complications arising if the iid assumption in collected data is not satisfied. However, a "one size fits all" approach has not been introduced, leaving the practitioner to experiment with the dozens of proposed methods available.

Despite this, both the block maxima and the peaks-over-threshold frameworks lead to useful methodological procedures that have applications in numerous fields. In fact, the similarities between the GEV distribution function (2.2) and the GP conditional survival function (2.4) is no coincidence. Through the use of Poisson point processes, it can be determined that the the block maxima and peaks-over-threshold frameworks are equivalent. By considering the number of points lying above some large value u in some interval and letting the sample size grow large, one recovers a Poisson point process with mean measure $\Lambda(u)$. The events of the recentered and rescaled maxima being below a certain value (see (2.1)) corresponds to the observing no values in a particular region. Using $\Lambda(u)$ and the void probability expression of a Poisson point process, we recover the GEV (2.2). Furthermore, if Λ is factorised into a time component Λ_1 and an amplitude component Λ_2, the conditional probability (2.4) in the peaks-over-threshold approach can be recovered via $\Lambda_2([x,\infty))/\Lambda_2([u,\infty))$ It is also worth nothing that extensions to both the GEV and the GP distributions in the case of nonstationary data have been proposed by several authors, but this is beyond the scope of this introduction.

2.2 Multivariate extremes

Suppose one wishes to model the extreme tail behaviour of multiple simultaneous processes. Can extreme value theory be extended to the multivariate setting? As it turns out, by making assumptions on the underlying multivariate distribution of the

processes, one could accurately make statements about the nature of the data beyond its observable domain. Classical approaches like multivariate regular variation make assumptions on the limiting behaviour of the joint distribution function. Though these assumptions hold often in known distributions and lead to rich theoretical guarantees, they result in inferential tools that are often not flexible enough to extrapolate into the entire tail region. Furthermore, the resulting extrapolation assumed a tail decay rate that is only valid if the variables obtain simultaneous extreme values. To remedy these issues, an approach was introduced in which a random vector conditioned on one variable being large is assumed to have a non-degenerate limiting distribution function. Again, this assumption tends to hold for a wide variety of multivariate distributions; however, in-practice this is modelled using a misspecified multivariate Gaussian distribution, and is not flexible if the underlying random vector has a complex mix of extremal dependencies. As an alternative, the work presented in this thesis instead makes an assumption on the regularity of the limiting joint density on the log-scale. Like the other approaches, this assumption holds for several known multivariate distributions. By making this assumption, a rich characterisation of the *entire* multivariate tail presents itself. This allows us to perform extrapolation in *any* direction for *any* extremal dependence structure.

2.2.1 Tail dependence of random vectors

To approach extreme value analysis in the multivariate setting, one needs to take into account the extremal dependence behaviour between the components of the multivariate process. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^\top$ be a random vector where each of the d components follow a univariate distribution with distribution function F_{Y_j}. An important characterisation of $\boldsymbol{Y}$ is to distinguish which collections of components experience simultaneous extremes. This is traditionally characterised via the following extremal dependence coefficient,

$$\chi_C(u) = \frac{1}{1-u} \Pr\left[F_{Y_j}(Y_j) > u\,,\, j \in C\right] \quad ; \quad C \subseteq \{1, \ldots, d\},\ |C| \geq 2$$

for $u > u_0$, where $u_0 \in [0,1]$ is close to 1, and $|C|$ denotes the cardinality of the set C. If $\lim_{u\to 1}\chi_C(u) > 0$, then we say that the variables indexed by C experience simultaneous extreme values, and are therefore *asymptotically dependent*. Conversely, if $\lim_{u\to 1}\chi_C(u) = 0$, we say that the variables indexed by C do not experience simultaneous extreme values, and are therefore *asymptotically independent*.

In trying to examine the relationship between components of a random vector in the tails, it is beneficial to have them on the same scale. If one component is orders of magnitude larger than the others, for example, it may obfuscate any dependence information. For this reason, *copulas* have become a useful tool in extremal dependence modelling. Given an unstandardised random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^\top$ with potentially unequal marginal distribution functions $F_{Y_1}, \ldots, F_{Y_d}$, the copula is the joint distribution function corresponding to $\boldsymbol{Y}$ in uniform margins,

$$\begin{aligned} C(u_1, \ldots, u_d) &= \Pr\left[F_{Y_1}(Y_1) \leq u_1, \ldots, F_{Y_d}(Y_d) \leq u_d\right] \\ &= \Pr\left[Y_1 \leq F_{Y_1}^{-1}(u_1), \ldots, Y_d \leq F_{Y_d}^{-1}(u_d)\right], \end{aligned}$$

where $F_{Y_j}^{-1}$ is the quantile function corresponding to Y_j. The copula is the joint distribution function of the vector $\boldsymbol{U} = (F_{Y_1}(Y_1), \ldots, F_{Y_d}(Y_d))^\top$ corresponding to $\boldsymbol{Y}$, but in standard uniform margins. Given an observed dataset $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\} \in \mathbb{R}^d$ with unknown marginal distributions, a common way of obtaining observations in uniform margins is through the empirical distribution function,

$$\hat{F}_{Y_j}(y) = \frac{1}{n+1}\sum_{i=1}^{n} \mathbf{1}_{(-\infty, y_{i,j}]}(y). \tag{2.6}$$

In the expression (2.6), $\mathbf{1}_A(x)$ is the indicator function, taking values $\mathbf{1}_A(x) = 1$ if $x \in A$ and $\mathbf{1}_A(x) = 0$ if $x \notin A$. Using the theoretical property from Pickands (1975) that univariate threshold exceedances follow a GP distribution, Coles and Tawn (1991) take the approach of fitting the GP model (2.5) above a high marginal threshold u. As points are more dense, the empirical distribution function is suitable in the bulk of the marginal data for approximating the underlying true distribution function.

Performing the marginal transformation this way ensures that the most extreme observations in a dataset preserve their relative magnitude, and the availability of a parametric marginal distribution function above a threshold is useful when transforming back to original margins after inference. This marginal transformation has since been adapted for a GPD fit in both the left and right tails of the marginal distribution (see, for example, Simpson and Tawn (2024a) and Papastathopoulos et al. (2025)). Define u_{L} and u_{R} to be thresholds in the left and right tails of the marginal distribution, respectively. Once GP parameters are estimated beyond these thresholds, the resulting distribution function estimate is given by

$$\hat{F}_{Y_j}(y) = \begin{cases} 1 - (1 - \zeta_{u_{\mathrm{L}}}) \left[1 + \hat{\xi}_{\mathrm{L}} \left(\frac{u_{\mathrm{L}} - y}{\hat{\sigma}_{\mathrm{L}}}\right)\right]^{-1/\hat{\xi}_{\mathrm{L}}} & ; \quad y < u_{\mathrm{L}} \\ \frac{1}{n+1} \sum_{i=1}^{n} \mathbf{1}_{(-\infty, y_{i,j}]}(y) & ; \quad y \in (u_{\mathrm{L}}, u_{\mathrm{R}}) \\ 1 - \zeta_{u_{\mathrm{R}}} \left[1 + \hat{\xi}_{\mathrm{R}} \left(\frac{y - u_{\mathrm{R}}}{\hat{\sigma}_{\mathrm{R}}}\right)\right]^{-1/\hat{\xi}_{\mathrm{R}}} & ; \quad y > u_{\mathrm{R}} \end{cases}$$

Note that, while $\zeta_{u_{\mathrm{R}}}$ is easily estimated using the empirical distribution function, $\zeta_{u_{\mathrm{L}}}$ is estimated using $1 - (n+1)^{-1} \sum_{i=1}^{n} \mathbb{1}[y_{ij} \geq u_{\mathrm{L}}]$, since few data points lie below u_{L}. Also note that in some cases, the left tail of the original marginal distribution may have high density; therefore, only using a GPD fit in the right tail is sufficient in approximating the underlying distribution function.

Using the copula directly, one could make statements about the tail behaviour between components of $\boldsymbol{Y}$; however, many statistical inference tools for multivariate extremes rely on other marginal forms. Suppose F_* is the univariate quantile function corresponding to the target marginal distribution, then $(F_*^{-1}(U_1), \ldots, F_*^{-1}(U_d))^\top$ is the vector associated to $\boldsymbol{Y}$ whose margins follow the desired distribution. The more common approaches benefit from components having a heavy-tailed distribution. Common choices are Pareto margins, whose quantile function is given by $F_*^{-1}(u) = 1/(1-u)$, or the Fréchet distribution, whose quantile function is given by $F_*^{-1}(u) = -1/\log(u)$. Recently introduced frameworks for multivariate extremes rely on light-tailed margins, such as exponential margins with quantile function $F_*^{-1}(u) = -\log(1-u)$ or Laplace margins with quantile

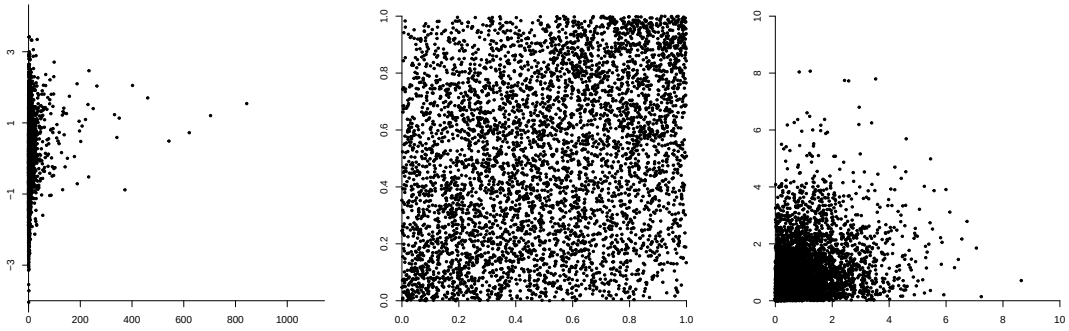

Figure 2.1: Left: Observations from a joint distribution with unequal margins. Centre: The same data, standardised to uniform margins. Right: The standardised data in exponential margins.

function $F_*^{-1}(u) = \log(2u)$ if $u \leq 1/2$ and $F_*^{-1}(u) = -\log(2(1-u))$ if $u > 1/2$. These marginal transformations are illustrated in Figure 2.1. The original data was constructed such that, while the joint dependence is Gaussian, the first component has standard Fréchet margins and the second component is standard normal. In this construction, it is not clear whether or not the two variables are dependent or independent in the tails. In uniform margins, it becomes obvious that data collects in the lower-left and upper-right corners. This is a sign that the two variables don't achieve extreme values simultaneously, a feature known to hold in the Gaussian dependence. Lastly, data presented in exponential margins can be used in a variety of frameworks to obtain information on the joint tail, such as estimating the probability of observing data in a range of extreme values. In the right-hand plot of Figure 2.1, for example, we see that the data does not show a tendency to obtain simultaneous extremes, confirming what is observed on the uniform scale. Throughout this work, we will see that standardising to marginal distributions with exponential-type behaviour leads to joint distributions with desirable mathematical properties.

2.2.2 Multivariate and hidden regular variation

In addition to gaining insight into a random vector's extremal dependence structure, we also wish to use this information to estimate probabilities of lying in the tails of the underlying joint distribution. A classical approach to going about this is by assuming its distribution function is multivariate regularly varying (MRV). In the MRV framework, it is assumed that, when appropriately scaled, the probability of random vectors lying in a region away from the origin converges vaguely to a Radon measure ν that lives on $[\mathbf{0}, \boldsymbol{\infty}] \setminus \{\mathbf{0}\}$. For some arbitrary random vector $\boldsymbol{Y}$ with common margins, this amounts to assuming that there exists a function $b(t) \to \infty$ as $t \to \infty$ such that

$$t \Pr\left[\frac{\boldsymbol{Y}}{b(t)} \in \cdot\right] \to \nu \quad ; \quad t \to \infty, \tag{2.7}$$

with convergence holding vaguely on the cone $[\mathbf{0}, \boldsymbol{\infty}] \setminus \{\mathbf{0}\}$. Resnick (2007) remark that, if the margins Y_j are identically distributed with distribution function F_Y, a good choice for the scaling function is $b(t) = \left(\frac{1}{1-F_Y}\right)^{\leftarrow}(t)$. Suppose we consider the random vector $\boldsymbol{Z}$ with standard Pareto margins. Further suppose that we consider a set $A \subset (\mathbf{1}, \boldsymbol{\infty}]$ and that operations on sets behave as expected, i.e., $h(A) = \{\boldsymbol{x} \in (\mathbf{1}, \boldsymbol{\infty}] \mid h^{-1}(\boldsymbol{x}) \in A\}$ for h invertible and the operation $h^{-1}(\boldsymbol{x})$ done componentwise. In this setting, the MRV assumption can be expressed as

$$\Pr[\boldsymbol{Z} \in tA] \sim t^{-1}\nu(A) \quad ; \quad t \to \infty, \tag{2.8}$$

where "$\sim$" denotes asymptotic equivalency: $a(t) \sim b(t)$ if $\lim_{t\to\infty} a(t)/b(t) = 1$. Equivalently, suppose $\boldsymbol{X} = \log(\boldsymbol{Z})$ has margins $X_j = \log(Z_j)$ following the lighter-tailed standard exponential distribution. Then assumption (2.8) amounts to

$$\Pr[\boldsymbol{X} \in B + t] \sim e^{-t}\nu(e^B) \quad ; \quad t \to \infty. \tag{2.9}$$

where $B = \log(A) \in (\mathbf{0}, \boldsymbol{\infty}]$. If the assumption of MRV is assumed to hold, then one can begin to estimate extremal probabilities in the multidimensional space. Given n independent observations of the random vector $\boldsymbol{X}$, suppose that the value of $t > 0$ is large, but the set $B + t$ lies within of the range of data. Suppose we have $v > 0$

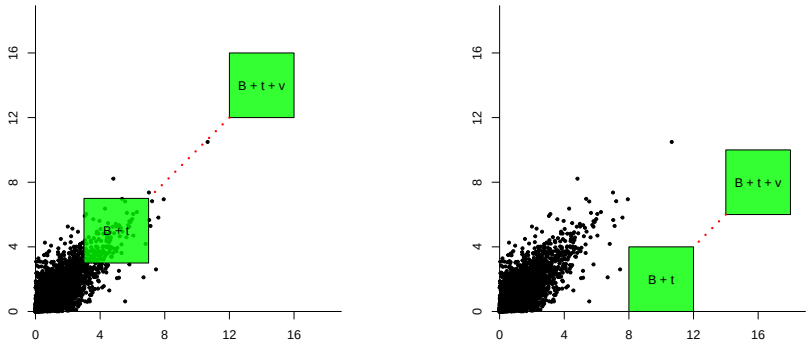

Figure 2.2: Two examples of extrapolation within the multivariate regular variation and hidden regular variation frameworks for bivariate data in standard exponential margins.

large such that the set of interest $B + t + v$ is outside of the range of observed data. By assumption (2.9), it also holds that

$$\begin{aligned} \Pr\left[\boldsymbol{X} \in B + t + v\right] &\sim e^{-(t+v)}\nu(e^B) \\ &\sim e^{-v}\Pr\left[\boldsymbol{X} \in B + t\right] \quad ; \quad v, t \to \infty. \end{aligned} \tag{2.10}$$

Therefore, under MRV, one can scale back the set linearly by a factor of v until it is in the range of given data and can be estimated empirically. The e^{-v} term is a correction factor accounting for the scaling. This scaling procedure is shown for exponential margin random vectors in Figure 2.2 for two different sets B.

In some settings, it may be more suitable to perform inference in a radial-angular setting rather than the usual Cartesian framework, and the assumption of MRV can equivalently be adapted for this. In the Pareto margin setting, define the radial and angular components corresponding to $\boldsymbol{Z}$ via

$$(R, \boldsymbol{W}) = (\|\boldsymbol{Z}\|, \boldsymbol{Z}/\|\boldsymbol{Z}\|) \in (0, \infty] \times \mathcal{S}_{d-1},$$

where $\|\cdot\|$ is a norm and $\mathcal{S}_{d-1} = \{\boldsymbol{w} \in \mathbb{R}^d \mid \|\boldsymbol{w}\| = 1\}$ is the $(d-1)$-dimensional simplex. Through Theorem 6.1 of Resnick (2007), the convergence assumption (2.8)

is equivalent to

$$t \Pr\left[\left(\frac{R}{b(t)}, \boldsymbol{W}\right) \in \cdot\right] \xrightarrow{v} \nu = \nu_\alpha \times H \quad ; \quad t \to \infty.$$

$b(t) = \left(\frac{1}{1-F_R}\right)^{\leftarrow}(t)$, ν_α is a Lévy measure of the radial component R on $(0, \infty]$. Pareto margins are a popular choice, because in this setting, $\alpha = 1$ and ν_α reduces to the simple form $\nu_\alpha((-\infty, x]) = 1/x$. The component H is the so-called *spectral measure*, the probability measure associated with the angular component $\boldsymbol{W}$ (de Haan, 1970; Resnick, 2007). The MRV assumption has found many practical applications when estimating the occurrence of rare events in real-world applications such as finance (Resnick, 2004; Cai et al., 2011; Das and Fasen-Hartmann, 2018; Kiriliouk et al., 2019), oceanography (Coles and Tawn, 1991, 1994; De Haan and De Ronde, 1998), and hydrology (Kiriliouk et al., 2019).

The MRV framework naturally leads to the notion of a multivariate analogue of the peaks-over-threshold approach introduced in Section 2.1. In standard exponential margins, this amounts to modelling for excesses in the domain $\{\boldsymbol{x} \in \mathbb{R}^d \mid \boldsymbol{x} > \boldsymbol{0}\}$. These excesses are modelled through the multivariate generalised Pareto (MGP) distribution (Tajvidi, 1995; Rootzén and Tajvidi, 2006; Rootzén et al., 2018a,b), defined through an exponent measure Λ that lives on $[-\boldsymbol{\infty}, \boldsymbol{\infty}) \setminus \{-\boldsymbol{\infty}\}$ and is finite on all sets bounded from below. The exponent measure is evaluated using the convention $\Lambda(\boldsymbol{x}) := \Lambda([-\boldsymbol{\infty}, \boldsymbol{x}) \setminus \{-\boldsymbol{\infty}\})$, and $\Lambda^c(\boldsymbol{x}) := \Lambda([-\boldsymbol{\infty}, \boldsymbol{\infty}) \setminus [-\boldsymbol{\infty}, \boldsymbol{x}))$ is the exponent measure evaluated at the complement of the set $[-\boldsymbol{\infty}, \boldsymbol{x})$. The distribution and density functions of MGP random vectors are respectively defined for random vectors in exponential margins:

$$\begin{aligned} \Pr(\boldsymbol{X} \leq x) &= \lim_{y \to \infty} \Pr(\boldsymbol{Y} - y\boldsymbol{1} \leq x | \boldsymbol{Y} > y\boldsymbol{1}) \\ &= \frac{\Lambda^c(\min(\boldsymbol{0}, \boldsymbol{x})) - \Lambda^c(\boldsymbol{x})}{\Lambda^c(\boldsymbol{0})}, \\ f(\boldsymbol{x}) &= \frac{\lambda(\boldsymbol{x})}{\Lambda^c(\boldsymbol{0})}, \end{aligned} \tag{2.11}$$

where $\boldsymbol{Y}$ is some random vector and λ is the intensity corresponding to Λ,

$$\Lambda(B) = \int_B \lambda(\boldsymbol{x}) d\boldsymbol{x},$$

where integrating over the set $B \subseteq \mathbb{R}^d$ denotes integrating over all elements $\boldsymbol{x} \in \mathbb{R}^d$ such that $\boldsymbol{x} \in B$. If the convergence in (2.11) holds, then we say that $\boldsymbol{Y}$ is in the domain of attraction of the MGP distribution. There are popular parametric choices for Λ, most notably the Hüsler-Reiss model (Engelke et al., 2015; Zhen Wai Olivier and Dombry, 2017) that has seen recent popularity in modelling sparse high-dimensional modelling tasks present in hydrology and complex airline networks (Engelke et al., 2022; Hentschel et al., 2024).

Despite being a powerful tool used for several decades in practical applications, MRV has some limitations. For one, the tail decay factor e^{-v} in the approximation (2.10) only characterises tail decay behaviour when all d components are asymptotically dependent. Furthermore, the direction of extrapolation is limiting; it is entirely possible that the less extreme set $B + t$ is also not in the range of the n observations (see Figure 2.2). To remedy the tail decay correction for variables that don't obtain simultaneous extremes, the notion of hidden regular variation (HRV) was introduced (Ledford and Tawn, 1996, 1997), leading to an update of the convergence (2.7). In Pareto margins with shape parameter $\alpha > 0$, this amounts to the assumption

$$t \Pr\left[\frac{\boldsymbol{Z}}{b_\eta(t)} \in \cdot\right] \to \nu_\eta \quad ; \quad t \to \infty \tag{2.12}$$

where the scaling function b_η is chosen to be regularly varying of order $\eta \in [0, 1/\alpha]$ and the limit measure ν_η is homogeneous of order $-1/\eta$. By the same intuition used in the MRV framework, (2.12) can be reformulated to perform the extrapolation, stated here in exponential margins:

$$\Pr[\boldsymbol{X} \in B + t + v] \sim e^{-v/\eta} \Pr[\boldsymbol{X} \in B + t] \quad ; \quad v, t \to \infty.$$

The added parameter η is called the *coefficient of tail dependence*, and accounts for the possibility of joint tails growing at a slower rate, allowing for statistical

modelling when the d components of $\boldsymbol{X}$ don't obtain simultaneous extremes. Through comparing the convergence (2.12) to that of (2.7), HRV can be thought of as a generalisation of MRV in the sense that, when all d components of $\boldsymbol{X}$ are asymptotically dependent, we recover the MRV setting and it follows that $\eta = 1$. Otherwise, we have $\eta < 1$. The HRV framework has successfully been used in the study of internet usage, hydrology, and extreme weather (Heffernan and Resnick, 2005; Ramos and Ledford, 2009).

2.2.3 The conditional approach

While the HRV framework solved the tail decay issues that MRV suffers from, it still only extrapolates to the joint tail region (Wadsworth and Tawn, 2013). To remedy the issues with extrapolation within MRV and HRV, and to be able to account for a wider range of extremal dependence structures, a conditional modelling approach (Heffernan and Tawn, 2004) was introduced. Given the d-dimensional random vector $\boldsymbol{X}$ in light-tailed margins and an index $j \in \{1, \ldots, d\}$, the conditional framework assumes the existence of normalising functions $\boldsymbol{a}_j, \boldsymbol{b}_j : \mathbb{R} \to \mathbb{R}^{d-1}$ such that

$$\Pr\left(\boldsymbol{X}_{-j} \leq \boldsymbol{a}_j(x_j) + \boldsymbol{b}_j(x_j)\boldsymbol{z}_j | X_j = x_j\right) \to G_j(\boldsymbol{z}_j) \tag{2.13}$$

as $x_j \to \infty$, where $\boldsymbol{X}_{-j}$ is the vector $\boldsymbol{X}$ with the j^{th} component removed, and G_j is a non-degenerate distribution function. Defining the *residual* vector as $\boldsymbol{Z}_j = \boldsymbol{b}_j(x_j)^{-1}\left(\boldsymbol{X}_{-j} - \boldsymbol{a}_j(x_j)\right)$ and considering a fixed value $x > 0$, the assumption (2.13) can be reformulated as

$$\Pr\left(\boldsymbol{Z}_j \leq \boldsymbol{z}_j, X_j - x_0 = x | X_j > x_0\right) \to e^{-x} G_j(\boldsymbol{z}_j)$$

as $x_0 \to \infty$. For $k \in \{1, \ldots, d\} \setminus \{j\}$, a common choice of normalising functions is $a_{j,k}(x) = \alpha_{j,k} x$ and $b_{j,k}(x) = x^{\beta_{j,k}}$. With this in mind, the main modelling assumption is that there exists parameters α_k and $\beta_k > 0$ such that, for large x_0,

$$Z_{j,k} | X_j > x_0 = \left.\frac{X_k - \alpha_{j,k} X_j}{X_j^{\beta_{j,k}}}\right| X_j > x_0 \sim \mathcal{N}(\mu_k, \sigma_k^2).$$

That is, after singling out component j from the random vector $\boldsymbol{X}$, it is assumed that a residual value constructed from each of the remaining components can independently be modelled using the normal distribution. Once the $2(d-1)$ parameter values are estimated, one can estimate the probability of lying in extremal regions using the relation

$$\Pr(\boldsymbol{X} \in B) = \Pr(X_j > x_0) \Pr(\boldsymbol{X} \in B | X_j > x_0) \tag{2.14}$$

The first probability in the right-hand side of (2.14) can be computed empirically, as the threshold level x_0 is chosen to be high, but still within the range of data. Alternatively, x_0 can be chosen such that $\Pr(X_j > x_0) = 1 - \tau$ for some value τ close to 1, but still low enough to allow for sufficient data when obtaining parameter estimates. To estimate the second probability, samples are drawn from $\boldsymbol{X} \mid X_j > x_0$, and an empirical probability is taken by counting the number of points in the extremal sample lying in region B. This sample is taken by first sampling points from the marginal standard exponential distribution corresponding to X_j. The memorylessness property of the exponential distribution is used to convert these to be observations from $X_j | X_j > x_0$. For the remaining margins, points are sampled with replacement from the empirical distribution of each of the $d-1$ residuals, i.e., components $Z_{j,k}$ of $\boldsymbol{Z}_j$. In using the sample from $X_j | X_j > x_0$ for each of the residuals $Z_{j,k}$, we induce dependence between the components of samples from $\boldsymbol{Z}_j$. Using the parameter estimates $(\widehat{\alpha}_{i,j}, \widehat{\beta}_{i.j})$ for $i \in \{1, \ldots, d\} \setminus \{j\}$ and the extremal exponential sample from $X_j | X_j > x_0$, we can convert the sampled residuals back to the marginal scale of the random vector. Joining the sample observations from each of the d margins, one ends up with vector observations from $\boldsymbol{X} \mid X_j > x_0$, and the probability $\Pr(\boldsymbol{X} \in B | X_j > x_0)$ can be empirically estimated. A convenient approach in the conditional extremes framework is to fit the residuals above a relatively high quantile x_0 and use the resulting parameter estimates to perform extrapolation above an arbitrarily large quantile $x_0^\star > x_0$. The probability of interest (2.14) can then be estimated by instead computing $\Pr(X_j > x_0^\star) \Pr(\boldsymbol{X} \in B | X_j > x_0^\star)$. For example, an extremal sample using the conditional approach is illustrated in Figure 2.3 for three

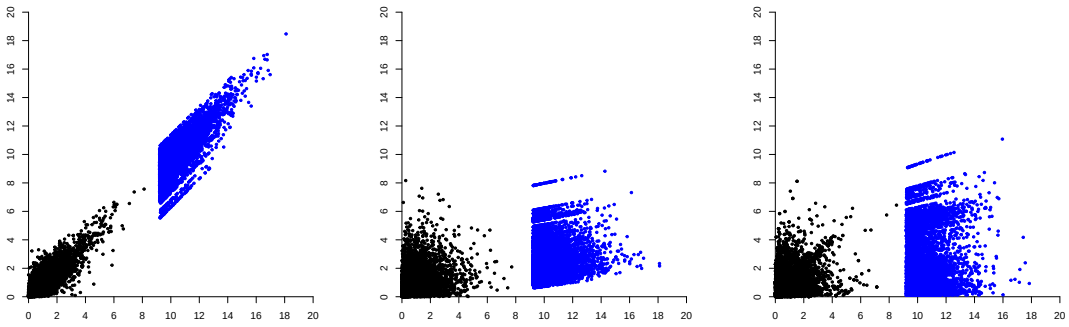

Figure 2.3: Examples of extrapolation (blue points) within the conditional extremes (Heffernan and Tawn, 2004) framework on bivariate data. From left to right, the data is asymptotically dependent, asymptotically independent, and a mixture of the two.

separate datasets by first fitting the model for $\boldsymbol{X} \mid X_1 > x_0$ with x_0 set to the 0.95 quantile of X_1, then performing extrapolation above $x_0^\star = x_0 + 4\log(10)$. Due to its flexibility, the conditional extremes approach has found numerous applications in areas such as oceanography (Ewans and Jonathan, 2014), finance (Nolde and Zhang, 2020), and hydrology (Richards et al., 2022).

While the conditional extremes approach provides a more flexible inference tool for multivariate extremes compared to MRV and HRV, it does have its own potential downsides. Particularly in higher dimensions, it isn't always obvious which X_j to condition upon being large in order to study the desired multivariate tail region. One could estimate the conditional extremes model d separate times, each time conditioning on a different variable being large; however, there is no theoretical link between the d models, meaning there is no way to join the d models to study the entire multivariate tail. There have been suggestions to bypass this difficulty in practice, such as through importance sampling (Wadsworth and Tawn, 2022). Another limitation with conditional extremes is the lack of flexibility in dependence structures on which modelling can be performed. Take for instance the data shown in Figure 2.3. The data on the right-hand plot was generated such that the joint tail grows simultaneously at a very fast rate, while each variable can obtain extreme

values on their own when the other is at intermediate values, but at a slower rate. In the conditional extremes framework, this results in two optimal values of the $\alpha_{2,1}$ parameter. We see in the extremal sample generated that, while the conditional model captured the behaviour of the asymptotically dependent and asymptotically independent data, it failed to capture both in the mixture model data, instead appearing to perform a form of interpolation between asymptotic dependence and asymptotic independence. This may lead to inaccurate estimation of the joint tail behaviour. One approach to handle this difficulty is through the use of mixture models (Tendijck et al., 2023). However, this is only feasible in low-dimensions, as the number of mixture components grows quickly as the dimension increases.

2.3 Geometric multivariate extremes

Given the potential downfalls of the aforementioned modelling approaches in multivariate extremes, it would be desirable to have an approach that was able to model across the *entire* tail region of random vectors when these vectors could have *any* dependence structure. Being able to model across the entire tail region refers to being able to make accurately predict the behaviour of a random vector when at least one of its any d components is large, and being able to model any dependence structure means that we can accurately make statements about this behaviour regardless of how any combination of the vector's components interact with each other in the tails. Recent literature has hinted towards a promising new avenue within multivariate extremes, using the geometry of the dataset's underlying data generating mechanism. Before this is discussed, it is important to define what exactly is meant by the geometry of data. Like in the approaches mentioned in Section 2.2, an assumption must be made on the data's underlying distribution before any statements can be made about its geometry. Suppose n independent copies $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ follow the same distribution as the random vector $\boldsymbol{X}$, where the margins X_j follow the standard exponential distribution. Further suppose the joint

density of $\boldsymbol{X}$ is given by $f_{\boldsymbol{X}}$. The main assumption in the geometric approach is that the log-scale of the density $f_{\boldsymbol{X}}$ is well-behaved in the tails, or

$$-\log f_{\boldsymbol{X}}(t\boldsymbol{x}) \sim tg(\boldsymbol{x}) \quad ; \quad t \to \infty. \tag{2.15}$$

The limiting function g is called the *gauge function*, is positive-valued, and is 1-homogeneous. If the density $f_{\boldsymbol{X}}$ corresponding to random copies $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ satisfies assumption (2.15), then as the sample size n grows arbitrarily large, the scaled sample cloud $N_n = \{\boldsymbol{X}_1/\log n, \ldots, \boldsymbol{X}_n/\log n\}$ converges in probably onto a *limit set*, defined by

$$G = \left\{\boldsymbol{x} \in \mathbb{R}^d \,\middle|\, g(\boldsymbol{x}) \leq 1\right\}. \tag{2.16}$$

That is, as n increases, the probability of observing points in N_n outside of G tends to 0. As an illustration, Figure 2.4 displays sample clouds of size $n = 5,000$ in dimensions $d = 2$ and 3, with the corresponding limit set and gauge function unit level set corresponding to the underlying distribution. By definition of G, the unit level set of the gauge function g is a key quantity in the geometric framework. Through its unit level set, the gauge function defines the limit set, and therefore the geometry of random vectors. This inherent link presents itself in the properties of both the limit set G and the gauge function g. A key property of G is that it is star-shaped, i.e., $\boldsymbol{x} \in G$ implies $t\boldsymbol{x} \in G$ for $t \in [0, 1]$, and is compact. The coordinatewise supremum of G depends on the choice of marginal distribution of $\boldsymbol{X}_i$ and of scaling in the sample cloud N_n, though can often be set to $(1, \ldots, 1)^\top$. The gauge function g is continuous, positively-valued, and 1-homogeneous. Given a gauge function g satisfying these conditions, one can obtain G through the definition (2.16). Furthermore, given a valid limit set G, one can obtain gauge function values through $g(\boldsymbol{x}) = \inf\{t \geq 0 : \boldsymbol{x} \in tG\}$. Throughout this work, we will largely restrict ourselves to non-degenerate limit sets and gauge functions. That is, rays from the origin intersect ∂G at a single point.

The limit set G and the corresponding gauge function g has been used to check for asymptotic independence in random vectors (Balkema and Nolde, 2010). This

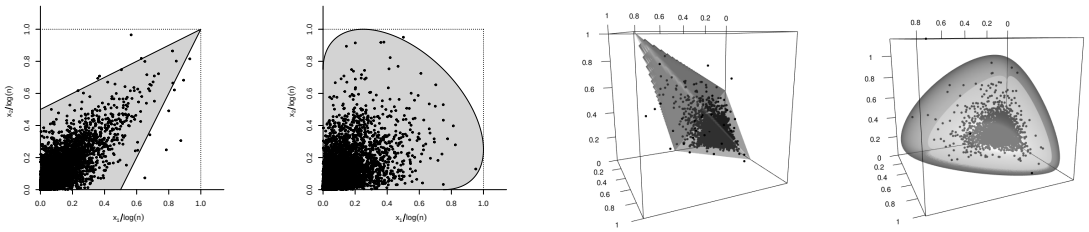

Figure 2.4: Scaled sample clouds N_n of size $n = 5,000$ and their corresponding limit set G in grey. From left to right, the underlying probability distributions are $d = 2$ max-stable with logistic dependence, $d = 2$ Gaussian, $d = 3$ max-stable with logistic dependence, and $d = 3$ Gaussian.

notion was extended in Nolde (2014), where it was shown that the value η from the HRV framework can be obtained from g. Through this relation, it is possible to determine whether or not the d components of $\boldsymbol{X}$ are asymptotically independent. This amounts to determining whether or not there is a "cusp" (or "point") at the corner $(1, \ldots, 1)^\top \in \mathbb{R}^d$, occurring when $(1, \ldots, 1)^\top \in G$ or $g(1, \ldots, 1) = 1$. We conclude that the d components of $\boldsymbol{X}$ do not observe simultaneous extreme values if $g(1, \ldots, 1) \neq 1$. This geometric interpretation of dependence leads to several boundary cases. When G is degenerate and lies only on the diagonal, or $g(\boldsymbol{x}) = 1$ only when $x_1 = \cdots = x_d$, then the random vector $\boldsymbol{X}$ is at the intersection between full dependence and asymptotic dependence. When G is the d-dimensional unit box, or $g(\boldsymbol{x}) = \max_j x_j$, then the random vector $\boldsymbol{X}$ is at the intersection between asymptotic asymptotic dependence and asymptotic independence. Lastly, when the $G = \{\boldsymbol{x} \in \mathbb{R}^d : \sum_{j=1}^d |x_j| = 1\}$, or $g(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$, then the random vector $\boldsymbol{X}$ is at the intersection between asymptotic independence and negative dependence.

With this in mind, one can examine the examples of Figure 2.4 and immediately conclude that the d components are not asymptotically independent in the max-stable logistic setting, but are in the Gaussian setting. Nolde and Wadsworth (2022) extended these ideas by linking several other extremal dependence coefficients to the gauge function, demonstrating the ability of the geometric framework to describe a wide variety of behaviours in the multivariate tail. Furthermore, the

assumption (2.15) was formalised in Nolde and Wadsworth (2022). In that work, given parametric forms of the joint density $f_{\boldsymbol{X}}$, the gauge function was derived using the limit (2.15), and several extremal coefficients were derived to described the tail behaviour of the random vector $\boldsymbol{X}$.

Recent advances of the geometric framework have solely focussed on the theoretical properties of limit sets and gauge functions for known distributions. What remains is how to estimate the limit set geometry from data. Furthermore, like in the MRV, HRV, and conditional extremes frameworks, it is also desirable to develop extrapolation techniques in the geometric approach. Given the flexibility the geometric setting has been shown to exhibit, the hope is that the geometry of limit sets could be estimated from data and then used to infer the *entire* d-dimensional tail for *any* extremal dependence structure, bypassing the limitations of previous methodology. Lastly, while recent literature uses the geometric approach the link several more recent notions of dependence in multivariate extremes, the question remains as to how the geometric approach is related to the classical approaches of MRV and HRV. One may notice similarities between the classical MRV convergence assumptions on the distribution function scale and the geometric convergence assumption on the log-density scale. Is this merely a coincidence? Are the two frameworks related in some way? What gains and losses does one take on when working in one framework versus the other? In this work, these questions will be examined through the lens of the multivariate generalised Pareto distribution, a commonly-used family of multivariate distributions used within the MRV framework.

Chapter 3

Statistical inference for multivariate extremes via a geometric approach

Abstract

A geometric representation for multivariate extremes, based on the shapes of scaled sample clouds in light-tailed margins and their so-called limit sets, has recently been shown to connect several existing extremal dependence concepts. However, these results are purely probabilistic, and the geometric approach itself has not been fully exploited for statistical inference. We outline a method for parametric estimation of the limit set shape, which includes a useful non/semi-parametric estimate as a pre-processing step. More fundamentally, our approach provides a new class of asymptotically-motivated statistical models for the tails of multivariate distributions, and such models can accommodate any combination of simultaneous or non-simultaneous extremes through appropriate parametric forms for the limit set shape. Extrapolation further into the tail of the distribution is possible via simulation from the fitted model. A simulation study confirms that our methodology is very competitive with existing approaches, and can successfully allow estimation

of small probabilities in regions where other methods struggle. We apply the methodology to two environmental datasets, with diagnostics demonstrating a good fit.

Disclaimer

The work presented in this chapter comes from a paper in which Jennifer Wadsworth was the lead author and was published in The Journal of the Royal Statistical Society, Series B (Statistical Methodology). My main contributions were the derivations in Section 3.3 and Appendix 3.8.2, as well as the real data applications in Sections 3.6.1, 3.6.2, and Appendix 3.8.8. I also addressed some possible theoretical and practical concerns when performing inference with the multivariate Gaussian gauge function in Appendix 3.8.3, and worked on the derivations for the general d-dimensional max-stable asymmetric logistic gauge function in Appendix 3.8.4. In Appendix 3.8.6, I considered the extent at which one could extrapolate in our models and the resulting bias-variance trade-off when not extrapolating far enough in the tails in the context of extreme probability estimation. In the process of submission to JRSSB, a reviewer asked if we could provide additional probability estimate simulations, comparing our model to parametrically-defined spectral densities within the MRV framework. I worked on this and it is presented in Section 3.5.1.

3.1 Introduction

3.1.1 Multivariate extreme value theory

Multivariate extreme value theory provides the basis for estimation of rare event probabilities that involve the effect of more than one variable. Applications are diverse and include estimating flood risk (Keef et al., 2013b; Engelke and Hitz, 2020), extreme air pollution levels (Heffernan and Tawn, 2004; Vettori et al., 2019), structural design (Coles and Tawn, 1994), dietary risk assessment (Chautru, 2015)

and financial risk assessment (Zhang and Huang, 2006; Hilal et al., 2014).

The study of multivariate extremes primarily began in the 1970s and 80s, with the theoretical study of multivariate regular variation (de Haan, 1970; de Haan and Resnick, 1977; Resnick, 1987). Multivariate regular variation is intrinsically tied up with the componentwise block maximum method for multivariate extremes. Suppose we have n independent replicates of a random vector $\boldsymbol{Y}_i \in \mathbb{R}^d$, $i = 1, \ldots, n$; the componentwise maximum vector is

$$\boldsymbol{M}_n = (M_{n,1}, \ldots, M_{n,d}) = \left(\max_{1 \leq i \leq n} Y_{1,i}, \ldots, \max_{1 \leq i \leq n} Y_{d,i} \right).$$

Univariate extreme value theory tells us that if, for each $j = 1, \ldots, d$, there exists $a_{n,j} > 0, b_{n,j}$ such that $(M_{n,j} - b_{n,j})/a_{n,j}$ converges to a non-degenerate random variable, then the distribution of this limiting variable is generalized extreme value (Fisher and Tippett, 1928; Gnedenko, 1943), which is the only univariate *max-stable* distribution. A distribution is max-stable if it is invariant to the operation of taking (componentwise) block maxima, up to marginal location and scale changes. The additional condition for joint convergence of the entire vector $(\boldsymbol{M}_n - \boldsymbol{b}_n)/\boldsymbol{a}_n$ to a multivariate max-stable distribution is multivariate regular variation. Since this represents an assumption on the dependence structure it can be expressed in standardized margins: a common choice is to set $X_{P,j} = 1/[1 - F_j(Y_j)]$, where $X_{P,j}$ follows a standard Pareto distribution if $Y_j \sim F_j$ has a continuous distribution, else it is asymptotically Pareto. A common way to express the multivariate regular variation assumption is

$$\lim_{t \to \infty} \Pr(\boldsymbol{X}_P / \|\boldsymbol{X}_P\| \in B, \|\boldsymbol{X}_P\| > ts \ \mid \ \|\boldsymbol{X}_P\| > t) = s^{-1} H(B), \qquad s > 1, \quad (3.1)$$

where $B \subset \mathcal{S}_{d-1} = \{\boldsymbol{v} \in [0,1]^d : \|\boldsymbol{v}\| = 1\}$ is a measurable set with $H(\partial B) = 0$. Assumption (3.1) shows that large values of the "radial" component $\|\boldsymbol{X}_P\|$ become independent of the "angular" component $\boldsymbol{X}_P/\|\boldsymbol{X}_P\|$, which follows some probability distribution H on $\mathcal{S}_{d-1}$, commonly referred to as the *spectral measure*. The choice of norm $\|\cdot\|$ is arbitrary, see, e.g., Beirlant et al. (2006, Chap. 8), but the most common choice is the L_1 norm $\|\cdot\|_1$, so that $\mathcal{S}_{d-1} = \{\boldsymbol{v} \in [0,1]^d : \sum_{j=1}^d v_j = 1\}$.

Statistical methodology for multivariate extremes followed shortly after this theoretical study, and focused initially on inference for data arising as componentwise block maxima through parametrized forms of multivariate max-stable distributions (Tawn, 1990). This was soon followed by more direct exploitation of the multivariate regular variation assumption (3.1), whereby parametric models were proposed for the spectral measure H, and inference performed on these (Coles and Tawn, 1991).

The study of componentwise maxima is a natural multivariate extension of the univariate block maximum approach, and the associated multivariate regular variation dependence condition (3.1) widely applicable. However, it has been known for a long time that while examples not satisfying (3.1) are rare, the number of examples for which this assumption forms a *useful* basis for statistical inference is very much smaller. This is because, for many dependence structures, mass of the spectral measure H accumulates on one or more regions of the form

$$\mathbb{B}_C = \{\boldsymbol{v} \in \mathbb{S}_{d-1} : v_j > 0, j \in C; v_k = 0, k \notin C\}, \qquad C \subset \{1, \ldots, d\}. \tag{3.2}$$

When this is the case, joint extremes of the random vector $\boldsymbol{Y}$ (or equivalently $\boldsymbol{X}_P$) may not always occur simultaneously; see, e.g., Goix et al. (2017) or Simpson et al. (2020) for a more detailed explanation. In practice, however, we never observe mass on such sets $\mathbb{B}_C$ at finite levels. This is illustrated in Figure 3.1, which displays the distribution of $\boldsymbol{X}_P/\|\boldsymbol{X}_P\|_1$ when the associated radial variable $\|\boldsymbol{X}_P\|_1$ exceeds its 0.98 quantile for two examples. In the left panel, the true limiting spectral measure H places mass only on the points $\mathbb{B}_{\{1\}} = \{(1, 0)\}$ and $\mathbb{B}_{\{2\}} = \{(0, 1)\}$, yet at observable levels, the distribution of angles is relatively evenly spread over $\mathbb{B}_{\{1,2\}}$, represented by the interval $(0, 1)$. In the right panel, the limiting spectral measure places mass only on $\mathbb{B}_{\{3\}}$ and $\mathbb{B}_{\{1,2\}}$, but once again, at observable levels we see all values in $\mathbb{B}_{\{1,2,3\}}$. A consequence of this mismatch between finite-sample and limiting distribution is a common modelling assumption that H places all mass on $\mathbb{B}_{\{1,\ldots,d\}} = \{\boldsymbol{v} \in \mathbb{S}_{d-1} : v_j > 0, j \in \{1, \ldots, d\}\}$, leading to overestimation of the probability of joint extremes. Moreover, even if one successfully detects the location of mass of the limiting object H, this does not lead to a practical strategy for

performing extrapolation beyond the observed values; to achieve this, more detailed information on the behaviour of $\boldsymbol{X}_P$ before the limiting regime is required.

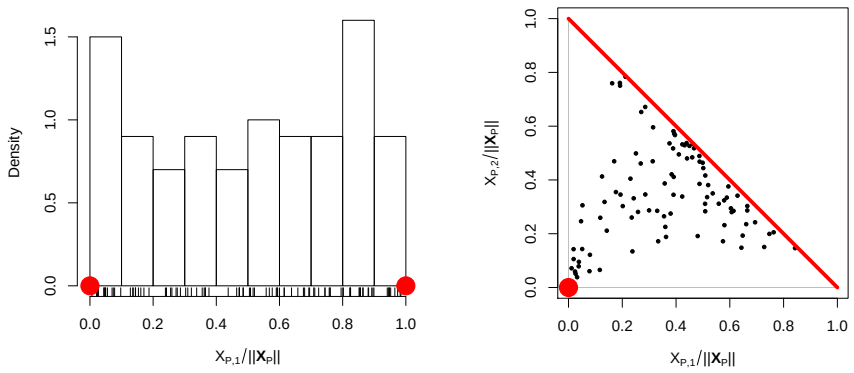

Figure 3.1: Illustration of the distribution of $\boldsymbol{X}_P/\|\boldsymbol{X}_P\|_1$, given that $\|\boldsymbol{X}_P\|_1$ is large, for two examples. Left panel: $\boldsymbol{X}_P$ has a bivariate Gaussian dependence structure with correlation parameter $\rho = 0.8$. The ticks correspond to observed angles $X_{P,1}/\|\boldsymbol{X}_P\|_1$ associated to large radii, along with a histogram of their density. The position of mass of the theoretical limiting H is illustrated by large filled circles (red). Right panel: $(X_{P,1}, X_{P,2})$ have an inverted Clayton dependence structure while $X_{P,3}$ is simulated conditional upon the value of $X_{P,2}$ so that these variables have an inverted logistic dependence structure. Points correspond to the observed distribution of $(X_{P,1}, X_{P,2})/\|\boldsymbol{X}_P\|$ with large radii. The position of mass of the theoretical limiting H is illustrated by the large filled circle and thick solid line (red).

3.1.2 Geometric approach to multivariate extremes

The early study of multivariate regular variation was followed by a smaller body of work that examined the convergence of light-tailed multivariate sample clouds onto so-called limit sets (Davis et al., 1988; Kinoshita and Resnick, 1991). These ideas did not have a clear link with multivariate max-stable models and did not lead to the

same proliferation of statistical methodology. More recently, several papers have revisited this geometric approach from a theoretical perspective (Balkema et al., 2010; Balkema and Nolde, 2010, 2012; Nolde, 2014; Nolde and Wadsworth, 2022) and in some cases shown how the shape of the limit set links to whether joint extremes of certain variables can occur.

To make ideas more concrete consider n independent copies of a random vector $\boldsymbol{X}_i$, $i = 1, \ldots, n$, with standard exponential margins; in practice, this will typically involve marginal transformation of the original vectors $\boldsymbol{Y}_i$. The *scaled n-point sample cloud* is defined as

$$N_n = \{\boldsymbol{X}_1/\log n, \ldots, \boldsymbol{X}_n/\log n\},$$

and we assume that this converges onto a *limit set* $G = \{\boldsymbol{x} \in \mathbb{R}^d_+ : g(\boldsymbol{x}) \leq 1\}$, where g is the 1-homogeneous *gauge function* of the limit set. This convergence is illustrated in Figure 3.2 for data with a logistic dependence structure, where the shape of the limit set can be seen to emerge in the scaled sample cloud as n becomes large. The precise sense of convergence of N_n onto G, and necessary and sufficient conditions for it, can be found in Balkema et al. (2010). Loosely, these conditions say that the expected number of points from N_n lying in sets that intersect with the limit set tends to infinity, whereas the expected number of points lying in sets that are disjoint from the limit set converges to zero. However, these specific conditions are rather unintuitive and make it difficult to determine the form of G for a given distribution, which led Nolde (2014) and Nolde and Wadsworth (2022) to consider alternative conditions in terms of the joint Lebesgue density of $\boldsymbol{X}$, when it exists. Denoting this joint density by $f_{\boldsymbol{X}}$, a sufficient condition for convergence of N_n onto G is

$$\lim_{t\to\infty} -\log f_{\boldsymbol{X}}(t\boldsymbol{x})/t = g(\boldsymbol{x}), \qquad \boldsymbol{x} \in [0, \infty)^d, \tag{3.3}$$

for a continuous gauge function g. Given that many statistical models have tractable joint densities and continuous gauge functions, equation (3.3) provides a simple way

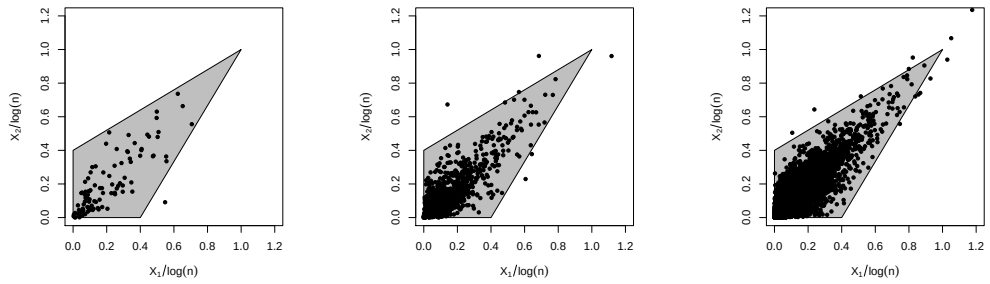

Figure 3.2: Illustration of the convergence of the scaled sample cloud N_n onto a limit set. From left to right, sample sizes are $n = 100, 1000, 10000$. The limit set G is depicted by the grey polygon.

to determine the form of g, and hence G, in several examples (Nolde and Wadsworth, 2022). Further illustrations of limit sets G are given for $d = 2, 3$ in Appendix 3.8.1.

The shape of G is important as a description of the extremal dependence of the underlying random vector. Limit sets exist for a much more general class of light-tailed marginal distributions than exponential, but we specialize to this case so that there is a clear correspondence between the shape of G and the dependence structure. In this case, the coordinatewise supremum of the limit set G is the point $(1, \ldots, 1)$, since for independent copies of an exponential variable X_i, the random variable $\max_{1 \leq i \leq n} X_i / \log n$ converges in probability to 1. Gauge functions, and therefore limit sets, of lower dimensional margins indexed by $J \subset \{1, \ldots, d\}$ can be found through the following minimization operation (Nolde and Wadsworth, 2022):

$$g_J(\boldsymbol{x}_J) = \min_{x_k \geq 0, k \notin J} g(\boldsymbol{x}),$$

where $\boldsymbol{x}_J = (x_j)_{j \in J}$. Exponential margins implies that for singleton sets $J = \{j\}$, $g_{\{j\}}(x_j) = x_j$.

Nolde and Wadsworth (2022) showed how G can be used to determine an array of extremal dependence measures which generally relate to representations of multivariate extremes that are more useful than multivariate regular variation when the spectral measure H places mass on one or more sets $\mathbb{B}_C$ as in equation (3.2).

These include expressions for determining the residual tail dependence coefficient (Ledford and Tawn, 1997), key elements of the conditional extremes model (Heffernan and Tawn, 2004), the angular dependence function (Wadsworth and Tawn, 2013), and the dependence coefficients of Simpson et al. (2020), which can be used to help determine the sets $\mathbb{B}_C$ on which H places mass.

Given the importance of the shape of G, a natural question that arises is how to estimate this from a sample of data. To date this question has been studied very little indeed; Jacob and Massé (1996) study estimation from a theoretical perspective but with no implementation. Very recently, Simpson and Tawn (2022) outlined an estimation approach in the bivariate case.

In this paper, we consider estimation of G as part of a wider new approach to the statistical analysis of extreme values. While G is an object of interest in itself, we direct our methodology more broadly at the question of statistical modelling and extrapolation for multivariate extreme values rather than focusing only on the descriptive aspects of extremal dependence that come from estimation of G. Our modelling approach allows in principle for any combination of joint extremes of sub-vectors of $\boldsymbol{Y}$ (equivalently, H may place mass on any valid combination of sets $\mathbb{B}_C$), and permits extrapolation in all directions, i.e., into the joint tail where all variables are large, or into other regions of the multivariate tail where only some variables are large. Existing alternatives to methodology based on multivariate regular variation do not capture these possibilities in a coherent manner.

To illustrate the potential importance of being able to capture complex structure in extremes, consider the dataset of river flow measurements from Simpson et al. (2020) that will also be analysed in Section 3.6.2. Their analysis showed that there are some events where all four rivers were extreme simultaneously, but that there were also extreme episodes involving single rivers, or groups of two or three rivers without the others. This might be explained physically by the weather patterns causing the extremes, and the relationships between catchments. While Simpson et al. (2020) introduced and estimated coefficients to help determine this structure,

they did not provide any modelling approach that could account for it, as we do here.

Section 3.2 outlines our statistical model and assumptions. Section 3.3 details theoretical examples that demonstrate applicability of the method. We focus on details of statistical inference in Section 4.3, and use simulation to show that our approach is very competitive for estimation of extreme set probabilities in a wide range of scenarios in Section 3.5. Section 3.6 contains applications to oceanographic and fluvial datasets, and we conclude in Section 6.6.

3.2 Model and assumptions

Here and throughout the rest of the paper, we assume that we have a random vector $\boldsymbol{X}$ with standard exponential margins and joint Lebesgue density denoted by $f_{\boldsymbol{X}}$. Marginal transformation can be applied as a standard step via estimation of each marginal distribution function. The assumption of a joint density is very common for statistical analysis, as it is required for most likelihood-based inference, for example. We further assume that the scaled sample cloud N_n converges onto a limit set G whose shape can either be described by a continuous gauge function g, or that we are only interested in the continuous part.

Assumption (3.3), which yields a sufficient condition for convergence of N_n onto G, can equivalently be expressed $f_{\boldsymbol{X}}(t\boldsymbol{x}) = \exp\{-tg(\boldsymbol{x})[1 + o(1)]\}$ for $g(\boldsymbol{x}) > 0$ as $t \to \infty$. The homogeneity of g suggests making the radial-angular transformation $R = \sum_{j=1}^{d} X_j, \boldsymbol{W} = \boldsymbol{X}/R$; such transformations are common in multivariate extremes, but normally on Pareto, rather than exponential, margins. The Jacobian of this transformation is r^{d-1}, which leads to joint density of $(R, \boldsymbol{W})$: $f_{R,\boldsymbol{W}}(r, \boldsymbol{w}) = r^{d-1} f_{\boldsymbol{X}}(r\boldsymbol{w}) = r^{d-1} \exp\{-rg(\boldsymbol{w})[1 + o(1)]\}$, as $r \to \infty$. This in turn means that the conditional density of $R \mid \boldsymbol{W} = \boldsymbol{w}$ satisfies

$$f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1} \exp\{-rg(\boldsymbol{w})[1 + o(1)]\}, \qquad r \to \infty. \tag{3.4}$$

If we were to ignore the $o(1)$ term, we recognize the form of the gamma kernel in

equation (3.4), suggesting that when $R \mid \boldsymbol{W} = \boldsymbol{w}$ is large, its distribution could potentially be well approximated by a gamma distribution. Indeed, if the $o(1)$ term in the exponent is negligible, this suggests a truncated gamma approximation above a high threshold $r_0(\boldsymbol{w})$ of the conditional distribution $R \mid \boldsymbol{W} = \boldsymbol{w}$.

A valid concern is whether the $o(1)$ term in the exponent of (3.4) is really negligible. In Section 3.3 we detail several examples which in fact have the more helpful asymptotic form $f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1} \exp[-rg(\boldsymbol{w})][1 + o(1)]$, i.e., with the $o(1)$ outside of the exponent, and give explicit rates for this term. Based on this latter asymptotic representation, we focus in this paper on the model

$$R \mid [\boldsymbol{W} = \boldsymbol{w}, R > r_0(\boldsymbol{w})] \dot{\sim} \text{truncGamma}(\alpha, g(\boldsymbol{w})), \tag{3.5}$$

where $\alpha > 0$ is the gamma shape, and $g(\boldsymbol{w})$ is the gamma rate parameter. In most examples, the theoretical shape parameter is $\alpha = d$, but for modelling purposes the flexibility of an estimated shape is desirable. By parametrizing flexible forms for the gauge function $g(\boldsymbol{w}) = g(\boldsymbol{w}; \boldsymbol{\theta})$, we can use approximation (3.5) to estimate these parameters. Full details of our approach are given in Section 4.3, including diagnostic plots for assessing approximation (3.5).

3.3 Examples

In this section we consider a variety of examples. The convergence onto a limit set G holds very broadly, and in many examples the gauge function for this limit set in exponential margins can be recovered fully or partly from convergence (3.3). The form of the gauge function and limit set for several examples, including multivariate t_ν, light-tailed elliptical, skew-normal, generalized hyperbolic, certain mixture distributions and multivariate generalized Pareto forms has been derived in Balkema et al. (2010), Nolde (2014), Nolde and Wadsworth (2022), and Zhang et al. (2022), for example, although not always in exponential margins.

The validity and quality of the truncated gamma approximation in (3.5) to the conditional density in (3.4) depends on the $o(1)$ term. Since this lies in the exponent,

it is not always guaranteed to be negligible. In this section, we explicitly calculate the density of $R \mid \boldsymbol{W} = \boldsymbol{w}$ for various theoretical examples, showing that most in fact have the form $f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1} \exp[-rg(\boldsymbol{w})][1 + o(1)]$, as $r \to \infty$. The exception to this is the Gaussian dependence structure, for which we find $f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{\alpha(\boldsymbol{w})-1} \exp[-rg(\boldsymbol{w})][1 + o(1)]$, as $r \to \infty$, i.e., the conditional gamma form is still applicable, but the shape parameter depends on the value of $\boldsymbol{w}$. Nonetheless, further investigations, described briefly below and in more detail in Appendix 3.8.3, show the assumption of a common shape in model (3.5) does not appear problematic in practice. This is also supported by our simulation study in Section 3.5. More generally, we will incorporate model checking of assumption (3.5) into our statistical analysis.

For each distribution, we provide the overall form of $f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w})$, with further calculations given in Appendix 3.8.2. We denote the ordered values of the vector $\boldsymbol{w}$ (and similarly $\boldsymbol{x}$) by $w_{(1)} \geq w_{(2)} \geq \cdots \geq w_{(d)} > 0$, assuming the minimum to be positive. In the convergence rates given below, we assume a strict ordering $w_{(1)} > w_{(2)} > \cdots > w_{(d)} > 0$; where this is not the case, following the derivations in the appendix, one usually observes improved rates, e.g., $O(e^{-r(w_{(d-2)}-w_{(d)})})$ replacing $O(e^{-r(w_{(d-1)}-w_{(d)})})$ if $w_{(d-2)} > w_{(d-1)} = w_{(d)}$.

Multivariate max-stable and generalized Pareto distributions Multivariate max-stable distributions are most readily expressed by their distribution functions. In exponential margins,

$$\Pr(\boldsymbol{X} \leq \boldsymbol{x}) = \exp\left(-V\left\{[-\log(1 - e^{-\boldsymbol{x}})]^{-1}\right\}\right),$$

where $V : \mathbb{R}_+^d \to \mathbb{R}_+$ is the homogeneous of order -1 *exponent function*, and operations are applied componentwise. The general asymptotic form of the density for a max-stable distribution in exponential margins is therefore

$$\begin{aligned} f_{\boldsymbol{X}}(t\boldsymbol{x}) = &\exp\{-V[e^{t\boldsymbol{x}} + 1/2 + O(e^{-t\boldsymbol{x}})]\} \sum_{\pi \in \Pi} \prod_{s \in \pi} V_s[e^{t\boldsymbol{x}} + 1/2 + O(e^{-t\boldsymbol{x}})] \\ &\times e^{t\sum_{j=1}^d x_j}[1 + O(e^{-2tx_{(d)}})], \end{aligned}$$

$t \to \infty$, where Π is the set of all partitions of $\{1, \ldots, d\}$, and $V_s(\boldsymbol{z}) = \partial^{|s|} V(\boldsymbol{z}) / \prod_{j \in s} \partial z_j$.

We focus on the d-dimensional logistic distribution, for which $V(\boldsymbol{z}) = \left(\sum_{j=1}^d z_j^{-1/\gamma} \right)^\gamma$ with parameter $\gamma \in (0, 1]$. This distribution has gauge function $g(\boldsymbol{x}) = \sum_{j=1}^d x_j/\gamma + (1 - d/\gamma) x_{(d)}$, and

$$f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1} e^{-r g(\boldsymbol{w})} [1 + O(e^{-r(w_{(d-1)} - w_{(d)})/\gamma}) + O(e^{-r w_{(d)}})], \qquad r \to \infty.$$

The simpler form of the densities make calculations more straightforward for corresponding multivariate generalized Pareto distributions (MGPDs), which are related to max-stable distributions (Rootzén and Tajvidi, 2006; Rootzén et al., 2018b). The support of MGPDs whose margins have unit scale and zero shape is contained in $\{\boldsymbol{x} \in \mathbb{R}^d : x_{(1)} > 0\}$. Densities for several models for which the spectral measure H places mass only on $\mathbb{B}_{\{1,\ldots,d\}}$ are given in Kiriliouk et al. (2019); in such cases the dependence structure can be determined by focusing on large values of $\boldsymbol{x} > \boldsymbol{0}$. Further details are in the Appendix 3.8.2.

For the MGPD associated to the negative logistic max-stable distribution (Galambos, 1975; Dombry et al., 2016), $g(\boldsymbol{x}) = (1 + d\gamma) x_{(1)} - \sum_{j=1}^d x_j \gamma$, $\gamma > 1$ and

$$f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1} e^{-r g(\boldsymbol{w})} \left[1 + O\left(e^{r(w_{(2)} - w_{(1)})\gamma} \right) \right], \qquad r \to \infty.$$

For the MGPD associated to the Dirichlet max-stable distribution (Coles and Tawn, 1991), $g(\boldsymbol{x}) = (1 + \sum_{j=1}^d \theta_j) x_{(1)} - \sum_{j=1}^d \theta_j x_j$, for all $\theta_j > 0$, and

$$f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1} e^{-r g(\boldsymbol{w})} \left[1 + O\left(e^{r(w_{(2)} - w_{(1)})} \right) \right], \qquad r \to \infty.$$

Inverted max-stable distributions Inverted max-stable distributions are derived by translating the joint lower tail of max-stable distributions to be the joint upper tail. This is achieved by applying a monotonically decreasing marginal transformation to a max-stable random vector. In exponential margins inverted max-stable distributions have density

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \exp[-l(\boldsymbol{x})] \sum_{\pi \in \Pi} \prod_{s \in \pi} l_s(\boldsymbol{x}),$$

where l is the *stable tail dependence function* of the corresponding max-stable distribution, obtained via $l(\boldsymbol{x}) = V(1/\boldsymbol{x})$, and $l_s(\boldsymbol{x}) = \partial^{|s|} l(\boldsymbol{x}) / \prod_{j \in s} \partial x_j$. The gauge function is always $g(\boldsymbol{x}) = l(\boldsymbol{x})$. Owing to the fact that $l_s(\boldsymbol{x})$ is homogeneous of order $1 - |s|$, we obtain

$$f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1} \exp[-rg(\boldsymbol{w})][1 + O(r^{-1})], \qquad r \to \infty.$$

Multivariate Gaussian distribution We consider the multivariate Gaussian dependence structure with correlation matrix Σ. When one or more correlation parameters is negative then the continuous convergence $-\log f_{\boldsymbol{X}}(t\boldsymbol{x})/t \to g(\boldsymbol{x})$ fails when components of $\boldsymbol{x}$ are zero, because the gauge function of the limit set is not continuous away from the strictly positive orthant (Nolde and Wadsworth, 2022). Since we are considering $w_{(d)} > 0$ this is not an issue here, but we note that to fully capture negative association it is ideal to reformulate ideas in terms of Laplace rather than exponential margins; see Nolde and Wadsworth (2022) and Section 6.6. For Σ with non-negative entries, $g(\boldsymbol{x}) = (\boldsymbol{x}^{1/2})^\top \Sigma^{-1} \boldsymbol{x}^{1/2}$, where $\boldsymbol{x}^{1/2} = (x_1^{1/2}, \ldots, x_d^{1/2})^\top$, and

$$f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{\alpha(\boldsymbol{w})-1} \exp[-rg(\boldsymbol{w})] \left[1 + O\left(\frac{(\log r)^2}{r}\right)\right]$$

as $r \to \infty$, where

$$\alpha(\boldsymbol{w}) = \frac{d}{2} + \frac{(\boldsymbol{w}^{1/2})^\top \Sigma^{-1} \boldsymbol{w}^{-1/2}}{2}.$$

In this case, the gamma shape parameter therefore depends on $\boldsymbol{w}$, and the region on which $\alpha(\boldsymbol{w}) > 0$ depends on the entries of Σ. We investigate this further in Appendix 3.8.3, showing that local estimates of α do not vary strongly with $\boldsymbol{w}$ and may reasonably be assumed constant. We also show that results from our model are useful even in the (typically small) regions where $\alpha(\boldsymbol{w}) \leq 0$.

Multivariate t_ν distribution We consider the multivariate t distribution with ν degrees of freedom, focusing only on positive dependence; see the appendix for further comment. The gauge function is $g(\boldsymbol{x}) = (1 + d/\nu)x_{(1)} - \sum_{j=1}^{d} x_j/\nu$, and

$$f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1} e^{-rg(\boldsymbol{w})} \left[1 + O(e^{r(w_{(2)} - w_{(1)})/\nu}) + O(e^{-2rw_{(d)}/\nu})\right], \qquad r \to \infty.$$

Clayton and inverted Clayton copulas We consider the Clayton and inverted Clayton copulas with parameter $\gamma > 0$. The Clayton copula has $g(\boldsymbol{x}) = \sum_{j=1}^{d} x_j$, and

$$f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1} e^{-rg(\boldsymbol{w})}[1 + O(e^{-rw_{(d)}})], \qquad r \to \infty.$$

The inverted Clayton copula has $g(\boldsymbol{x}) = (1 + d\gamma)x_{(1)} - \sum_{j=1}^{d} x_j\gamma$, and

$$f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1} e^{-rg(\boldsymbol{w})}[1 + O(e^{-r(w_{(2)} - w_{(1)})})], \qquad r \to \infty.$$

In the appendix, we also calculate $f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w})$ for a trivariate vine copula example.

3.4 Statistical inference

3.4.1 Calculating the threshold $r_0(\boldsymbol{w})$

To implement model (3.5), we firstly need to calculate $r_0(\boldsymbol{w})$, which represents a high threshold of the conditional distribution $R \mid \boldsymbol{W} = \boldsymbol{w}$. A natural approach to calculating this threshold is quantile regression, treating $\boldsymbol{W}$ as the covariate. A similar approach has been taken in the context of establishing covariate-dependent thresholds in univariate extreme value analysis (Northrop and Jonathan, 2011). When data are bivariate, so that $\boldsymbol{W} \in \mathcal{S}_1$ is equivalent to $W \in [0, 1]$, this approach is straightforward. However, standard parametric quantile regression requires a high degree of manual tuning to ensure that the model form captures the relation between R and $\boldsymbol{W}$ well. We therefore suggest using additive quantile regression (Fasiolo et al., 2021) via the corresponding R package `qgam`.

When $\boldsymbol{W} \in \mathcal{S}_{d-1}$, $d > 2$, then both parametric and additive quantile regression become more difficult due to the specific support of $\boldsymbol{W}$ on the simplex. A simple alternative is to calculate quantiles of $R \mid \boldsymbol{W} = \boldsymbol{w}$ from overlapping blocks of $\boldsymbol{W}$ values, which is feasible for relatively low dimensions, but becomes more laborious as d grows. The top row of Figure 3.3 illustrates the concepts for $d = 2, 3$. In each case, $r_0(\boldsymbol{w})$ is calculated as the 0.95 quantile of $R \mid \boldsymbol{W} = \boldsymbol{w}$.

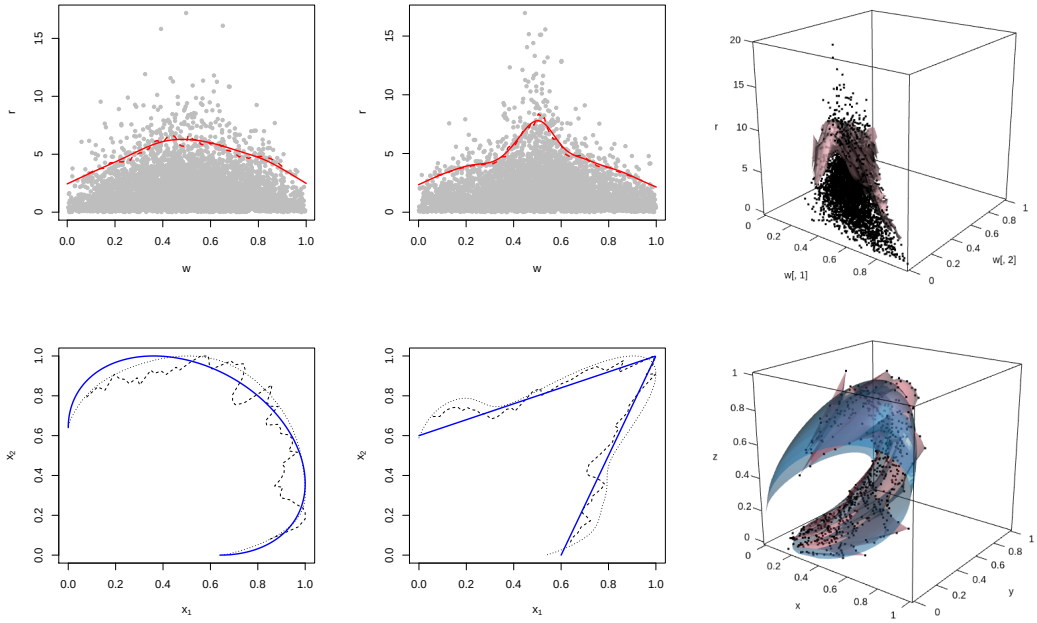

Figure 3.3: Top row: R against $\boldsymbol{W}$, with the estimated 0.95 quantile of $R \mid \boldsymbol{W} = \boldsymbol{w}$ in red. In the left and centre ($d = 2$) plots, solid lines represent the output from `qgam`, and dashed lines from rolling-windows quantiles. In the right ($d = 3$) plot, the surface is calculated through a rolling-windows technique. Bottom row: Plots of $\boldsymbol{x} = \boldsymbol{v} r_0(\boldsymbol{v})$, rescaled to lie in $[0, 1]^d$ as per equation (3.7). In the left and centre plots, dotted black lines represent the output from `qgam`, and dashed black lines from rolling-windows quantiles. The solid lines (blue) are the unit level sets of $g(\boldsymbol{x})$, with g the true gauge function. In the right plot, the jagged surface (red) comes from the rolling-windows technique, and the smooth surface (blue) is the unit level set of the true gauge function.

In the second row of Figure 3.3 we demonstrate that the threshold $r_0(\boldsymbol{w})$, suitably rescaled, can be viewed as a non-/semi-parametric estimate of g. The reason for this can roughly be explained by considering the case where the gamma approximation is exact. Let $\bar{F}(r \mid \boldsymbol{w})$ be the (gamma) survival function of $R \mid \boldsymbol{W} = \boldsymbol{w}$, then for quantile regression at level $\tau \in (0, 1)$, $\bar{F}[r_0(\boldsymbol{w}) \mid \boldsymbol{w}] = 1 - \tau$. We have

$$\begin{aligned} \bar{F}[r_0(\boldsymbol{w}) \mid \boldsymbol{w}] &= 1 - \int_0^{r_0(\boldsymbol{w})} \frac{g(\boldsymbol{w})^\alpha}{\Gamma(\alpha)} v^{\alpha-1} e^{-vg(\boldsymbol{w})} \, \mathrm{d}v \\ &= 1 - \int_0^{r_0(\boldsymbol{w})g(\boldsymbol{w})} \frac{s^{\alpha-1}}{\Gamma(\alpha)} e^{-s} \, \mathrm{d}s = 1 - \tau, \end{aligned} \tag{3.6}$$

using the change of variables $s = g(\boldsymbol{w})v$. Equation (3.6) is solved by taking $r_0(\boldsymbol{w}) = C_\tau / g(\boldsymbol{w})$, with C_τ the solution to the equation $\int_0^{C_\tau} \frac{s^{\alpha-1}}{\Gamma(\alpha)} e^{-s} \, \mathrm{d}s = \tau$. Since the gamma approximation is only asymptotically valid, we have in practice that $r_0(\boldsymbol{w}) \approx C_\tau / g(\boldsymbol{w})$ for τ close to 1. To depict unit level sets of the gauge function g, we plot points $\boldsymbol{x} = \boldsymbol{v}/g(\boldsymbol{v})$, where $\boldsymbol{v}$ is a sequence of points covering the simplex $\mathcal{S}_{d-1}$. Consequently, we can compare $r_0(\boldsymbol{w})$ to g by plotting points $\boldsymbol{x} = \boldsymbol{v} r_0(\boldsymbol{v}) / C_\tau$. However, since the gamma approximation is not exact, we instead scale each margin so that the coordinatewise supremum exactly equals one, by plotting

$$\boldsymbol{x} = \left\{ v_1 r_0(\boldsymbol{v}) / \max_{j \in \{1,\ldots,d\}} [v_j r_0(\boldsymbol{v})], \ldots, v_d r_0(\boldsymbol{v}) / \max_{j \in \{1,\ldots,d\}} [v_j r_0(\boldsymbol{v})] \right\}. \tag{3.7}$$

We will use the observation that links $r_0(\boldsymbol{w})$ and $g(\boldsymbol{w})$ later to assist with model checking, but note that, combined with extension of additive quantile regression to higher dimensions, this presents a very interesting avenue for future work.

3.4.2 Likelihood

In order to fit model (3.5), we use likelihood-based inference. For n_0 independent observations of $R_i \mid [\boldsymbol{W}_i = \boldsymbol{w}_i, R_i > r_0(\boldsymbol{w}_i)]$, $i = 1, \ldots, n_0$, we maximize the likelihood

$$L(\boldsymbol{\psi}) = \prod_{i=1}^{n_0} \frac{g(\boldsymbol{w}_i; \boldsymbol{\theta})^\alpha}{\Gamma(\alpha)} \frac{r_i^{\alpha-1} e^{-r_i g(\boldsymbol{w}_i; \boldsymbol{\theta})}}{\bar{F}[r_0(\boldsymbol{w}_i); \alpha, g(\boldsymbol{w}_i; \boldsymbol{\theta})]}, \tag{3.8}$$

where $\boldsymbol{\psi} = (\alpha, \boldsymbol{\theta})^\top$ and $\bar{F}[\cdot; \alpha, g(\boldsymbol{w}; \boldsymbol{\theta})]$ represents the gamma survival function with shape parameter α, and rate parameter $g(\boldsymbol{w}; \boldsymbol{\theta})$. Estimates of uncertainty in the maximum likelihood estimators may be obtained through the inverse Hessian matrix, subject to model validity and independence checks, or via the bootstrap. In practice, many datasets exhibit weak-to-moderate temporal dependence, so that while likelihood (3.8) may be used for parameter estimation (e.g., Chandler and Bate, 2007), block-bootstrap techniques will be preferable for estimation of uncertainty.

3.4.3 Gauge functions and model selection

3.4.3.1 Gauge functions from specific distributions

Key to a successful fit of model (3.5) via likelihood (3.8) are flexible parametrized forms of g that are able to capture a wide variety of limit set shapes. In Section 3.3, we detail various forms of gauge function that come from different underlying distributions, some of which are illustrated in Appendix 3.8.1. Further forms can also be found in Nolde and Wadsworth (2022). Any of these parametric forms could be fitted as a candidate model, and standard model-selection techniques, such as information criteria, used to establish a best choice; we will demonstrate this in our simulation study of Section 3.5.

A key attraction of our new approach to inference for multivariate extremes is the ability to capture the complex dependence structures that arise when different subgroups of variables can potentially be co-extreme while the others are small. Under multivariate regular variation, this corresponds to the spectral measure H placing mass on sets $\mathbb{B}_C$ as described in Section 6.1. In order to capture these scenarios, we consider the gauge function corresponding to the asymmetric logistic distribution (Tawn, 1990), which can place mass on any valid combination of sets $\mathbb{B}_C$. The full expression for this involves minimization over several components, and is given in Appendix 3.8.4. Figure 3.4 depicts some of the potential limit sets arising from this structure when $d = 3$.

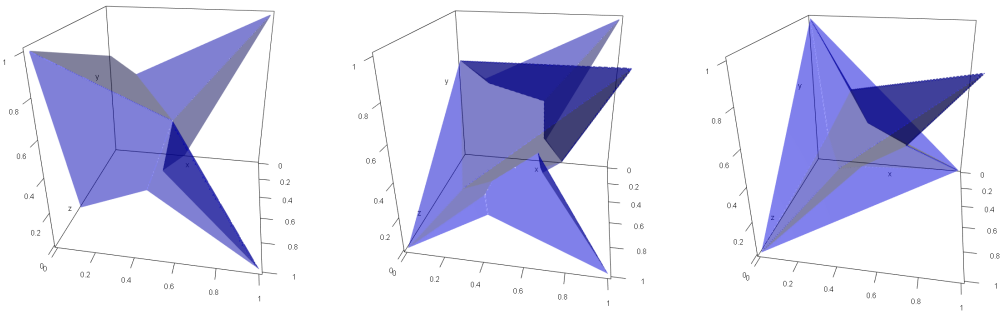

Figure 3.4: Example limit sets G (area between surface and planes $x_j = 0$) from the asymmetric logistic distribution. Left: mass of H on $\mathbb{B}_{\{1,2\}}$, $\mathbb{B}_{\{1,3\}}$, $\mathbb{B}_{\{2,3\}}$ with parameters $\gamma_{\{1,2\}} = 0.5$, $\gamma_{\{1,3\}} = 0.2$, $\gamma_{\{2,3\}} = 0.7$. Centre: mass on $\mathbb{B}_{\{3\}}$, $\mathbb{B}_{\{1,2\}}$, $\mathbb{B}_{\{1,3\}}$, $\mathbb{B}_{\{1,2,3\}}$ with $\gamma_{\{1,2\}} = 0.5$, $\gamma_{\{1,3\}} = 0.5$, $\gamma_{\{1,2,3\}} = 0.7$. Right: mass on $\mathbb{B}_{\{1\}}$, $\mathbb{B}_{\{2\}}$, $\mathbb{B}_{\{3\}}$, $\mathbb{B}_{\{1,2,3\}}$ with $\gamma_{\{1,2,3\}} = 0.5$.

3.4.3.2 Additively mixing gauge functions

The gauge functions described in Section 3.3 provide a starting point for inference on model (3.5), but may not always be flexible enough to capture the structures of observed data. We now consider how to mix gauge functions to generate more flexible models. As mentioned in Section 6.1, the limit sets G for data with exponential margins have coordinatewise supremum equal to $(1, \ldots, 1)$; equivalently, the one-dimensional marginal gauge functions are $g_{\{j\}}(x_j) = x_j$. Each form of g given in Section 3.3 satisfies this constraint, and we require any scheme for mixing gauge functions to also satisfy this, since they will be applied to data in exponential margins.

A simple way to mix that retains the marginal condition $g_{\{j\}}(x_j) = x_j$ is via minimization: $g(\boldsymbol{x}) = \min[g^{[1]}(\boldsymbol{x}), \ldots, g^{[m]}(\boldsymbol{x})]$, for $g^{[1]}, \ldots, g^{[m]}$ each satisfying this marginal condition. The resulting gauge function is the one that would correspond to a mixture density $f_{\boldsymbol{X}}(\boldsymbol{x}) = \sum_{k=1}^{m} \pi_k f_{\boldsymbol{X}}^{[k]}(\boldsymbol{x})$ with $\sum_{k=1}^{m} \pi_k = 1$ and $\pi_k \in (0, 1)$ for each k; the mixture weights do not appear in g because as $n \to \infty$ there would be infinitely many points in the sample cloud from each mixture component. However, such an

approach has the effect of retaining the most protruding part of each limit set and may not yield the most realistic shapes; some examples are given in Appendix 3.8.5. Instead we focus on additive mixing, defining

$$\tilde{g}(\boldsymbol{x}) = a_1 g^{[1]}(\boldsymbol{x}) + \cdots + a_{m-1} g^{[m-1]}(\boldsymbol{x}) + g^{[m]}(\boldsymbol{x}), \qquad a_1, \ldots, a_{m-1} > 0. \quad (3.9)$$

The resulting function is denoted by $\tilde{g}$ as in general it will not satisfy the marginal condition, and will need to be rescaled to do so. Suppose that the coordinatewise supremum of the set $\tilde{G} = \{\boldsymbol{x} : \tilde{g}(\boldsymbol{x}) \leq 1\}$ is $\tilde{\boldsymbol{c}} = (\tilde{c}_1, \ldots, \tilde{c}_m)$. Then the rescaled gauge function $g(\boldsymbol{x}) = \tilde{g}(\tilde{c}_1 x_1, \ldots, \tilde{c}_d x_d)$ satisfies $g_{\{j\}}(x_j) = x_j$. Some examples of limit sets from additively mixed functions are depicted in Figure 3.5. Interestingly, we observe for $d = 2$ that this process is able to interpolate between limit sets for which $g(1, 1) = 1$ and have a "pointy" shape, to those with $g(1, 1) < 1$ and are described by Balkema and Nolde (2012) as "blunt". The former arise for dependence structures representing joint extremes (H places mass only on $\mathbb{B}_{\{1,2\}}$), while the latter arise for those representing separate extremes (H places mass only on $\mathbb{B}_{\{1\}}$ and $\mathbb{B}_{\{2\}}$). Figures in Appendix 3.8.5 also show that for $d = 3$ we retain the ability to move between "pointy" limit set shapes representing joint occurrence of extremes in some components and "blunt" shapes representing separate extremes. We focus in the figures only on the case $m = 2$, and leave theoretical study of this phenomenon for any m to future work.

Note that when using additive mixing, the component gauge functions $g^{[k]}(\boldsymbol{x})$ need not satisfy the marginal condition $g_{\{j\}}(x_j) = x_j$ due to the rescaling. This allows, for example, one to include the Gaussian gauge function $g(\boldsymbol{x}) = (\boldsymbol{x}^{1/2})^\top \Sigma^{-1} \boldsymbol{x}^{1/2}$ when Σ has negative entries, and increases the flexibility of this approach. In practice, we use numerical methods to find the vector $\tilde{\boldsymbol{c}}$ for rescaling, by finding the coordinatewise supremum of $\tilde{G}$ on a grid.

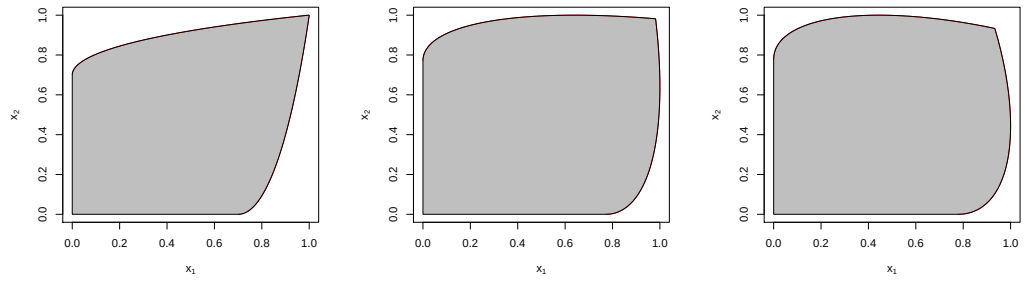

Figure 3.5: Examples of limit sets (grey shaded region) from additively mixed gauge functions according to equation (3.9), with subsequent rescaling. In each case the component gauge functions are $g^{[1]}(x_1, x_2) = [x_1 + x_2 - 2\rho(x_1 x_2)^{1/2}]/(1-\rho^2)$ with $\rho = 0.5$ and $g^{[2]}(x_1, x_2) = (x_1 + x_2)/\gamma + (1 - 2/\gamma)\min(x_1, x_2)$ with $\gamma = 0.5$; these gauges correspond to the Gaussian and logistic max-stable distributions, respectively. From left to right, the weights are $a_1 = 1, 2, 3$.

3.4.4 Model checking

We propose checking the fitted model from likelihood (3.8) via probability-probability (PP) plots. The fitted distribution function (df) of the truncated gamma model is

$$\widehat{F}_{\text{tg}}[r \mid \boldsymbol{w}, r_0(\boldsymbol{w})] := \Pr[R \leq r \mid \boldsymbol{W} = \boldsymbol{w}, R > r_0(\boldsymbol{w})] = 1 - \frac{\bar{F}[r; \widehat{\alpha}, g(\boldsymbol{w}; \widehat{\boldsymbol{\theta}})]}{\bar{F}[r_0(\boldsymbol{w}); \widehat{\alpha}, g(\boldsymbol{w}; \widehat{\boldsymbol{\theta}})]},$$

with $\bar{F}$ as in likelihood (3.8), and $\widehat{\alpha}, \widehat{\boldsymbol{\theta}}$ representing the maximum likelihood estimates of the parameters. The PP plot for n_0 observations with $R_i > r_0(\boldsymbol{w}_i)$ is the set of points: $[i/(n_0 + 1), u_{(n_0 - i + 1)}]$, where $u_i = \widehat{F}_{\text{tg}}[r_i; \boldsymbol{w}_i, r_0(\boldsymbol{w}_i)]$, and $u_{(1)} \geq u_{(2)} \geq \cdots \geq u_{(n_0)}$ represent the ordered sample of u_i. This diagnostic will be demonstrated in Section 3.6.

Comparison of the "empirical" estimate of the gauge function $\hat{g}(\boldsymbol{w}) \approx \hat{C}/r_0(\boldsymbol{w})$, as outlined in Section 3.4.1, provides another check on the form of the fitted model. As was seen in Section 3.4.1, while we do not expect perfect correspondence between $\hat{g}(\boldsymbol{w})$ and $g(\boldsymbol{w}; \widehat{\boldsymbol{\theta}})$, we can expect to see broad similarities in shape. Again we use this in Section 3.6.

3.4.5 Prediction

A key aspect of our proposed geometric framework for statistical inference is that we can use simulation from the fitted model to estimate probabilities of lying in extreme regions, enabling extrapolation outside the range of the observed data. Up to this point, we have focused on the conditional distribution of $R \mid [\boldsymbol{W} = \boldsymbol{w}, R > r_0(\boldsymbol{w})]$. In order to perform extrapolation and estimate multivariate tail probabilities, we need realizations of the distribution of $\boldsymbol{X}$ in some suitably extreme region. Notationally it is helpful to introduce an alternative radial variable, $R' = R/r_0(\boldsymbol{W})$, so that $\boldsymbol{X} = R\boldsymbol{W} = R'r_0(\boldsymbol{W})\boldsymbol{W}$. Given a particular value of $\boldsymbol{W} = \boldsymbol{w}$, our extreme region to date has been $\{R > r_0(\boldsymbol{w})\}$. Now considering our extreme region across all $(R, \boldsymbol{W})$ values, this corresponds to $\{R' > 1\}$. In Figure 3.3, all points above the red line / surface in the top row are those with $\{R' > 1\}$.

We focus initially on simulating an arbitrary number of points satisfying the conditioning event $\{R' > 1\}$, and discuss below adaptations for simulating above higher thresholds. To get draws from the distribution of $\boldsymbol{X} \mid R' > 1$, we multiply simulations from two components:

(i) Draw $\boldsymbol{w}^\star$ from the distribution of $\boldsymbol{W} \mid R' > 1$;

(ii) Conditional upon $\boldsymbol{w}^\star$, draw $r^\star$ from the distribution of $R \mid [\boldsymbol{W} = \boldsymbol{w}^\star, R > r_0(\boldsymbol{w}^\star)]$.

The sampled value is then $\boldsymbol{x}^\star = r^\star \boldsymbol{w}^\star$. The second of these steps is a simple case of simulating from the fitted truncated gamma distribution, which can be done via the inverse probability integral transform. For the first of these, we may either resample from the empirical distribution of $\boldsymbol{W} \mid R' > 1$, or we could fit a parametric model to such samples and simulate from this. We opt for the former in this work, and note the latter as a potential line of future investigation. Figure 3.6 shows 5000 draws simulated from $\boldsymbol{X} \mid R' > 1$, based on a model fitted to 2500 data points.

To estimate the probability of lying in extreme sets, we exploit the simple

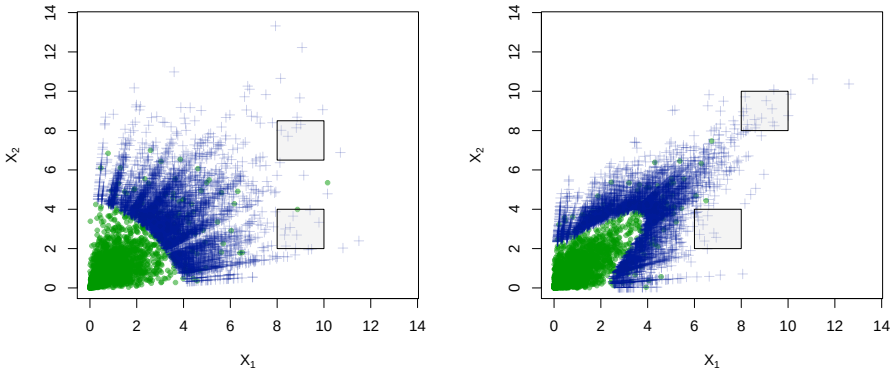

Figure 3.6: Example of 5000 points simulated from $\boldsymbol{X} \mid R' > 1$, depicted by crosses (blue), for the inverted logistic (left) and logistic (right) distributions. Models with the true gauge were fitted to threshold exceedances of 2500 data points, with all data depicted by dots (green). Light grey squares represent potential sets B in equation (3.10).

equation

$$\Pr(\boldsymbol{X} \in B) = \Pr(\boldsymbol{X} \in B \mid R' > 1) \Pr(R' > 1), \tag{3.10}$$

for any set B lying entirely within the region $\{\boldsymbol{x} \in \mathbb{R}_+^d : \sum_{j=1}^d x_j > r_0(\boldsymbol{x}/\sum_{j=1}^d x_j)\}$; some examples are given in Figure 3.6. The first probability on the right-hand side of (3.10) can be estimated empirically from the simulated draws. The second probability may be estimated from the dataset as the proportion of points R' exceeding 1. When quantile regression at level τ has been used to estimate $r_0(\boldsymbol{w})$, we expect the proportion of points above the threshold to be near $1 - \tau$.

The fact we can simulate an arbitrary number of points from our model with the condition $\{R' > 1\}$ means that in principle we can extrapolate quite a way beyond the observed data. Nonetheless, such an approach may be computationally demanding for very extreme sets that require a large number of simulations. We consider now how to simulate given the condition $\{R' > k\}$, with $k > 1$; results will be illustrated in Section 3.5.

Simulation from the truncated gamma distribution of $R \mid [\boldsymbol{W} = \boldsymbol{w}, R > kr_0(\boldsymbol{w})]$ is again straightforward, but simulation from the distribution of angles $\boldsymbol{W} \mid R' > k$ is more challenging if k is sufficiently high that there are few or no empirical samples available. However, we have the relation

$$\begin{aligned} f_{\boldsymbol{W}}(\boldsymbol{w} \mid R' > k) &= \frac{\int_k^\infty f_{R',\boldsymbol{W}}(r', \boldsymbol{w} \mid R' > 1)\,\mathrm{d}r'}{\int_{\mathcal{S}_{d-1}} \int_k^\infty f_{R',\boldsymbol{W}}(r', \boldsymbol{v} \mid R' > 1)\,\mathrm{d}r'\,\mathrm{d}\boldsymbol{v}} \\ &= \frac{f_{\boldsymbol{W}}(\boldsymbol{w} \mid R' > 1) \int_k^\infty f_{R'|\boldsymbol{W}}(r' \mid \boldsymbol{w}, R' > 1)\,\mathrm{d}r'}{\int_{\mathcal{S}_{d-1}} \int_k^\infty f_{R'|\boldsymbol{W}}(r' \mid \boldsymbol{v}, R' > 1) f_{\boldsymbol{W}}(\boldsymbol{v} \mid R' > 1)\,\mathrm{d}r'\,\mathrm{d}\boldsymbol{v}}, \end{aligned} \tag{3.11}$$

where $f_{\boldsymbol{U}}(\cdot \mid V > v)$ denotes the density of a random vector $\boldsymbol{U} \mid V > v$, and integrating over $\mathcal{S}_{d-1}$ refers to integrating over the $(d-1)$-dimensional submanifold of $\mathbb{R}^d$ corresponding to the $(d-1)$ unit sphere, here and throughout. Note that

$$\int_k^\infty f_{R'|\boldsymbol{W}}(r' \mid \boldsymbol{w}, R' > 1)\,\mathrm{d}r' = \int_{kr_0(\boldsymbol{w})}^\infty f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}, R > r_0(\boldsymbol{w}))\,\mathrm{d}r,$$

so that under the truncated gamma approximation (3.5) for $R \mid [\boldsymbol{W} = \boldsymbol{w}, R > r_0(\boldsymbol{w})]$, we have the proportionality statement

$$f_{\boldsymbol{W}}(\boldsymbol{w} \mid R' > k) \propto f_{\boldsymbol{W}}(\boldsymbol{w} \mid R' > 1) \frac{\bar{F}[kr_0(\boldsymbol{w}); \alpha, g(\boldsymbol{w})]}{\bar{F}[r_0(\boldsymbol{w}); \alpha, g(\boldsymbol{w})]}. \tag{3.12}$$

The ratio of gamma survival functions in (3.12) can therefore be used as importance weights to derive an approximate sample from the distribution of $\boldsymbol{W} \mid R' > k$, using a sample from the distribution of $\boldsymbol{W} \mid R' > 1$.

Finally, to estimate $\Pr(R' > k)$, so that we can calculate extreme probabilities as in equation (3.10), note that the constant of proportionality in (3.12) is $\Pr(R' > k \mid R' > 1)$, from the denominator of equation (3.11). An estimate of this is therefore

$$\widehat{\Pr}(R' > k \mid R' > 1) = \frac{1}{n_0} \sum_{i=1}^{n_0} \frac{\bar{F}[kr_0(\boldsymbol{w}_i); \alpha, g(\boldsymbol{w}_i)]}{\bar{F}[r_0(\boldsymbol{w}_i); \alpha, g(\boldsymbol{w}_i)]},$$

where $\boldsymbol{w}_i, i = 1, \ldots, n_0$ are the angles corresponding to the values for which $R' > 1$. Lastly, $\widehat{\Pr}(R' > k) = \widehat{\Pr}(R' > k \mid R' > 1)\widehat{\Pr}(R' > 1)$, where $\widehat{\Pr}(R' > 1)$ is estimated empirically, as previously. We note that another alternative to this procedure is to fit the generalized Pareto distribution to $R' \mid R' > 1$ and use this fitted model to

estimate $\Pr(R' > k \mid R' > 1)$. Our investigation into this found that both options perform similarly for relatively small k, but the generalized Pareto model introduces extra uncertainty for larger k, and so we stick to the first approach in Section 3.5.

In our experience we have found that estimates of $\Pr(\boldsymbol{X} \in B)$ are relatively insensitive to the precise choice of k, provided both that k is large enough to ensure that several sample points lie in B, and that $B \subset \{\boldsymbol{x} \in \mathbb{R}_+^d : \sum_{j=1}^d x_j > kr_0(\boldsymbol{x}/\sum_{j=1}^d x_j)\}$, as is required for the analogue of equation (3.10) to hold. The simplicity of checking this latter condition depends on the shape of B and of $r_0(\boldsymbol{w})$, but it is easy to check visually for $d = 2$, and it may crudely be checked by ensuring that $k < \sum_{j=1}^d \tilde{x}_{l,j}/r_0(\tilde{\boldsymbol{x}}_l/\sum_{j=1}^d \tilde{x}_{l,j})$ for a sample of points $\tilde{\boldsymbol{x}}_l, l = 1, \ldots, m$, along the boundary of B. See Appendix 3.8.6 for further discussion and summary of our investigation. We recommend taking an intermediate k that is slightly smaller than the maximum for which this series of m inequalities holds, to safeguard against the crudeness of this check.

3.4.6 Summary of inference and prediction procedures

For convenience, we briefly summarize the procedures for inference and prediction via the geometric framework:

1. Determine a high threshold $r_0(\boldsymbol{w})$ of the distribution of $R \mid \boldsymbol{W} = \boldsymbol{w}$ for all $\boldsymbol{w} \in \mathcal{S}_{d-1}$ using either additive quantile regression or a rolling-windows approach.

2. Select a set of candidate parametric gauge functions $g(\cdot; \boldsymbol{\theta})$ and for each one fit the truncated gamma likelihood (3.8) to the n_0 values of $R \mid [\boldsymbol{W} = \boldsymbol{w}, R > r_0(\boldsymbol{w})]$.

3. Compare model fits using selection criteria such as the Akaike or Bayesian information criterion.

4. Use diagnostics such as the PP plot and comparison with the empirically-estimated gauge function to confirm acceptable fit of the best model(s).

5. Letting $R' = R/r_0(\boldsymbol{W})$, simulate new realizations from the distribution of $\boldsymbol{X} \mid R' > 1$ by drawing from the empirical distribution of $\boldsymbol{W} \mid R' > 1$ and multiplying by draws of $R \mid [\boldsymbol{W} = \boldsymbol{w}, R > r_0(\boldsymbol{w})]$ from the fitted truncated gamma distribution. If required, adapt these steps to simulate from the distribution of $\boldsymbol{X} \mid R' > k$ with $k > 1$.

6. Estimate $\Pr(\boldsymbol{X} \in B)$ using equation (3.10), or suitable adaptation if $R' > k$.

3.5 Simulation study

We now demonstrate the performance of our methods against existing approaches for analysing multivariate extremes. Our focus lies on estimation of probabilities $\Pr(\boldsymbol{X} \in B)$ for three sets B that lie in different parts of the region where $\boldsymbol{X}$ may be considered extreme.

We begin with the bivariate case, which is well-established and understood, demonstrating that our methodology gives estimates with low bias in each situation, performing competitively with other methods across a range of scenarios. Specifically, we compare with estimation methodology based on multivariate regular variation (MRV), hidden regular variation (Ledford and Tawn, 1997) (HRV) and the conditional extreme value model (CE) of Heffernan and Tawn (2004). The simplest approach to implementing MRV methodology is to use the approximation $\Pr(\boldsymbol{X} \in v + B') \approx e^{-v} \Pr(\boldsymbol{X} \in B')$, where we take as the set of interest $B = v + B'$, and B' is extreme, but in the range of the data so can be estimated empirically. This is a nonparametric implementation, but parametric assumptions are possible as well. Specifically we can also assume that equation (3.1) holds at finite levels and choose a parametric form for the angular measure H. We adopt both techniques below. HRV is a refinement of MRV that allows for situations where the spectral measure H places no mass on $\mathbb{B}_{\{1,\ldots,d\}}$. Implementation of this methodology relies on exploiting the relation $\Pr(\boldsymbol{X} \in v + B') \approx e^{-v/\eta} \Pr(\boldsymbol{X} \in B')$, where $\eta \in (0, 1]$ is the residual tail dependence coefficient; this is estimated using the Hill estimator (Ledford and

Tawn, 1997). Parametric models based on HRV exist (Ramos and Ledford, 2009), but are generally poorly-justified since the so-called "hidden angular measure" is often not a finite measure over the unit simplex; we therefore do not consider these here. Like MRV however, the asymptotics of HRV are suited only to extrapolating into regions where all variables are large simultaneously. Implementation of the CE methodology to model the distribution of $\boldsymbol{X} \mid X_j > u$, for large u, follows the original approach suggested in Heffernan and Tawn (2004), adapted to exponential margins. Probability estimation in this case is performed by simulating from the distribution of $\boldsymbol{X} \mid X_j > u'$, $u' \geq u$. Following the bivariate case, we move on to the more difficult case of $d = 3$, and show that we can substantially outperform the CE model in this setting, which is the only other viable approach for providing an estimate of the probabilities of interest.

3.5.1 Dimension $d = 2$

For the bivariate case, we perform estimation based on 5000 datapoints simulated from four different dependence structures: (I) logistic distribution with parameter $\gamma = 0.4$; (II) Gaussian distribution with $\rho = 0.8$; (III) inverted logistic distribution with $\gamma = 0.7$; (IV) logistic distribution with $\gamma = 0.8$. Distributions (I) and (IV) represent moderately strong and weak logistic dependence structures, respectively. In Appendix 3.8.7 we show examples of the four datasets, and three sets of interest $B_1 = (10, 12) \times (10, 12)$, $B_2 = (10, 12) \times (6, 8)$, and $B_3 = (10, 12) \times (2, 4)$.

In each case we fit model (3.5) to the data using four different gauge functions: those corresponding to the unique distributions (I)–(III), where the parameter is to be estimated from the data, and the function $g(\boldsymbol{x}; \theta) = \max[(x_1 - x_2)/\theta, (x_2 - x_1)/\theta, (x_1 + x_2)/(2 - \theta)]$. We select the model that yields the lowest value of the Akaike information criterion (AIC) for the prediction step, thereby avoiding using knowledge of the true data-generating process. Recall that before fitting model (3.5), we need to calculate a high threshold $r_0(\boldsymbol{w})$. In Section 3.4.1, we described using either additive quantile regression or a rolling-windows quantile calculation for

this. We used both techniques in the simulation study, setting $\tau = 0.95$, finding relatively little difference in the performance of the resulting inference, particularly in comparison to differences across extreme-value methodologies. Therefore, to keep presentation focused, we detail only the results where $r_0(\boldsymbol{w})$ was found using the simpler rolling-windows quantile method. Although our focus is on extreme probability estimation, we also display (non-)parametric estimates of G, obtained via $\hat{g}(\boldsymbol{w})$ and $g(\boldsymbol{w}; \widehat{\boldsymbol{\theta}})$, in Section 3.8.7.

For the parametric MRV approach, we employ a similar strategy to our geometric approach. After transforming to radial-angular coordinates $\|\boldsymbol{X}_P\|$ and $\boldsymbol{X}_P/\|\boldsymbol{X}_P\|_1$ from Pareto margins, we take all angles for which the corresponding radius exceeds the 0.95 quantile of radii, and fit a parametric form for the density of H via maximum likelihood. We choose between five parametric models for H using AIC. The five bivariate models are the logistic (Émile and Gumbel, 1960, corresponding to distributions (I) and (IV)), negative logistic (Galambos, 1975), bilogistic (Smith, 1990), negative bilogistic (Coles and Tawn, 1994), and the so-called Coles-Tawn distribution (Coles and Tawn, 1991). Probabilities are estimated using numerical integration over B using the fitted model for angles, combined with Pareto density for radii.

Figure 3.7 displays boxplots of the estimated probabilities for 200 repetitions across different methodologies:

- the geometric approach with different k (G1, G2);
- the conditional extremes model with different simulation thresholds (CE1, CE2);
- the nonparametric hidden regular variation approach (HRV), and
- nonparametric and parametric multivariate regular variation approaches (MRV, MRV2);

see the caption for further details. For distribution (I), all methods estimate $\Pr(\boldsymbol{X} \in B_1)$ with little bias; the smallest variance is attributed to the MRV approaches,

which is as expected since we are looking at a distribution where H places mass on $\mathbb{B}_{\{1,2\}}$ and estimating a probability in the joint tail. The geometric approach and CE estimate $\Pr(\boldsymbol{X} \in B_2)$ relatively well, with the smallest variance attributable to the geometric approach based on $\boldsymbol{X} \mid R' > k$ for suitable $k > 1$. HRV and MRV start to exhibit some bias because B_2 lies outside the joint tail region. For $\Pr(\boldsymbol{X} \in B_3)$, all estimates based on the nonparametric HRV and MRV approaches are equal to zero. For the geometric approach, we are able to estimate this probability well when selecting a suitable k. Specifically, in each repetition, we select one of the largest values of k such that $B_3 \subset \{\boldsymbol{x} : (x_1 + x_2) > kr_0[\boldsymbol{x}/(x_1 + x_2)]\}$. This results in all probabilities having a non-zero estimate, compared to 0% for nonparametric HRV/MRV, and 4.5% for CE (at each of two thresholds). This probability can be estimated as non-zero by parametric MRV, but with a little bias. A boxplot of this case is included in the left panel of Figure 3.8.

Distribution (IV) also represents the case where H places mass on $\mathbb{B}_{\{1,2\}}$, yet interestingly, MRV gives biased estimates in for all probabilities. This is likely due to the practical rate of convergence to the limiting angular measure H being slower under this weaker dependence scenario. Indeed we see differing estimates from the two MRV approaches, which are based on different effective "thresholds" for defining extremes. MRV changes from appreciably over-estimating the probabilities $\Pr(\boldsymbol{X} \in B_1)$ and $\Pr(\boldsymbol{X} \in B_2)$ to hugely under-estimating $\Pr(\boldsymbol{X} \in B_3)$. The geometric approach suggests a small under-estimation of $\Pr(\boldsymbol{X} \in B_1)$ and $\Pr(\boldsymbol{X} \in B_2)$ and good performance for $\Pr(\boldsymbol{X} \in B_3)$. CE shifts from large under-estimation to over-estimation moving from $\Pr(\boldsymbol{X} \in B_1)$ to $\Pr(\boldsymbol{X} \in B_3)$.

For distributions (II) and (III), the geometric approach and CE exhibit quite similar performance, although CE has a smaller variance for estimates of $\Pr(\boldsymbol{X} \in B_2)$ under distribution (II), and of $\Pr(\boldsymbol{X} \in B_3)$ under distribution (III). MRV is not an appropriate method for these distributions and always performs badly; HRV is appropriate in the joint tail, where it exhibits similar performance to other methods for (II) and better performance for (III), while it leads to poor estimates in other

regions. Additional boxplots in the right panel of Figure 3.8 display more detailed information for the estimates of $\Pr(\boldsymbol{X} \in B_1)$ under distribution (III). As described for distribution (I), we also used a suitable $k > 1$ for estimating this probability. The geometric approach outperforms CE in this case. This is because, using an appropriate k, we are able to simulate points to generate non-zero estimates of the probabilities (93% and 100% of estimates are positive for the two thresholds shown). In contrast, only 45.5% and 46% of estimates are positive for CE.

3.5.2 Dimension $d = 3$

We again perform estimation based on 5000 data points from three different data structures: (I) asymmetric logistic distribution, for which the spectral measure H places mass on $\mathbb{B}_{\{1,2\}}, \mathbb{B}_{\{1,3\}}, \mathbb{B}_{\{2,3\}}$, with parameters $\gamma_{\{1,2\}} = \gamma_{\{1,3\}} = \gamma_{\{2,3\}} = 0.4$; (II) asymmetric logistic distribution where H places mass on $\mathbb{B}_{\{1\}}, \mathbb{B}_{\{1,2\}}, \mathbb{B}_{\{2,3\}}$ and with parameters $\gamma_{\{1,2\}} = \gamma_{\{2,3\}} = 0.4$; (III) distribution constructed by taking an inverted Clayton copula with parameter 2 for (X_1, X_2), with $X_3 \mid X_2 = x_2$ drawn from an inverted logistic dependence structure with parameter 0.5. Such a distribution is in the domain of attraction of a spectral measure H placing mass on $\mathbb{B}_{\{1,2\}}, \mathbb{B}_{\{3\}}$. In Appendix 3.8.7 we display examples of the three datasets along with sets of interest $B_1 = (8, 10) \times (8, 10) \times (0.01, 3)$, $B_2 = (8, 10) \times (5, 7) \times (0.01, 3)$ and $(8, 10) \times (2, 4) \times (0.01, 3)$.

For the $d = 3$ case we consider only two methodologies: the geometric approach and CE, as HRV/MRV only perform well when considering sets B where all variables are of a similar magnitude, and the sets that we are considering are all small in x_3. Moreover for MRV we require mass on $\mathbb{B}_{\{1,2,3\}}$ for good performance of this method.

For the geometric approach we fit model (3.5) to the data after identifying potential suitable forms for the gauge function g. For this initial step, we calculate the coefficients $\tau_C(\delta)$, and associated estimates of the probability of mass on $\mathbb{B}_C$ as in Simpson et al. (2020), for $\delta = 0.4, 0.5, 0.6$ and $C \subseteq \{1, 2, 3\}$. These estimates help to identify potential faces of the simplex on which the limiting spectral measure

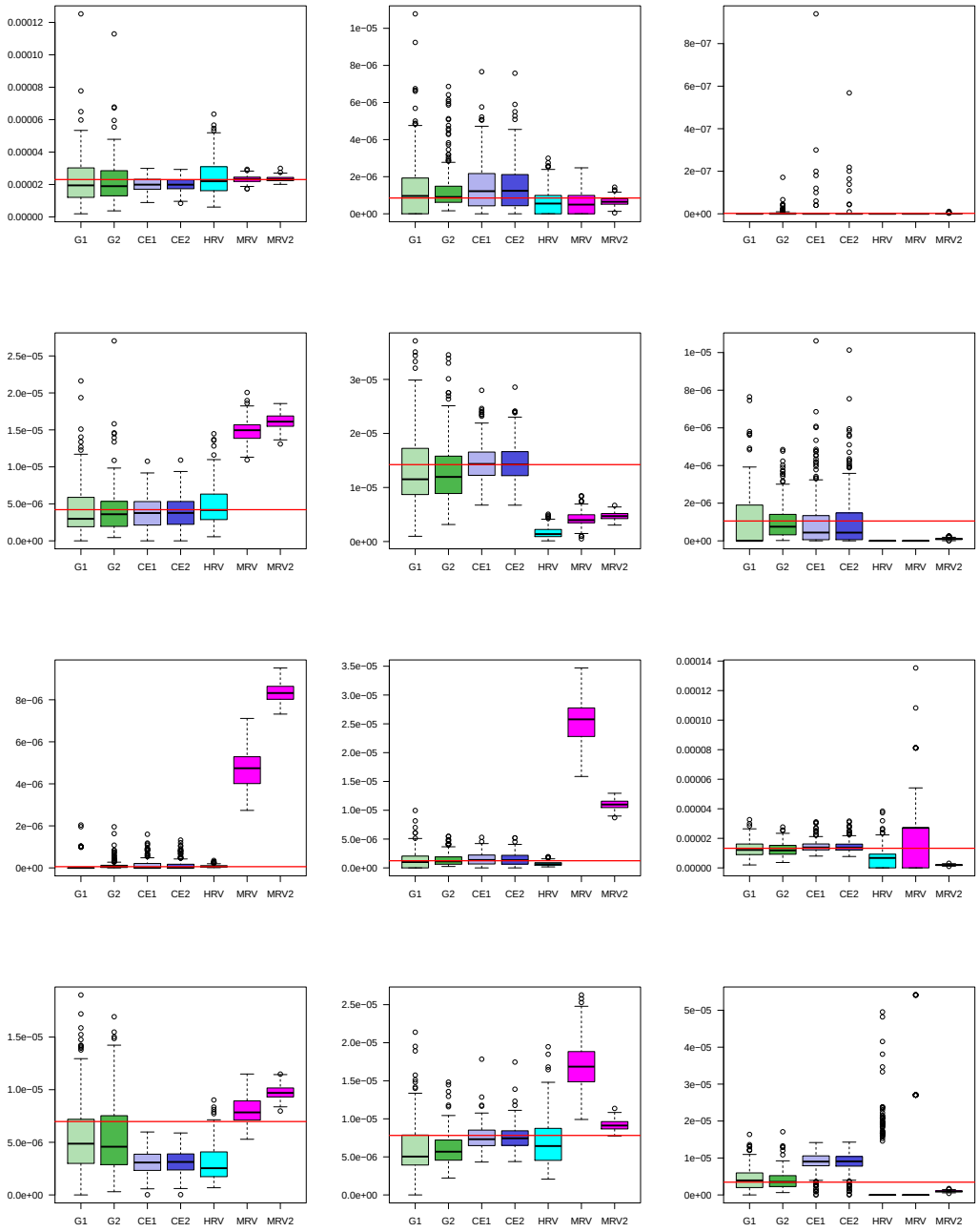

Figure 3.7: Boxplots of the 200 estimated probabilities for $d = 2$. From left to right columns represent $\Pr(\boldsymbol{X} \in B_1), \Pr(\boldsymbol{X} \in B_2), \Pr(\boldsymbol{X} \in B_3)$, respectively. From top to bottom, datasets are (I), (II), (III), (IV) respectively. Green boxplots, labelled G1, G2, give results from our geometric approach: G1 is calculated from $\boldsymbol{X} \mid R' > 1$; G2 is calculated from $\boldsymbol{X} \mid R' > k$, where k is determined as the maximum value such that all sets B_1, B_2, B_3 lie in the region $\{\boldsymbol{x} : x_1 + x_2 > kr_0(\boldsymbol{x}/(x_1 + x_2))\}$. Dark blue boxplots, labelled CE1, CE2 give results from the conditional extremes model: CE1 is calculated from $\boldsymbol{X} \mid X_1 > 6.9$; CE2 is calculated from $\boldsymbol{X} \mid X_1 > 10$. Turquoise boxplots, labelled HRV, give results from hidden regular variation methodology; purple boxplots, labelled MRV and MRV2, represent nonparametric and parametric multivariate regular variation, respectively. True values of the probabilities are indicated by horizontal red lines.

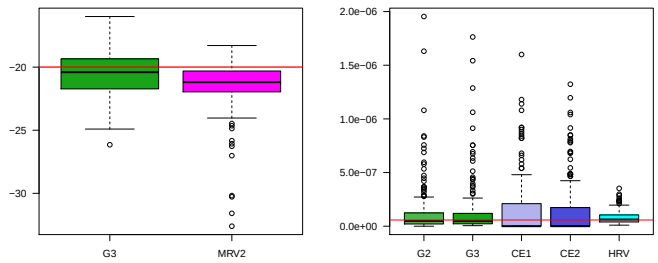

Figure 3.8: Left: boxplot of log estimates of $\Pr(\boldsymbol{X} \in B_3)$ for distribution (I), $d = 2$, using the geometric approach at a high threshold as described in the text (G3), and parametric MRV (MRV2). Right: estimates of $\Pr(\boldsymbol{X} \in B_1)$ for distribution (III), $d = 2$, using two different thresholds for the geometric approach, labelled G2, G3 (green). Estimates for two different thresholds from the conditional approach are labelled CE1, CE2 (dark blue), and from hidden regular variation, HRV (turquoise).

H places mass, and hence a suitable structure for the form of the gauge function. Specifically, where they exist, the coefficients $\tau_C(\delta) \in (0, 1]$ should be equal to 1 if H places mass on $\mathbb{B}_C$, for all values of the tuning parameter $\delta \in [\delta^\star, 1]$ and some $\delta^\star \in [0, 1]$. However, because of difficulties in estimating these coefficients precisely in the presence of nuisance parameters, Simpson et al. (2020) use them as part of a broader strategy to estimate of the probability of mass on $\mathbb{B}_C$. If all estimates for the three values of δ suggest the same extremal dependence structure in terms of where H places mass, then a single model is fitted, where the gauge function corresponds to that of the asymmetric logistic distribution for the identified structure. Otherwise, up to three different models are fitted, and the model with the lowest AIC is selected. We note that, for distributions (I) and (II), this means that we have the potential to fit the correct model form to the data, subject to its identification via the Simpson et al. (2020) methodology, although for distribution (III), we always have a misspecified model.

Figure 3.9 displays boxplots of the estimated probabilities across 200 repetitions using the two methods. In most cases, the geometric approach exhibits relatively

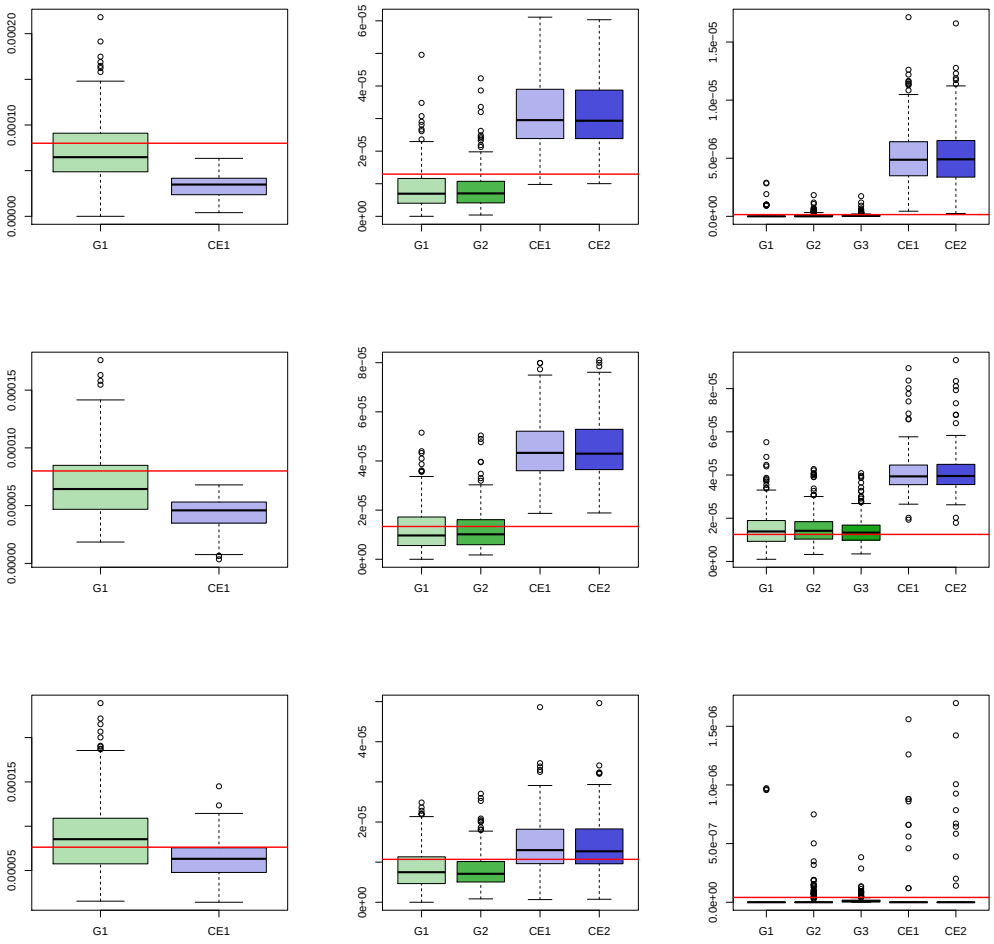

Figure 3.9: Boxplots of the 200 estimated probabilities for $d = 3$. From left to right columns represent $\Pr(\boldsymbol{X} \in B_1), \Pr(\boldsymbol{X} \in B_2), \Pr(\boldsymbol{X} \in B_3)$, respectively. From top to bottom, datasets are (I), (II), (III), respectively. Green boxplots, labelled G1, G2, G3 give results from the geometric approach: G1 is calculated from $\boldsymbol{X} \mid R' > 1$; G2 and G3 are calculated from $\boldsymbol{X} \mid R' > k_j$, $j = 1, 2$, where k_j is determined as a large value such that the sets B_2 or B_3 lie in the region $\{\boldsymbol{x} : x_1 + x_2 + x_3 > k_j r_0[\boldsymbol{x}/(x_1 + x_2 + x_3)]\}$. Dark blue boxplots, labelled CE1, CE2 give results from the conditional extremes model: CE1 is calculated from $\boldsymbol{X} \mid X_1 > 6.9$; CE2 is calculated from $\boldsymbol{X} \mid X_1 > 8$. True values of the probabilities are indicated by horizontal red lines.

low bias, particularly in comparison to CE, which is typically biased down for $\Pr(\boldsymbol{X} \in B_1)$ and up for $\Pr(\boldsymbol{X} \in B_2), \Pr(\boldsymbol{X} \in B_3)$. In conditional extreme value modelling, dependence structures are defined pairwise, so while any pair of variables (X_i, X_j) can theoretically have mass on $\mathbb{B}_{\{i,j\}}$ or $\mathbb{B}_{\{i\}}$ and $\mathbb{B}_{\{j\}}$, the methodology cannot usually capture more complex higher-order structures well. The structure of distribution (III) is the simplest, with only variables X_1, X_2 exhibiting simultaneous extremes, and CE is correspondingly more successful in this case. For $\Pr(\boldsymbol{X} \in B_3)$ and distributions (I) and (III), additional boxplots are provided in Appendix 3.8.7. These demonstrate that the geometric approach labelled G3 provides the best estimate in both cases, but underestimates the probability. In contrast we can see from Figure 3.9 that estimates of this probability for distribution (I) are biased strongly upwards for CE, while for distribution (III) only 5.5% of estimates for CE are positive at either threshold.

3.6 Data analyses

We use our new modelling approach to analyse two multivariate environmental datasets. The first is wave data from Newlyn, UK, included because of its extensive previous analysis in the literature. The second is a set of river flow data from Simpson et al. (2020).

3.6.1 Newlyn wave data

This dataset of 2894 measurements of wave height (metres), surge (metres) and period (seconds), denoted here as (X_H, X_S, X_P), was originally analysed in Coles and Tawn (1994) using a model that assumed multivariate regular variation with all mass of the spectral measure on $\mathbb{B}_{\{H,S,P\}}$. The full trivariate dataset has subsequently been analysed in Bortot et al. (2000), who assumed a censored multivariate Gaussian model, and Coles and Pauli (2002), whose model was able to accommodate the situation where the spectral measure places mass on some faces of the simplex, but

was otherwise quite restrictive.

The first step is to transform each marginal to exponential, which is done using a semi-parametric estimate of the distribution function for each variable X_j:

$$\widehat{F}_j(x) = \begin{cases} \tilde{F}_j(x), & x \leq u_j, \\ 1 - \phi_{u,j}\left[1 + \xi_j(x - u_j)/\sigma_j\right]_+^{-1/\xi_j}, & x > u, \end{cases} \tag{3.13}$$

where $\tilde{F}_j$ is the empirical df, u_j is a high threshold, $\phi_{u,j} = \Pr(X_j > u_j)$, and the form above u_j is the generalized Pareto distribution with scale $\sigma_j > 0$ and shape ξ_j. We take the thresholds u_H, u_S and u_P to be the 95% quantiles of the respective distributions.

To get an initial idea of the extremal dependence structure, we use the Simpson et al. (2020) methodology and calculate $\tau_C(\delta)$ for a range of values of δ. These estimates suggest that the spectral measure places mass on the faces $\mathbb{B}_{\{H\}}, \mathbb{B}_{\{S\}}, \mathbb{B}_{\{P\}}, \mathbb{B}_{\{H,S\}}$, which fits with the assessment in Bortot et al. (2000) and Coles and Pauli (2002).

To calculate the threshold $r_0(\boldsymbol{w})$, we use the rolling-windows procedure described in Section 3.4.1, with $\tau = 0.95$. We then fit model (3.5) with three forms for g: (i) the asymmetric logistic gauge function with the structure given by $\tau_C(\delta)$, (ii) gauge corresponding to the Gaussian distribution, and (iii) an additive mixture of the Gaussian and asymmetric logistic gauges, as described in Section 3.4.3.2. The respective AIC values are 374.9, 365.5 and 369.5.

In spite of the structure suggested by the estimated $\tau_C(\delta)$ values, the AIC indicates a preference for the Gaussian gauge function. The maximum likelihood estimates are $(\widehat{\alpha}, \widehat{\theta}_{HP}, \widehat{\theta}_{HS}, \widehat{\theta}_{PS}) = (0.79, 0.70, 0.65, 0.30)$, where θ_{jk} are the Gaussian correlation parameters in the gauge function. The data have been filtered to give approximate temporal independence, so we estimate Hessian-based standard errors as $(0.74, 0.12, 0.12, 0.18)$. Figure 3.10 displays the PP plot for this fit as described in Section 3.4.4, as well as the same plot transformed onto the exponential scale to emphasize the upper tail, indicating no lack of fit. We also compare the empirical gauge $\hat{g}(\boldsymbol{w})$ and the fitted Gaussian gauge function $g(\boldsymbol{w}; \widehat{\boldsymbol{\theta}})$ in the right panel of

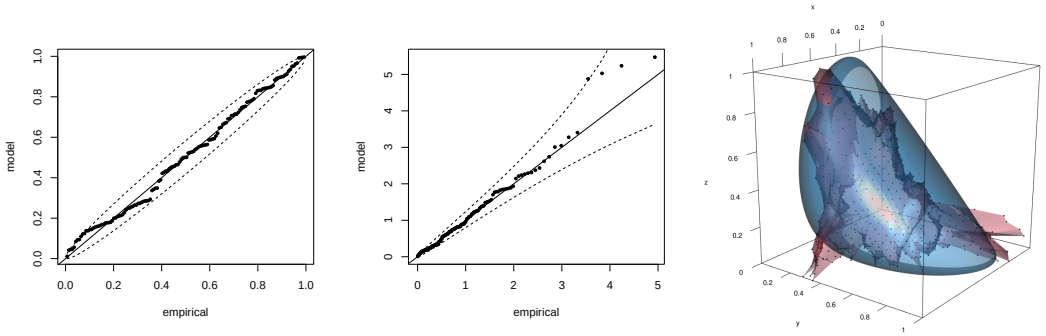

Figure 3.10: Left and centre: PP and exponential QQ plots for the fitted truncated gamma model with the Gaussian gauge function. Right: unit level set of the empirical gauge function (jagged, red) and fitted Gaussian gauge function (smooth, blue) for the Newlyn wave data.

Figure 3.10. The empirical gauge is relatively "jagged" and variable due to the manner of its calculation, but there is broad correspondence between its overall shape and that of the fitted gauge. Interestingly, the fit of the asymmetric logistic gauge returns a parameter estimate of $\widehat{\gamma}_{HS} = 1$, which is on the boundary of the parameter space. This could be indicative of wave height and surge not displaying exceptionally strong dependence, but also because there are restrictions on the shape of the limit set arising from the asymmetric logistic distribution, and this parameter estimate provides the best overall fit to all data simultaneously.

As a further diagnostic, we compare empirical and model-based estimates of the sub-asymptotic joint *tail dependence coefficient*. For $X_j \sim F_j$, this is defined by

$$\chi_C(u) = \frac{1}{1-u} \Pr\left[F_j(X_j) > u, \forall j \in C\right], \qquad u \in (0,1), \quad C \subseteq \{H, S, P\}. \tag{3.14}$$

The empirical estimator of $\chi_C(u)$ is obtained by replacing each distribution function and joint probability with its empirical counterpart, while the model-based estimate is calculated using simulation from the fitted model as described in Section 4.3.5, and suitable sets B. In Figure 3.11 we consider $\chi_{HSP}(u)$ and $\chi_{HS}(u)$, meaning $B = (-\log(1-u), \infty)^3$ and $B = (-\log(1-u), \infty)^2 \times (0, \infty)$, respectively. The range over which the model-based tail dependence coefficients can be calculated

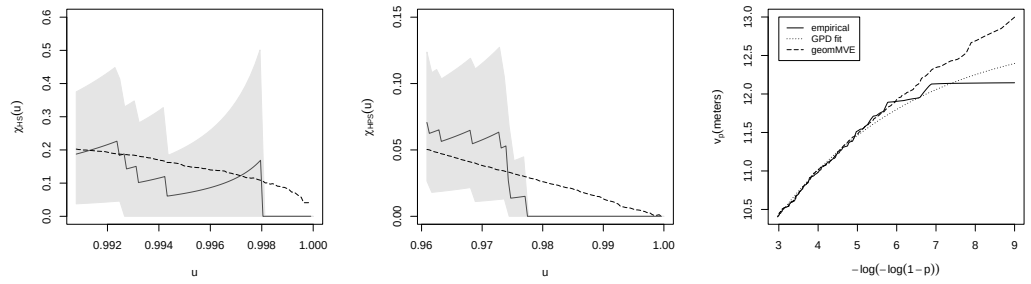

Figure 3.11: Left: Estimates of $\chi_{HS}(u)$, $u \in (0.99, 1)$; centre: estimates of $\chi_{HSP}(u)$, $u \in (0.965, 1)$. Black solid lines represent empirical estimates, grey shaded regions 95% pointwise confidence intervals, and dashed lines the model-based estimate for the Newlyn wave dataset. Right: plot of quantiles v_p of the structure variable V, representing sea wall height.

depends on the values of $\boldsymbol{X}$ constituting the extreme region $\{R' > 1\}$. There is good agreement with the empirical estimates, with the model-based estimates allowing extrapolation beyond the range of the data.

Finally we consider analysis of the structure variable outlined in Coles and Tawn (1994). They introduce the overtopping discharge rate $Q(v; \boldsymbol{X}_{HSP})$ for a sea wall of height v as

$$Q(v; \boldsymbol{X}_{HSP}) = a_1 X_S X_P \exp\left[-\frac{a_2\,(v - X_S - l)}{X_P {X_H^*}^{1/2}}\right],$$

where

$$X_H^* = X_H \left\{1 - \exp\left[-\frac{(l + X_S)^2}{2X_H^2}\right]\right\}^{1/2}.$$

The value X_H^* is introduced to approximate the actual off-shore wave height, since measurements are taken on-shore. The goal is to estimate the sea wall height v_p (in metres) for which the overtopping discharge rate is expected to exceed $0.002 m^3 s^{-1}$ per metre of sea wall with probability p. That is, setting $V = Q^{-1}(0.002; \boldsymbol{X}_{HSP})$, we solve $\Pr(V > v_p) = p$ for v_p using realizations of V generated through simulation and reverse marginal transformation. Specifically, we generate new realizations of $\boldsymbol{X}_{HSP}$, and hence V, in the tail region of our model by simulating on exponential

margins and inverting equation (3.13). Outside of the tail region, we use the empirical distribution of V. As in Bortot et al. (2000), we fix $a_1 = 0.25$, $a_2 = 26$, and $l = 4.3$. The right panel of Figure 3.11 displays the obtained values v_p, with empirical quantiles and those calculated from fitting the generalized Pareto distribution directly to the tail of V (the so-called "structure variable approach") for comparison. For very small p, the return levels obtained from the geometric model are larger than those from the generalized Pareto fit. They are comparable to those obtained in Bortot et al. (2000), but much lower than those in Coles and Tawn (1994), whose model incorrectly assumes that the spectral measure places mass on $\mathbb{B}_{\{H,S,P\}}$.

3.6.2 River flow data

We now apply our modelling approach on 12,327 measurements of daily mean river flow (m^3/s) from four gauging stations in the north west of England. The data were previously explored in Simpson et al. (2020), where focus lay on determining the support of the spectral measure, but not subsequent modelling of the variables, due to lack of suitable models that could account for complex structures. We opt to consider four out of the five locations initially used in order to keep the number of parameters reasonable; further discussion on dimensionality can be found in Section 6.6. The four stations, labelled 1, 2, 3, 4, correspond to those labelled A, B, C, D in Simpson et al. (2020).

Margins are standardized using equation (3.13). We then use the Simpson et al. (2020) methodology, which suggests that the spectral measure may place mass on the faces $\mathbb{B}_{\{2\}}, \mathbb{B}_{\{4\}}, \mathbb{B}_{\{1,4\}}, \mathbb{B}_{\{1,3,4\}}$, and $\mathbb{B}_{\{1,2,3,4\}}$ of the simplex $\mathcal{S}_3$. We fit the model with the corresponding asymmetric logistic gauge function, a Gaussian gauge function, and an additive mixture of the two. The AIC values are 2666, 2601 and 2609, respectively. Once again, the model with the Gaussian gauge is preferred, in apparent conflict with the estimated structure of the spectral measure, though we note this is also subject to uncertainty. Parameter estimates and approximate

standard errors are given in Table 3.1. To account for temporal dependence of river flows, standard errors are found via use of a block bootstrap on the original data series, with block length 20. The asymmetric logistic gauge, while able to capture the structure of different groups of variables being co-extreme, appears too inflexible to capture other aspects of the dependence. The additively mixed model is an attempt to alleviate this problem, but leads to a large number of parameters without a sufficient improvement in fit to compensate for them.

Parameter	α	θ_{12}	θ_{13}	θ_{14}	θ_{23}	θ_{24}	θ_{34}
Estimate	2.46	0.83	0.90	0.80	0.90	0.57	0.62
Standard error	0.62	0.11	0.14	0.14	0.14	0.16	0.14

Table 3.1: Parameter estimates and approximate block bootstrap-based standard errors for the river flow data. Parameter θ_{jk} represents the Gaussian gauge correlation parameter between sites j, k.

Figure 4.8 displays coefficients $\chi_{123}(u)$, $\chi_{134}(u)$ and $\chi_{1234}(u)$, defined analogously to (3.14). If H places mass on $\mathbb{B}_{\{1,2,3,4\}}$, then each of these coefficients has a positive limit as $u \to 1$, but at observable levels, the model-based estimates from the Gaussian gauge all represent a good fit to the data. Plots of $\chi_C(u)$ for the remaining groups of variables are given in Appendix 3.8.8, along with the PP plot, showing no lack of fit. In the limit as $u \to 1$, estimates of $\chi_C(u)$ from the geometric model with Gaussian gauge will all be zero. However, the inference that H places mass on $\mathbb{B}_{\{1,2,3,4\}}$, and other faces, is subject to uncertainty. From the plots in Figure 4.8, it is difficult to determine whether the limits of $\chi_{123}(u)$, $\chi_{134}(u)$ and $\chi_{1234}(u)$ as $u \to 1$ are indeed positive or zero, and as a consequence whether a gauge function that reflects $H(\mathbb{B}_{\{1,2,3,4\}}) > 0$ is truly preferable. Nonetheless, this framework offers the chance to test these models and assumptions in a way that was not previously possible.

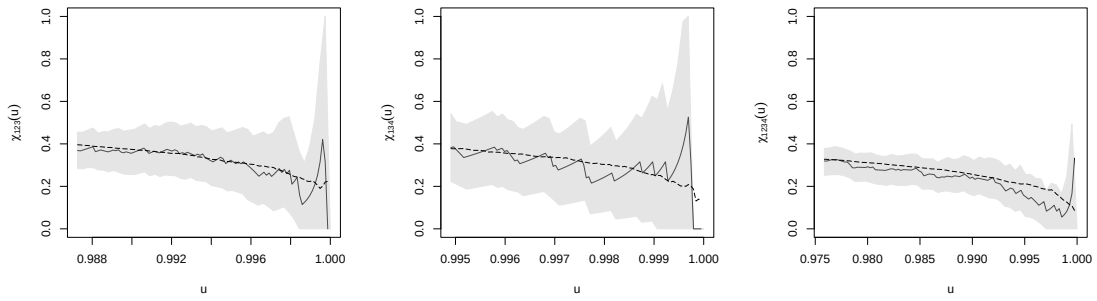

Figure 3.12: Empirical (solid black) estimates of $\chi_C(u)$ with 95% pointwise confidence interval (grey shaded region), and model-based estimate (dashed black) for $C = \{1, 2, 3\}$, $\{1, 3, 4\}$ and $\{1, 2, 3, 4\}$ (left to right).

3.7 Discussion

We have presented a new approach to multivariate extreme value modelling, based on estimation of the shape of the limit set of a sample cloud of data points in light-tailed margins. The methodology allows for modelling datasets with complicated extremal dependence structures, whereby different groups of variables may be co-extreme, as well as extrapolation into parts of the multivariate tail where only some variables are large.

By offering models for complex dependence structures with non-simultaneous extremes, this approach paves the way for more useful higher dimensional extreme value modelling. Recent literature on multivariate extremes that is targeted at higher dimensions typically involves making strong simplifying assumptions on the dependence structure. For example, the extremal graphical models outlined in Engelke and Hitz (2020) require an assumption that the spectral measure H places all mass on $\mathbb{B}_{\{1,\ldots,d\}}$.

In this work, we demonstrated the methods up to dimension $d = 4$. The main challenges for routine application of the methods for d larger than 3 or 4 lie in calculation of the threshold function $r_0(\boldsymbol{w})$, and specification of flexible gauge functions. The former could potentially be addressed by adapting the additive

quantile regression approach of Fasiolo et al. (2021) to incorporate basis functions whose support is the simplex $\mathcal{S}_{d-1}$. Addressing the latter challenge requires ways to build flexible and parsimonious gauge functions, which is a topic of current work. In particular, we note that models fitted in Section 3.6 had the ability to capture the complex structures suggested by the Simpson et al. (2020) methodology, but the best fits were obtained through models that were more flexible in other aspects. This led to the conclusion that the model with the Gaussian gauge function was preferred for both datasets, which is likely a consequence of being able to capture a range of strengths of dependences across different groups of variables; in contrast, the asymmetric logistic gauge function treats groups of variables that do not exhibit simultaneous extremes as effectively independent. We note also that estimates of the faces $\mathbb{B}_C$ on which H places mass are themselves subject to uncertainty, which is not easily quantifiable thanks to the requirement to select tuning parameters. Conflicts between the estimated structure and the selected gauge function may therefore not be too concerning, provided the diagnostics for the model are adequate.

A further challenge with our methodology for dimensions $d \geq 5$ is the use of the empirical distribution for the angles $\boldsymbol{W}$. We anticipate that considering (semi-)parametric forms for this distribution will be needed as part of adapting the methods to higher dimensions.

We have focused here primarily on positive dependence as it is common in many datasets and simplifies the presentation. For datasets exhibiting any form of negative dependence, the limit set shapes are more descriptive in Laplace, rather than exponential, margins. For example, we mentioned for the multivariate Gaussian case that the continuous convergence to $g(\boldsymbol{x})$ fails when some component of $\boldsymbol{x}$ is zero; this is not an issue in Laplace margins, where the limit set lies in the region $[-1, 1]^d$ rather than $[0, 1]^d$, and similarly for the t_ν distribution. Moving from the positive quadrant to $\mathbb{R}^d$ requires defining the angles $\boldsymbol{W}$ differently, but otherwise a similar approach could be applied, and represents a natural next step in developing this methodology.

Data availability

Data for the analyses in Section 3.6 are available as supplementary material in the JRSSB submission (Wadsworth and Campbell, 2024). Acknowledgement: Data from the UK National River Flow Archive (Section 3.6.2).

Computer code

Code for the analyses in Section 3.6 is available as supplementary material in the JRSSB submission (Wadsworth and Campbell, 2024). An R package `geometricMVE` for implementing the methodology presented in the article is also available as supplementary material and at `http://www.lancaster.ac.uk/~wadswojl/geometricMVE.html`. Interactive versions of 3d plots are available at the same URL.

3.8 Appendix

3.8.1 Example limit sets

Figures 3.13 and 3.14 display example illustrations of limit sets in exponential margins for three dependence structures in dimension $d = 2$ and $d = 3$ respectively. Equations for the gauge functions of these limit sets can be found in Section 3.3, or Appendix 3.8.2.

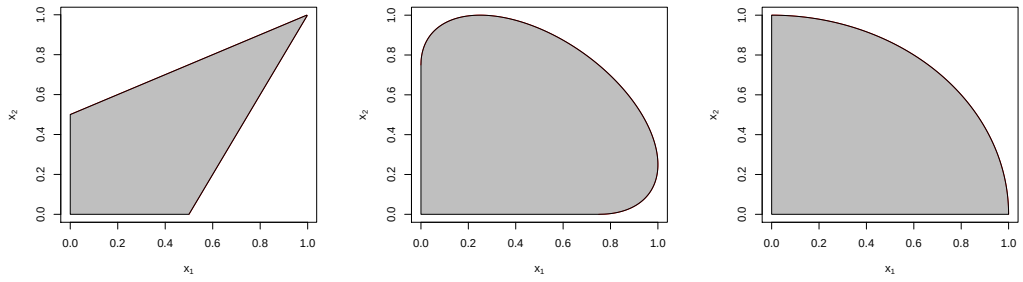

Figure 3.13: Illustration of limit sets (grey shaded region) arising from the logistic, Gaussian, and inverted logistic distributions (L-R) in dimension $d = 2$. Red lines represent unit level sets of the gauge function. Each dependence parameter is equal to 0.5.

3.8.2 Gauge functions and conditional distributions of $R \mid \boldsymbol{W} = \boldsymbol{w}$

Here we provide detailed calculations of the gauge functions such that we can establish the asymptotic behaviour of $f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w})$ as $r \to \infty$. In each case, we begin with the relevant density in exponential margins and calculate the asymptotic behaviour of $f_{\boldsymbol{X}}(t\boldsymbol{x})$ as $t \to \infty$; this is subsequently used to establish results for $f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w})$ as $r \to \infty$. We recall that the notation for ordered values is $x_{(1)} \geq x_{(2)} \geq \cdots \geq x_{(d)} > 0$, and similarly $w_{(1)} \geq w_{(2)} \geq \cdots \geq w_{(d)} > 0$.

Logistic distribution The d-dimensional logistic distribution with unit Fréchet margins has density

$$f_{\boldsymbol{Z}_F}(\boldsymbol{z}) = \exp\{-V(\boldsymbol{z})\} \sum_{\pi \in \Pi} \prod_{s \in \pi} -V_s(\boldsymbol{z}),$$

where $V : \mathbb{R}_+^d \to \mathbb{R}_+$ is the homogeneous of order -1 *exponent function*. For the logistic distribution, this is

$$V(\boldsymbol{z}) = \left(\sum_{j=1}^{d} z_j^{-1/\gamma} \right)^{\gamma}, \qquad \gamma \in (0, 1].$$

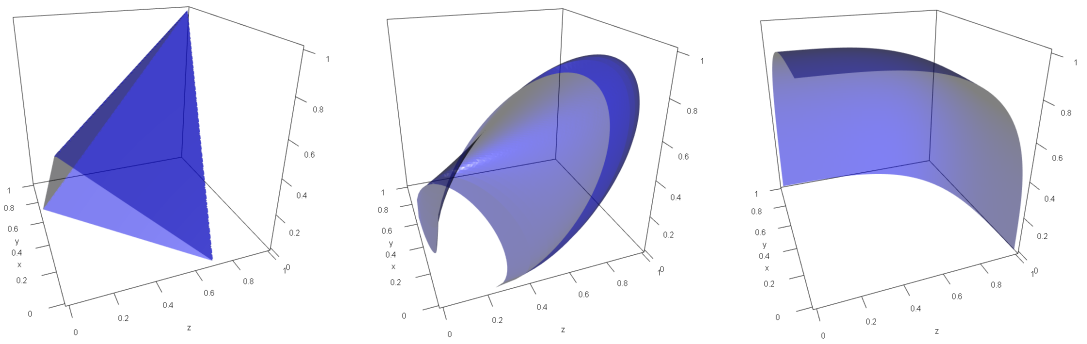

Figure 3.14: Illustration of limit sets (region between blue surface and planes $x_j = 0$) arising from the logistic, Gaussian, and inverted logistic distributions (L-R) in dimension $d = 3$. Blue surfaces represent unit level sets of the gauge function. Dependence parameters are set to: $\gamma = 0.7$, $\boldsymbol{\rho} = (0.2, 0.5, 0.8)$ and $\gamma = 0.3$, respectively.

The transformation to exponential margins is given by $z_j = [-\log(1 - e^{-x_j})]^{-1}$. Expanding this to give the asymptotic behaviour for large x_j yields $z_j(x_j) = e^{x_j} + 1/2 + O(e^{-x_j})$. We can therefore express the density in exponential margins as

$$\begin{aligned} f_{\boldsymbol{X}}(t\boldsymbol{x}) =& \exp\{-V(e^{t\boldsymbol{x}} + 1/2 + O(e^{-t\boldsymbol{x}}))\} \sum_{\pi \in \Pi} \prod_{s \in \pi} V_s(e^{t\boldsymbol{x}} + 1/2 + O(e^{-t\boldsymbol{x}})) \\ & \times e^{t\sum_{j=1}^{d} x_j} [1 + O(e^{-2tx_{(d)}})]. \end{aligned}$$

Firstly consider the contribution $\exp\{-V(e^{t\boldsymbol{x}}+1/2+O(e^{-t\boldsymbol{x}}))\}$. By homogeneity

$$\begin{aligned} V(e^{t\boldsymbol{x}} + 1/2 + O(e^{-t\boldsymbol{x}})) &= e^{-tx_{(d)}} V(e^{t(\boldsymbol{x} - x_{(d)})} + e^{-tx_{(d)}}/2 + O(e^{-t(\boldsymbol{x} + x_{(d)})})) \\ &= e^{-tx_{(d)}} [c + o(1)], \qquad t \to \infty, \end{aligned}$$

where c is a constant that equals 1 if $x_{(d)} < x_{(d-1)}$. Consequently,

$$\exp\{-V(e^{t\boldsymbol{x}} + 1/2 + O(e^{-t\boldsymbol{x}}))\} = 1 + O(e^{-tx_{(d)}}).$$

Next consider the partial derivatives $V_s(\boldsymbol{z})$. We have

$$V_s(\boldsymbol{z}) \propto \left(\prod_{j \in s} z_j \right)^{-1/\gamma - 1} \left(\sum_{j=1}^{d} z_j^{-1/\gamma} \right)^{\gamma - |s|},$$

and therefore

$$
\begin{aligned}
&V_s(e^{t\boldsymbol{x}} + 1/2 + O(e^{-t\boldsymbol{x}})) \\
&\propto \left(\prod_{j\in s}[e^{tx_j} + 1/2 + O(e^{-tx_j})]\right)^{-1/\gamma-1} \left(\sum_{j=1}^{d}[e^{tx_j} + 1/2 + O(e^{-tx_j})]^{-1/\gamma}\right)^{\gamma-|s|} \\
&= e^{-t(1/\gamma+1)\sum_{j\in s} x_j}[1 + O(e^{-t\min_{j\in s} x_j})] \\
&\qquad \times e^{-tx_{(d)}(1-|s|/\gamma)}[1 + O(e^{-t(x_{(d-1)}-x_{(d)})/\gamma}) + O(e^{-tx_{(d)}/\gamma})] \\
&= e^{-t(1/\gamma+1)\sum_{j\in s} x_j - tx_{(d)}(1-|s|/\gamma)}[1 + O(e^{-t(x_{(d-1)}-x_{(d)})/\gamma}) + O(e^{-tx_{(d)}/\gamma})].
\end{aligned}
$$

For each partition $\pi \in \Pi$,

$$
\begin{aligned}
&\prod_{s\in\pi} V_s(e^{t\boldsymbol{x}} + 1/2 + O(e^{-t\boldsymbol{x}})) \\
&\propto e^{-t(1/\gamma+1)\sum_{j=1}^{d} x_j - t\sum_{s\in\pi} x_{(d)}(1-|s|/\gamma)}[1 + O(e^{-t(x_{(d-1)}-x_{(d)})/\gamma}) + O(e^{-tx_{(d)}/\gamma})] \\
&= e^{-t[(1/\gamma+1)\sum_{j=1}^{d} x_j + x_{(d)}(|\pi|-d/\gamma)]}[1 + O(e^{-t(x_{(d-1)}-x_{(d)})/\gamma}) + O(e^{-tx_{(d)}/\gamma})].
\end{aligned}
$$

Combining all of these results yields

$$
f_{\boldsymbol{X}}(t\boldsymbol{x}) \propto \sum_{\pi\in\Pi} e^{-t[(1/\gamma)\sum_{j=1}^{d} x_j + x_{(d)}(|\pi|-d/\gamma)]}[1 + O(e^{-t(x_{(d-1)}-x_{(d)})/\gamma}) + O(e^{-tx_{(d)}})].
$$

The gauge function comes from taking $\min_{\pi\in\Pi}[(1/\gamma)\sum_{j=1}^{d} x_j + x_{(d)}(|\pi| - d/\gamma)]$, which clearly occurs for π with $|\pi| = 1$, i.e. $\pi = \{\{1, \ldots, d\}\}$. Hence $g(\boldsymbol{x}) = (1/\gamma)\sum_{j=1}^{d} x_j + x_{(d)}(1 - d/\gamma)$.

Turning to the conditional distribution of $R \mid \boldsymbol{W} = \boldsymbol{w}$, we have

$$
f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1}e^{-rg(\boldsymbol{w})}[1 + O(e^{-r(w_{(d-1)}-w_{(d)})/\gamma}) + O(e^{-rw_{(d)}})], \qquad r \to \infty.
$$

Negative logistic MGPD MGPDs have marginal scale and shape parameters, and when these are all set to 1 and 0 respectively, the marginal distributions are exponential conditionally upon being positive. That is, if $\boldsymbol{Z}$ follows a MGPD with unit-scale and zero-shape parameters, the marginal distribution functions $F_j(z)$ are

$$
F_j(z) = \begin{cases} \Pr(Z_j \le z), & z < 0, \\ c_j + (1 - c_j)(1 - e^{-z}), & z \ge 0, \end{cases}
$$

with $c_j = \Pr(Z_j \leq 0)$. To translate to exponential margins we solve $x_j = -\log(1 - F_j(z_j))$, which leads to $z_j(x_j) = x_j + \log(1 - c_j)+$, $x_j > -\log(1 - c_j)$.

Following calculations in Kiriliouk et al. (2019), the unit-scale zero-shape MGPD density associated to the negative logistic max-stable distribution is

$$f(\boldsymbol{z}) \propto e^{\gamma \sum_{j=1}^d z_j} \left(\sum_{j=1}^d e^{\gamma z_j} \right)^{-(d+1/\gamma)}, \qquad \gamma > 1,$$

and so on the region $\{\boldsymbol{x} : x_j > -\log(1 - c_j)/t, j = 1, \ldots, d\}$

$$\begin{aligned} f_{\boldsymbol{X}}(t\boldsymbol{x}) &\propto e^{\gamma \sum_{j=1}^d tx_j + \gamma \sum_{j=1}^d \log(1-c_j)} \left(\sum_{j=1}^d e^{\gamma t x_j + \gamma \log(1-c_j)} \right)^{-(d+1/\gamma)}, \\ &\propto e^{t\gamma \sum_{j=1}^d x_j - t(1+d\gamma)x_{(1)}} \left[1 + O(e^{t(x_{(2)} - x_{(1)})\gamma}) \right], \end{aligned}$$

so that $g(\boldsymbol{x}) = (1 + d\gamma)x_{(1)} - \gamma \sum_{j=1}^d x_j$, and

$$f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1} e^{-rg(\boldsymbol{w})} \left[1 + O(e^{r(w_{(2)} - w_{(1)})\gamma}) \right],$$

on the region $\{r > \min_{1 \leq j \leq d} -\log(1 - c_j)/w_j\}$. Outside of this region we require knowledge of the distribution of $Z_j \mid Z_j < 0$, $j = 1, \ldots, d$, which is harder to summarize in general.

Dirichlet MGPD In this case

$$f(\boldsymbol{z}) \propto e^{\sum_{j=1}^d \theta_j z_j} \left(\sum_{j=1}^d e^{z_j} \right)^{-\left(\sum_{j=1}^d \theta_j + 1\right)}, \qquad \theta_1, \ldots, \theta_d > 0,$$

and so on the region $\{\boldsymbol{x} : x_j > -\log(1 - c_j)/t, j = 1, \ldots, d\}$

$$\begin{aligned} f_{\boldsymbol{X}}(t\boldsymbol{x}) &\propto e^{\sum_{j=1}^d t\theta_j x_j + \sum_{j=1}^d \theta_j \log(1-c_j)} \left(\sum_{j=1}^d e^{tx_j + \log(1-c_j)} \right)^{-\left(\sum_{j=1}^d \theta_j + 1\right)}, \\ &\propto e^{t \sum_{j=1}^d \theta_j x_j - t(1 + \sum_{j=1}^d \theta_j)x_{(1)}} \left[1 + O(e^{t(x_{(2)} - x_{(1)})}) \right], \end{aligned}$$

so that $g(\boldsymbol{x}) = (1 + \sum_{j=1}^d \theta_j)x_{(1)} - \sum_{j=1}^d \theta_j x_j$, and

$$f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1} e^{-rg(\boldsymbol{w})} \left[1 + O(e^{r(w_{(2)} - w_{(1)})}) \right],$$

on the region $\{r > \min_{1 \leq j \leq d} -\log(1 - c_j)/w_j\}$.

Inverted max-stable distributions Recall the form of the density is

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \exp\{-l(\boldsymbol{x})\} \sum_{\pi \in \Pi} \prod_{s \in \pi} l_s(\boldsymbol{x}).$$

The derivatives $l_s(\boldsymbol{x})$ are homogeneous of order $1 - |s|$, and so

$$f_{\boldsymbol{X}}(t\boldsymbol{x}) = \exp\{-tl(\boldsymbol{x})\} \sum_{\pi \in \Pi} \prod_{s \in \pi} t^{1-|s|} l_s(\boldsymbol{x}).$$

The leading-order term in the summation therefore comes from the partition $\pi = \{\{1\}, \{2\}, \ldots, \{d\}\}$, with $|s| = 1$ for all $s \in \pi$. Second-order behaviour comes from the d partitions containing one set with $|s| = 2$ and all others with $|s| = 1$. We therefore have

$$f_{\boldsymbol{X}}(t\boldsymbol{x}) = \exp\{-tl(\boldsymbol{x})\} l_{\{1\}}(\boldsymbol{x}) \cdots l_{\{d\}}(\boldsymbol{x})[1 + O(t^{-1})],$$

so that $g(\boldsymbol{x}) = l(\boldsymbol{x})$, and

$$f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1} \exp\{-rg(\boldsymbol{w})\}[1 + O(r^{-1})].$$

Multivariate Gaussian distribution The multivariate Gaussian density with Gaussian margins is

$$f_{\boldsymbol{Z}_G}(\boldsymbol{z}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{z}^\top \Sigma^{-1} \boldsymbol{z}\right\} = \exp\left\{-\frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{d} z_j z_k \omega_{jk}\right\},$$

where $\Omega = (\omega_{jk})_{j,k} = \Sigma^{-1}$ is the precision matrix. Transforming to exponential margins, we obtain

$$f_{\boldsymbol{X}}(\boldsymbol{x}) \propto \exp\left\{-\frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{d} z_j(x_j) z_k(x_k) \omega_{jk}\right\} \prod_{j=1}^{d} \frac{1 - \Phi(z_j(x_j))}{\phi(z_j(x_j))}, \tag{3.15}$$

where $z_j(x_j)$ is found through solving $x_j = -\log(1 - \Phi(z_j))$, with ϕ, Φ the standard univariate normal density and df, respectively. Since we are interested in tx_j, $t \to \infty$, we exploit Mills' ratio for the solution. Dropping the component index, and writing $z_t = z(tx)$, this gives

$$tx = -\log\left\{\frac{\phi(z_t)}{z_t}[1 + O(z_t^{-2})]\right\} = \frac{z_t^2}{2} + \frac{1}{2}\log 2\pi + \log z_t + O(z_t^{-2}),$$

which is rearranged to give

$$z_t = (2tx)^{1/2} - (2tx)^{-1/2}\frac{\log 4\pi tx}{2} + O\left(\frac{(\log t)^2}{t^{3/2}}\right).$$

We firstly deal with the Jacobian expression in density (3.15). Again via Mills' ratio,

$$\begin{aligned}\prod_{j=1}^{d}\frac{1-\Phi(z_j(x_j))}{\phi(z_j(x_j))} &= \prod_{j=1}^{d} z_j(tx_j)^{-1}[1+O(z_j(tx_j)^{-2})] \\ &= 2^{-d/2}t^{-d/2}\prod_{j=1}^{d} x_j^{-1/2}[1+O(\log t/t)].\end{aligned}$$

Next consider the terms in the quadratic form of the exponent:

$$\begin{aligned}\omega_{jk}z_j(tx_j)z_k(tx_k) =& \omega_{jk}\left\{2t(x_jx_k)^{1/2} - \left(\frac{x_j}{x_k}\right)^{1/2}\frac{\log 4\pi tx_k}{2} - \left(\frac{x_k}{x_j}\right)^{1/2}\frac{\log 4\pi tx_j}{2}\right\} \\ &+ O\left(\frac{(\log t)^2}{t}\right) \\ =& \omega_{jk}\left\{2t(x_jx_k)^{1/2} - \left[\left(\frac{x_j}{x_k}\right)^{1/2} + \left(\frac{x_k}{x_j}\right)^{1/2}\right]\frac{\log t}{2}\right. \\ &\left. - \left(\frac{x_j}{x_k}\right)^{1/2}\frac{\log 4\pi x_k}{2} - \left(\frac{x_k}{x_j}\right)^{1/2}\frac{\log 4\pi x_j}{2}\right\} + O\left(\frac{(\log t)^2}{t}\right).\end{aligned}$$

Therefore

$$\begin{aligned}&-\frac{1}{2}\sum_{j,k}\omega_{jk}z_j(tx_j)z_k(tx_k) \\ &= -t(\boldsymbol{x}^{1/2})^\top\Sigma^{-1}\boldsymbol{x}^{1/2} + (\boldsymbol{x}^{1/2})^\top\Sigma^{-1}\boldsymbol{x}^{-1/2}\frac{\log t}{2} + k(\boldsymbol{x}) + O\left(\frac{(\log t)^2}{t}\right),\end{aligned}$$

where $k(\boldsymbol{x})$ does not depend on t. Putting everything together, with $g(\boldsymbol{x}) = (\boldsymbol{x}^{1/2})^\top\Sigma^{-1}\boldsymbol{x}^{1/2}$,

$$\begin{aligned}f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) &\propto r^{d/2-1+\frac{1}{2}(\boldsymbol{w}^{1/2})^\top\Sigma^{-1}\boldsymbol{w}^{-1/2}}[1+O(\log r/r)]\exp\left\{-rg(\boldsymbol{w}) + O\left(\frac{(\log r)^2}{r}\right)\right\} \\ &= r^{d/2-1+\frac{1}{2}(\boldsymbol{w}^{1/2})^\top\Sigma^{-1}\boldsymbol{w}^{-1/2}}\exp\left\{-rg(\boldsymbol{w})\right\}\left[1+O\left(\frac{(\log r)^2}{r}\right)\right].\end{aligned}$$

We therefore observe that the conditional distribution of $R \mid \boldsymbol{W} = \boldsymbol{w}$ has the gamma form, but with shape parameter $\alpha(\boldsymbol{w}) = d/2 + (\boldsymbol{w}^{1/2})^\top\Sigma^{-1}\boldsymbol{w}^{-1/2}/2$, rather than d as in all other examples calculated here.

Multivariate t_ν distribution (positive dependence) The multivariate t distribution with ν degrees of freedom exhibits both positive and negative dependence. After transformation to exponential marginals, the negative dependence is manifested in the limit set by inclusion of sections on the planes $\{x_j = 0\}$, $j = 1, \ldots, d$. We focus here on the shape of the limit sets for $\boldsymbol{x} > \mathbf{0}$ only, which captures the positive dependence in the tail.

The density with centred t_ν margins and dispersion matrix Σ is

$$f_{\boldsymbol{Z}_T}(\boldsymbol{z}) \propto \left[1 + \frac{\boldsymbol{z}^\top \Sigma^{-1} \boldsymbol{z}}{\nu}\right]^{-(\nu+d)/2} = \left[1 + \frac{\sum_{j=1}^d \sum_{k=1}^d \omega_{jk} z_j z_k}{\nu}\right]^{-(\nu+d)/2},$$

with $\Omega = (\omega_{jk})_{j,k}$ the inverse dispersion matrix. Transforming to exponential margins gives

$$f_{\boldsymbol{X}}(\boldsymbol{x}) \propto \left[1 + \frac{\sum_{j=1}^d \sum_{k=1}^d \omega_{jk} z_j(x_j) z_k(x_k)}{\nu}\right]^{-(\nu+d)/2} \prod_{j=1}^d \frac{1 - F_{Z_T}(z_j(x_j))}{f_{Z_T}(z_j(x_j))}, \quad (3.16)$$

where $z_j(x_j)$ is the solution to $x_j = -\log(1 - F_{Z_T}(z_j))$, and f_{Z_T}, F_{Z_T} represent the marginal density and distribution function of the t_ν distribution. Again, we are interested in large values of x_j and z_j: Soms (1976) gave an expansion for the ratio of the univariate survival function to density, from which we can deduce that

$$\frac{1 - F_{Z_T}(z_j(tx_j))}{f_{Z_T}(z_j(tx_j))} = \frac{z_j(tx_j)}{\nu} + O(z_j(tx_j)^{-1}).$$

Dropping the component index and writing $z_t = z(tx)$, we have

$$tx = -\log\left\{f_{Z_T}(z_t)\left[\frac{z_t}{\nu} + O(z_t^{-1})\right]\right\} = c + \nu \log z_t + O(z_t^{-2}),$$

where c is a constant depending on ν. For a new constant c', this is rearranged to give

$$z_t = c' e^{tx/\nu}[1 + O(e^{-2tx/\nu})].$$

To find the asymptotic behaviour of (3.16) we firstly consider the Jacobian term:

$$\prod_{j=1}^d \frac{1 - F_{Z_T}(z_j(tx_j))}{f_{Z_T}(z_j(tx_j))} = \prod_{j=1}^d \frac{z_j(tx_j)}{\nu}[1 + O(z_j(tx_j)^{-2})] \propto e^{t\sum_{j=1}^d x_j/\nu}[1 + O(e^{-2tx_{(d)}/\nu})].$$

Considering the kernel, we have

$$\begin{aligned}
&\left[1+\sum_{j=1}^{d}\sum_{k=1}^{d}\frac{\omega_{jk}}{\nu}z_j(tx_j)z_k(tx_j)\right]^{-\frac{\nu+d}{2}} \\
&= \left[1+\sum_{j=1}^{d}\sum_{k=1}^{d}\frac{\omega_{jk}c'^2}{\nu}e^{t(x_j+x_k)/\nu}+O(e^{t(x_{(1)}-x_{(d)})/\nu})\right]^{-\frac{\nu+d}{2}} \\
&\propto e^{-\frac{2tx_{(1)}}{\nu}\frac{\nu+d}{2}}\left[1+O(e^{t(x_{(2)}-x_{(1)})/\nu})\right] \\
&= e^{-tx_{(1)}(1+d/\nu)}\left[1+O(e^{t(x_{(2)}-x_{(1)})/\nu})\right].
\end{aligned}$$

Combining both expressions,

$$f_{\boldsymbol{X}}(t\boldsymbol{x}) \propto e^{-t[(1+d/\nu)x_{(1)}-\sum_{j=1}^{d}x_j/\nu]}\left[1+O(e^{t(x_{(2)}-x_{(1)})/\nu})+O(e^{-2tx_{(d)}/\nu})\right].$$

Therefore $g(\boldsymbol{x}) = (1+d/\nu)x_{(1)} - \sum_{j=1}^{d}x_j/\nu$ and

$$f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1}e^{-rg(\boldsymbol{w})}\left[1+O(e^{r(w_{(2)}-w_{(1)})/\nu})+O(e^{-2rw_{(d)}/\nu})\right].$$

Clayton and inverted Clayton copulas The Clayton copula with parameter $\gamma > 0$ has distribution function in uniform margins

$$F_{\boldsymbol{U}}(\boldsymbol{u}) = \left(\sum_{j=1}^{d}u_j^{-\gamma}-d+1\right)^{-1/\gamma}.$$

The corresponding density is

$$f_{\boldsymbol{U}}(\boldsymbol{u}) \propto \left(\prod_{j=1}^{d}u_j\right)^{-\gamma-1}\left(\sum_{j=1}^{d}u_j^{-\gamma}-d+1\right)^{-1/\gamma-d}.$$

The density in exponential margins is $f_{\boldsymbol{X}}(\boldsymbol{x}) = e^{-\sum_{j=1}^{d}x_j}f_{\boldsymbol{U}}(1-e^{-\boldsymbol{x}})$, and so

$$\begin{aligned}
f_{\boldsymbol{X}}(t\boldsymbol{x}) &\propto e^{-\sum_{j=1}^{d}tx_j}\left(\prod_{j=1}^{d}[1-e^{-tx_j}]\right)^{-(\gamma+1)}\left(\sum_{j=1}^{d}(1-e^{-x_j})^{-\gamma}-d+1\right)^{-1/\gamma-d} \\
&\propto e^{-\sum_{j=1}^{d}tx_j}\left[1+O(e^{-tx_{(d)}})\right].
\end{aligned}$$

The gauge function is therefore $g(\boldsymbol{x}) = \sum_{j=1}^{d}x_j$, and

$$f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1}e^{-rg(\boldsymbol{w})}[1+O(e^{-rw_{(d)}})].$$

If the random vector $\boldsymbol{U}$ follows a Clayton copula with uniform margins, then the random vector $1 - \boldsymbol{U}$ follows an inverted Clayton copula with uniform margins. Its density is $f_{\boldsymbol{U}}(1 - \boldsymbol{u})$. In exponential margins

$$\begin{aligned} f_{\boldsymbol{X}}(t\boldsymbol{x}) &\propto e^{t(\gamma+1)\sum_{j=1}^d x_j - t\sum_{j=1}^d x_j} \left(\sum_{j=1}^d e^{tx_j/\gamma} - d + 1 \right)^{-1/\gamma - d} \\ &= e^{t\gamma\sum_{j=1}^d x_j - t(1+d\gamma)x_{(1)}} \left(\sum_{j=1}^d e^{t(x_j - x_{(1)})/\gamma} + (1-d)e^{-tx_{(1)}/\gamma} \right)^{-1/\gamma - d} \\ &= e^{t\gamma\sum_{j=1}^d x_j - t(1+d\gamma)x_{(1)}} \left(1 + O(e^{t(x_{(2)} - x_{(1)})/\gamma})\right). \end{aligned}$$

The gauge function is therefore $g(\boldsymbol{x}) = (1 + d\gamma)x_{(1)} - \sum_{j=1}^d x_j\gamma$, and

$$f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1}e^{-rg(\boldsymbol{w})}[1 + O(e^{-r(w_{(2)} - w_{(1)})})].$$

Vine copula from Nolde and Wadsworth (2022) Nolde and Wadsworth (2022) give an example of a gauge function derived from a particular vine copula construction. Vine copulas are specified by pairs of bivariate copulas: in this case we take the two base pairs to be independence (between (X_1, X_2)) and inverted Clayton with parameter $\beta > 0$ (between (X_2, X_3)), and use the inverted Clayton with parameter $\gamma > 0$ to model the dependence between $(X_3 \mid X_2, X_1 \mid X_2)$.

Let $c_{1,2}, c_{2,3}$ and $c_{1|2,1|3}$ denote the densities of the respective copulas in standard uniform margins. The joint density in exponential margins is

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x}) =& e^{-(x_1+x_2+x_3)} c_{1,2}\left(1 - e^{-x_1}, 1 - e^{-x_2}\right) c_{2,3}\left(1 - e^{-x_2}, 1 - e^{-x_3}\right) \\ & \times c_{1|2,3|2}\left(F_{1|2}(x_1 \mid x_2), F_{3|2}(x_3 \mid x_2)\right), \end{aligned}$$

where $F_{1|2}(x_1 \mid x_2) = \Pr(X_1 \leq x_1 \mid X_2 = x_2) = 1 - e^{-x_1}$, and

$$F_{3|2}(x_3 \mid x_2) = \Pr\left(X_3 \leq x_3 \mid X_2 = x_2\right) = 1 - e^{(\beta+1)x_2}\left[e^{\beta x_2} + e^{\beta x_3} - 1\right]^{-\left(\frac{1}{\beta}+1\right)},$$

so that

$$F_{3|2}(tx_3 \mid tx_2) = 1 - e^{-t(\beta+1)[\max(x_2,x_3) - x_2]}[1 + O(e^{t\beta[\min(x_2,x_3) - \max(x_2,x_3)]})].$$

For the copula densities, $c_{1,2} = 1$, while

$$c_{2,3}(u_2, u_3) = (1+\beta)\,(1-u_2)^{-(\beta+1)}(1-u_3)^{-(\beta+1)}\left[(1-u_2)^{-\beta} + (1-u_3)^{-\beta} - 1\right]^{-\left(\frac{1}{\beta}+2\right)},$$

and $c_{1|2,3|2}$ is of the same form but with parameter γ. We have

$$\begin{aligned} c_{2,3}\left(1-e^{-tx_2}, 1-e^{-tx_3}\right) &\propto e^{(\beta+1)tx_2}e^{(\beta+1)tx_3}\left[e^{\beta tx_2} + e^{\beta tx_3} - 1\right]^{-\left(\frac{1}{\beta}+2\right)} \\ &= e^{-\beta t\max(x_2,x_3)+(\beta+1)t\min(x_2,x_3)}[1+O(e^{\beta t(\min(x_2,x_3)-\max(x_2,x_3))})], \end{aligned}$$

and

$$\begin{aligned} &c_{1|2,3|2}(1-e^{-tx_1}, 1-e^{-t(\beta+1)[\max(x_2,x_3)-x_2]}[1+O(e^{t\beta[\min(x_2,x_3)-\max(x_2,x_3)]})]) \\ &\quad\propto e^{(\gamma+1)tx_1+(\gamma+1)(\beta+1)t[\max(x_2,x_3)-x_2]}[1+O(e^{t\beta[\min(x_2,x_3)-\max(x_2,x_3)]})] \\ &\qquad\times\left\{e^{\gamma tx_1} + e^{\gamma(\beta+1)t[\max(x_2,x_3)-x_2]}[1+O(e^{t\beta[\min(x_2,x_3)-\max(x_2,x_3)]})] - 1\right\}^{-\left(\frac{1}{\gamma}+2\right)} \\ &\quad= e^{(\gamma+1)tx_1+(\gamma+1)(\beta+1)t[\max(x_2,x_3)-x_2]-(2\gamma+1)t\max(x_1,(\beta+1)[\max(x_2,x_3)-x_2])} \\ &\qquad\times[1+O(e^{t\beta[\min(x_2,x_3)-\max(x_2,x_3)]}) \\ &\qquad\quad+O(e^{t\gamma\{\min(x_1,(\beta+1)[\max(x_2,x_3)-x_2])-\max(x_1,(\beta+1)[\max(x_2,x_3)-x_2])\}})] \end{aligned}$$

Combining all components,

$$\begin{aligned} &f_{\boldsymbol{X}}(t\boldsymbol{x}) \\ &\propto \exp\{-t[x_1+x_2+x_3+\beta\max(x_2,x_3)-(\beta+1)\min(x_2,x_3) \\ &\qquad -(\gamma+1)\{x_1+(\beta+1)[\max(x_2,x_3)-x_2]\} \\ &\qquad +(2\gamma+1)\max(x_1,(\beta+1)[\max(x_2,x_3)-x_2])]\} \\ &\quad\times[1+O(e^{t\beta[\min(x_2,x_3)-\max(x_2,x_3)]}) + O(e^{t\gamma\{\min(x_1,(\beta+1)[\max(x_2,x_3)-x_2])-\max(x_1,(\beta+1)[\max(x_2,x_3)-x_2])\}})]. \end{aligned}$$

Simplifying the gauge function we get

$$\begin{aligned} g(\boldsymbol{x}) =&(1+\beta)\max(x_2,x_3) - \beta\min(x_2,x_3) - \gamma x_1 - (\gamma+1)(\beta+1)\,(\max(x_2,x_3)-x_2) \\ &+(2\gamma+1)\max(x_1,(\beta+1)(\max(x_2,x_3)-x_2)), \end{aligned}$$

and

$$f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1}e^{-rg(\boldsymbol{w})}$$

$$\times [1 + O(e^{r\beta[\min(w_2,w_3)-\max(w_2,w_3)]}) + O(e^{r\gamma\{\min(w_1,(\beta+1)[\max(w_2,w_3)-w_2])-\max(w_1,(\beta+1)[\max(w_2,w_3)-w_2])\}})].$$

3.8.3 Multivariate Gaussian case

In this section we consider the conditional distribution $R \mid \boldsymbol{W} = \boldsymbol{w}$ for the multivariate Gaussian dependence structure in more detail. In Section 3.3, and Appendix 3.8.2, the asymptotic form of this distribution is shown to have a gamma form with shape parameter given by a function of the angle $\boldsymbol{w}$,

$$\alpha(\boldsymbol{w}) = \frac{d}{2} + \frac{1}{2}\left(\boldsymbol{w}^{1/2}\right)^{\top} \Sigma^{-1} \boldsymbol{w}^{-1/2}.$$

There are two concerns with this shape parameter: (i) whether its variation with $\boldsymbol{w}$ indicates the need for a more complex model than that in equation (3.5), where the shape is assumed constant, and (ii) the fact that $\alpha(\boldsymbol{w}) \leq 0$ for some values of $\boldsymbol{w}$. We investigate these issues in turn, using a single correlation across all pairwise variables to define our covariance matrix,

$$\Sigma_{ij} = \begin{cases} 1 & ; \ i = j \\ \rho & ; \ i \neq j \end{cases}, \qquad i, j = 1, \ldots, d.$$

Figures 3.15 and 3.16 display local estimates of the shape parameter α under the truncated gamma model (3.5) for $R \mid [\boldsymbol{W} = \boldsymbol{w}, R > r_0(\boldsymbol{w})]$ for $d = 2, 3$, respectively. In each case the angles $\boldsymbol{w}$ are restricted to a small section of the simplex, and the rate parameter is fixed at the "true" value $g(\boldsymbol{w})$. For comparison, we also perform the same procedure for the logistic and inverted logistic dependence structures, over a variety of dependence strengths. For each distribution the total sample size, over all angular subsections, is 500,000. In Figure 3.15 we plot the median estimates and pointwise 95% CIs based on 100 repetitions. In Figure 3.16 we plot the median estimates based on 100 repetitions.

For the Gaussian case, we observe that estimates $\widehat{\alpha}$ remain relatively constant on the simplex $\mathcal{S}_{d-1}$ in practice for $d = 2, 3$. In particular, the shape parameter

estimates generally vary no more with the angle $\boldsymbol{w}$ when compared to the logistic and inverted logistic setting. The evidence suggests that for many distributions the shape parameter will vary with the angle $\boldsymbol{w}$ in practice to some degree. This is likely due to the fact that the rate of convergence of $R \mid \boldsymbol{W} = \boldsymbol{w}$ to the gamma form can depend on $\boldsymbol{w}$ either explicitly (as in the logistic case), or practically through a constant term (as in the inverted logistic case). When the dependence is strong for the inverted logistic distribution there are large parts of the simplex where there is insufficient data to estimate the local model.

The second issue with the shape parameter is that there will be values of $\boldsymbol{w}$ on the simplex $\mathcal{S}_{d-1}$ such that $\alpha(\boldsymbol{w}) \leq 0$. Figure 3.17 illustrates these regions, demonstrating that they appear to be small in volume and thus not important, especially as the dimension d increases. In both the $d = 2$ and $d = 3$ cases, we see that the region is almost negligible for $\rho < 0.7$. Furthermore, simulations from our models often do not produce points in these regions because when the dependence is high, there are very few points $\boldsymbol{W}$ near these boundaries that are accompanied by large values of R.

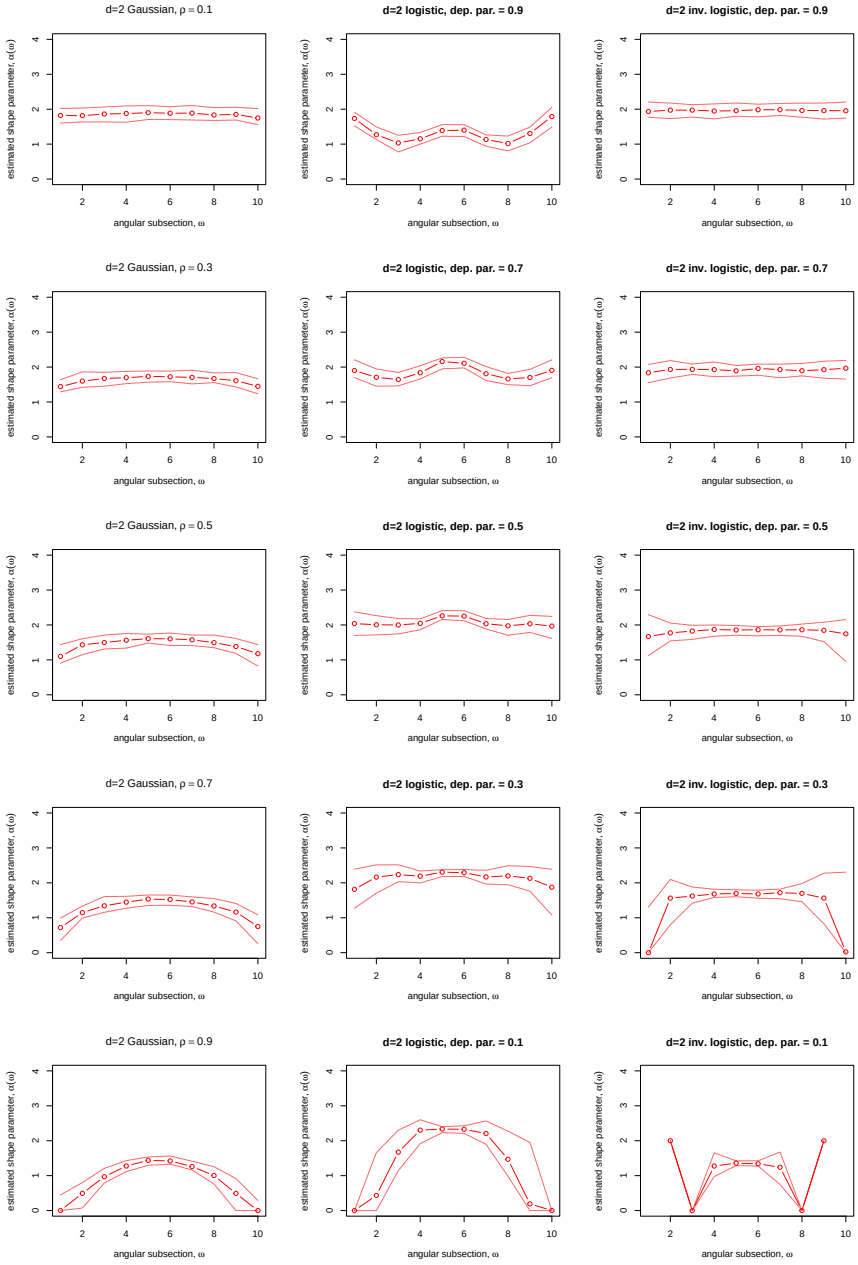

Figure 3.15: Shape parameter estimates across non-overlapping sections of the simplex $\mathcal{S}_1$ in the $d = 2$ case. Gaussian data and gauge function are used for column 1, logistic in column 2, inverted logistic in column 3. From top to bottom, dependence parameters are such that joint dependence is increasing. Dots represent median point estimates, and outer solid lines approximate 95% pointwise confidence intervals.

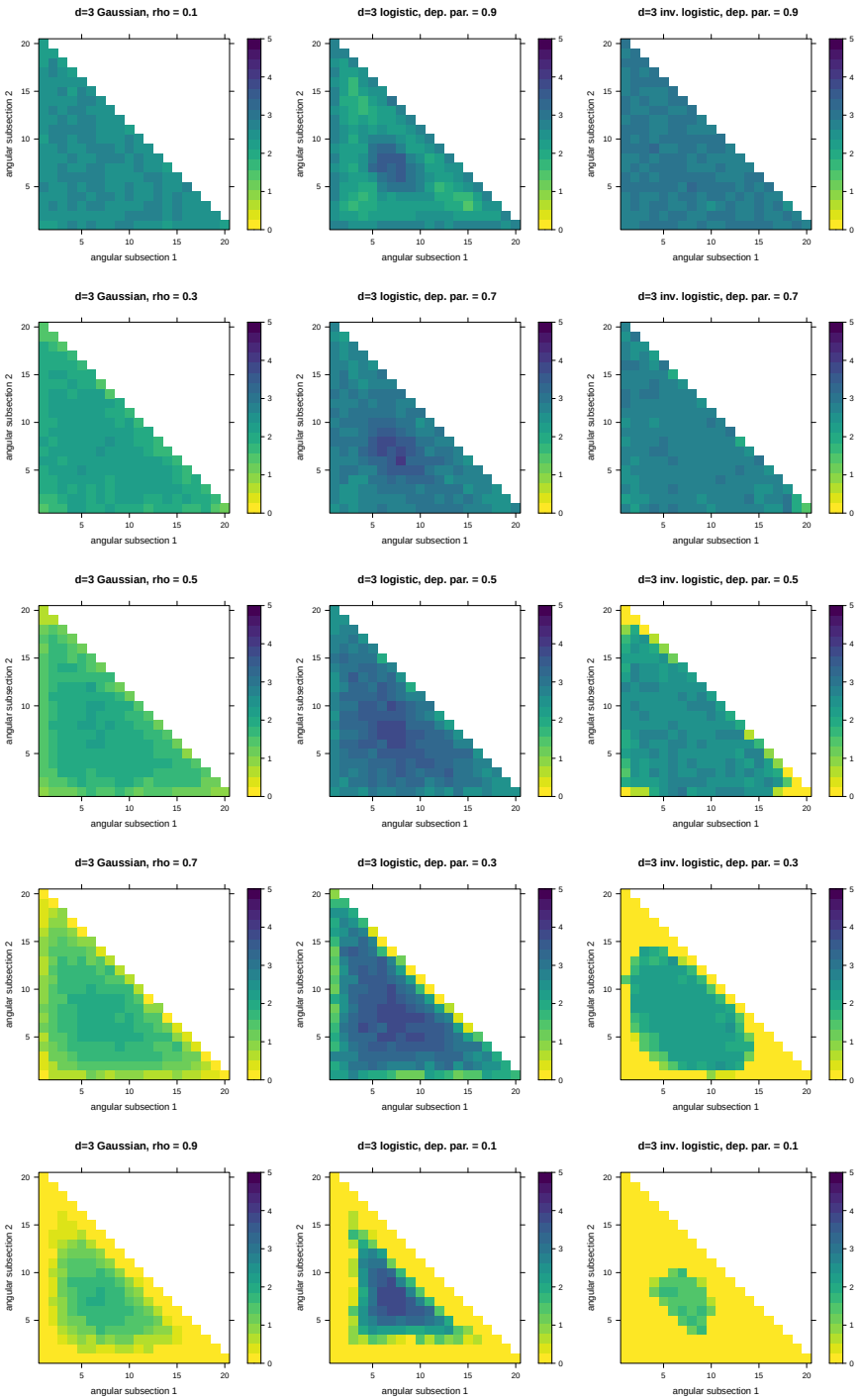

Figure 3.16: Median shape parameter estimates across non-overlapping sections of the simplex $\mathcal{S}_2$ in the $d = 3$ case. Gaussian data and gauge function are used for column 1, logistic in column 2, inverted logistic in column 3. From top to bottom, the dependence parameters are such that joint dependence is increasing. Pixels in yellow represent estimates $\widehat{\alpha} = 0$.

Finally, we offer evidence that extrapolation and probability estimation in these potentially problematic regions is not an issue when compared to other contemporary multivariate extremes methods. In both the $d = 2$ and $d = 3$ setting, we fix $\rho = 0.7$ and consider rectangular sets which overlap with the regions where $\alpha(\boldsymbol{w}) \leq 0$. For $d = 2$ we estimate the probability of lying in the set $(5, 7) \times (0, 0.75)$, and for $d = 3$ we estimate the probability of lying in the set $(5, 10) \times (0, 2) \times (0, 2)$ (see Figure 3.18). The resulting probability estimates are presented in the boxplots provided in Figure 3.19. We compare our geometric approach to the conditional extremes model of Heffernan and Tawn (2004). For $d = 3$ the results are comparable in terms of bias and variance, while for $d = 2$, our method is unbiased but with slightly higher variability than conditional extremes.

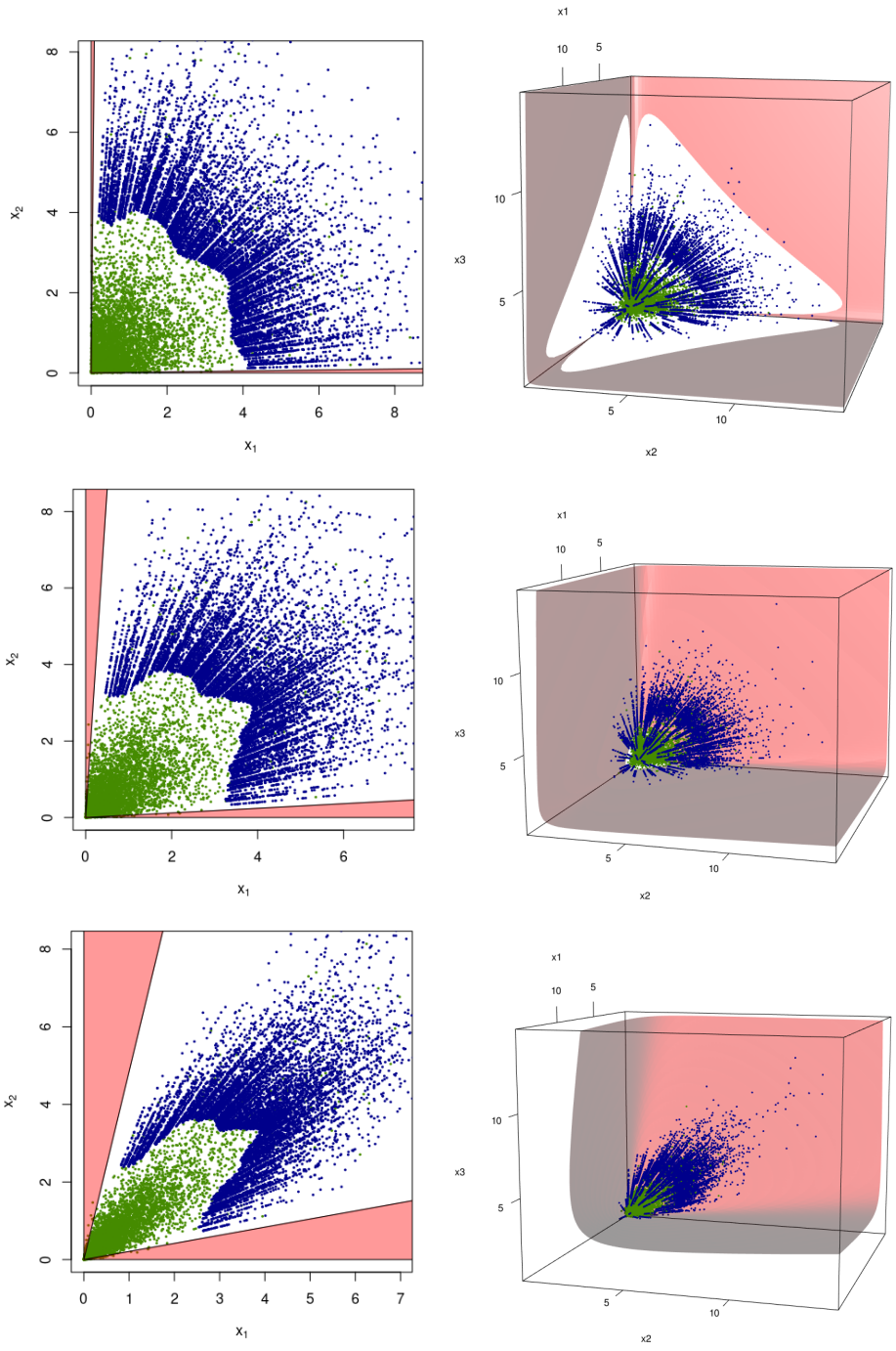

Figure 3.17: The region $\{(r, \boldsymbol{w}) \in (0, \infty) \times \mathcal{S}_{d-1} : \alpha(\boldsymbol{w}) < 0\}$ for correlation $\rho = 0.4$, 0.7, and 0.9 with $d = 2$ on the left (red region) and $d = 3$ on the right (volume between coordinate planes and red surface). Interactive versions of the plots for $d = 3$ can be found at `www.lancaster.ac.uk/~wadswojl/geometricMVE.html`.

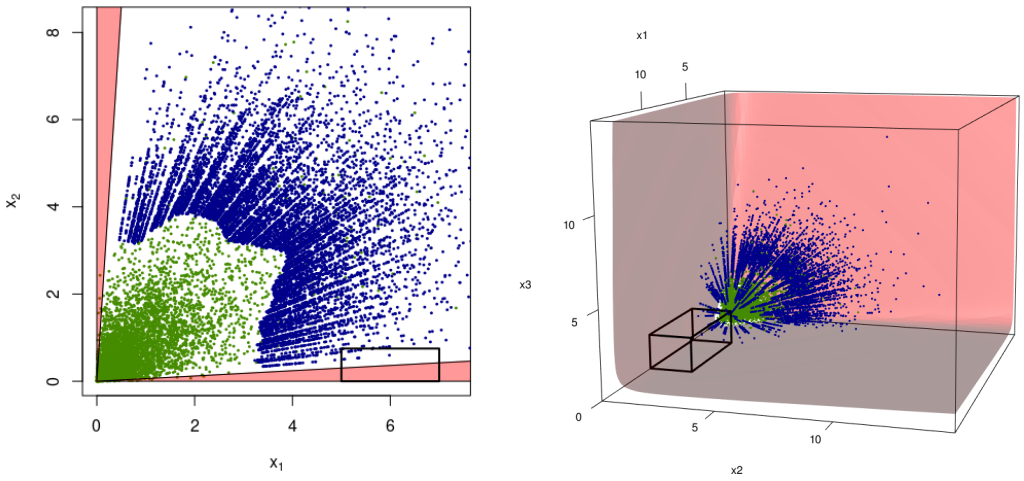

Figure 3.18: Rectangular regions with black borders are the extremal sets that we estimate the probability of lying in. Left: $d = 2$, right: $d = 3$.

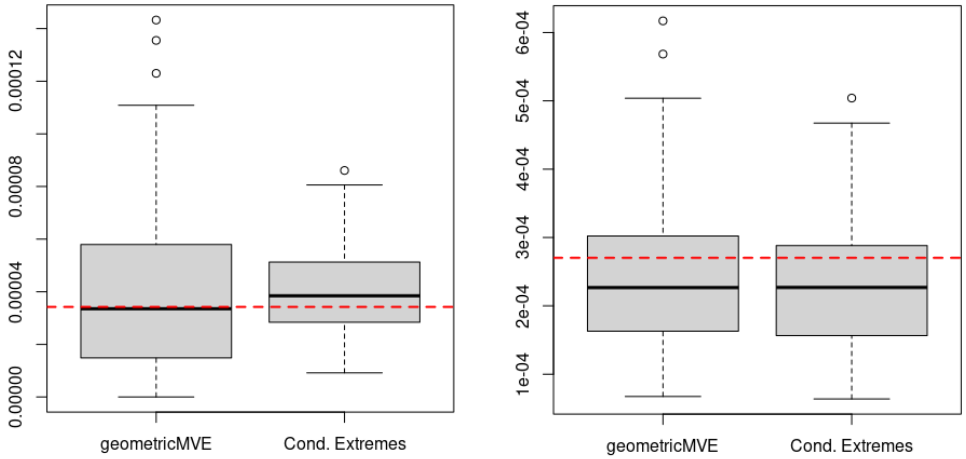

Figure 3.19: Probability estimates of lying in a predefined extremal region. Left: $d = 2$, right: $d = 3$. True probability values are displayed by the red dashed line.

3.8.4 Gauge function corresponding to the asymmetric logistic distribution

The asymmetric logistic distribution (Tawn, 1990) is a max-stable distribution with exponent function

$$V(\boldsymbol{z}) = \sum_{C \in 2^D \setminus \emptyset} \left[\sum_{j \in C} \left(\frac{\theta_{j,C}}{z_j} \right)^{1/\gamma_C} \right]^{\gamma_C}, \tag{3.17}$$

where $2^D \setminus \emptyset$ is the power set of $D = \{1, \ldots, d\}$. The parameters satisfy $\gamma_C \in (0, 1]$ and for each j there is a marginal condition that $\sum_{C \in 2^D \setminus \emptyset} \theta_{j,C} = 1$, with $\theta_{j,C} = 0$ if $j \notin C$. Notice that for singleton sets the values $\gamma_{\{1\}}, \ldots, \gamma_{\{d\}}$ are irrelevant; for convenience we assume that each of these is equal to 1.

The parameters $\theta_{j,C}$ play no role in determining the structure of the gauge function, except for where they lead to certain sets of variables being discounted as not taking extreme values simultaneously. We therefore consider a modification of V that allows us to derive the associated gauge function in a simpler manner. Define

$$V^*(\boldsymbol{z}) = \sum_{C \in 2^D \setminus \emptyset} \theta_C \left[\sum_{j \in C} \left(\frac{1}{z_j} \right)^{1/\gamma_C} \right]^{\gamma_C}, \tag{3.18}$$

with γ_C as before, and

$$\theta_C = \begin{cases} 1 & \text{Variables in } C \text{ can be simultaneously extreme} \\ 0 & \text{Variables in } C \text{ cannot be simultaneously extreme.} \end{cases}$$

In other words, $\theta_C = 1$ when the corresponding spectral measure H places mass on $\mathbb{B}_C$. For each variable $j = 1, \ldots, d$ there must be at least one $\theta_C = 1$ for $j \in C$. The distribution function $\exp\{-V^*(e^{\boldsymbol{x}})\}$ is a multivariate max-stable distribution with asymmetric logistic type dependence and Gumbel margins with non-zero location and unit scale. Distributions with unit-scale Gumbel margins have the same limit sets as distributions with exactly unit exponential margins (Nolde and Wadsworth, 2022). There is one further difference between the models defined by exponent

functions (3.17) and (3.18): when $\theta_C = 0$ in (3.18), this "switches off" the group of variables corresponding to C with no effect on other groups. With (3.17), if $\theta_{j,C} = 0$ for a single $j \in C$, then the set of variables that can be simultaneously extreme corresponds to $\theta_{C\setminus\{j\}}$, meaning that there could be two (or more) different γ parameters corresponding to the same set of variables. The function V^* in (3.18) is restricted to a single γ parameter per group of variables.

We define the collection of sets $\mathcal{C}_J$ to be all of those containing the index set $J \subset \{1, \ldots, d\}$. For example, with $d = 3$, $\mathcal{C}_{\{1\}} = \{\{1\}, \{1,2\}, \{1,3\}, \{1,2,3\}\}$ and $\mathcal{C}_{\{1,2\}} = \{\{1,2\}, \{1,2,3\}\}$. Here we outline the important steps in derivation of the gauge function for $d = 3$.

The partial derivatives of $V^*(\boldsymbol{z})$ are, for $k, l \in \{1,2,3\}$:

$$
\begin{aligned}
V^*_{\{k\}}(\boldsymbol{z}) &= \sum_{C_k \in \mathcal{C}_{\{k\}}} \theta_{C_k} \kappa_{C_k} z_k^{-1/\gamma_{C_k}-1} \left[\sum_{j \in C_k} z_j^{-1/\gamma_{C_k}}\right]^{\gamma_{C_k}-1} \\
V^*_{\{k,l\}}(\boldsymbol{z}) &= \sum_{C_{kl} \in \mathcal{C}_{\{k,l\}}} \theta_{C_{k,l}} \kappa_{C_{k,l}} z_k^{-1/\gamma_{C_{kl}}-1} z_l^{-1/\gamma_{C_{kl}}-1} \left[\sum_{j \in C_{kl}} z_j^{-1/\gamma_{C_{kl}}}\right]^{\gamma_{C_{kl}}-2} \\
V^*_{\{1,2,3\}}(\boldsymbol{z}) &= \sum_{C_{123} \in \mathcal{C}_{\{1,2,3\}}} \theta_{C_{123}} \kappa_{C_{123}} z_1^{-1/\gamma_{C_{123}}-1} z_2^{-1/\gamma_{C_{123}}-1} z_3^{-1/\gamma_{C_{123}}-1} \left[\sum_{j \in C_{123}} z_j^{-1/\gamma_{C_{123}}}\right]^{\gamma_{C_{123}}-3} \\
&= \theta_{\{1,2,3\}} \kappa_{\{1,2,3\}} z_1^{-1/\gamma_{\{1,2,3\}}-1} z_2^{-1/\gamma_{\{1,2,3\}}-1} z_3^{-1/\gamma_{\{1,2,3\}}-1} \left[\sum_{j=1}^{3} z_j^{-1/\gamma_{\{1,2,3\}}}\right]^{\gamma_{\{1,2,3\}}-3},
\end{aligned}
$$

with $\kappa_C \neq 0$ representing constant terms. Consequently,

$$
\begin{aligned}
&V^*_{\{1\}}(e^{t\boldsymbol{x}}) V^*_{\{2\}}(e^{t\boldsymbol{x}}) V^*_{\{3\}}(e^{t\boldsymbol{x}}) \\
&= \sum_{C_1 \in \mathcal{C}_{\{1\}}} \sum_{C_2 \in \mathcal{C}_{\{2\}}} \sum_{C_3 \in \mathcal{C}_{\{3\}}} \{\theta_{C_1} \theta_{C_2} \theta_{C_3} \kappa_{C_1} \kappa_{C_2} \kappa_{C_3} \\
&\quad e^{-tx_1\left(\frac{1}{\gamma_{C_1}}+1\right) - tx_2\left(\frac{1}{\gamma_{C_2}}+1\right) - tx_3\left(\frac{1}{\gamma_{C_3}}+1\right)} \\
&\quad e^{-t\min_{j \in C_1} x_j\left(1-\frac{1}{\gamma_{C_1}}\right) - t\min_{j \in C_2} x_j\left(1-\frac{1}{\gamma_{C_2}}\right) - t\min_{j \in C_3} x_j\left(1-\frac{1}{\gamma_{C_3}}\right)}\Bigg\} [1 + o(1)],
\end{aligned}
$$

$$
\begin{aligned}
&V^*_{\{j\}}(e^{t\boldsymbol{x}})V^*_{\{k,l\}}(e^{t\boldsymbol{x}}) \\
&= \sum_{C_j\in\mathcal{C}_{\{j\}}}\sum_{C_{kl}\in\mathcal{C}_{\{k,l\}}}\left\{\theta_{C_j}\theta_{C_{kl}}\kappa_{C_j}\kappa_{C_{kl}}e^{-tx_j\left(\frac{1}{\gamma_{C_j}}+1\right)-tx_k\left(\frac{1}{\gamma_{C_{kl}}}+1\right)-tx_l\left(\frac{1}{\gamma_{C_{kl}}}+1\right)}\right. \\
&\quad\left. e^{-t\min_{i\in C_j}x_i\left(1-\frac{1}{\gamma_{C_j}}\right)-t\min_{i\in C_{kl}}x_i\left(1-\frac{2}{\gamma_{C_{kl}}}\right)}\right\}[1+o(1)],
\end{aligned}
$$

$$
\begin{aligned}
&V^*_{\{1,2,3\}}(e^{t\boldsymbol{x}}) \\
&= \sum_{C_{123}\in\mathcal{C}_{\{1,2,3\}}}\theta_{C_{123}}\kappa_{C_{123}}e^{-tx_1\left(\frac{1}{\gamma_{C_{123}}}+1\right)-tx_2\left(\frac{1}{\gamma_{C_{123}}}+1\right)-tx_3\left(\frac{1}{\gamma_{C_{123}}}+1\right)+t\min_{j\in C_{123}}x_j\left(1-\frac{3}{\gamma_{C_{123}}}\right)} \\
&= \theta_{\{1,2,3\}}\kappa_{\{1,2,3\}}e^{-tx_1\left(\frac{1}{\gamma_{\{1,2,3\}}}+1\right)-tx_2\left(\frac{1}{\gamma_{\{1,2,3\}}}+1\right)-tx_3\left(\frac{1}{\gamma_{\{1,2,3\}}}+1\right)+t\min_{j\in\{1,2,3\}}x_j\left(1-\frac{3}{\gamma_{\{1,2,3\}}}\right)}.
\end{aligned}
$$

The density of the distribution in non-centred Gumbel margins is

$$
\exp\{-V^*(e^{t\boldsymbol{x}})\}e^{t\sum_{j=1}^d x_j}\sum_{\pi\in\Pi}\prod_{s\in\pi}-V_s(e^{t\boldsymbol{x}}),
$$

and the gauge function is derived through the minimum of all non-zero terms coming from the partial derivatives. To this end, we define $\mathcal{C}_J^+$ to be the collection of all sets that contain the index set J and for which $\theta_J = 1$. For example, when $d = 3$ and $\theta_{\{1\}} = \theta_{\{1,2,3\}} = 0$ and all other $\theta_J = 1$ then $\mathcal{C}^+_{\{1\}} = \{\{1,2\},\{1,3\}\}$. For $d = 3$ this yields the following expression for $g(\boldsymbol{x})$:

$$
\begin{aligned}
\min\Bigg\{&\min_{\substack{C_1\in\mathcal{C}^+_{\{1\}},C_2\in\mathcal{C}^+_{\{2\}},\\ C_3\in\mathcal{C}^+_{\{3\}}}}\left[\frac{x_1}{\gamma_{C_1}}+\frac{x_2}{\gamma_{C_2}}+\frac{x_3}{\gamma_{C_3}}+\min_{j\in C_1}x_j\left(1-\frac{1}{\gamma_{C_1}}\right)+\min_{j\in C_2}x_j\left(1-\frac{1}{\gamma_{C_2}}\right)\right. \\
&\qquad\left.+\min_{j\in C_3}x_j\left(1-\frac{1}{\gamma_{C_3}}\right)\right], \\
&\min_{C_1\in\mathcal{C}^+_{\{1\}},C_{23}\in\mathcal{C}^+_{\{2,3\}}}\left[\frac{x_1}{\gamma_{C_1}}+\frac{x_2}{\gamma_{C_{23}}}+\frac{x_3}{\gamma_{C_{23}}}+\min_{j\in C_1}x_j\left(1-\frac{1}{\gamma_{C_1}}\right)+\min_{j\in C_{23}}x_j\left(1-\frac{2}{\gamma_{C_{23}}}\right)\right], \\
&\min_{C_2\in\mathcal{C}^+_{\{2\}},C_{13}\in\mathcal{C}^+_{\{1,3\}}}\left[\frac{x_1}{\gamma_{C_{13}}}+\frac{x_2}{\gamma_{C_2}}+\frac{x_3}{\gamma_{C_{13}}}+\min_{j\in C_2}x_j\left(1-\frac{1}{\gamma_{C_2}}\right)+\min_{j\in C_{13}}x_j\left(1-\frac{2}{\gamma_{C_{13}}}\right)\right],
\end{aligned}
$$

$$\min_{C_3\in\mathcal{C}^+_{\{3\}},C_{12}\in\mathcal{C}^+_{\{1,2\}}}\left[\frac{x_1}{\gamma_{C_{12}}}+\frac{x_2}{\gamma_{C_{12}}}+\frac{x_3}{\gamma_{C_3}}+\min_{j\in C_3}x_j\left(1-\frac{1}{\gamma_{C_3}}\right)+\min_{j\in C_{12}}x_j\left(1-\frac{2}{\gamma_{C_{12}}}\right)\right],$$
$$\left.\min_{C_{123}\in\mathcal{C}^+_{\{1,2,3\}}}\left[\frac{x_1}{\gamma_{C_{123}}}+\frac{x_2}{\gamma_{C_{123}}}+\frac{x_3}{\gamma_{C_{123}}}+\min_{j\in C_{123}}x_j\left(1-\frac{3}{\gamma_{C_{123}}}\right)\right]\right\}.$$

From the derivation and form of this gauge function, we observe that the general form for any dimension d is

$$g(\boldsymbol{x})=\min_{\pi\in\Pi}\min_{C_s\in\mathcal{C}^+_s:s\in\pi}\left[\sum_{s\in\pi}\left(\sum_{j\in s}\frac{x_j}{\gamma_{C_s}}+\min_{j\in C_s}x_j\left(1-\frac{|s|}{\gamma_{C_s}}\right)\right)\right].$$

3.8.5 Example limit sets obtained by mixing gauge functions

3.8.5.1 Mixing via minimization

Figure 3.20 displays illustrations of limit sets that arise from mixing two gauge functions by minimization: $g(x_1, x_2) = \min\{g^{[1]}(x_1, x_2), g^{[2]}(x_1, x_2)\}$.

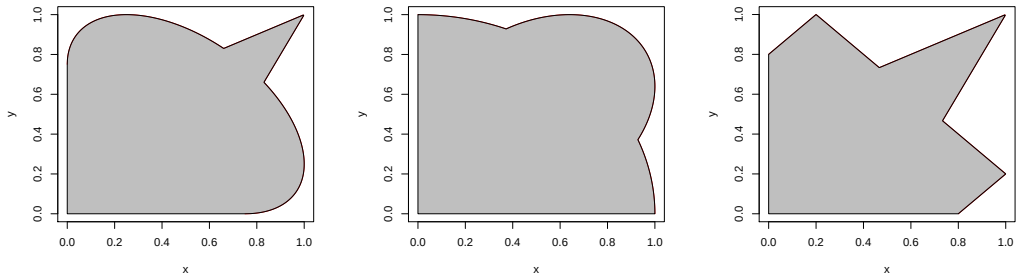

Figure 3.20: Illustration of limit sets (grey shaded region) arising from taking the minimum of two component gauge functions. Left: $g^{[1]}(x_1, x_2) = (x_1 + x_2)/\gamma + (1 - 2/\gamma)\min(x_1, x_2)$ with $\gamma = 0.5$, $g^{[2]}(x_1, x_2) = (x_1 + x_2 - 2\rho(x_1x_2)^{1/2})/(1 - \rho^2)$ with $\rho = 0.5$; centre: $g^{[1]}(x_1, x_2) = (x_1^{1/\gamma} + x_2^{1/\gamma})^\gamma$ with $\gamma = 0.5$, $g^{[2]}(x_1, x_2) = (x_1 + x_2 - 2\rho(x_1x_2)^{1/2})/(1 - \rho^2)$ with $\rho = 0.8$; right: $g[1](x_1, x_2) = (x_1 + x_2)/\gamma + (1 - 2/\gamma)\min(x_1, x_2)$ with $\gamma = 0.5$, $g^{[2]}(x_1, x_2) = \max((x_1 - x_2)/\theta, (x_2 - x_1)/\theta, (x_1 + x_2)/(2 - \theta))$ with $\theta = 0.8$.

3.8.5.2 Additive mixing

Some examples of shapes obtainable by additively mixing gauge functions with $d = 2$ are presented in Figure 3.5. Figure 3.21 displays some examples with $d = 3$.

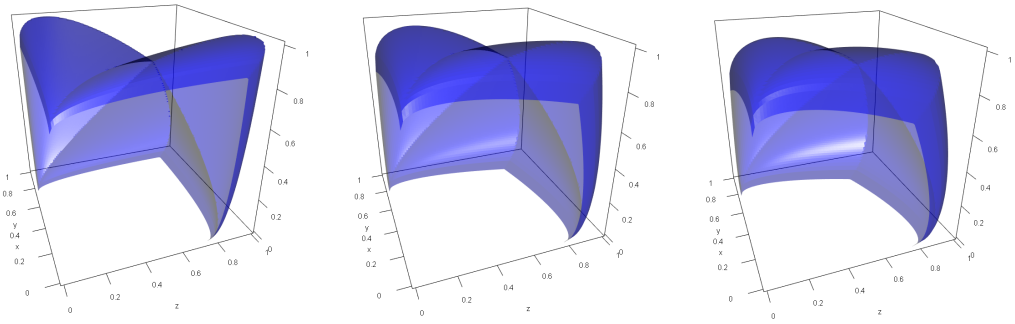

Figure 3.21: Illustration of limit sets (region between blue surface and planes $x_j = 0$) arising from additively mixing gauge functions corresponding to the Gaussian distribution with $\boldsymbol{\rho} = (0.5, 0.5, 0.5)$, and the asymmetric logistic distribution with spectral measure placing mass on $\mathbb{B}_{\{1,2\}}, \mathbb{B}_{\{1,3\}}, \mathbb{B}_{\{2,3\}}$, and parameters $\gamma_{\{1,2\}} = \gamma_{\{1,3\}} = \gamma_{\{2,3\}} = 0.5$. From left to right the weights are $a_1 = 1, 2, 3$, with $g^{[1]}$ the Gaussian gauge.

3.8.6 Choice of k

We describe an investigation into the choice of k, as discussed in Section 4.3.5, illustrating that overall that the methodology is relatively insensitive to the precise value selected, as long as it is in the feasible range. Since we require $k \geq 1$ such that $B \subset \{\boldsymbol{x} \in \mathbb{R}^d_+ : \sum_{j=1}^d x_j > k r_0(\boldsymbol{x}/\sum_{j=1}^d x_j)\}$, the feasible range for k is $[1, k^\star]$, where $k^\star$ is the maximum value such that the inequality $k < r_0(\boldsymbol{x}/\sum_{j=1}^d x_j)/(\sum_{j=1}^d x_j)$ holds for all $\boldsymbol{x} \in B$. In practice we may approximate $k^\star$ by k^*, where for a fixed number of points $\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_m$ on the boundary of B, and grid of values for k, we check the inequality

$$k < \frac{\sum_{j=1}^d \tilde{x}_{l,j}}{r_0\left(\tilde{\boldsymbol{x}}_l / \sum_{j=1}^d \tilde{x}_{l,j}\right)}$$

and take k^* to be the largest value k such that this holds for all $l = 1, \ldots, m$. We summarize the steps we used to find k^* below: in particular, for $d = 2$ and a rectangular region B, we use the vertices as $\tilde{\boldsymbol{x}}_l$, $l = 1, 2, 3, 4$.

1. For the region $B = [x_1, x_2] \times [y_1, y_2]$, consider the corner points $\tilde{\boldsymbol{x}}_1 = (x_1, y_1)$,

$\tilde{\boldsymbol{x}}_2 = (x_1, y_2)$, $\tilde{\boldsymbol{x}}_3 = (x_2, y_1)$, and $\tilde{\boldsymbol{x}}_4 = (x_2, y_2)$.

2. On a mesh $\boldsymbol{w}$ of the unit simplex $\mathcal{S}_1$, find the point $\widetilde{\boldsymbol{w}}_l$ that is closest to $\tilde{\boldsymbol{x}}_l / \sum_{j=1}^{d} \tilde{x}_{l,j}$ for $l = 1, 2, 3, 4$ for which we have a value of the threshold $r_0(\widetilde{\boldsymbol{w}}_l)$.

3. Beginning with $k = 1$, increase k by increments of 0.10, and find the largest value of k such that $k < \sum_{j=1}^{d} \tilde{x}_{l,j} / r_0(\widetilde{\boldsymbol{w}}_l)$ is satisfied for all $l \in \{1, 2, 3, 4\}$. Denote this largest k value by k^*.

Setting $k = k^*$ ensures that we sample close to the region B. Figure 3.22 below shows samples of $\boldsymbol{X} \mid R' > k^*$ for regions B_1, B_2, and B_3, as defined in Section 3.5.1 for a single simulated dataset and fitted threshold $r_0(\boldsymbol{w})$. We observe that the procedure used to obtain k^* ensures that extremal samples start very close to the boundary of the region of interest.

To assess the sensitivity of probability estimates to the choice of k, we consider the first three examples from the $d = 2$ simulation study in Section 3.5.1. For 200 repetitions, we fit our geometric models and estimate $\Pr(\boldsymbol{X} \in B_1)$, $\Pr(\boldsymbol{X} \in B_2)$ and $\Pr(\boldsymbol{X} \in B_3)$ by simulating from $\boldsymbol{X} \mid R' > k$ on a regularly spaced grid of 30 k values in the range $[1, k^*]$. Figure 3.23 below shows that probability estimates remain relatively unchanged as k varies from 1 to k^*. It is worth noting that confidence intervals are wider for k closer to 1, and we sometimes lose accuracy as k reaches k^*. We believe this is because of the approximation involved in finding k^* rather than $k^\star$, which means that the inequality $k^* < r_0(\boldsymbol{x} / \sum_{j=1}^{d} x_j) / (\sum_{j=1}^{d} x_j)$ may fail for a small set of points $\boldsymbol{x} \in B$. As such, we recommend taking an intermediate k that is slightly smaller than k^*.

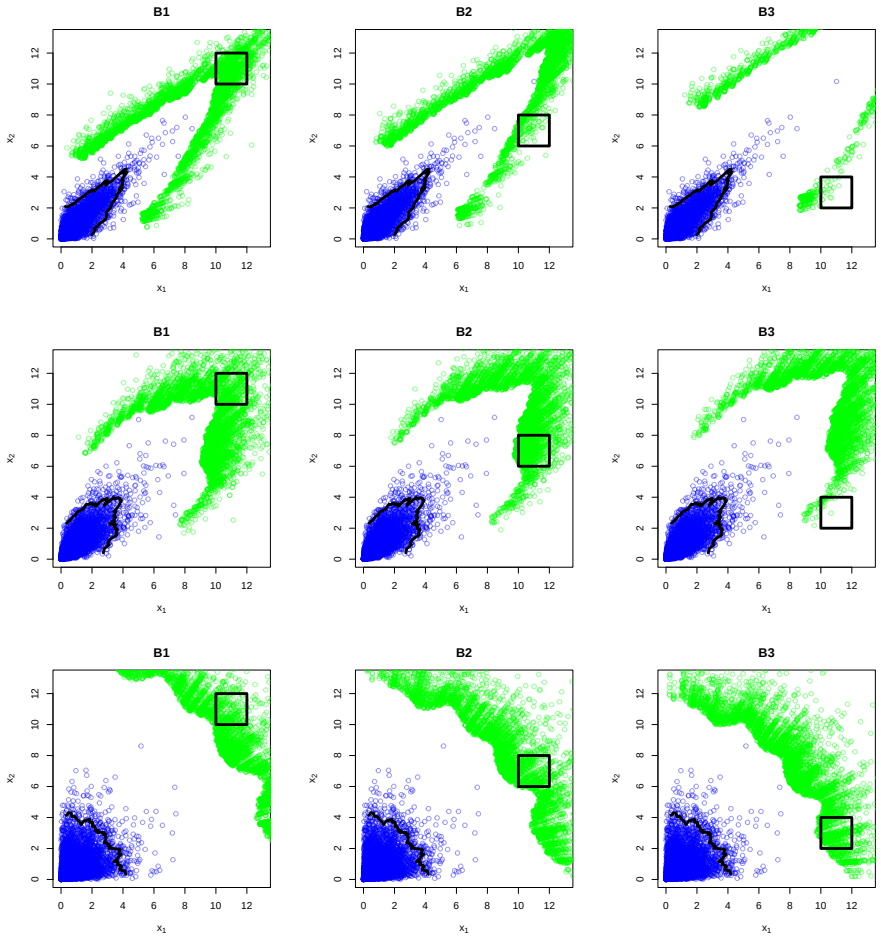

Figure 3.22: Samples of $\boldsymbol{X} \mid R' > k^*$ for k^* obtained by regions B_1, B_2, and B_3 (left to right) generated from models fitted on bivariate logistic (top row), Gaussian (middle row), and inverted logistic (bottom row) data. Blue points represent the original sample, the black line represents the threshold $r_0(w)$ found using rolling-windows method, and the green points are samples of $\boldsymbol{X} \mid R' > k^*$.

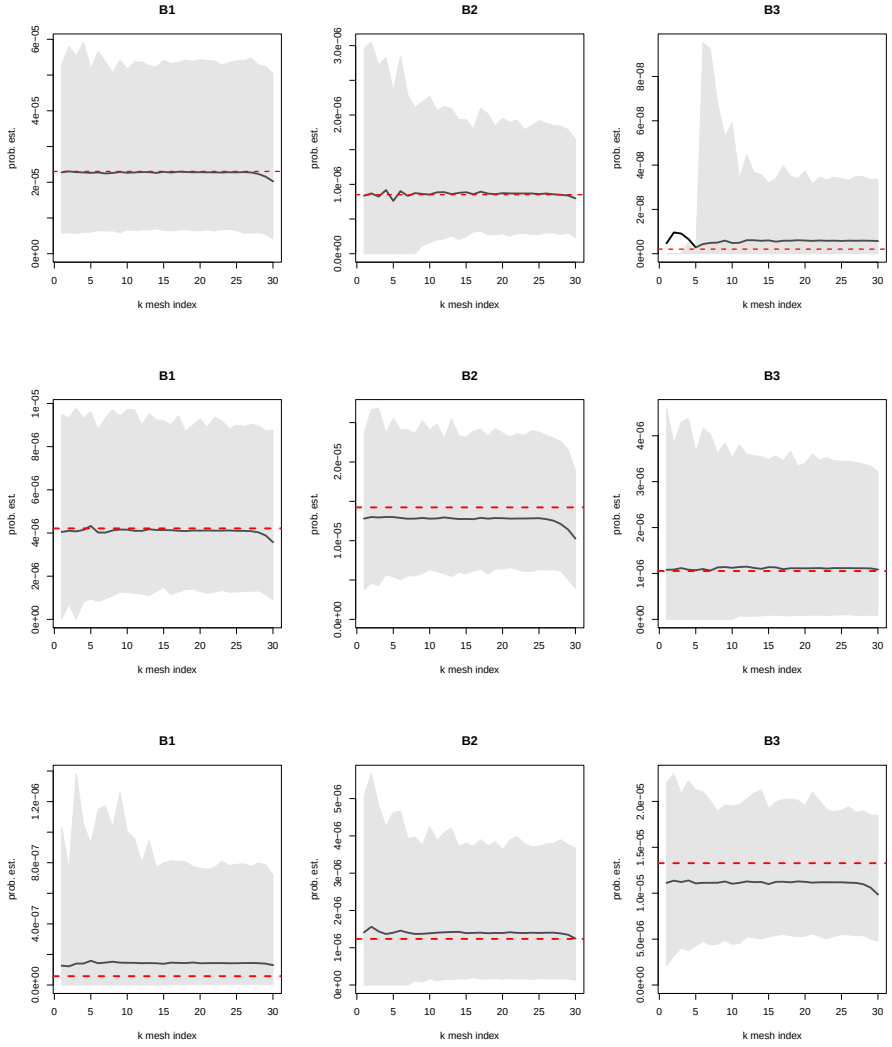

Figure 3.23: Probability estimates for increasing $k \in [1, k^*]$ for models fitted on bivariate logistic (top row), Gaussian (middle row), and inverted logistic (bottom row) data. Solid black line represents the mean over the 200 repetitions, grey shaded regions represent 95% pointwise sample confidence intervals. True probabilities are given by the red dashed line. From left to right, estimates are of $\Pr(\boldsymbol{X} \in B_1), \Pr(\boldsymbol{X} \in B_2)$ and $\Pr(\boldsymbol{X} \in B_3)$, respectively. For the first few k values, most estimates of $\Pr(\boldsymbol{X} \in B_3)$ are zero for the logistic distribution.

3.8.7 Additional simulation study figures

3.8.7.1 $d = 2$

Figure 3.24 shows examples of the four datasets for the $d = 2$ simulation study, and three sets of interest $B_1 = (10, 12) \times (10, 12)$, $B_2 = (10, 12) \times (6, 8)$, and $B_3 = (10, 12) \times (2, 4)$.

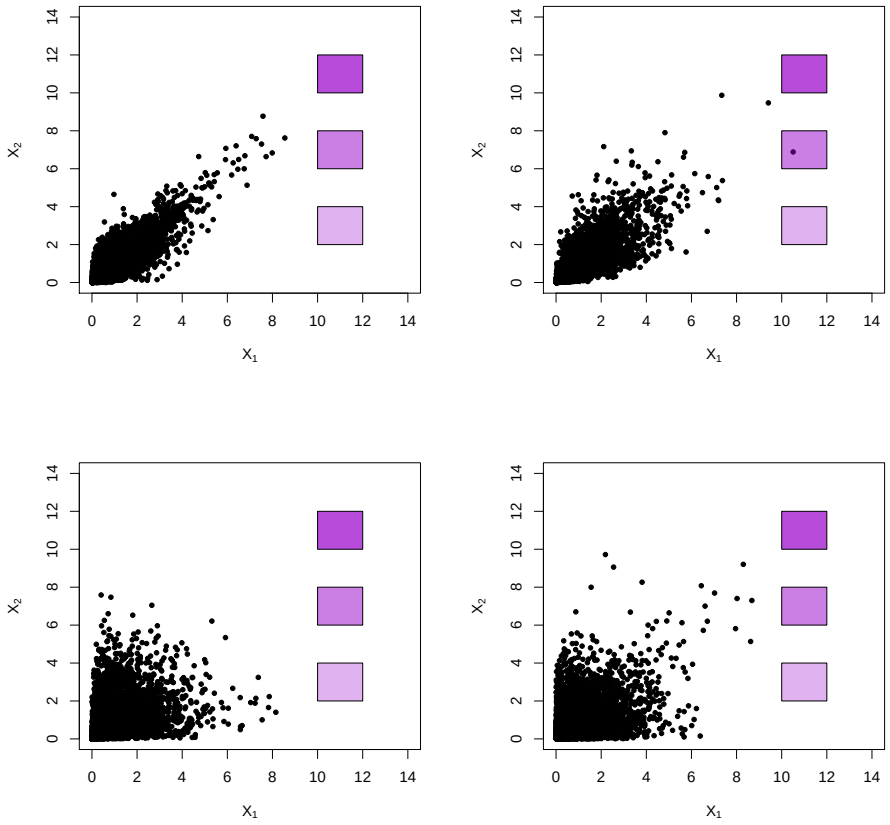

Figure 3.24: $d = 2$: Example data from distributions (I) (top left), (II) (top right), (III) (bottom left) and (IV) (bottom right), and illustration of sets B_1, B_2, B_3 (purple shading).

Figure 3.25 displays estimates of the limit set shape via non-parametric estimation of g using rolling-windows quantiles, as described in Section 3.4.1, and parametric estimation from the maximum likelihood estimates of the gauge function

parameters. We display parametric estimates using both all fits from the correct gauge function, and only the fits where the correct gauge function returned the minimum AIC. For distributions (I), (II), (III) and (IV), this is 82.5%, 41.5%, 39.5% and 82%, respectively. The non-parametric estimates do not quite join to the axes because we ascribe the rolling-windows estimate of $r_0(\boldsymbol{w})$ to the centre of each window for $w \in [0, 1]$. For distribution (III) we note that the non-parametric estimates display lower variability than the parametric ones. In spite of this, the performance of the method for probability estimation, which relies both on non-parametric estimation of $r_0(\boldsymbol{w})$ and parametric estimation of g, appears quite reasonable.

3.8.7.2 $d = 3$

Figure 3.26 depicts examples of the three datasets for the $d = 3$ simulation study, along with sets of interest $B_1 = (8, 10) \times (8, 10) \times (0.01, 3)$, $B_2 = (8, 10) \times (5, 7) \times (0.01, 3)$ and $(8, 10) \times (2, 4) \times (0.01, 3)$. Figure 3.27 displays boxplots relating only to the geometric approach for $d = 3$ from Figure 3.9, with a clearer vertical scale. Although there is a downwards bias in the estimation, the geometric approach still provides reasonable estimates in these difficult cases.

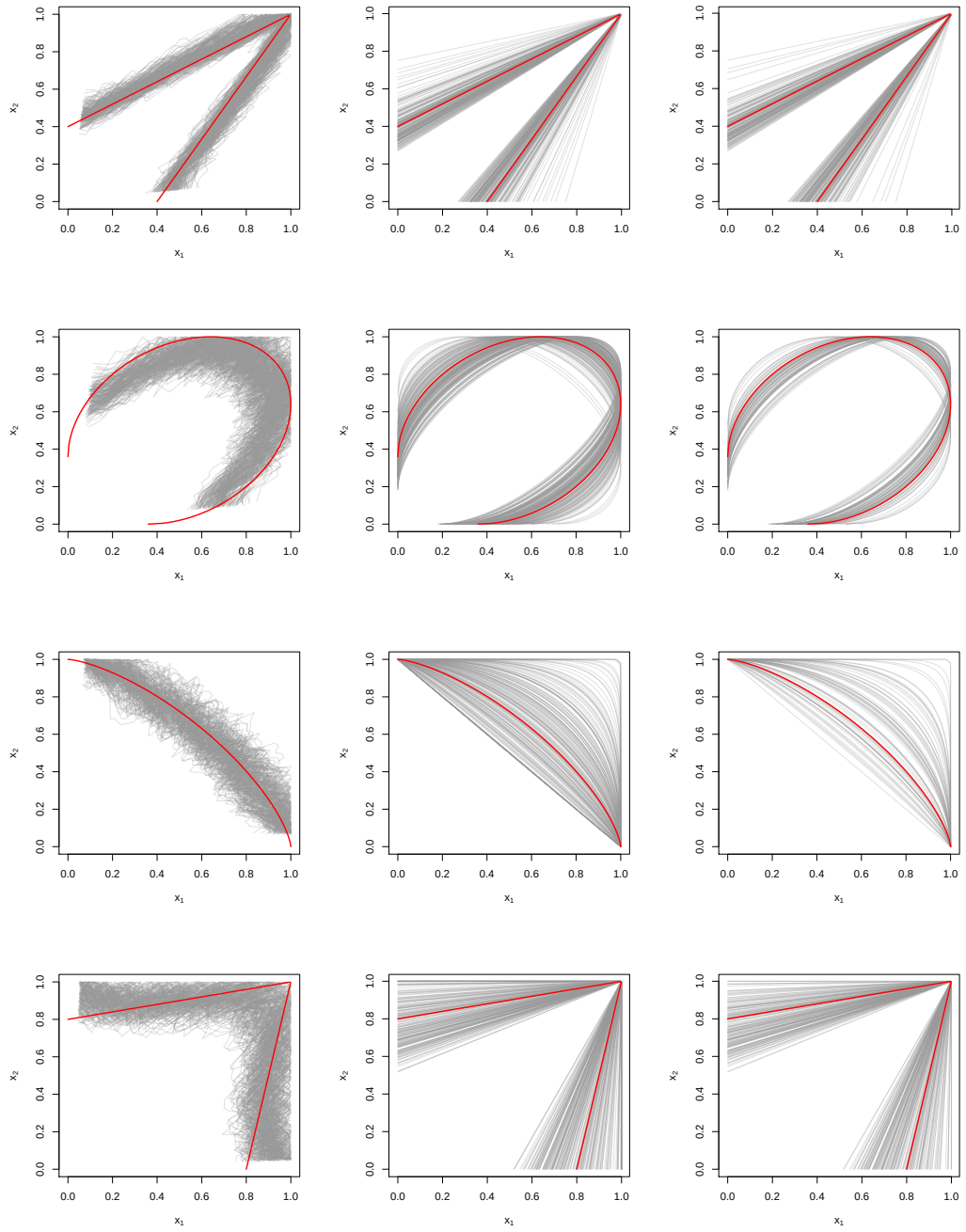

Figure 3.25: Non-parametric (left column) and parametric estimates (centre and right columns) of unit level sets of g. The unit level sets of the true g are shown in red. Top-bottom: distributions (I), (II), (III) and (IV), respectively. The centre column includes all parametric estimates from the correct gauge function; the right column includes only those parametric estimates where the true model had the lowest AIC.

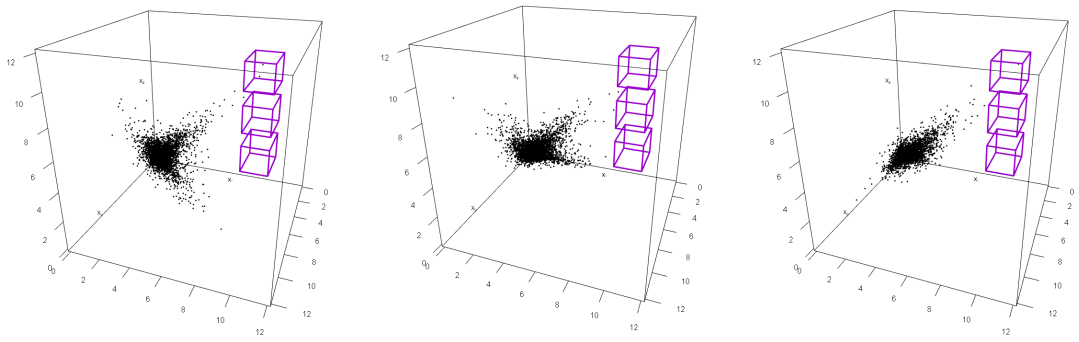

Figure 3.26: $d = 3$: Example data from distributions (I), (II) and (III) (left to right), and illustration of sets B_1, B_2, B_3 (purple boxes).

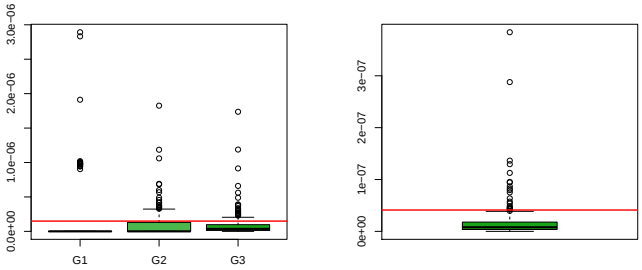

Figure 3.27: Boxplots of probability estimates for $\Pr(\boldsymbol{X} \in B_3)$ for $d = 3$ under the geometric approach. Left: for distribution (I); right: for distribution (III). For clarity only a single boxplot, corresponding to that labelled G3 in Section 3.5.2, is included in the right panel.

3.8.8 Additional figures for river flow analysis

Figures 3.28 and 3.29 display plots of $\chi_C(u)$ for the remaining groups of variables. The bottom row of Figure 3.29 displays the PP and exponential QQ plots for the fit of model (3.5).

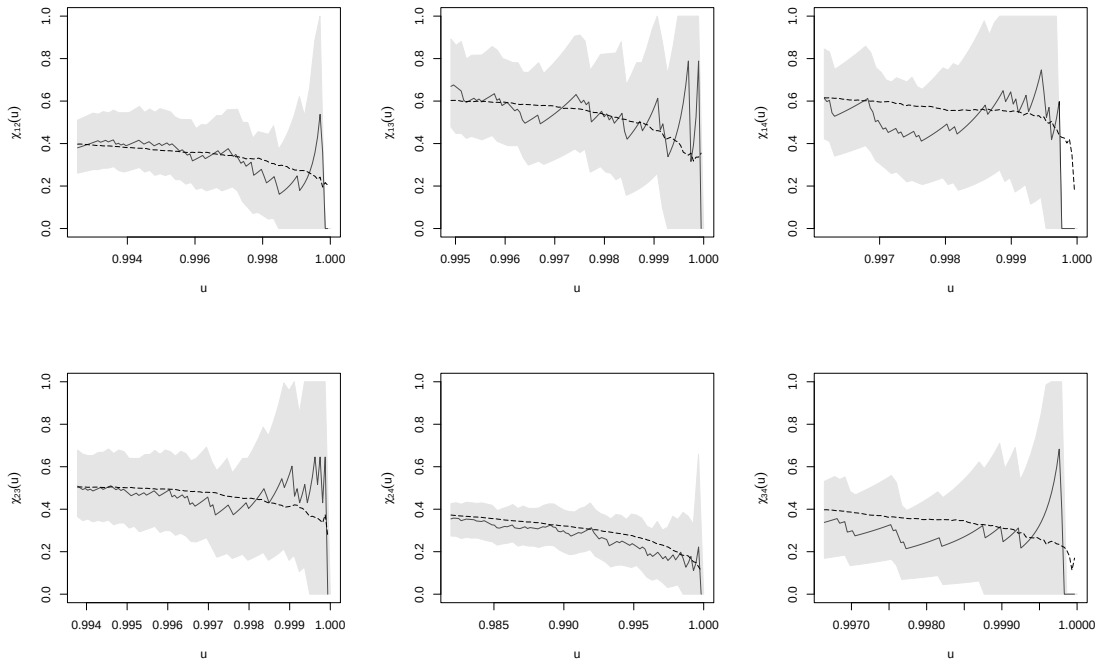

Figure 3.28: Empirical (solid line) estimates of $\chi_C(u)$ with approximate pointwise 95% confidence intervals (grey shaded region), and model-based estimate (dashed line) for groups $C = \{1,2\}$, $\{1,3\}$, $\{1,4\}$, $\{2,3\}$, $\{2,4\}$, and $\{3,4\}$ (top left–bottom right).

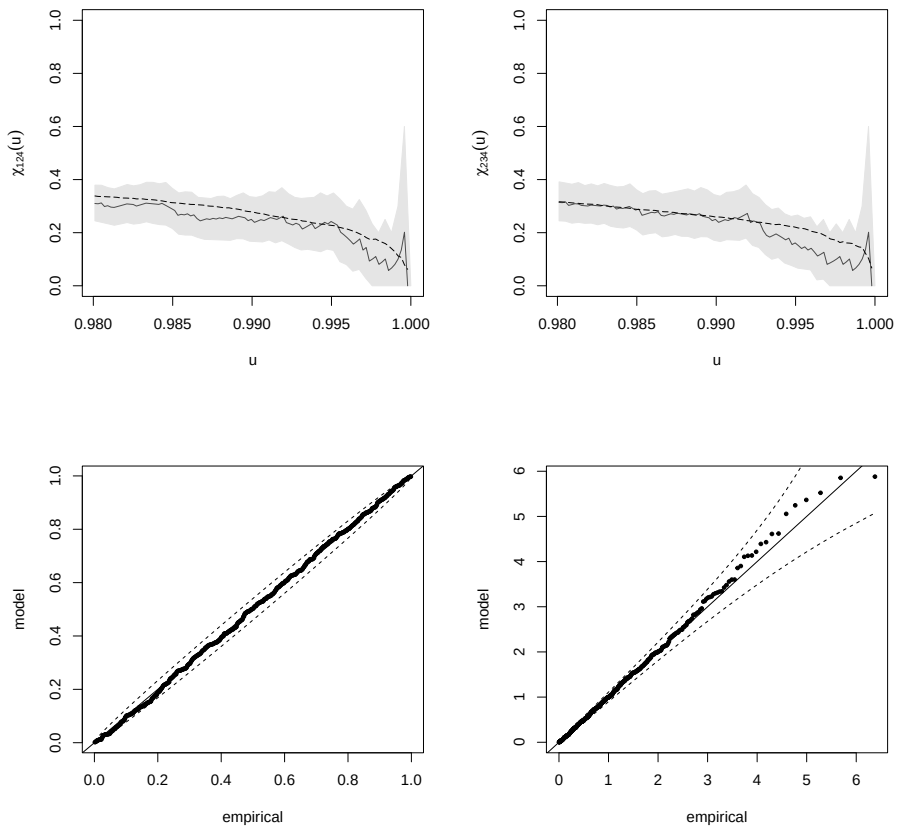

Figure 3.29: Top row: Empirical (solid line) estimates of $\chi_C(u)$ with approximate pointwise 95% confidence interval (grey shaded region), and model-based estimate (dashed line) for groups $C = \{1, 2, 4\}$ and $\{2, 3, 4\}$. Bottom row: PP and exponential QQ plots for the fitted geometric model with Gaussian gauge function.

Chapter 4

On the stability and Bayesian semi-parametric modelling of multivariate geometric exceedances

Abstract

Under a pseudo-radial directional decomposition of d-dimensional random vectors in light-tailed margins, we show that the point process of recentered and renormalised radial components and directions have a Poisson point process representation, where the radial component follows an exponential distribution. The memoryless property of the exponential distribution allows for the estimation of *return sets*, defined by the complement of probability sets and expressed in terms of return periods. Building on the limit Poisson point process likelihood, we develop parsimonious statistical models that leverage theoretical links between the parameters of the limit distribution. These models facilitate Bayesian inference for return sets with arbitrarily large return period and for probabilities of unobserved rare events, and include directional information from observations outside probability sets. The framework supports

efficient computations in dimensions two and three, along with diagnostics for assessing convergence to the limiting distribution and quality of fit. We validate the methods through illustrative simulations and demonstrate their utility via case studies involving hydrological and oceanographic data, showcasing the potential for robust and interpretable analysis of multivariate extremes.

Disclaimer

This chapter represents work completed with Ioannis Papastathopoulos and Lambert De Monte at the University of Edinburgh, primarily from January to September 2023, with periodic updates in the following months. My main contributions were in developing the directional model in Section 4.2.2, developing the standardisation of margins in Section 4.3.1, high quantile estimation in Section 4.3.2, model selection and validation techniques in Section 4.3.7, the return sets and their illustrative examples in Section 4.3.6, the real data examples in Sections 4.4 and 4.5 with their accompanying Appendices 4.7.6 and 4.7.7, the derivations of the radial and angular densities in Appendices 4.7.4 and 4.7.5, respectively, and writing the introduction and concluding discussion. I also contributed to the remaining material, but in more of an advisory role. Additional material that I did not work on was left out, but the reader is referred to Papastathopoulos et al. (2025) when necessary. The notation used in this chapter may vary somewhat from the rest of the thesis. For example, random vectors rescaled by their norm are called *directions* in this chapter but are called *angles* in the rest of the thesis.

4.1 Introduction

The multivariate nature of extreme events casts a shadow of potentially devastating consequences upon ecosystems, infrastructures, as well as financial, economic, and insurance sectors. Knowledge of the frequency and magnitude of extreme events

is crucial in enhancing planning strategies and adaptation efforts. The statistical properties of univariate extremes are well-established (Balkema and de Haan, 1974; Pickands, 1975; Davison and Smith, 1990), but statistical inference for multivariate random processes is much more intricate: one must analyse how random processes interact with and influence each other. A common way to describe the extremal dependence structure of a real-valued random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ is through the coefficient of tail dependence

$$\chi_q(A) = \mathbb{P}\Big[\bigcap_{j \in A} \{F_j(X_j) > q\} \Big] / (1-q), \quad q \in (0,1), \quad A \subseteq \{1, \ldots, d\}, |A| > 1, \tag{4.1}$$

where F_j is the cumulative distribution function of X_j. When $\lim_{q \to 1} \chi_q(A) = 0$, the variables in A are unlikely to grow large together. Conversely, when $\lim_{q \to 1} \chi_q(A) > 0$, the variables in A are likely to simultaneously exhibit extreme values. The different dependence structures that can be present within subgroups of the marginal variables of $\boldsymbol{X}$ can make inference for rare events challenging and lead to inaccuracies in extrapolation.

Classical approaches to multivariate extreme events often rely on the framework of multivariate regular variation (MRV) (de Haan and Resnick, 1977), which posits that the point processes of exceedances of a random vector over a high threshold, when suitably renormalised, converge in distribution to a non-degenerate non-homogenous Poisson point process (de Haan, 1984). This provides a framework for understanding the joint behaviour of extreme observations and leads to meaningful limit distributions that can be used for statistical inference of multivariate extremes. In practice, MRV is applied in a way that does not adequately describe relationships among asymptotically independent random variables (Nolde and Wadsworth, 2022). This is due to the type of the renormalisation that is employed: given a random sample of n d-dimensional observations, all components are normalised by the same amount (see Figure 4.1). This leads to only considering the dependence structure when all d variables are large.

Another drawback of MRV is the limited set of directions in the multivariate

space in which one can extrapolate the model. Under MRV, a common way to estimate the probability of lying in an extremal set is to shift the set linearly and perform empirical probability estimation on the translated set. When the translated set does not contain observations from the initial dataset, the estimate the probability of interest is inevitably 0. To correct the joint rate of tail decay in the case of weaker extremal dependence, the notion of hidden regular variation (HRV) was introduced (Ledford and Tawn, 1996, 1997). However, HRV also suffers from the drawback that it does not allow extrapolation along directions where not all variables are simultaneously large. To extrapolate to a wider range of extremal regions with a wider array of dependence structures, the frameworks of conditional extremes (Heffernan and Tawn, 2004) and angular dependence (Wadsworth and Tawn, 2013) have been introduced, but statistical methodology based on these frameworks suffers from drawbacks. Despite its wide applicability and widespread adoption, the conditional extremal inference method of Heffernan and Tawn (2004) is based on composite likelihood methods and on gluing separate models post-fit, making statistical inference and computations unwieldy. While the angular dependence method of Wadsworth and Tawn (2013) permits extrapolation in regions where variables are not simultaneously extreme, it is only useful for joint survival regions. Recently, a characterisation of extremal dependence through the limiting geometry of observations from $\boldsymbol{X}$ has become of interest. The *limit set*, denoted by $\mathcal{G}$, whose boundary arises as the limiting hull of appropriately scaled sample clouds $N_n = \{\boldsymbol{X}_1/r_n, \ldots, \boldsymbol{X}_n/r_n\}$ as $n \to \infty$, provides insight into the extremal dependence structure of $\boldsymbol{X}$. The *gauge function*, denoted by $g_{\mathcal{G}}$, whose unit level set is in one-to-one correspondence with the boundary of the limit set, has been shown to connect several known coefficients describing extremal dependence of known copulas (Nolde, 2014; Nolde and Wadsworth, 2022).

Wadsworth and Campbell (2024) proposed a framework for performing statistical inference for multivariate extremes using this geometric approach. Using a radial-directional decomposition, this framework treats the gauge function evaluated at

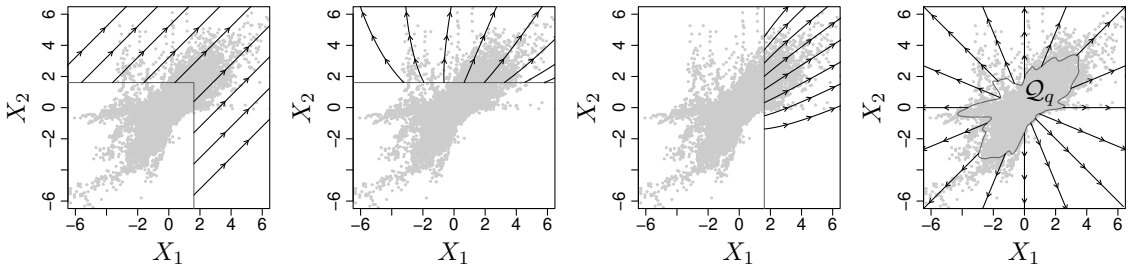

Figure 4.1: Directions along which different frameworks allow for extrapolation to tail regions of a bivariate river flow dataset in Laplace margins: MRV (left), conditional extremes given X_2 and X_1 are large (centre-left and centre-right), and geometric extremes (right), with $\mathcal{Q}_q$ illustrating the posterior mean of the quantile set at $q = 0.95$ (details presented in Section 4.3). The support of the distribution of exceedances is inscribed by the region containing the arrows.

directions as a rate parameter of a left-truncated gamma model for the distribution of radii conditioned on directions on the unit simplex. Inference for the gauge function and its associated limit set is based on parametric models derived from known copulas in d-dimensional random vectors with standard exponential margins, and a maximum likelihood approach is implemented within the rate parameter of the truncated gamma distribution. The result is a new statistical inference method for estimating extremal probabilities with great flexibility relative to state-of-the-art methods in multivariate extremes. Also in exponential margins, but in a bivariate setting, Simpson and Tawn (2024a) model the conditional distribution of the excess radii given directions on the simplex via the generalised Pareto (GP) distribution (Pickands, 1975). In this framework, the gauge function is seen as a rate parameter of the GP distribution and is modelled via generalised additive models (Wood, 2017). Majumder et al. (2025) also propose a statistical method to estimate the bivariate gauge function and the shape of its associated limit set using Bézier splines.

In this work, we develop a framework that extends beyond the cone $\mathbb{R}^d_+$ by

leveraging the structure of weak limits of suitably radially renormalized *multivariate exceedances*. Multivariate exceedances are observations that lie within a *return set*, defined as the complement of a *probability set*. Given a d-dimensional distribution, a probability set is a set in $\mathbb{R}^d$ where the the probability of lying within is prespecified. Among the infinite class of probability, we identify *quantile sets*, defined through the quantile of the radial distribution at a given direction. The quantile set derives its name from its role as a d-dimensional generalisation of quantiles, addressing challenges in multivariate extreme value theory that arise from the absence of natural ordering in $\mathbb{R}^d$ (Barnett, 1976). This concept parallels the quantile regions introduced by Hallin et al. (2021), but our definition is specifically designed for extrapolation beyond the observed data.

By specifying an appropriate sequence of quantile sets, we show under mild conditions that the framework of MRV can be extended to the cone $\mathbb{R}^d \setminus \{\mathbf{0}\}$ for random vectors with light-tailed margins. In this multivariate setting, our framework enables the modelling of the entire joint tail, accommodating scenarios where subsets of components are extreme. This approach can capture behaviours across the entire spectrum of multivariate space and reveal hidden dependencies, thereby bridging the gap between theory and practice. Using radially recentred and rescaled exceedances over high quantile sets, we characterise non-trivial limit distributions on $\mathbb{R}^d$, termed *radially-stable exponential distributions*. The radial stability properties of these families of distributions permit extrapolation beyond the range of observed data along any direction (see Figure 4.1) and naturally lead to the notion of *return sets*, a geometric d-dimensional extension of the univariate (upper-tail) return level. This yields an interpretable way to communicate the risk associated with extreme multivariate events allowing decision-makers, policymakers, and the general public to understand the likelihood of experiencing impact from joint extreme events. Throughout our work, inference is done in a Bayesian manner, allowing us us to obtain inferences for any functional of the joint tail distribution. Our methods account for multiple sources of uncertainty in the estimation, including

that of uncertainty in the estimation of the marginal tails as well as the multivariate threshold, which is typically not accounted for in previous statistical methods for multivariate extremes. We use use prediction intervals for functionals of interest, allowing us to quantify uncertainty via simultaneous prediction bands when the functional of interest consists of the boundary of a distribution-dependent set.

The work is organised as follows. In Section 4.2, we introduce the main results, which includes the weak convergence of radially recentred and rescaled exceedances to a Poisson point process. Our findings lead to a novel class of limit multivariate distributions that are presented in Section 4.2.3. Section 4.3 details methodology for statistical inference of extremes using hierarchical Bayesian models with latent Gaussian random effects on Euclidean spheres. Finally, Sections 4.4 and 4.5 illustrates the merits of our approach on a two-dimensional hydrological dataset and a three-dimensional oceanographic dataset, respectively. Publicly available code can be accessed via our `R` package `geometricExtremes` on `GitHub`.

4.2 Theory

4.2.1 Scaling and limit sets

Given a d-dimensional random vector $\boldsymbol{X}$ with joint density $f_{\boldsymbol{X}}$, interest lies in characterising its tail region. While some methods focus on modelling the Cartesian representation of $\boldsymbol{X}$ directly, it can sometimes be more convenient to model radii $R = \|\boldsymbol{X}\| \in \mathbb{R}_+$ and directions $\boldsymbol{W} = \boldsymbol{X}/\|\boldsymbol{X}\| \in \mathbb{S}^{d-1}$. Here, $\mathbb{S}^{d-1} := \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| = 1\}$ is the $(d-1)$-dimensional simplex corresponding to some choice of norm $\|\cdot\|$. The approach of Wadsworth and Campbell (2024) uses the L_1 norm and, under certain assumptions on the distribution of $\boldsymbol{X}$, they show that the conditional distribution $R \mid \{\boldsymbol{W}, R > r_0(\boldsymbol{W})\}$ for some high radial threshold $r_0(\cdot)$ can be modelled used a left-truncated Gamma distribution whose rate parameter is given by values of the gauge function $g_{\mathcal{G}}(\boldsymbol{w})$. Their statistical inference methodology rests on regularity assumptions on $f_{\boldsymbol{X}}$, which itself depends on the margins of $\boldsymbol{X}$ following a light-

tailed von Mises distribution. That is, X_j having distribution function satisfying $1 - F_j(x) \sim e^{-\varphi(x)}$ as $x \to \infty$, where φ is a C^2 function with a positive derivative such that $(1/\varphi'(x))' \to 0$ for $x \to \infty$. In practice, their methodology is implemented with standard exponential margins, meaning the domain of $\boldsymbol{X}$ is the positive orthant, $\mathbb{R}^d_+$. We extend on thesis ideas by allowing the domain of $\boldsymbol{X}$ to be the the entire space $\mathbb{R}^d$, and show under regularity assumptions that the random point measure

$$P_n(\cdot) := \sum_{i=1}^{n} \delta_{(a_n(\boldsymbol{W}_i)^{-1}(R_i - b_n(\boldsymbol{W}_i)), \boldsymbol{W}_i)}(\cdot) \quad ; \quad n = 1, 2, \ldots, \tag{4.2}$$

converges to a Poisson point process (PPP) as $n \to \infty$. In the expression (4.2), δ denotes the Dirac measure,

$$\delta_x(A) = \begin{cases} 1 & ; \quad x \in A \\ 0 & ; \quad x \notin A \end{cases}.$$

For more information on the convergence of P_n see Proposition 2 of Papastathopoulos et al. (2025). The intensity measure of this Poisson point process has two components: one corresponding to directions and the other to renormalised radial exceedances conditioned on these directions. The component corresponding to renormalised radial exceedances is given by the density of the generalised Pareto distribution (GPD). Papastathopoulos et al. (2025) therefore suggest that radial exceedances can be modelled via the GPD, whose shape parameter changes based on the margins of $\boldsymbol{X}$. In the case of light-tailed von Mises margins, this is simplified to modelling radial exceedances via the exponential distribution.

Verifying the proposition of convergence to a PPP directly can be challenging. In this section, we introduce a proposition that provides easily verifiable sufficient conditions for this convergence. With a view for more straightforward statistical modelling of extremes, this work will focus on the special case where the margins of $\boldsymbol{X}$ are von Mises. We note that these results extend to more general margins (see Sections 2.1 and 2.2 of Papastathopoulos et al. (2025)), but this is beyond the required scope for this chapter.

In order to establish convergence to a PPP, the notion of *radial functions* corresponding to limit sets and quantile sets needs to be defined. Suppose that $\mathcal{G}$ is a non-degenerate limit set with gauge function $g_{\mathcal{G}}$. Then we define the radial function of $g_{\mathcal{G}}$ to be $r_{\mathcal{G}}(\boldsymbol{w}) = 1/g_{\mathcal{G}}(\boldsymbol{w})$ for $\boldsymbol{w} \in \mathbb{S}^{d-1}$. This can be thought of as the distance from the origin to the boundary of the limit set $\mathcal{G}$ at the direction $\boldsymbol{w}$. Gauge functions can in principle be zero-valued; however, throughout this work we will assume $g_{\mathcal{G}}(\boldsymbol{w})$ is strictly positive on the domain $\mathbb{S}^{d-1}$. This ensures that $r_{\mathcal{G}}$ is always well-defined. In addition to limit sets, the geometric framework considers exceedances above high quantiles of the radii R given the direction $\boldsymbol{W}$. Given some value $q \in (0,1)$ close to 1, define the radial quantile function $r_{\mathcal{Q}_q}$ to be the radial value such that the probability of lying below it is q at a given direction $\boldsymbol{w} \in \mathbb{S}^{d-1}$,

$$r_{\mathcal{Q}_q}(\boldsymbol{w}) := \inf\left\{r \in \mathbb{R} : F_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \geq q\right\}.$$

Under this framework, the limit set $\mathcal{G}$ and the quantile set $\mathcal{Q}_q$ are defined through their respective radial functions via

$$\begin{aligned}\mathcal{G} &:= \bigcup_{\boldsymbol{w}\in\mathbb{S}^{d-1}} \{t r_{\mathcal{G}}(\boldsymbol{w})\boldsymbol{w} : t \in [0,1]\},\\ \mathcal{Q}_q &:= \bigcup_{\boldsymbol{w}\in\mathbb{S}^{d-1}} \left\{t r_{\mathcal{Q}_q}(\boldsymbol{w})\boldsymbol{w} : t \in [0,1]\right\}.\end{aligned}$$

Let $\bigstar$ denote the collection of star-shaped compact subsets of $\mathbb{R}^d \setminus \{\mathbf{0}\}$ (or "star-bodies"), then by definition, we have $\mathcal{G}, \mathcal{Q}_q \in \bigstar$. Further let $\mathbb{P}_{\boldsymbol{W}}(d\boldsymbol{w})$ denote the intensity measure associated with directions $\boldsymbol{W}$. Under a limiting regularity condition on the joint density of $\boldsymbol{X}$, the radial functions $r_{\mathcal{Q}_q}$ and $r_{\mathcal{G}}$ can be used as the location and scale functions respectively in the convergence of radial exceedances to a Poisson point process.

Proposition 4.1. *Suppose that the random vector $\boldsymbol{X}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$, has von Mises margins, and joint density $f_{\boldsymbol{X}}$. Let $\mathcal{G} \in \bigstar$ be described by continuous 1-homogeneous gauge function $g_{\mathcal{G}} : \mathbb{R}^d \to \mathbb{R}_+$. Suppose that there exists $\psi : \mathbb{R}_+ \to \mathbb{R}_+$, and a $\rho > 0$ such that*

$$-\frac{\log f_{\boldsymbol{X}}(t\boldsymbol{x})}{\psi(t)} \to g_{\mathcal{G}}(\boldsymbol{x})^{\rho}, \qquad \boldsymbol{x} \in \mathbb{R}^d \setminus \{\mathbf{0}\}. \tag{4.3}$$

as $t \to \infty$ *uniformly on* $\mathbb{S}^{d-1}$*, then*

$$P_{1/(1-q)}(\cdot) := \sum_{i=1}^{\lceil 1/(1-q) \rceil} \delta_{\left(a_{1/(1-q)}(\boldsymbol{W}_i)^{-1}\left(R_i - b_{1/(1-q)}(\boldsymbol{W}_i)\right), \boldsymbol{W}_i\right)}(\cdot)$$

converges to a Poisson point process with intensity measure $e^{-z}dz\mathbb{P}_{\boldsymbol{W}}(d\boldsymbol{w})$ *as* $q \to 1$ *with normalising functions* $b_{1/(1-q)}(\boldsymbol{w}) = r_{\mathcal{Q}_q}(\boldsymbol{w})$ *and* $a_{1/(1-q)}(\boldsymbol{w}) = r_{\mathcal{G}}(\boldsymbol{w})^{\rho}/\psi'(r_{\mathcal{Q}_q}(\boldsymbol{w}))$*.*

A proof is given in Appendix 4.7.1. Proposition 4.1 states the radial functions $r_{\mathcal{Q}_q}$ and $r_{\mathcal{G}}$ can act as appropriate standardising functions such that the limiting distribution of exceedances of $\mathcal{Q}_q$ follow a standard exponential distribution as $q \to 1$. Note that the assumptions in Proposition 4.1 guarantees that a suitable sequence of scaling constants r_n can be found so that the random point-set $N_n = \{\boldsymbol{X}_1/r_n, \ldots, \boldsymbol{X}_n/r_n\}$ converges in probability onto a limit set $\mathcal{G}$ having radial function $r_{\mathcal{G}}$ (Nolde and Wadsworth, 2022). While this is holds for random vectors $\boldsymbol{X}$ with von Mises margins, we assume through the remaining work that the margins are standardised to follow the standard Laplace distribution. Following from Nolde and Wadsworth (2022), this setting has scaling function $\psi(t) = t$ and $\rho = 1$ in the convergence (4.3). Therefore, radii are recentred using $b_{1/(1-q)} := r_{\mathcal{Q}_q}$ and exceedances are renormalised using $a_{1/(1-q)} = r_{\mathcal{G}}$. Using this, renormalised exceedances are expressed as $(R - r_{\mathcal{Q}_q}(\boldsymbol{w}))/r_{\mathcal{G}}(\boldsymbol{w}) \mid \{\boldsymbol{W} = \boldsymbol{w}, R > r_{\mathcal{Q}_q}(\boldsymbol{W})\}$, which will converge to an exponential distribution as $q \to 1$. This is established in more detail in Section 4.2.3, where a joint distribution of renormalised exceedances and directions is presented. For convenience, this work will assume Further note that throughout this work, we define directions using the L_2, (or "Euclidean") norm. Therefore, for our purposes, $\mathbb{S}^{d-1}$ is the d-dimensional unit sphere.

4.2.2 Probability density function of directions

The point process framework presented in Section 4.2.1 has the important advantage that it yields a distribution of direction of exceedances of $\mathcal{Q}_q$. Let $F_{\boldsymbol{W}}$ be the joint

distribution function of the directional variable $\boldsymbol{W}$. Here we present additional properties of $\boldsymbol{W}$, including a connection between the distribution of directions and star-bodies, even in cases where $F_{\boldsymbol{W}}$ partially depends on the limit set $\mathcal{G}$. For simplicity, we assume throughout that $F_{\boldsymbol{W}}$ is absolutely continuous with respect to the spherical Lebesgue measure and that it admits a probability density $f_{\boldsymbol{W}}$. Denote the radial function $r_{\mathcal{W}} = f_{\boldsymbol{W}}$ such that it defines the directional star body,

$$\mathcal{W} := \bigcup_{\boldsymbol{w} \in \mathbb{S}^{d-1}} \{ t r_{\mathcal{W}}(\boldsymbol{w}) \boldsymbol{w} \, : \, t \in [0,1] \} \in \bigstar. \tag{4.4}$$

Through (4.4), we observe that any continuous probability density function on the sphere $f_{\boldsymbol{W}}$ is in one-to-one correspondence with a strongly starshaped set $\mathcal{W}$ with radial function $r_{\mathcal{W}}$, where "strongly starshaped" denotes the halfline $\{ \, t \, \boldsymbol{w} \quad : \quad t \, \in \, [0,1) \, \}$ not intersecting $\partial\mathcal{W}$ more than once for every $\boldsymbol{w} \in \mathbb{S}^{d-1}$. For $\mathcal{W} \in \bigstar$, define the set $\mathcal{W}^{1/d}$ with radial functions $r_{\mathcal{W}^{1/d}}(\boldsymbol{w}) := r_{\mathcal{W}}(\boldsymbol{w})^{1/d}$. From the definition of star-shaped sets, for $f_{\boldsymbol{W}}$ to be a valid density, $\mathcal{W}$ must satisfy

$$1 = \int_{\mathbb{S}^{d-1}} f_{\boldsymbol{W}}(\boldsymbol{w}) \, \mathrm{d}\boldsymbol{w} = d \left[\frac{1}{d} \int_{\mathbb{S}^{d-1}} r_{\mathcal{W}^{1/d}}(\boldsymbol{w})^d \, \mathrm{d}\boldsymbol{w} \right] = d|\mathcal{W}^{1/d}|$$

(Klain, 1997, Proposition 1.13). Hence, for some star body $\mathcal{L}$, we construct a valid directional star body $\mathcal{W}$ using

$$\mathcal{W} = \mathcal{L}^d / (d|\mathcal{L}|). \tag{4.5}$$

For modelling purposes, we consider three approaches in defining $\mathcal{L}$, and hence, $\mathcal{W}$. A first possibility for the form of the set $\mathcal{L}$ is motivated from the properties of probability density functions that are homothetic with respect to $\mathcal{G}$ (Balkema and Nolde, 2010). Recall that a joint density $f_{\boldsymbol{X}}$ is termed *homothetic* if it has level-sets that are scaled copies of $\mathcal{G}$. This setting is used in the following proposition.

Proposition 4.2. *Suppose that* $f_{\boldsymbol{X}}(\boldsymbol{x}) = f_0\left(r_{\mathcal{G}}(\boldsymbol{x})^{-1}\right)$ *for a decreasing, positive, continuous function* $f_0 : [0,\infty) \to [0,\infty)$ *and a radial function* $r_{\mathcal{G}}$ *characterizing a set* $\mathcal{G} \in \bigstar$. *Then* $\mathcal{L} = \mathcal{G}$.

The proof of Proposition 4.2 is given in Appendix 4.7.2. Proposition 4.2 states that homothetic densities have one-to-one correspondence between the density of directions and $\mathcal{G}$. As a result, under the homothetic framework, an analytic form of the density of directions can be obtained. Consider the three examples below, all of which have homothetic densities $f_{\boldsymbol{X}}$ with common radial function, $r_{\mathcal{G}}(\boldsymbol{x}) = (\boldsymbol{x}^\top \mathsf{Q}\boldsymbol{x})^{-1/2}$ for $\mathsf{Q} \in \mathbb{R}^{d\times d}$ positive definite, but different generators, f_0:

Example 4.1. *Suppose $\boldsymbol{X}$ follows the multivariate Gaussian distribution with precision matrix Q and standard normal margins. The density $f_{\boldsymbol{X}}$ is homothetic with respect to $r_{\mathcal{G}}$ and $f_0(r) = C_1 \exp(-r^{-2}/2)$, $r > 0$, where $C_1 = [|\mathsf{Q}|^{1/2}/(2\pi)^{d/2}]$.*

Example 4.2. *Suppose $\boldsymbol{X}$ follows the multivariate Laplace distribution with precision matrix Q and standard Laplace margins. The density $f_{\boldsymbol{X}}(\boldsymbol{x})$ is homothetic with respect to $r_{\mathcal{G}}$ and $f_0(r) = C_2 r^{-\nu} K_\nu(r^{-1})$, $r > 0$, where $C_2 = [|\mathsf{Q}|^{1/2}/(2\pi)^{d/2}]$, K_ν is the modified Bessel function of the second kind and $\nu = (2-d)/2$.*

Example 4.3. *Suppose $\boldsymbol{X}$ follows the multivariate Student's t distribution with ν degrees of freedom, precision matrix Q, and Student-t_ν margins. The density $f_{\boldsymbol{X}}(\boldsymbol{x})$ is homothetic with respect to $r_{\mathcal{G}}$ and*
$f_0(r) = C_3 (1 + r^{-2}/\nu)^{-(\nu+d)/2}$, $r > 0$, where $C_3 = [\Gamma\{(\nu+d)/2\}|\mathsf{Q}|^{1/2}/\{(\nu\pi)^{d/2}\Gamma(\nu/2)\}]$.

More details on these derivations are presented in Appendix 4.7.5. In each of the above examples, Proposition 4.2 gives an analytic form for the star set $\mathcal{W}$. The set $\mathcal{W}$ is common to all examples as it is defined by $r_{\mathcal{G}}^d/(d|\mathcal{G}|)$ for the same $r_{\mathcal{G}}$ in all examples.

The class of homothetic densities, although rich, serves at best as an idealistic setting. A second option for the form of $\mathcal{L}$ is revealed when standardising the margins of $\boldsymbol{X}$, a common practice in analysing its tail behaviour of random vectors. Consider the map $T : \mathbb{R}^d \to \mathbb{R}^d, \boldsymbol{x} \mapsto \left(\Psi^{\leftarrow}(F_{X_i}(x_i)) : i = 1, \ldots, d\right)$ performing a transformation of the marginal distributions of $\boldsymbol{X}$, so that the i^{th} element of $T(\boldsymbol{X})$

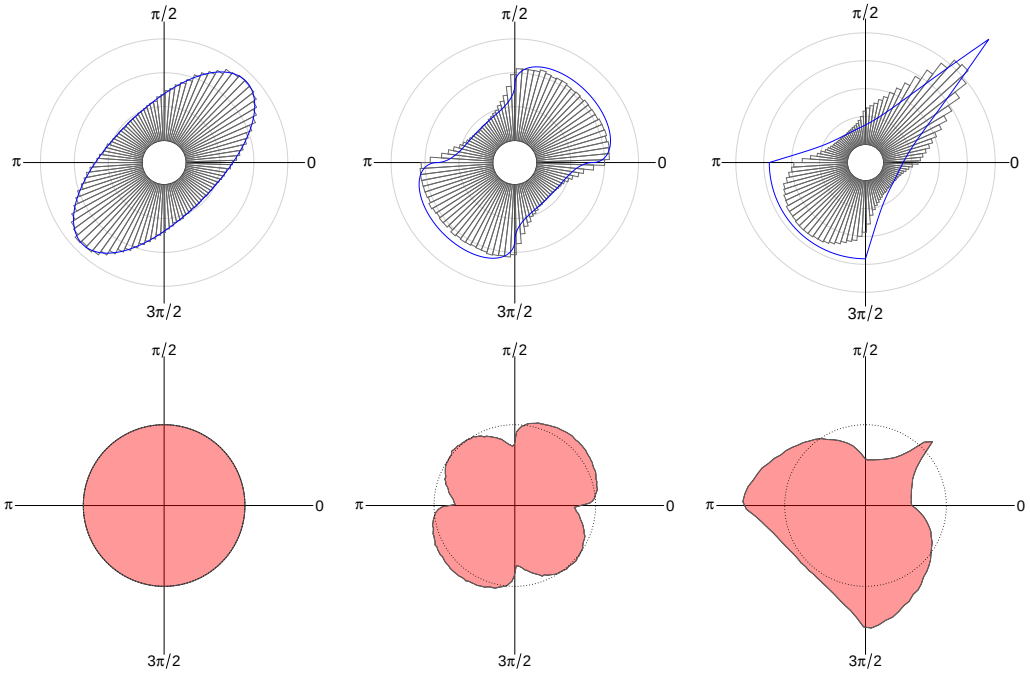

Figure 4.2: Top row: Circular histogram of 10^6 directions sampled from $\mathcal{W}$ for the bivariate Gaussian distribution with standard normal margins (left), the bivariate Gaussian distribution with standard Laplace margins (centre), and from a bivariate max-stable logistic distribution with Laplace margins (right). A concentric circle corresponds to a leap of 0.1 in the density of directions. Solid blue lines correspond to the boundary of $\mathcal{G}^d/(d|\mathcal{G}|)$, for $\mathcal{G}$ of the respective distributions. Bottom row: Sets $[\mathcal{G}^d/(d|\mathcal{G}|)] \cdot \mathcal{W}^{-1}$ in red corresponding to discrepancy between the empirical distribution of $\mathcal{W}$ (given by the above histograms) and $\mathcal{G}^d/(d|\mathcal{G}|)$ (given by the blue curves). If the boundary of the red set matches the unit circle (dotted line), then $\mathcal{W}$ and $\mathcal{G}^d/(d|\mathcal{G}|)$ are in perfect agreement.

follows Ψ, a continuous cumulative distribution function on $\mathbb{R}$. The transformation preserves the cardinality of the set of input vectors $\boldsymbol{X}/\|\boldsymbol{X}\|$, ensuring a one-to-one correspondence with the output vector $T(\boldsymbol{X})/\|T(\boldsymbol{X})\|$. From a geometric perspective, this means that the densities of $\boldsymbol{X}/\|\boldsymbol{X}\|$ and $T(\boldsymbol{X})/\|T(\boldsymbol{X})\|$ can be defined over the same set of points. The distribution of directions, however, may change when T introduces nonlinearities. For example, suppose that $\boldsymbol{X}_N$

is distributed according to a multivariate normal distribution with zero mean and precision matrix Q such that each marginal follows the standard Gaussian distribution. Let $\boldsymbol{X} = F_L^{-1}[\Phi(\boldsymbol{X}_N)]$, were F_L and Φ denote the cumulative distribution functions of standard Laplace and standard normal random variables, respectively. While the density of $\boldsymbol{X}_N$ is homothetic, thus having uniform rate of convergence is across directions, the density of $\boldsymbol{X}$ is not homothetic. For the multivariate Gaussian distribution in Laplace margins, we prove in Appendix 4.7.4.1 that convergence of Proposition 4.1 holds uniformly, with $\mathcal{G}$ determined by

$$r_{\mathcal{G}}(\boldsymbol{w}) = \left[\left\{ \operatorname{sgn}(\boldsymbol{w}) |\boldsymbol{w}|^{1/2} \right\}^\top \mathsf{Q} \left\{ \operatorname{sgn}(\boldsymbol{w}) |\boldsymbol{w}|^{1/2} \right\} \right]^{-1}, \qquad \boldsymbol{w} \in \mathbb{S}^{d-1}.$$

where $\operatorname{sgn}(\boldsymbol{w})|\boldsymbol{w}|^{1/2} = (\operatorname{sgn}(w_i)|w_i|^{1/2} \,:\, i = 1, \ldots, d)$ and $\operatorname{sgn}(x) = x/|x|$ denotes the sign function. From Figure 4.2, there is empirical evidence that the star set $\mathcal{W}$ describing the density of $\boldsymbol{X}/\|\boldsymbol{X}\|$ is no longer a constant scale multiple of $\mathcal{G}$, but instead a radial product of $\mathcal{G}$ with another star-body. Figure 4.2 also shows that, after standardising the margins to Laplace, solely using $\mathcal{G}$ is not sufficient to capture the distribution of directions corresponding max-stable logistic random vectors. The radial function $r_{\mathcal{G}}$ corresponding to $\mathcal{G}$ is given here by

$$r_{\mathcal{G}}(\boldsymbol{w}) = \begin{cases} \left[\frac{1}{\theta} \sum_{j=1}^{d} w_j + \left(1 - \frac{d}{\theta}\right) \min(\boldsymbol{w}) \right]^{-1} & ; \quad \min(\boldsymbol{w}) > 0 \\ \left[\frac{1}{\theta} \sum_{j:w_j>0} w_j + \left\{ \sum_{j:w_j<0} (-w_j)^{1/\theta} \right\}^{\theta} \right]^{-1} & ; \quad \text{otherwise} \end{cases}, \qquad \boldsymbol{w} \in \mathbb{S}^{d-1},$$

where $\theta \in (0, 1)$ is the dependence parameter of the max-stable logistic distribution. The full derivation in Appendix 4.7.4.5.

The empirical analysis of the marginal transformation to Laplace margins and its effect on the distribution of the directions reveals a case where $\mathcal{W}$ can depend both on $\mathcal{G}$ and on some additional star-body $\mathcal{B}$ independent of $\mathcal{G}$. The star-body $\mathcal{B}$ captures residual directional variation that is not explained by a homothetic density. For the practical case where both $r_{\mathcal{W}}$ and r_G are strictly positive on $\mathbb{S}^{d-1}$, then we can consider the form $\mathcal{L} = \mathcal{B} \cdot \mathcal{G}$, where $\mathcal{B}, \mathcal{G} \in \bigstar$ such that $\mathcal{B}$ is independent of $\mathcal{G}$. In

the specific context where f is homothetic with respect to $r_{\mathcal{G}}$, we have $\mathcal{B} = B_1(\mathbf{0})$, the Euclidean ball with radius 1 centred at the origin $\mathbf{0}$.

During inference, it may be that we wish to model directions completely separately to the limit set $\mathcal{G}$. Therefore, a third possibility for the set $\mathcal{L}$, and consequently the set $\mathcal{W}$, is to set $\mathcal{L} = \mathcal{B}$ for some set $\mathcal{B}$ independent of $\mathcal{G}$, implying a case where $\boldsymbol{W}$ is independent of $\mathcal{G}$. This can occur within the class of the radial exponential distribution, introduced in the following section. In Appendix 4.7.5, we derive closed-form expressions of the distribution of directions $f_{\boldsymbol{W}}$ for a number of distributions when the marginal distributions are common and prespecified across all d components of $\boldsymbol{X}$. All three setups for the form of $\mathcal{L}$ will be considered during model fitting, as is outlined in Section 4.3.

4.2.3 Radial exponential distributions

We now present a novel family of multivariate distributions, termed *radial exponential (rExp) distributions* and detail their threshold stability property that makes them suitable for extrapolation. They arise as the only non-trivial limits of radially renormalised exceedances above a threshold $\mathcal{Q}_q$, as detailed in Section 4.2.1, and enable the modelling of rare events in a much wider set of joint-tail regions of $\mathbb{R}^d$ than other well-established frameworks of extreme value theory. The rExp class of distributions are parameterised via members of the class of star bodies $\bigstar$. Before defining the rExp distribution, note that a star body $\mathcal{A} \in \bigstar$ can be written in terms of its radial function, $r_{\mathcal{A}} : \mathbb{S}^{d-1} \to \mathbb{R}_+$, through

$$\mathcal{A} = \bigcup_{\boldsymbol{w} \in \mathbb{S}^{d-1}} \{t r_{\mathcal{A}}(\boldsymbol{w}) \boldsymbol{w} : t \in [0, 1]\}.$$

Definition 4.1 (Multivariate radially-stable exponential distributions)**.** *A random vector $\boldsymbol{Z} \in \mathbb{R}^d$ is said to follow the multivariate radially-stable exponential (rExp) distribution with location $\mathcal{M} \in \bigstar$, scale $\Sigma \in \bigstar$, and directional shape $\mathcal{W} \in \bigstar$, if*

for any Borel set $B \subseteq \mathbb{R}^d \setminus \mathcal{M}$,

$$\mathbb{P}\left[\boldsymbol{Z} \in B\right] = \int_{S_B} \int_{I_B(\boldsymbol{w})} \exp\left\{-\left(\frac{r - r_{\mathcal{M}}(\boldsymbol{w})}{r_{\Sigma}(\boldsymbol{w})}\right)\right\} r_{\mathcal{W}}(\boldsymbol{w})\, \mathrm{d}r\, \mathrm{d}\boldsymbol{w},$$

where $S_B = \{\boldsymbol{w} \in \mathbb{S}^{d-1} : \exists\, t > 0 \text{ such that } t\boldsymbol{w} \in B\}$ *and* $I_B(\boldsymbol{w}) = \{t \in \mathbb{R}_+ : t\boldsymbol{w} \in B\}$.

The rExp distribution presented in Definition 4.1 possesses a stability property that we here detail through Proposition 4.3.

Proposition 4.3. *Let* $\boldsymbol{Z} \in \mathbb{R}^d$ *have an rExp distribution with location* $\mathcal{M}$*, scale* Σ*, and directional shape* $\mathcal{W}$*, then* $\boldsymbol{Z}$ *satisfies the radial memoryless property:*

$$\mathbb{P}\left[\boldsymbol{Z} \in \{\mathcal{M} + B_{r_1+r_2}(\boldsymbol{0}) \cdot \Sigma\}' \mid \boldsymbol{Z} \in \{\mathcal{M} + B_{r_1}(\boldsymbol{0}) \cdot \Sigma\}'\right] = \mathbb{P}\left[\boldsymbol{Z} \in \{\mathcal{M} + B_{r_2}(\boldsymbol{0}) \cdot \Sigma\}'\right], \tag{4.6}$$

where A' *denotes the complement of the set* A.

Proof of the stability property in Proposition 4.3 is given in Appendix 4.7.3. In Section 4.3, we propose a statistical inference method utilising the family of rExp distributions to extrapolate to extreme regions lying beyond the range of observed data.

4.3 Statistical inference

4.3.1 Standardisation of margins

Suppose that $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$, are observations drawn randomly from the distribution of the random variable $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^\top$. When the tails of the original distribution decay exponentially, we have the flexibility to model a broader range of extremal dependence structures, not only allowing for both asymptotic independence and asymptotic dependence, but also allowing for more complex types of dependencies such as when some coordinates exhibit positive extremal dependence while others exhibit negative extremal dependence (Keef et al., 2013a; Nolde and

Wadsworth, 2022). When $\boldsymbol{Y}$ does not have exponential decay, we choose to apply the transformation $X_j = F_L^{-1}(\widehat{F}_j(Y_j))$ for $j = 1, \ldots, d$, where F_L^{-1} is the distribution function of the standard Laplace distribution, and

$$\widehat{F}_j(y) = \begin{cases} \left[1 - \left\{1 - \widehat{\xi}_{j,\mathrm{L}} \left(\dfrac{u_{j,\mathrm{L}} - y}{\widehat{\sigma}_{j,\mathrm{L}}}\right)\right\}_+^{-1/\widehat{\xi}_{j,\mathrm{L}}}\right] \widetilde{\widetilde{F}}_j(u_{j,\mathrm{L}}) & y \leq u_{j,\mathrm{L}} \\ \widetilde{F}_j(y) & u_{j,\mathrm{L}} < y \leq u_{j,\mathrm{R}} \\ 1 - [1 - \widetilde{F}(u_{j,\mathrm{R}})] \left\{1 + \widehat{\xi}_{j,\mathrm{R}} \left(\dfrac{y - u_{j,\mathrm{R}}}{\widehat{\sigma}_{j,\mathrm{R}}}\right)\right\}_+^{-1/\widehat{\xi}_{j,\mathrm{R}}} & y > u_{j,\mathrm{R}} \end{cases} . \quad (4.7)$$

$\widetilde{F}_j(y) = (n+1)^{-1} \sum_{i=1}^n \mathbb{1}[y_{ij} \leq y]$ and $\widetilde{\widetilde{F}}_j(y) = (n+1)^{-1} \sum_{i=1}^n \mathbb{1}[y_{ij} \geq y]$. The quantities $(\widehat{\sigma}_{j,\mathrm{L}}, \widehat{\xi}_{j,\mathrm{L}})$, $(\widehat{\sigma}_{j,\mathrm{R}}, \widehat{\xi}_{j,\mathrm{R}})$ are scale and shape parameters of generalised Pareto distributions for the lower (left) and upper (right) tail of the j^{th} margin, which are obtained using frequentist maximum likelihood estimation and are used to model the tail decay below and above the thresholds $u_{j,-}$ and $u_{j,+}$, respectively.

As we perform Bayesian modelling for the multivariate extremal dependence structure (see Sections 4.3.2 and 4.3.4), modelling the GP parameters ($\sigma_{j,\mathrm{L}}$,$\xi_{j,\mathrm{L}}$,$\sigma_{j,\mathrm{R}}$, $\xi_{j,\mathrm{R}}$) in (4.7) would ideally also be done in a Bayesian manner. In this setting, a posterior distribution is assumed on the joint distribution of the quantiles $u_{j,\mathrm{L}}$ and $u_{j,\mathrm{R}}$. This posterior would then enable for the sampling of candidate thresholds, above which separate GP models are implemented for the tails. This approach would more naturally account for uncertainty in the marginal distributions' parameters since the posterior distribution of the quantiles propagates this uncertainty forward into the tail modelling process. However, for simplicity, the thresholds $u_{j,\mathrm{L}}$ and $u_{j,\mathrm{R}}$ are presently obtained using high quantiles of the j^{th} marginal distribution, and we adopt a frequentist maximum likelihood approach in estimating the GP parameters. In this framework, we propagate the uncertainty in the quantile estimate and the corresponding GP parameters forward using a bootstrap approach.

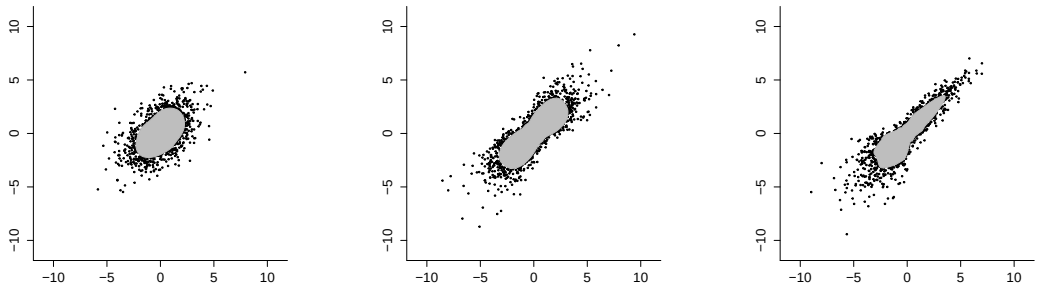

Figure 4.3: Estimated posterior mean sets $\mathcal{Q}_q$ (grey) using $q = 0.90$ for three different bivariate datasets (black points) in Laplace margins. From left to right: Laplace, Gaussian, and max-stable with logistic dependence. The boundaries of the sets $\mathcal{Q}_q$ are the estimated posterior mean values.

4.3.2 Quantile regression

Quantile regression methods are typically implemented using a pinball loss function constructed using the check function (Koenker, 2005) and often without a distributional model for the density. Yu and Moyeed (2001) propose the use of the asymmetric Laplace distribution for the model density due to the equivalence of the negative log density of the asymmetric Laplace with the pinball loss function; however, the asymmetric Laplace as an adequate model for the data is precarious in a Bayesian setting. For example, Waldmann et al. (2013) show that the resulting posterior prediction intervals are not well-calibrated, and this is especially pronounced for tail quantiles, which are essential in our setup. Fasiolo et al. (2021) confirm the poor posterior prediction intervals, and show that the scale parameter of the asymmetric Laplace distribution is arbitrary in a Bayesian framework and can therefore lead to inaccurate quantile estimates. They instead propose a novel method for selecting the scale parameter; however, accompanying `R` code is only suitable for the bivariate setting, or when modelling for quantiles of $R \mid \{W = w\}$ for $w \in [0, 2\pi]$.

We instead adopt a generalised linear model based approach for quantile

regression, requiring an adequate distributional model for the density of $R \mid \boldsymbol{W}$. Wadsworth and Campbell (2024) show that the distribution of $R \mid \boldsymbol{W}$ above a high threshold can often be approximated by the left-truncated gamma distribution. For modelling purposes, we therefore choose the gamma distribution for the density in the tail region of $R \mid \boldsymbol{W}$. Within the gamma likelihood fititng, we model the logarithm of its conditional q-quantile, $\log r_{\mathcal{Q}_q}(\boldsymbol{w})$, using a finite-dimensional continuously specified Gaussian process prior on $\mathbb{S}^{d-1}$ using the stochastic partial differential equation approach by Lindgren et al. (2011). This is described in detail in Section 4.3.4, and is achieved using the default settings in the `INLA` package within `R` (see `www.r-inla.org`).

Since estimators of high quantiles are not influenced by the bulk of the distribution, a likely misspecification between our choice and the true density of $R \mid \boldsymbol{W}$ in the body of the distribution is not of concern. Our model choice exploits the form of the decay of the conditional density of $R \mid \boldsymbol{W}$, and allows for Bayesian inference. Given i.i.d. observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ from a random vector $\boldsymbol{Y}$, we obtain the standardised data $\boldsymbol{x} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ via $\boldsymbol{x}_i = (F^{-1}(\widehat{F}_1(y_{i,1})), \ldots, F^{-1}(\widehat{F}_d(y_{i,d})))^\top$ as described in Section 4.3.1. To infer the quantile set $\mathcal{Q}_q$ of $\boldsymbol{X}$, we treat $\boldsymbol{x}_{1:n} = \{(r_i, \boldsymbol{w}_i) : i = 1, \ldots, n\}$, with $(r_i, \boldsymbol{w}_i) = (\|\boldsymbol{x}_i\|, \boldsymbol{x}_i/\|\boldsymbol{x}_i\|)$, as observations from $(R, \boldsymbol{W})$ and apply the gamma quantile regression method detailed above. Figure 4.3 shows posterior mean estimates for $\mathcal{Q}_q$ on three different bivariate datasets. In it, we see that we can accurately capture the tail behaviour of $R \mid \{W = w\}$ as w varies in $\mathbb{S}^1 = [0, 2\pi]$.

While the gamma model is simple, we remark that it appears to be a good enough approximation in estimating the high tail quantiles for use in further inference. In this work, diagnostics for $\mathcal{Q}_q$ estimated using the gamma model remain purely visual. Numerical diagnostics are possible (see Campbell and Wadsworth (2024) or Chapter 5); however, we leave this to future work in the context of the gamma model and declare it sufficient to use in the estimation the star bodies $\mathcal{G}$ and $\mathcal{L}$.

4.3.3 Conditional likelihood function

Suppose $\boldsymbol{X}$ has been standardised to have standard Laplace margins, we assume that exceedances of $\mathcal{Q}_q$ follow an rExp distribution with location $\mathcal{Q}_q$ and scale $\mathcal{G}$, and directional shape $\mathcal{W}$. Given the quantile set $\mathcal{Q}_q$ obtained from the quantile regression outlines in Section 4.3.2, we use the rExp and distribution distribution to build a likelihood that will allow inference for $\mathcal{G}$ and $\mathcal{W}$. We recall from equation (4.5) that the set $\mathcal{W}$ can be expressed via $f_{\boldsymbol{W}}(\boldsymbol{w}) = r_{\mathcal{L}}(\boldsymbol{w})^d/(d|\mathcal{L}|)$, where $\mathcal{L}$ is a star-body. Based on Section 4.2.2, we consider three possible models for the directional shape $\mathcal{W}$:

$$\mathrm{M}_1 : \mathcal{L} = \mathcal{B}, \qquad \mathrm{M}_2 : \mathcal{L} = \mathcal{G}, \qquad \mathrm{M}_3 : \mathcal{L} = \mathcal{B} \cdot \mathcal{G}, \tag{4.8}$$

for $\mathcal{B} \in \bigstar$ independent of $\mathcal{G}$. That is, in M_1, we model the directional distribution completely separate from the rExp distribution. In M_2, we use the same gauge function in the rExp and directional distributions. In M_3, we allow for the directional distribution to be modelled by a star body related to the gauge function used in the rExp distribution, but we allow for greater flexibility through modification with another star body. The nested structure of model M_2 within the parameter space of model M_3 translates into a bias-variance trade-off as the latter offers additional flexibility at the cost of a possibly increased variance for the latent set $\mathcal{G}$. A similar trade-off occurs for $\mathcal{G}$ between models M_1 and M_3 since the former ignores possible information contained in the observed directions.

Conditionally on the quantile set $\mathcal{Q}_q$ and the data $\boldsymbol{x}_{1:n} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, the likelihood of $\boldsymbol{\theta} = (\mathcal{G}, \mathcal{L})$, is given by

$$L(\boldsymbol{\theta} \mid \mathcal{Q}_q, \boldsymbol{x}_{1:n}) := \prod_{i=1}^{n} f_{R,\boldsymbol{W}}(r_i, \boldsymbol{w}_i \mid \boldsymbol{\theta}, \mathcal{Q}_q), \tag{4.9}$$

where $\boldsymbol{\theta}$ represents a vector of parameters defining $\mathcal{G}$ and $\mathcal{L}$. These star bodies can either be defined through parametric families, or in a semiparametric manner. Letting $\mathcal{S}_q := \{i \in \{1, \ldots, n\} : r_i > r_{\mathcal{Q}_q}(\boldsymbol{w}_i)\}$ and $\mathcal{S}_q' := \{1, \ldots, n\} \backslash \mathcal{S}_q$, the likelihood (4.9) can be expressed in terms of contributions of the radii $\{r_1, \ldots, r_n\}$

above and below the radial threshold $r_{\mathcal{Q}_q}$,

$$L(\boldsymbol{\theta} \mid \mathcal{Q}_q, \boldsymbol{x}_{1:n}) = \prod_{i \in \mathcal{S}_q'} f_{R|R \leq r_{\mathcal{Q}_q}(\boldsymbol{W}), \boldsymbol{W}}(r_i \mid \boldsymbol{w}_i) f_{\boldsymbol{W}}(\boldsymbol{w}_i) \times \prod_{i \in \mathcal{S}_q} f_{R|R > r_{\mathcal{Q}_q}(\boldsymbol{W}), \boldsymbol{W}}(r_i \mid \boldsymbol{w}_i) f_{\boldsymbol{W}}(\boldsymbol{w}_i). \tag{4.10}$$

Under the assumption that $\boldsymbol{x}_{1:n}$ is a random sample, a consequence of Proposition 5 of Papastathopoulos et al. (2025) is that the distribution of directions corresponding to exceedance radii is the same as the distribution of all directions. This suggests that all directions $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$ may be used in the likelihood for the inference of $f_{\boldsymbol{W}} = r_{\mathcal{L}}^d/(d|\mathcal{L}|)$ used to model $\boldsymbol{W} \mid R > r_{\mathcal{Q}_q}(\boldsymbol{W})$. In particular, when models M_1 or M_3 are used, substantial gains in the inference for $\mathcal{G}$ can be attained by also including in inference the directions at which non-exceedances occur, as illustrated in the simulation study of Papastathopoulos et al. (2025). As we assume an rExp distribution is followed in the exceedances only, non-exceedance radii $\{r_i : i \in \mathcal{S}_q'\}$ are assumed not to carry information about $\mathcal{G}$ and $\mathcal{L}$; we pose that $\mathcal{L}$ is constant with respect to them. Denoting by $\mathcal{S}_{\boldsymbol{w}}$ the set of indices of at least all exceedances and at most all observations — either $\mathcal{S}_{\boldsymbol{w}} = \mathcal{S}_q$ or $\mathcal{S}_{\boldsymbol{w}} = \{1, \ldots, n\}$ — the likelihood (4.10) reduces to

$$\begin{aligned} &L(\boldsymbol{\theta} \mid \mathcal{Q}_q, \boldsymbol{x}_{1:n}) \\ &\propto \prod_{i \in \mathcal{S}_q} f_{R|R > r_{\mathcal{Q}_q}(\boldsymbol{W}), \boldsymbol{W}}(r_i \mid \boldsymbol{w}_i) \prod_{i \in \mathcal{S}_{\boldsymbol{w}}} f_{\boldsymbol{W}}(\boldsymbol{w}_i) \qquad (4.11) \\ &= \exp\{-|\mathcal{S}_{\boldsymbol{w}}| \log(d|\mathcal{L}|)\} \prod_{i \in \mathcal{S}_q} f_{R|\boldsymbol{W}}\left((r_i - r_{\mathcal{Q}_q}(\boldsymbol{w}_i))/r_{\mathcal{G}}(\boldsymbol{w}_i) \mid \boldsymbol{w}_i\right) \prod_{j \in \mathcal{S}_{\boldsymbol{w}}} r_{\mathcal{L}}(\boldsymbol{w}_j)^d, \end{aligned}$$

where $f_{R|\boldsymbol{W}}(z \mid \boldsymbol{w}) \propto e^{-z}/r_{\mathcal{G}}(\boldsymbol{w})$. The likelihood function (4.11) is amenable to standard likelihood based inference using either frequentist or Bayesian methods when parametric models are selected for $\mathcal{W}$ and $\mathcal{G}$. When interest is in semi-parametric models, evaluating the likelihood function (4.11) requires computing the constant $d|\mathcal{L}|$, which makes inference difficult. However, using the *Poisson transform* (Baker, 1994), we can map the likelihood into an equivalent likelihood function

$L(\boldsymbol{\theta}, \gamma \mid \mathcal{Q}_q, \boldsymbol{x}_{1:n})$ of a Poisson point process defined in an expanded space given by

$$L(\boldsymbol{\theta}, \gamma) = \exp\left[-|\mathcal{S}_{\boldsymbol{w}}|e^{\gamma}(d|\mathcal{L}|)\right] \prod_{i \in \mathcal{S}_q} f_{R|\boldsymbol{W}}\left((r_i - r_{\mathcal{Q}_q}(\boldsymbol{w}_i))/r_{\mathcal{G}_q}(\boldsymbol{w}_i) \mid \boldsymbol{w}_i\right) \prod_{i \in \mathcal{S}_{\boldsymbol{w}}} e^{\gamma} r_{\mathcal{L}}(\boldsymbol{w}_i)^d, \tag{4.12}$$

where the latent variable γ estimates the normalising constant $d|\mathcal{L}|$, inferred as another parameter at no loss of information (Barthelmé and Chopin, 2015). It is worth noting that through this approach, we still need to compute the volume of $\mathcal{L}$. However, when taking the logarithm of (4.12) for maximization, the volume $|\mathcal{L}|$ is now on the linear scale and can therefore be estimated numerically in a more stable and efficient manner based on the method introduced by Simpson et al. (2016); see also Yuan et al. (2017), Arce Guillen et al. (2023), and Papastathopoulos et al. (2023) for further applications of this method. Inference can be performed either using frequentist methods or in a fully Bayesian manner, that is, by assigning suitable prior distributions on $\mathcal{G}$, $\mathcal{L}$, and γ (Arce Guillen et al., 2023). More details on statistical inference for the latent variables and on how the fitted models can be used to perform rare event probability estimation are found in Sections 4.3.4 and 4.3.5.

4.3.4 Inference for latent variables

Given n independent observations $\boldsymbol{x}_{1:n} := \{r_1\boldsymbol{w}_1, \ldots, r_n\boldsymbol{w}_n\}$ from $\boldsymbol{X} = R\boldsymbol{W} \in \mathbb{R}^d$, the goal is to estimate the quantile set $\mathcal{Q}_q$, the limit set $\mathcal{G}$, and the set $\mathcal{L}$ defining the directional distribution. To do this, we model the logarithms of the random radial functions $r_{\mathcal{Q}_q}$, $r_{\mathcal{G}}$ and $r_{\mathcal{L}}$ as Matérn (Gaussian) fields on $\mathbb{S}^{d-1}$ using the stochastic partial differential equation (SPDE) approach by Lindgren et al. (2011), which is also the default option in the `INLA` package in `R` (see `www.r-inla.org`). In this setting, the simplex $\mathbb{S}^{d-1}$ is partitioned using p equally-spaced vertices, and each radial function is evaluated for some point $\boldsymbol{w} \in \mathbb{S}^{d-1}$ using

$$r_{\bullet}(\boldsymbol{w}) = r_{\bullet}(\boldsymbol{w}; \beta_{\bullet}, \boldsymbol{z}_{\bullet}) = \exp\left\{-\left[\beta_{\bullet} + \sum_{k=1}^{p} z_{\bullet,k} h_k(\boldsymbol{w})\right]\right\}, \tag{4.13}$$

where h_k is a piecewise-linear basis function of order 1, and $\bullet$ is one of either $\mathcal{Q}_q, \mathcal{G}, \mathcal{L}$ (Lindgren et al., 2011). In (4.13), $\beta_{\mathcal{Q}_q}, \beta_{\mathcal{G}}, \beta_{\mathcal{L}} \in \mathbb{R}$ are random intercepts, and $\boldsymbol{z}_{\mathcal{Q}_q}, \boldsymbol{z}_{\mathcal{G}}, \boldsymbol{z}_{\mathcal{L}} \in \mathbb{R}^p$ are stochastic weights modelled using a p-variate Gaussian distribution with zero mean and a sparse $p \times p$ precision matrix whose exact form is given in Lindgren et al. (2011) and again in Papastathopoulos et al. (2023). In this construction, the collection of parameters used is represented by $\boldsymbol{\theta} = (\beta_{\mathcal{Q}_q}, \boldsymbol{z}_{\mathcal{Q}_q}, \beta_{\mathcal{G}}, \boldsymbol{z}_{\mathcal{G}}, \beta_{\mathcal{L}}, \boldsymbol{z}_{\mathcal{L}}) \in \mathbb{R}^{3p+3}$.

It follows that the joint posterior distribution of $\boldsymbol{\theta}$ fully determines that of $\mathcal{Q}_q$, $\mathcal{G}$ and $\mathcal{L}$. Due to the hierarchical structure of all proposed models detailed in Sections 4.3.2 and 4.3.3, the joint posterior density of $\boldsymbol{\theta}$ factorises according to $\pi[\boldsymbol{\theta} \mid \boldsymbol{x}_{1:n}] = \pi[\beta_{\mathcal{G}}, \boldsymbol{z}_{\mathcal{G}}, \beta_{\mathcal{L}}, \boldsymbol{z}_{\mathcal{L}} \mid \beta_{\mathcal{Q}_q}, \boldsymbol{z}_{\mathcal{Q}_q}, \boldsymbol{x}_{1:n}]\pi[\beta_{\mathcal{Q}_q}, \boldsymbol{z}_{\mathcal{Q}_q} \mid \boldsymbol{x}_{1:n}]$. During inference, we begin by fitting the Bayesian gamma log-linear quantile regression model described in Section 4.3.2 for a predetermined probability q to all observations $\boldsymbol{x}_{1:n}$. Samples $\{(\beta_{\mathcal{Q}_q,i}, \boldsymbol{z}_{\mathcal{Q}_q,i}) : i = 1, \ldots, n_{\mathcal{Q}_q}\}$ from the posterior density $\pi[\beta_{\mathcal{Q}_q}, \boldsymbol{z}_{\mathcal{Q}_q} \mid \boldsymbol{x}_{1:n}]$ map to a set of radial functions $\{r_{\mathcal{Q}_q,i} : i = 1, \ldots, n_{\mathcal{Q}_q}\}$, each interpreted as candidate radial functions of a q^{th} quantile set.

Given a radial function $r_{\mathcal{Q}_q,i}$, we define a set of exceedances through

$$\mathcal{X}_i = \left\{(r_j, \boldsymbol{w}_j) \in (0, \infty) \times \mathbb{S}^{d-1} \ : \ r_j > r_{\mathcal{Q}_q,i}(\boldsymbol{w}_j),\, r_j \boldsymbol{w}_j \in \mathbb{R}^d,\, j = 1, \ldots, n\right\}. \quad (4.14)$$

We then fit model M_1, M_2, or M_3 to each collection of exceedances $\mathcal{X}_i$ using the likelihood detailed in equation (4.12). This is either done with all angles or exceedance angles only, leaving with a total of six possible modelling setups. This procedure yields a conditional posterior density $\pi[\beta_{\mathcal{G}}, \boldsymbol{z}_{\mathcal{G}}, \beta_{\mathcal{L}}, \boldsymbol{z}_{\mathcal{L}} \mid \beta_{\mathcal{Q}_q,i}, \boldsymbol{z}_{\mathcal{Q}_q,i}, \boldsymbol{x}_{1:n}]$ for each $i = 1, \ldots, n_{\mathcal{Q}_q}$. Sampling $n_{\mathcal{G}\mathcal{L}}$ realisations jointly from each $\pi[\beta_{\mathcal{G}}, \boldsymbol{z}_{\mathcal{G}}, \beta_{\mathcal{L}}, \boldsymbol{z}_{\mathcal{L}} \mid \beta_{\mathcal{Q}_q,i}, \boldsymbol{z}_{\mathcal{Q}_q,i}, \boldsymbol{x}_{1:n}]$ provides an assembled sample of $n_{\mathcal{Q}_q} \cdot n_{\mathcal{G}\mathcal{L}}$ realisations from the joint posterior distribution of $\boldsymbol{\theta}$,

$$\left\{\boldsymbol{\theta}_{i,j} = \left(\beta_{\mathcal{Q}_q,i}, \boldsymbol{z}_{\mathcal{Q}_q,i}, \beta_{\mathcal{G},(i,j)}, \boldsymbol{z}_{\mathcal{G},(i,j)}, \beta_{\mathcal{L},(i,j)}, \boldsymbol{z}_{\mathcal{L},(i,j)}\right) \in \boldsymbol{\Theta} \ : \ i = 1, \ldots, n_{\mathcal{Q}_q}, j = 1, \ldots, n_{\mathcal{G}\mathcal{L}}\right\}. \quad (4.15)$$

For simplicity and without loss, we re-index the sample (4.15) from the posterior

distribution of $\boldsymbol{\theta}$ to $\{\boldsymbol{\theta}_i : i = 1, \ldots, n_{\boldsymbol{\theta}}\}$ with $n_{\boldsymbol{\theta}} := n_{\mathcal{Q}_q} \cdot n_{\mathcal{GL}}$ and use this notation henceforth.

Uncertainty estimation for the latent fields $r_{\mathcal{Q}_q}$, $r_{\mathcal{G}}$, and $r_{\mathcal{L}}$ is obtained via *prediction intervals* (see Bolin and Lindgren, 2018). A $(1-\alpha)$ prediction interval for the process $r_\bullet$ defined on $\mathbb{S}^{d-1}$ consists of the strip $R_{1-\alpha} := \cup_{\boldsymbol{w} \in \mathbb{S}^{d-1}} R_{1-\alpha}(\boldsymbol{w})$ defined through $R_{1-\alpha}(\boldsymbol{w}) := [q_\rho(\boldsymbol{w})\boldsymbol{w} : q_{1-\rho}(\boldsymbol{w})\boldsymbol{w}]$ for some ρ such that $q_\rho(\boldsymbol{w})$ and $q_{1-\rho}(\boldsymbol{w})$ satisfy

$$\mathbb{P}\left[q_\rho(\boldsymbol{w}) \leq r_\bullet(\boldsymbol{w}) \leq q_{1-\rho}(\boldsymbol{w}), \boldsymbol{w} \in \mathbb{S}^{d-1}\right] = 1 - \alpha.$$

In the context of our latent fields $r_{\mathcal{Q}_q}$, $r_{\mathcal{G}}$, and $r_{\mathcal{L}}$, $(1-\alpha)$ prediction intervals consist of sets within which the true functions lie entirely with probability $1-\alpha$. Prediction intervals can be obtained from a sample from the posterior distribution of a parameter of interest using the `excursions` package in `R` (Bolin and Lindgren, 2015, 2017). Throughout the rest of this work, we present prediction intervals for $\alpha = 0.05$.

4.3.5 Rare event probability estimation

In multivariate extremes, interest often lies in estimating $\mathbb{P}[\boldsymbol{X} \in B]$ for some set $B \in \mathbb{R}^d \setminus \{\boldsymbol{0}\}$ lying outside the range of observable data. The geometric approach in Wadsworth and Campbell (2024) performs this task by first sampling directions, then using them to extrapolate radii beyond a high threshold. The radii and directions are multiplied together to obtain an extremal sample in Cartesian coordinates, and a Monte Carlo approximation of $\mathbb{P}[\boldsymbol{X} \in B]$ is obtained. While shown to be very successful, some difficulties arise. Mainly, the further B lies in the tails of the distribution of $\boldsymbol{X}$, the less obvious it becomes on how far one needs to extrapolate in order to get non-zero probability estimates.

In this work, we bypass this by staying in radial-directional coordinates, and using the rExp distribution and a sample of directions to give a closed-form expression for the probability of radii lying in the set of interest given a direction. In order to do this, we restrict ourselves to starshaped sets B. This covers a wide

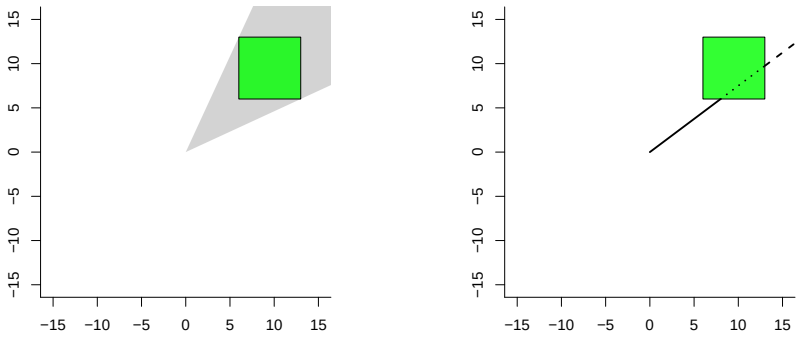

Figure 4.4: Illustration of the proposed probability estimation technique for the set $B = [6, 13] \times [6, 13]$ (in green). Left: the region $\mathbb{R}_+ \times S_B$ in grey. Right: The lines $I_{\text{inf}}(\boldsymbol{w})\boldsymbol{w}$ (solid line), $I_B(\boldsymbol{w})\boldsymbol{w}$ (dotted line), and $I_{\text{sup}}(\boldsymbol{w})\boldsymbol{w}$ (dashed line) at a fixed direction $\boldsymbol{w} = (0.8, 0.6)^\top \in S_B$.

variety of cases, as common types of sets of interest are boxes $\{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{a} \leq \boldsymbol{x} \leq \boldsymbol{b},\ \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d\}$ and sets of the form $\{r\boldsymbol{w} : r > h(\boldsymbol{w}) > 0, r \in (0, \infty), \boldsymbol{w} \in \mathbb{S}^{d-1}\}$ for some positive function h defined on $\mathbb{S}^{d-1}$, both starshaped at $\mathbf{0}$. This is useful in our setting as it allows for an exact probability calculation with respect to our model specification for exceedances.

For a set $B \in \mathbb{R}^d \setminus \{\mathbf{0}\}$ starshaped at $\mathbf{0}$, define $S_B := \{\boldsymbol{w} \in \mathbb{S}^{d-1} : r\boldsymbol{w} \in B, r \in (0, \infty]\} \subseteq \mathbb{S}^{d-1}$ and consider, for any $\boldsymbol{w} \in S_B$, the partition $I_{\text{inf}}(\boldsymbol{w}) \cup I_B(\boldsymbol{w}) \cup I_{\text{sup}}(\boldsymbol{w})$ of $(0, \infty)$, where $I_{\text{inf}}(\boldsymbol{w}) := (0, r_{B_{\text{inf}}}(\boldsymbol{w}))$, $I_B(\boldsymbol{w}) := [r_{B_{\text{inf}}}(\boldsymbol{w}), r_B(\boldsymbol{w})]$, and $I_{\text{sup}}(\boldsymbol{w}) := (r_B(\boldsymbol{w}), \infty)$ for the radial function $r_B(\boldsymbol{w})$ of B and the function $r_{B_{\text{inf}}}(\boldsymbol{w}) = \inf\{r > 0 : r\boldsymbol{w} \in B\}$. An illustration of these subsets is given by Figure 4.4.

The posterior predictive distribution of

$$\mathbb{P}_{B|\boldsymbol{x}_{1:n}} = \mathbb{P}[R \in I_B(\boldsymbol{W}), \boldsymbol{W} \in S_B \mid \boldsymbol{x}_{1:n}], \tag{4.16}$$

is given from the posterior density of $\boldsymbol{\theta}$ via

$$\mathbb{P}_{B|\boldsymbol{x}_{1:n}} = \int_{\mathbb{R}^{3p+3}} \mathbb{P}[R \in I_B(\boldsymbol{W}), \boldsymbol{W} \in S_B \mid \boldsymbol{\theta}] \pi[\boldsymbol{\theta} \mid \boldsymbol{x}_{1:n}]\, d\boldsymbol{\theta}, \tag{4.17}$$

where $\pi[\boldsymbol{\theta} \mid \boldsymbol{x}_{1:n}]$ is obtained following the procedure described in Section 4.3.4. Since $\mathcal{Q}_q$ is a latent variable, $r_{\mathcal{Q}_q}(\boldsymbol{w})$ intersects with $I_{\inf}(\boldsymbol{w})$, $I_B(\boldsymbol{w})$, or $I_{\sup}(\boldsymbol{w})$ with non-zero probability for all $\boldsymbol{w} \in S_B$. Hence, from total probability, $\mathbb{P}_{B|\boldsymbol{\theta}} := \mathbb{P}[R \in I_B(\boldsymbol{W}), \boldsymbol{W} \in S_B \mid \boldsymbol{\theta}]$ of (4.17) satisfies

$$\mathbb{P}_{B|\boldsymbol{\theta}} = \mathbb{P}[R\boldsymbol{W} \in B, R > r_{\mathcal{Q}_q}(\boldsymbol{W}) \mid \boldsymbol{\theta}] + \mathbb{P}[R\boldsymbol{W} \in B, R \leq r_{\mathcal{Q}_q}(\boldsymbol{W}) \mid \boldsymbol{\theta}], \quad (4.18)$$

and our model is best suited for sets B such that the second term in the equation (4.18) is small. The first term in equation (4.18) decomposes into

$$\begin{aligned}\mathbb{P}[R\boldsymbol{W} \in B, R > r_{\mathcal{Q}_q}(\boldsymbol{W}) \mid \boldsymbol{\theta}] = \mathbb{P}[R \in I_B(\boldsymbol{W}) \mid \boldsymbol{W} \in S_B, R > r_{\mathcal{Q}_q}(\boldsymbol{W}), \boldsymbol{\theta}] \times \\ \times \mathbb{P}[\boldsymbol{W} \in S_B \mid R > r_{\mathcal{Q}_q}(\boldsymbol{W}), \boldsymbol{\theta}] \mathbb{P}[R > r_{\mathcal{Q}_q}(\boldsymbol{W}) \mid \boldsymbol{\theta}]. \quad (4.19)\end{aligned}$$

By the assumptions detailed in Section 4.3.2, the last term $\mathbb{P}[R > r_{\mathcal{Q}_q}(\boldsymbol{W}) \mid \boldsymbol{\theta}]$ in equation (4.19) equals $1 - q$ for all $\boldsymbol{\theta} \sim \pi[\boldsymbol{\theta} \mid \boldsymbol{x}_{1:n}]$. Following our model formulation for $\boldsymbol{W}$, the second term in equation (4.19) corresponds to the predictive distribution of the angles given $\boldsymbol{\theta}$ and is given by

$$\begin{aligned}&\mathbb{P}[\boldsymbol{W} \in S_B \mid R > r_{\mathcal{Q}_q}(\boldsymbol{W}), \boldsymbol{\theta}] \\ &= \int_{S_B} \widetilde{f}_{\boldsymbol{W}|\boldsymbol{\theta}}(\boldsymbol{w} \mid \boldsymbol{\theta}) d\boldsymbol{w} \Big/ \int_{\mathbb{S}^{d-1}} \widetilde{f}_{\boldsymbol{W}|\boldsymbol{\theta}}(\boldsymbol{w} \mid \boldsymbol{\theta}) d\boldsymbol{w}, \quad , \quad (4.20)\end{aligned}$$

where $\widetilde{f}_{\boldsymbol{W}}$ is a directional density defined by M_1, M_2, or M_3 (see Section 4.2.2). The integrals in (4.20) are computed efficiently via numerical integration for $d = 2$ or 3. The first term in (4.19) is obtained through

$$\begin{aligned}&\mathbb{P}[R \in I_B(\boldsymbol{W}) \mid \boldsymbol{W} \in S_B, R > r_{\mathcal{Q}_q}(\boldsymbol{W}), \boldsymbol{\theta}] \\ &= \mathbb{E}_{\boldsymbol{W}|\boldsymbol{W} \in S_B, \boldsymbol{x}_{1:n}}(\mathbb{P}[R \in I_B(\boldsymbol{W}) \mid R > r_{\mathcal{Q}_q}(\boldsymbol{W}), \boldsymbol{\theta}]), \quad (4.21)\end{aligned}$$

where

$$\begin{aligned}&\mathbb{P}[R \in I_B(\boldsymbol{w}) \mid \boldsymbol{W} = \boldsymbol{w}, R > r_{\mathcal{Q}_q}(\boldsymbol{w}), \boldsymbol{\theta}] \\ &= F_{R|\boldsymbol{W}} \left[\frac{\max\left\{r_B(\boldsymbol{w}), r_{\mathcal{Q}_q}(\boldsymbol{w})\right\} - r_{\mathcal{Q}_q}(\boldsymbol{w})}{r_{\mathcal{G}_q}(\boldsymbol{w})} \middle| \boldsymbol{w} \right]\end{aligned}$$

$$
\begin{aligned}
&- F_{R|\boldsymbol{W}}\left[\frac{\max\left\{r_{B_{\inf}}(\boldsymbol{w}), r_{\mathcal{Q}_q}(\boldsymbol{w})\right\} - r_{\mathcal{Q}_q}(\boldsymbol{w})}{r_{\mathcal{G}_q}(\boldsymbol{w})}\middle|\boldsymbol{w}\right] \\
&= \exp\left\{-\left[\frac{\max\left\{r_{B_{\inf}}(\boldsymbol{w}), r_{\mathcal{Q}_q}(\boldsymbol{w})\right\} - r_{\mathcal{Q}_q}(\boldsymbol{w})}{r_{\mathcal{G}_q}(\boldsymbol{w})}\right]\right\} \\
&\quad - \exp\left\{-\left[\frac{\max\left\{r_{B}(\boldsymbol{w}), r_{\mathcal{Q}_q}(\boldsymbol{w})\right\} - r_{\mathcal{Q}_q}(\boldsymbol{w})}{r_{\mathcal{G}_q}(\boldsymbol{w})}\right]\right\}.
\end{aligned}
$$

We approximate expectation (4.21) through Monte-Carlo integration by sampling angles $\{\widetilde{\boldsymbol{w}}_1, \ldots, \widetilde{\boldsymbol{w}}_{n_w}\}$ from $f_{\boldsymbol{W}}$ restricted to the set S_B via

$$
n_w^{-1} \sum_{i=1}^{n_w} \mathbb{P}[R \in I_B(\widetilde{\boldsymbol{w}}_i) \mid R > r_{\mathcal{Q}_q}(\widetilde{\boldsymbol{w}}_i), \boldsymbol{\theta}]
$$

as $n_w \to \infty$. The second term in equation (4.18), $\mathbb{P}[R\boldsymbol{W} \in B, R \leq r_{\mathcal{Q}_q}(\boldsymbol{W}) \mid \boldsymbol{\theta}] =: p$, corresponds to a probability over a region in which our model is not specified. This region occurs where the data is more dense, and is therefore estimated empirically. For practical reasons, we restrict ourselves to sets B that are starshaped at with respect to their centrepoint. We acknowledge that this integration-based approach is less convenient if B is not star-shaped. For more general regions, a sampling-based approach akin to that of Wadsworth and Campbell (2024) could be used, but this is omitted in this work.

4.3.6 Return sets

A nice property of the rExp distribution is that one could derive *return sets*. Analogously to a univariate return level corresponding to a value expected to be exceeded by a random variable X once in a certain return period, return sets correspond to a specific set expected to be attained by a random vector $\boldsymbol{X}$ once in a certain return period T. For any Bernoulli experiment with probability $1 - T^{-1}$ of success, a return period T defines the expected number of experiments needed to obtain one success. Given a random vector $\boldsymbol{X} \in \mathbb{R}^d$ and an arbitrary set $B \in \mathbb{R}^d$, the event $\boldsymbol{X} \in B$ can be interpreted as a Bernoulli experiment with some return period T. Depending on the properties of $\boldsymbol{X}$, there can be infinitely many distinct sets in $\mathbb{R}^d$ satisfying this property.

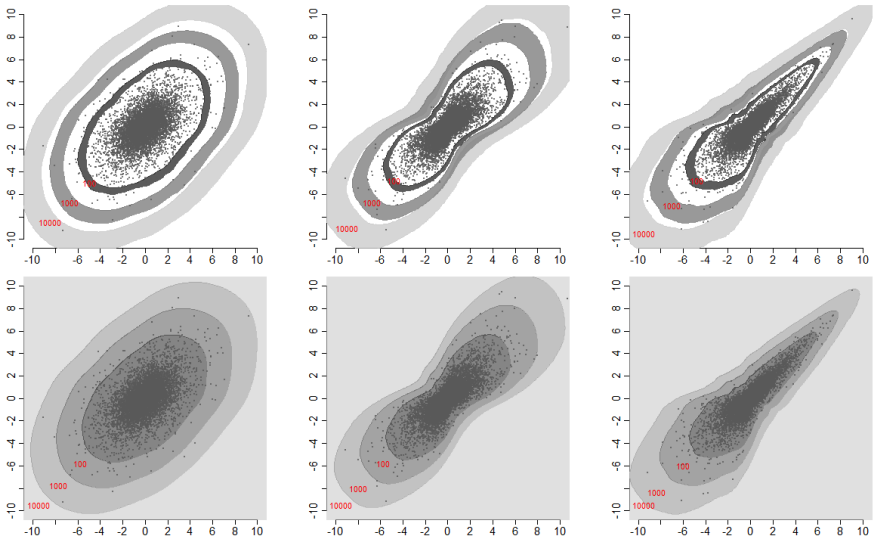

Figure 4.5: Top row: 0.95 prediction intervals for the boundaries $r_{1-T^{-1}}$ of the return set X_T with associated return period T displayed in red. Bottom row: Posterior mean of the return sets with associated return period (red) defined by the angle-wise posterior mean of $r_{1-T^{-1}}$. Every return set contains all lighter-grey sets. Columns from left to right: Bivariate Laplace, normal, and max-stable logistic distributions in Laplace margins.

In this work, given a return period T, we focus on sets of the form

$$\mathsf{X}_T = \left\{ \boldsymbol{x} \in \mathbb{R}^d \,:\, \|\boldsymbol{x}\| \geq r_{\mathcal{Q}_{1-T^{-1}}}(\boldsymbol{x}/\|\boldsymbol{x}\|) \right\},$$

where $r_{\mathcal{Q}_{1-T^{-1}}}(\boldsymbol{w})$ is the solution of

$$\Pr\left(R \geq r_{\mathcal{Q}_{1-T^{-1}}}(\boldsymbol{w}) \middle| \boldsymbol{W} = \boldsymbol{w} \right) = T^{-1}, \tag{4.22}$$

for $\boldsymbol{w} \in \mathbb{S}^{d-1}$. That is, $\mathsf{X}_T \subset \mathbb{R}^d$ is a set in which we expect to observe observations with probability $1/T$. As we model the rExp distribution above the q^{th} radial quantile, we consider return periods such that $T > (1-q)^{-1}$. By marginalising the rExp distribution from Definition 4.1, we assume rescaled exceedances conditioned on the direction follows a standard exponential distribution, i.e., $(R - r_{\mathcal{Q}_q}(\boldsymbol{w}))/r_{\mathcal{G}}(\boldsymbol{w}) \mid \{\boldsymbol{W} = \boldsymbol{w}, R > r_{\mathcal{Q}_q}(\boldsymbol{w})\} \sim \text{Exp}(1)$. Therefore, using the rExp distribution for radial exceedances and solving for $r_{\mathcal{Q}_{1-T^{-1}}}(\boldsymbol{w})$ in equation

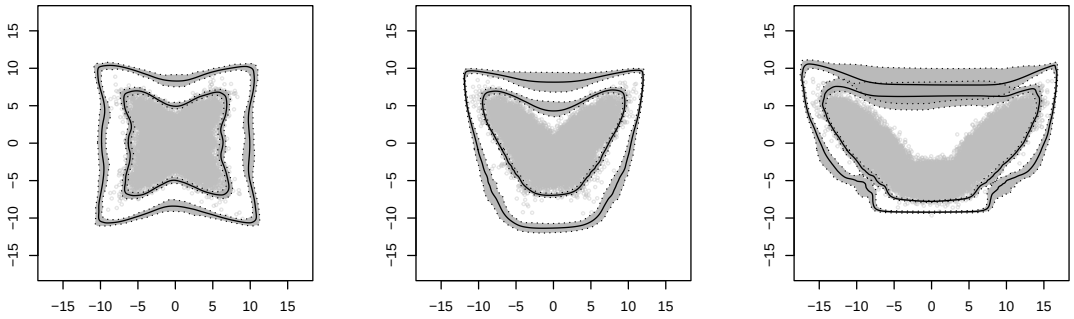

Figure 4.6: Posterior return set boundaries $\partial \mathsf{X}_T$ for return periods $T = 200$ (inner) and $T = 10^6$ (outer) for 3 different bivariate Gaussian mixture distributions. Grey points are the data in original margins, solid black lines indicate the posterior mean return set boundaries, with dark grey regions indicating the corresponding 0.95 prediction intervals, bordered by dotted lines.

(4.22) yields

$$r_{\mathcal{Q}_{1-T^{-1}}}(\boldsymbol{w}) = r_{\mathcal{Q}_q}(\boldsymbol{w}) + r_{\mathcal{G}}(\boldsymbol{w}) \log((1-q)T)$$

and the T return set is

$$\mathsf{X}_T = \left\{ \boldsymbol{x} \in \mathbb{R}^d \,:\, \boldsymbol{x} = \left[r_{\mathcal{Q}_q}(\boldsymbol{w}) + r_{\mathcal{G}}(\boldsymbol{w}) \log((1-q)T) \right] \boldsymbol{w}, \boldsymbol{w} \in \mathbb{S}^{d-1} \right\}.$$

In the star body notation, this is equivalent to

$$\mathsf{X}_T = [\mathcal{Q}_q + \mathcal{G} \cdot B_{\{\log(T(1-q))\}}(\mathbf{0})]'.$$

In practice, we obtain return sets X_T given data by first fitting the rExp distribution in the Bayesian semi-parametric manner described in Section 4.3.4. Then, by sampling the functions $r_{\mathcal{Q}_q,i}$ and $r_{\mathcal{G},i}$, $i = 1, \ldots, n_{\boldsymbol{\theta}}$ from the posterior distribution of $\boldsymbol{\theta}$, we obtain posterior mean estimates for X_T along with the associated prediction intervals. Figure 4.5 displays return sets return sets of period $T \in \{100, 1000, 10000\}$ estimated using samples drawn from the bivariate Laplace, Gaussian, and max-stable logistic distributions in standard Laplace margins.

As an illustrative example, we recreate the task of return set boundary estimation in the case of bivariate Gaussian mixture data presented in Section 4 of Hallin et al. (2021). Our approach differs in that it is non-empirical, and we are able to extrapolate to higher levels than observed in the data with uncertainty computed for our estimates. This is exemplified in Figure 4.6, in which the return level curves $\partial \mathsf{X}_T$ are shown for return periods $T = 200$ and $T = 10^6$. For each mixture distribution, we generate $n = 50,000$ datapoints and fit a quantile regression model with threshold at the $q = 0.9$ level. For Bayesian inference, we used the M_3 hierarchical structure. For display purposes, 20 posterior observations of $r_{\mathcal{Q}_q}$ were drawn and a joint radial-directional model was fit at each of these observations. There were 50 draws of posterior return level curves drawn for each of these models for a total of 1,000 return level curves for which to obtain posterior mean and credible intervals for each plot in Figure 4.6.

4.3.7 Model selection and validation

Our model formulation gives rise to various modelling choices needing to be assessed and validated. For instance, as discussed in Section 4.3.3, selecting the most appropriate model within the set of candidate models M_1, M_2, and M_3 given observed data amounts to analysing the properties of the distribution of $(R, \boldsymbol{W}) \mid R > r_{\mathcal{Q}_q}(\boldsymbol{W})$ with respect to the association between $g_{\mathcal{G}}$ and $f_{\boldsymbol{W}}$. Another modelling choice requiring assessment is that of the set of angles contributing to likelihood (4.12) since it translates into a bias-variance trade-off for $\mathcal{G}$ and $\mathcal{L}$. Further, hyper-parameters values imposed on the prior distributions of $\mathcal{Q}_q$, $\mathcal{G}$, and $\mathcal{L}$ imply *a priori* information on the differentiability properties of their boundaries (see Section 4.3.4), and their impact on the posterior may need assessment. Finally, the usual concern of the sensitivity of the posterior distribution of the latent variables to the return period for the latent threshold function $r_{\mathcal{Q}_q}$ (Section 4.3.2) remains. We detail possible methods for model selection and validation below.

The nested structure of model M_2 within the parameter space of model M_3 raises

the question of whether model M_2 can serve as a sensible simplification of model M_3 given observed data $\boldsymbol{x}_{1:n} = \{r_1\boldsymbol{w}_1, \ldots, r_n\boldsymbol{w}_n\}$. As discussed in Section 4.3.3, evidence for a constant $r_{\mathcal{B}}$ in model M_3 points to evidence that M_2 can serve as an sensible simplification of M_3, or equivalently that the observed data $\boldsymbol{x}_{1:n}$ may come from a process following a homothetic density with respect to its gauge function $g_{\mathcal{G}}$. The sensitivity of the posterior distributions of the latent parameters to the remaining modelling choices can be assessed through the quality of the calibration of the posterior predictive distribution to the observed data. Given a set of observed exceedances $\mathcal{X}_i$ of a sampled latent function $r_{\mathcal{Q}_q,i}$ (see expression (4.14)), we wish to assess the calibration of the predictive distribution for the excess variable $R - r_{\mathcal{Q}_q,i}(\boldsymbol{W}) \mid R > r_{\mathcal{Q}_q,i}(\boldsymbol{W})$. We thus compare each observed excess $r_j - r_{\mathcal{Q}_q,i}(\boldsymbol{w}_j)$ with the quantile of the Exponential$(r_{\mathcal{G}}(\boldsymbol{w}_j))$, with $\log r_{\mathcal{G}}(\boldsymbol{w}) = \beta_{\mathcal{G}} + \boldsymbol{\psi}(\boldsymbol{w})^\top \boldsymbol{z}_{\mathcal{G}}$ and $(\beta_{\mathcal{G}}, \boldsymbol{z}_{\mathcal{G}}) \sim \pi[\beta_{\mathcal{G}}, \boldsymbol{z}_{\mathcal{G}} \mid \beta_{\mathcal{Q}_q,i}, \boldsymbol{z}_{\mathcal{Q}_q,i}, \boldsymbol{x}_{1:n}]$. We draw exceedance radii $\{\widetilde{r}_1, \ldots, \widetilde{r}_{n_s}\}$ from their predictive distribution $\mathrm{Exp}(r_{\mathcal{G}}(\boldsymbol{w}_j))$, and define the empirical distribution function $F_{\boldsymbol{w}_j,n_s}(r) := {n_s}^{-1} \sum_{k=1}^{n_s} \mathbb{1}[\widetilde{r}_k \leq r]$, for $r \in (0, \infty)$. A probability-probability (PP) plot for the exceedances $\mathcal{X}_i$ is then given by

$$\left\{ \left(\frac{j}{|\mathcal{X}_i| + 1}, p_{(j)} \right) : j = 1, \ldots, |\mathcal{X}_i| \right\}, \tag{4.23}$$

where $p_{(j)}$ denotes the j-th order statistic of the sample $\{F_{\boldsymbol{w}_j,n_s}(r_j - r_{\mathcal{Q}_q,i}(\boldsymbol{w}_j)) : j = 1, \ldots, |\mathcal{X}_i|\}$. A quantile-quantile (QQ) plot in unit exponential margins for the exceedances $\mathcal{X}_i$ is then easily obtained via probability-probability (PP) plot for the exceedances $\mathcal{X}_i$ via

$$\left\{ \left(-\log\left(1 - \frac{j}{|\mathcal{X}_i| + 1}\right), -\log\left(1 - p_{(j)}\right) \right) : j = 1, \ldots, |\mathcal{X}_i| \right\}. \tag{4.24}$$

In a similar fashion, we can obtain PP plots for the predictive distribution of $\boldsymbol{W} \mid R > r_{\mathcal{Q}_q,i}(\boldsymbol{W})$. This is achieved by sampling angles $\{\widetilde{\boldsymbol{w}}_1, \ldots, \widetilde{\boldsymbol{w}}_{n_s}\}$ from the predictive density $f_{\boldsymbol{W}|\beta_{\mathcal{G}},\boldsymbol{z}_{\mathcal{G}},\beta_{\mathcal{B}},\boldsymbol{z}_{\mathcal{B}},\boldsymbol{x}_{1:n}}$ where $(\beta_{\mathcal{G}}, \boldsymbol{z}_{\mathcal{G}}, \beta_{\mathcal{B}}, \boldsymbol{z}_{\mathcal{B}}) \sim \pi[\beta_{\mathcal{G}}, \boldsymbol{z}_{\mathcal{G}}, \beta_{\mathcal{B}}, \boldsymbol{z}_{\mathcal{B}} \mid \beta_{\mathcal{Q}_q,i}, \boldsymbol{z}_{\mathcal{Q},i}, \boldsymbol{x}_{1:n}]$. We transform the sampled and the observed directions from $\mathcal{X}$ to spherical coordinates (or "angles"), respectively denoted $\{\widetilde{\boldsymbol{\varphi}}_j = (\widetilde{\varphi}_{1,j}, \ldots, \widetilde{\varphi}_{d-1,j}) \in \Phi : j = 1, \ldots, n_s\}$ and $\{\boldsymbol{\varphi}_j = (\varphi_{1,j}, \ldots, \varphi_{d-1,j}) \in \Phi : j = 1, \ldots, |\mathcal{X}_i|\}$. Using these,

we consider the empirical distribution function $F_{n_s}(\boldsymbol{\varphi}) := {n_s}^{-1} \sum_{k=1}^{n_s} \mathbb{1}[\widetilde{\boldsymbol{\varphi}}_k \leq \boldsymbol{\varphi}]$. A PP plot for the calibration of the predictive distribution of angles to the observed angles in $\mathcal{X}_i$ is then given by expression (4.23) with $p_{(j)}$ denoting the j-th order statistic of the sample $\{F_{n_s}(\boldsymbol{\varphi}_j) : j = 1, \ldots, |\mathcal{X}_i|\}$. Note that in dimensions $d > 2$, we cannot obtain QQ plots corresponding to these directional PP plots, as there is no obvious inverse of the multivariate empirical distribution function F_{n_s}.

In addition to assessing the agreement of the fitted posterior with the underlying model via PP and QQ plots, it is of interest to see how well the posterior model approximates extreme probabilities. To do this, we estimate values of the coefficient of tail dependence, $\chi_q(A)$, defined in (4.1), on models fitted with data drawn from distributions of dimensions $d = 2$ and 3. These values are compared to their empirical values. For a value of q close to 1, estimating $\chi_q(A)$ amounts to estimating the probability of lying in the region

$$B_q := \{\boldsymbol{x} \in \mathbb{R}^d \, : \, x_j \in (-\log(2(1-q)), \infty) \text{ if } j \in A, \, x_j \in (-\infty, \infty) \text{ if } j \notin A\}.$$

Using posterior models fitted on data, the posterior probability estimation procedure presented in Section 4.3.5 is used with region B_q, we are able to obtain an estimate $\widehat{\mathbb{P}}(\boldsymbol{X} \in B_q)$. When dividing by $(1-q)$, a posterior estimate $\widehat{\chi}_q(A)$ for the coefficient of tail dependence is obtained. This is done on a fine grid of values $q \in [0.95, 1]$. Furthermore, posterior confidence intervals of the models' estimates are compared to bootstrap confidence intervals of the empirical estimates. In addition to being a diagnostic for how well posterior models can estimate large values, $\chi_q(A)$ estimates could also give insight on the tail dependence structure of the data's underlying distribution. For example, if $\widehat{\chi}_q(A) \to 0$ as $q \to 1$, then the posterior model concludes that the components in A do not achieve simultaneous extremes. If values of $\widehat{\chi}_q(A)$ agree with their empirical counterparts as $q \to 1$, it is a good indication that the posterior captures the tail behaviour when all d components of the random vector grow large.

4.4 River flow data analysis

The flexibility and accuracy of our Bayesian modelling approach is illustrated by performing inference on a bivariate dataset of river flow measurements (m^3/s) displaying a complicated dependence structure. First analysed in Keef et al. (2013a), we focus on daily measurements of the rivers Pang and Windrush, each tributaries of the River Thames in southern England, from 1968 to 2008. For the purposes of this data analysis, we make the simplifying assumption of temporal stationarity between daily measurements. Once standardised to Laplace margins, we use the data $\boldsymbol{x}_{1:n}$ in the hierarchical modelling approach outlined in Section 4.3. For the threshold model, we fit the hierarchical Bayesian gamma quantile regression model for the quantile set $\mathcal{Q}_q$ at the $q = 0.95$ probability level, and sample $n_{\mathcal{Q}} = 20$ posterior observations of the radial threshold, $r_{\mathcal{Q}_q,1}, \ldots, r_{\mathcal{Q}_q,20}$, from $\pi[\beta_{\mathcal{Q}}, \boldsymbol{z}_{\mathcal{Q}} | \boldsymbol{x}_{1:n}]$. The assumption of exponential radial excesses (rExp) is made, and models M_1, M_2, M_3 are fitted, each using all directions and exceedance directions. For each model and posterior threshold $r_{\mathcal{Q}_q,i}$, $n_{\mathcal{GL}} = 50$ samples are taken from the posterior distribution $\pi[\beta_{\mathcal{G}}, \boldsymbol{z}_{\mathcal{G}}, \beta_{\mathcal{L}}, \boldsymbol{z}_{\mathcal{L}} | \beta_{\mathcal{Q},i}, \boldsymbol{z}_{\mathcal{Q},i}, \boldsymbol{x}_{1:n}]$, resulting in $n_{\boldsymbol{\theta}} = 1000$. In Appendix 4.7.6, we compare observed posterior boundaries of the scaling set $\mathcal{G}$ obtained from $r_{\mathcal{G},1}, \ldots, r_{\mathcal{G},1000}$ to $\log(n/2)$-scaled data, and posterior angular densities for $\mathcal{W} = \mathcal{L}^d/(d|\mathcal{L}|)$ obtained from $r_{\mathcal{L},1}, \ldots, r_{\mathcal{L},1000}$ to their empirical counterparts. This is done for all six fitted models. Though this visual check, it is determined that model M_2 is preferred when fitted with exceedance directions only, and will henceforth be considered for this dataset.

Figure 4.7 displays a posterior mean boundary of the scaling set $\mathcal{G}$, or gauge function unit level set, that seems to accurately capture the behaviour of scaled data. This leads to the belief that we can accurately capture the tail dependence behaviour across the entire bivariate tail region of the data's underlying distribution. Also in Figure 4.7, we show our ability to extrapolate information far beyond the reach of observable data by displaying return sets in Laplace margins, with increasing uncertainty as the return period T increases, which is sensible. Lastly, Figure 4.7

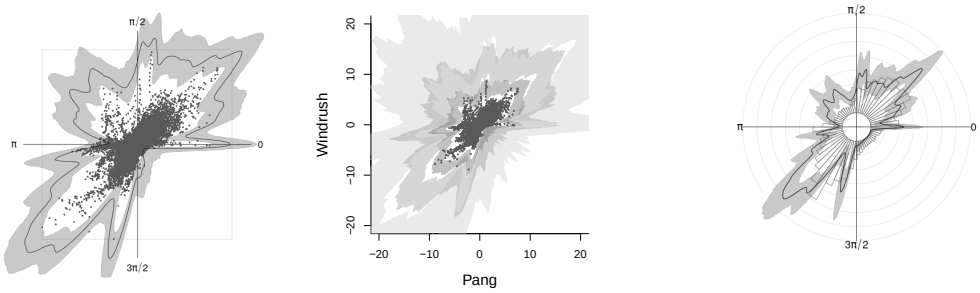

Figure 4.7: Fitted posterior star bodies on Thames tributary data. Left: Posterior mean boundary of the scaling set $\mathcal{G}$ (solid line) with 0.95 prediction intervals (in grey), with $\log(n/2)$-scaled data in Laplace margins. Centre: Prediction intervals for the boundary of river flow return sets. From dark to light grey: $T = 10^3$, 10^5, and $T = 10^9$ days for data in standard Laplace margins. Right: Posterior mean boundary of the directional distribution set $\mathcal{W}$ (solid line) with 0.95 prediction intervals (in grey), empirical estimate (histogram).

shows good agreement between the estimated posterior density of directions, $f_{\boldsymbol{W}}$, with the corresponding directional histogram. To check posterior model fit's predictive performance, values of χ_q, whose expression is presented in (4.1), at high levels q increasing to 1. In it, we see that the posterior model's probability estimates closely follows the empirical counterpart, with good coverage of confidence intervals. This shows the posterior model's good ability at capturing the joint tail behaviour of river flow measurements at these particular sites.

Additional posterior model goodness-of-fit diagnostics are presented in Appendix 4.7.6. The QQ plots in Figures 4.19 show that there is good agreement with the empirical and model-based estimates in their abilities to extrapolate exceedances to extreme values. Figure 4.20 shows PP plots for directions generated from the posterior distribution of $r_{\mathcal{W}}$ have good agreement with the empirical distribution of the directions.

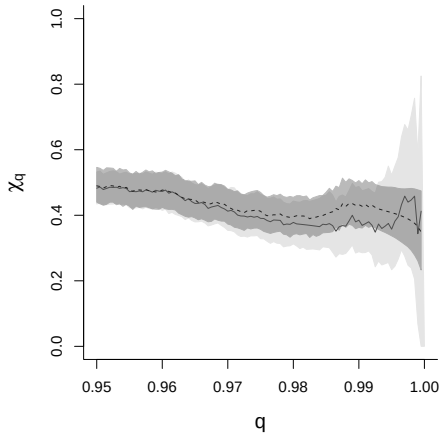

Figure 4.8: χ_q values for the pair of marginal variables corresponding to the rivers Pang and Windrush. Solid line is an empirical estimate with 95% bootstrap confidence intervals in light grey, and the dashed line is a model-based estimate with 95% posterior confidence intervals in dark grey.

4.5 Newlyn wave heights data analysis

We apply our methodology to a dataset of dimension $d = 3$ consisting of hourly measurements of wave height Y_H, in meters, wave period Y_P, in seconds, and surge Y_S, in meters, measured over the period 1971–1977 at the Newlyn port in south-west England. The dataset was first analysed in Coles and Tawn (1994) in the case where asymptotic dependence was assumed between all three variables. Although a typical assumption, Wadsworth and Campbell (2024) show that asymptotic independence is a more reasonable assumption. Here, we revisit this data with a more flexible approach also enabling the modelling of negative dependence. Following previous literature, we analyse componentwise maxima over 15-hour periods, resulting in a dataset of $n = 2,894$ observations. The margins Y_H, Y_P, Y_S of the data are unknown and are therefore standardised to standard Laplace using methods from Section 4.3.1. This results in observations $\boldsymbol{x}_{1:n} := \{\boldsymbol{x}_i = (x_{H,i}, x_{P,i}, x_{S,i})^\top : i = 1, \ldots, n\}$ interpreted as random draws from $\boldsymbol{X} = (X_H, X_P, X_S)$.

We begin by fitting the hierarchical Bayesian gamma quantile regression model

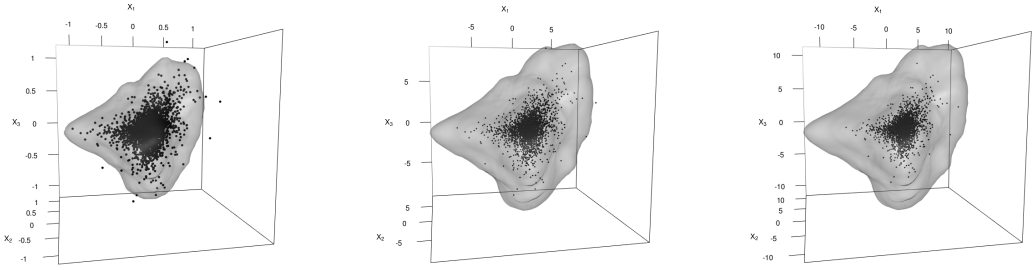

Figure 4.9: Newlyn wave data fitted posterior star bodies. Left: Posterior mean boundary of the scaling set $\mathcal{G}$, with $\log(n/2)$-scaled data in Laplace margins. Centre: Posterior mean return set with $T = 10^3$ and data in Laplace margins. Right: Posterior mean return set with $T = 10^4$ and data in Laplace margins. X_1, X_2, and X_3 refer to the wave height, period, and surge variables, respectively.

to $\boldsymbol{x}_{1:n}$ for the quantile set $\mathcal{Q}_q$ at the $q = 0.9$ probability level. For each of $n_{\mathcal{Q}_q} = 20$ samples $r_{\mathcal{Q}_q,1}, \ldots, r_{\mathcal{Q}_q,20}$ from the posterior distribution $\pi[\beta_{\mathcal{Q}_q}, \boldsymbol{z}_{\mathcal{Q}_q} | \boldsymbol{x}_{1:n}]$ of $\mathcal{Q}_q$. Using these posterior samples, we assume that the exceedances of $\mathcal{Q}_{q,i}$ follow an rExp distribution exactly, and fit the models M_1, M_2, and M_3 to them using both exceedance directions and all directions. and posterior threshold $r_{\mathcal{Q}_q,i}$, $n_{\mathcal{GL}} = 50$ samples are taken from the posterior distribution $\pi[\beta_{\mathcal{G}}, \boldsymbol{z}_{\mathcal{G}}, \beta_{\mathcal{L}}, \boldsymbol{z}_{\mathcal{L}} | \beta_{\mathcal{Q},i}, \boldsymbol{z}_{\mathcal{Q},i}, \boldsymbol{x}_{1:n}]$, resulting in $n_{\boldsymbol{\theta}} = 1000$. In Appendix 4.7.7, we compare observed posterior boundaries of the scaling set $\mathcal{G}$ obtained from $r_{\mathcal{G},1}, \ldots, r_{\mathcal{G},1000}$ to $\log(n/2)$-scaled data for all six fitted models. This visual check leads to the belief that model M_2 is preferred when fitted with exceedance directions only, and will henceforth be considered for this dataset. Figure 4.9 shows the posterior mean of the boundary of $\mathcal{G}$ obtained from model M_2, as well as the posterior means of the boundary of the return sets $\partial \mathsf{X}_T$ for $T = 10^3$ and 10^4. Though Figure 4.9, we see that the model captures the extremal dependence structure of our dataset well as observed through the correspondence between the posterior mean of $\mathcal{G}$ and the scaled sample cloud.

These diagnostics suggest a well-performing posterior model. To check its

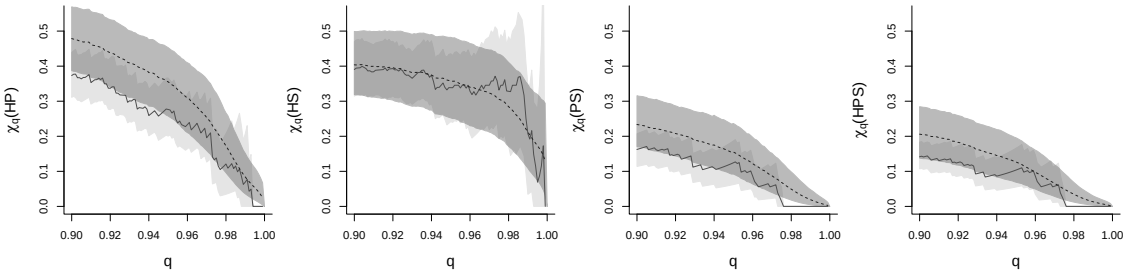

Figure 4.10: χ_q values for all pairwise combinations of marginal variables and for all three marginal variables of the Newlyn wave dataset. Solid line is an empirical estimate with 95% bootstrap confidence intervals in light grey, and the dashed line is a model-based estimate with 95% posterior confidence intervals in dark grey.

performance in predicting tail behaviour, $\chi_q(A)$ values are computed for large q approaching 1 for all combinations of variables $A \subseteq \{H, P, S\}$ of length at least 2 (pairs and the entire triplet). Figure 4.10 displays posterior mean estimates of $\chi_q(A)$ with their empirically-computed counterparts. Results show agreement between the estimates, with good coverage of confidence intervals. This suggests the ability of the posterior model fit in capturing the joint tail behaviour of the Newlyn data's underlying distribution.

Coles and Tawn (1994) introduce a structure variable $Q(v; \boldsymbol{Y})$, interpreted as the volume of water (in cubic meters, m^3) overtopping the sea-wall per unit length (in meters, m) over a fixed duration (in seconds, s), and measured in $m^3 s^{-1} m^{-1}$ for a sea-wall v meters in height. More precisely,

$$Q(v; \boldsymbol{Y}) = a_1 Y_S Y_P \exp\{a_2 (v - Y_S - l)/(Y_P Y_H^{*1/2})\}.$$

The wave height component Y_H^* is a calibration of the wave height marginal variable Y_H to approximate the off-shore wave height, since measurements are taken on-shore. We estimate the sea-wall height structure variable v_p (in meters, m) for which the discharge rate value $Q(v_p; \boldsymbol{Y})$ is expected to exceed the design standard of $0.002 \mathrm{m}^3 \mathrm{s}^{-1} \mathrm{m}^{-1}$ with probability p. Setting $V = Q^{-1}(0.002; \boldsymbol{Y})$, v_p is the solution to $\mathbb{P}(V > v_p) = p$. As in Bortot et al. (2000), we fix the sea-wall design feature constants to $a_1 = 0.25$, $a_2 = 26$, and tidal level relative to the seabed to $l = 4.3$.

We obtain three separate estimates for v_p at a range of p values. First, an empirical estimate of v_p is obtained by using the empirical quantile function on values of the structure variable V computed from the dataset. Second, a univariate generalised Pareto (GP) model is fitted using the dataset's computed values of V obtained from the dataset, and quantiles are then obtained for v_p. Finally, we compare these with prediction intervals for v_p obtained from our fitted model M_2. This consists of first sampling a collection of new datasets $\{\boldsymbol{Y}_{1:n,k} = \{\boldsymbol{Y}_{1,k}, \ldots, \boldsymbol{Y}_{n,k}\} : k = 1, \ldots, 200\}$, each comprising $n = 2,894$ observations,. These are used to collect a sample of wall heights $\{v_{p,k} : k = 1, \ldots, 200\}$ where each $v_{p,k}$ is the p-quantile of $\{V_{j,k} = Q^{-1}(0.002; \boldsymbol{Y}_{j,k}) : j = 1, \ldots, n\}$. For a sequence of $p \in (0.9, 1)$, we consider the p-wise mean and the 95% confidence intervals based on equal-tailed quantiles of the sample of sea wall heights.

To generate a new data set $\widetilde{\boldsymbol{Y}}_{1:n,k}$, we randomly sample one of the $n_{\mathcal{Q}_q}$ realisations from the posterior distribution $\pi[\beta_{\mathcal{Q}_q}, \boldsymbol{z}_{\mathcal{Q}} | \boldsymbol{x}_{1:n}]$, label it $r_{\mathcal{Q},k}$, as well as one of the $n_{\mathcal{G},\mathcal{L}} = 50$ realisations from the conditional posterior distribution of $\pi[\beta_{\mathcal{G}}, \boldsymbol{z}_{\mathcal{G}}, \beta_{\mathcal{L}}, \boldsymbol{z}_{\mathcal{L}} | \beta_{\mathcal{Q}_q}, \boldsymbol{z}_{\mathcal{Q},k}, \boldsymbol{x}_{1:n}]$, and label it $(r_{\mathcal{G},k}, r_{\mathcal{L},k})$. The new sample $\boldsymbol{X}_{1:n,k}$ is then constructed by sampling with replacement $\lceil q \cdot n \rceil$ observations from the original observations falling in $\mathcal{Q}_{q,k}$, and $\lceil (1-q) \cdot n \rceil$ observations from an rExp distribution with location $\mathcal{Q}_{q,k}$, scale $\mathcal{G}_k$, and directional shape $\mathcal{W}_k = \mathcal{L}_k^d / (d|\mathcal{L}_k|)$, before transforming the sample to original margins using the probability integral transform via the standard Laplace distribution function and then inverse transform via the inverse marginal model $\widehat{F}_j$ specified in Section 4.3.1. Figure 4.11 includes the resulting estimates of v_p plotted against $-\log(-\log(1-p))$ for a range of p values. Compared to the empirical and GP distribution fit approach, our method accurately estimates the sea-wall height variable v_p across all values of $p \in (0, 0.10)$. The larger prediction intervals corresponding to our method can be attributed to our more holistic account of uncertainty via a joint model for extreme events, in contrast with a structured variable approach. Our method hence reveals that risk may have been underestimated by such previous methods.

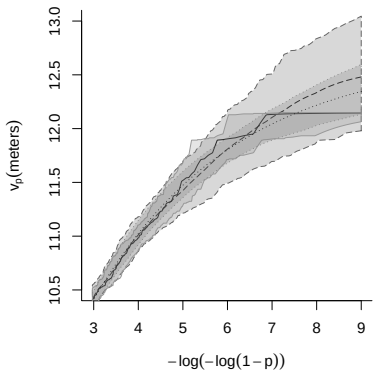

Figure 4.11: Estimated return levels for sea-wall height. Presented are results from empirical fits (solid black line), GP (dotted black line), and our semi-parametric method (dashed black line). Grey regions correspond to 95% confidence intervals.

Additional posterior model goodness-of-fit diagnostics are presented in Appendix 4.7.7. The QQ plots in Figures 4.22 show that there is good agreement with the empirical and model-based estimates in their abilities to extrapolate exceedances to extreme values. Figure 4.23 shows PP plots for directions generated from the posterior distribution of directions $\mathcal{W}$ have good agreement with the empirical counterpart.

4.6 Discussion

This work introduces a framework for defining quantile sets and return sets in multivariate extreme value analysis, emphasizing the role of these sets in capturing the geometry of extreme events. Central to this framework is the identification of return sets under a radial-directional decomposition, which describes exceedances across all directions and provides a natural representation of the geometry of a sample cloud. Importantly, in our framework these return sets arise through the use of normalising functions that lead to the weak convergence of radially renormalised sample clouds to a novel Poisson point process. This builds upon recent results on the convergence of scaled sample clouds and allowing for extrapolation of return

sets, with Bayesian inference allowing for the estimation of uncertainty.

In appendices, we show that this point process convergence on exceedance radii holds for a number of multivariate distributions whose margins are pre-specified. This wide applicability is further demonstrated through the modelling of real-world datasets. Using the assumed limiting distributions, we are able to accurately characterise the multivariate tail of these datasets. While the theory on radial exponential exceedances presented here holds for any dimension $d \geq 2$, our Bayesian semiparametric inference procedure is currently only supported for data of at most three variables. This is due to the difficulty in obtaining a sparse precision matrix for the stochastic partial differential equation approximation of Matérn Gaussian fields on $\mathbb{S}^{d-1}$ for $d > 3$. Mainly, on these hyperspheres, it is left to future work to triangulate the space and define appropriate bases functions. Once this is done, then work can begin on updating the methods shown in this chapter to higher dimensions.

In parallel independent work, Simpson and Tawn (2024b) introduced methodology for obtaining environmental contours, which are similar to our notion of return sets, further demonstrating the utility of the geometric approach for representing risk. The recent developments show the broad applicability of return sets in the geometric framework, as evidenced by their use in recent works, including Mackay and Jonathan (2024), Murphy-Barltrop et al. (2024a), Mackay et al. (2024), Campbell and Wadsworth (2024), and De Monte et al. (2025). This emerging research highlights promising directions and the potential for significant impact in risk assessment and the communication of risk.

Acknowledgments

We thank Jennifer L. Wadsworth for insightful discussions on the framework of geometric extremes, and Finn Lindgren for guidance with the `inlabru` and `excursions` packages in the `R` computing language.

4.7 Appendix

4.7.1 Density convergence of rescaled radial excess variable

Proof of Proposition 4.1. Begin by noticing that

$$\begin{aligned}
&\mathbb{P}\left[\frac{R - r_{\mathcal{Q}_q}(\boldsymbol{w})}{r_{\mathcal{G}_q}(\boldsymbol{w})} \leq z \;\middle|\; R > r_{\mathcal{Q}_q}(\boldsymbol{W}), \boldsymbol{W} = \boldsymbol{w}\right] \\
&= \frac{\mathbb{P}\left[\frac{R - r_{\mathcal{Q}_q}(\boldsymbol{w})}{r_{\mathcal{G}_q}(\boldsymbol{w})} \leq z, R > r_{\mathcal{Q}_q}(\boldsymbol{w}), \boldsymbol{W} = \boldsymbol{w}\right]}{\mathbb{P}\left[R > r_{\mathcal{Q}_q}(\boldsymbol{w}), \boldsymbol{W} = \boldsymbol{w}\right]} \\
&= \frac{\mathbb{P}\left[R \leq r_{\mathcal{Q}_q}(\boldsymbol{w}) + r_{\mathcal{G}_q}(\boldsymbol{w})z, R > r_{\mathcal{Q}_q}(\boldsymbol{w}), \boldsymbol{W} = \boldsymbol{w}\right]}{\int_{r_{\mathcal{Q}_q}(\boldsymbol{w})}^{\infty} f_{R,\boldsymbol{W}}(s, \boldsymbol{w}) ds} \\
&= \frac{\mathbb{P}\left[R \leq z_q(\boldsymbol{w}), \boldsymbol{W} = \boldsymbol{w}\right] - \mathbb{P}\left[R \leq r_{\mathcal{Q}_q}(\boldsymbol{w}), \boldsymbol{W} = \boldsymbol{w}\right]}{\int_{r_{\mathcal{Q}_q}(\boldsymbol{w})}^{\infty} f_{R,\boldsymbol{W}}(s, \boldsymbol{w}) ds}
\end{aligned}$$

where $z_q(\boldsymbol{w}) = r_{\mathcal{Q}_q}(\boldsymbol{w}) + r_{\mathcal{G}_q}(\boldsymbol{w})z$ and $\boldsymbol{w} \in \mathbb{S}^{d-1}$. Taking the partial derivative with respect to z, obtain the density of radial exceedances at a given direction,

$$\begin{aligned}
&\frac{\partial}{\partial z}\mathbb{P}\left[\frac{R - r_{\mathcal{Q}_q}(\boldsymbol{w})}{r_{\mathcal{G}_q}(\boldsymbol{w})} \leq z \;\middle|\; R > r_{\mathcal{Q}_q}(\boldsymbol{W}), \boldsymbol{W} = \boldsymbol{w}\right] \\
&= \frac{r_{\mathcal{G}_q}(\boldsymbol{w}) f_{R,\boldsymbol{W}}(z_q(\boldsymbol{w}), \boldsymbol{w})}{\int_{r_{\mathcal{Q}_q}(\boldsymbol{w})}^{\infty} f_{R,\boldsymbol{W}}(s, \boldsymbol{w}) \, \mathrm{d}s} \\
&= \frac{r_{\mathcal{G}_q}(\boldsymbol{w}) z_q(\boldsymbol{w})^{d-1} f_{\boldsymbol{X}}(z_q(\boldsymbol{w})\boldsymbol{w})}{\int_{r_{\mathcal{Q}_q}(\boldsymbol{w})}^{\infty} f_{R,\boldsymbol{W}}(s, \boldsymbol{w}) \, \mathrm{d}s}.
\end{aligned} \tag{4.25}$$

Rewrite the integral in the denominator of (4.25) as

$$\begin{aligned}
\int_{r_{\mathcal{Q}_q}(\boldsymbol{w})}^{\infty} f_{R,\boldsymbol{W}}(s, \boldsymbol{w}) \, \mathrm{d}s &= \int_{r_{\mathcal{Q}_q}(\boldsymbol{w})}^{\infty} s^{d-1} f_{\boldsymbol{X}}(s\boldsymbol{w}) \, \mathrm{d}s \\
&= \int_{r_{\mathcal{Q}_q}(\boldsymbol{w})}^{\infty} s^{d-1} \exp[-\psi(s)\{-\{\log f_{\boldsymbol{X}}(s\boldsymbol{w})\}/\psi(s)\}]\mathrm{d}s \\
&\sim \int_{r_{\mathcal{Q}_q}(\boldsymbol{w})}^{\infty} s^{d-1} \exp[-\psi(s) g_{\mathcal{G}}(\boldsymbol{w})^{\rho}]\mathrm{d}s.
\end{aligned}$$

as $q \to 1$, where the asymptotic equivalence holds by the uniform convergence assumption in (4.3). Without loss, ψ can be taken smooth (Lemma 1.4, Seneta (1976)) so that $d^k\psi(t)/dt^k$ exists for all $t > 0$ and $k \in \mathbb{N}$. Using this,

$$\int_{r_{\mathcal{Q}_q}(\boldsymbol{w})}^{\infty} s^{d-1} \exp[-\psi(s) g_{\mathcal{G}}(\boldsymbol{w})^{\rho}]\mathrm{d}s = \int_{r_{\mathcal{Q}_q}(\boldsymbol{w})}^{\infty} s^{d-1} \left[\frac{\frac{d}{ds}\exp[-\psi(s) g_{\mathcal{G}}(\boldsymbol{w})^{\rho}]}{-g_{\mathcal{G}}(\boldsymbol{w})^{\rho}\psi'(s)}\right] \mathrm{d}s$$

$$\sim r_{\mathcal{Q}_q}(\boldsymbol{w})^{d-1}\left[\frac{\exp[-\psi(r_{\mathcal{Q}_q}(\boldsymbol{w}))g_{\mathcal{G}}(\boldsymbol{w})^\rho]}{g_{\mathcal{G}}(\boldsymbol{w})^\rho\psi'(r_{\mathcal{Q}_q}(\boldsymbol{w}))}\right]$$

as $q \to 1$, where the asymptotic equivalence holds by recursively performing integration by parts. Putting this limiting result into (4.25), obtain

$$\begin{aligned}
&\frac{\partial}{\partial z}\,\mathbb{P}\left[\frac{R - r_{\mathcal{Q}_q}(\boldsymbol{w})}{r_{\mathcal{G}_q}(\boldsymbol{w})} \le z \;\middle|\; R > r_{\mathcal{Q}_q}(\boldsymbol{W}), \boldsymbol{W} = \boldsymbol{w}\right] \\
&\sim \frac{r_{\mathcal{G}_q}(\boldsymbol{w})z_q(\boldsymbol{w})^{d-1}}{r_{\mathcal{Q}_q}(\boldsymbol{w})^{d-1}} f_{\boldsymbol{X}}(z_q(\boldsymbol{w})\boldsymbol{w}) g_{\mathcal{G}}(\boldsymbol{w})^\rho \psi'(r_{\mathcal{Q}_q}(\boldsymbol{w}))\exp[\psi(r_{\mathcal{Q}_q}(\boldsymbol{w}))g_{\mathcal{G}}(\boldsymbol{w})^\rho] \\
&\sim \frac{r_{\mathcal{G}_q}(\boldsymbol{w})z_q(\boldsymbol{w})^{d-1}}{r_{\mathcal{Q}_q}(\boldsymbol{w})^{d-1}} g_{\mathcal{G}}(\boldsymbol{w})^\rho \psi'(r_{\mathcal{Q}_q}(\boldsymbol{w}))\exp[-g_{\mathcal{G}}(\boldsymbol{w})^\rho\{\psi(z_q(\boldsymbol{w})) - \psi(r_{\mathcal{Q}_q}(\boldsymbol{w}))\}] \\
&\sim r_{\mathcal{G}_q}(\boldsymbol{w}) g_{\mathcal{G}}(\boldsymbol{w})^\rho \psi'(r_{\mathcal{Q}_q}(\boldsymbol{w}))\exp[-g_{\mathcal{G}}(\boldsymbol{w})^\rho\{\psi(z_q(\boldsymbol{w})) - \psi(r_{\mathcal{Q}_q}(\boldsymbol{w}))\}]
\end{aligned}$$

as $q \to 1$, where the third line holds by the convergence assumption (4.3) of $f_{\boldsymbol{X}}$ and the fourth line holds by factoring out $r_{\mathcal{Q}_q}(\boldsymbol{w})$ from $z_q(\boldsymbol{w})^{d-1} := [r_{\mathcal{Q}_q}(\boldsymbol{w}) + r_{\mathcal{G}_q}(\boldsymbol{w})z]^{d-1}$. A Taylor expansion of $\psi(z_q(\boldsymbol{w}))$ about $r_{\mathcal{Q}_q}(\boldsymbol{w})$ and local uniform convergence give $\psi(z_q(\boldsymbol{w})) = \psi(r_{\mathcal{Q}_q}(\boldsymbol{w})) + \psi'(r_{\mathcal{Q}_q}(\boldsymbol{w}))r_{\mathcal{G}_q}(\boldsymbol{w})z + o(1)$ as $q \to 1$. Thus,

$$\begin{aligned}
&\frac{\partial}{\partial z}\,\mathbb{P}\left[\frac{R - r_{\mathcal{Q}_q}(\boldsymbol{w})}{r_{\mathcal{G}_q}(\boldsymbol{w})} \le z \;\middle|\; R > r_{\mathcal{Q}_q}(\boldsymbol{W}), \boldsymbol{W} = \boldsymbol{w}\right] \\
&\sim r_{\mathcal{G}_q}(\boldsymbol{w}) g_{\mathcal{G}}(\boldsymbol{w})^\rho \psi'(r_{\mathcal{Q}_q}(\boldsymbol{w}))\exp[-g_{\mathcal{G}}(\boldsymbol{w})^\rho\{\psi'(r_{\mathcal{Q}_q}(\boldsymbol{w}))r_{\mathcal{G}_q}(\boldsymbol{w})z + o(1)\}] \\
&\sim r_{\mathcal{G}_q}(\boldsymbol{w}) g_{\mathcal{G}}(\boldsymbol{w})^\rho \psi'(r_{\mathcal{Q}_q}(\boldsymbol{w}))\exp[-g_{\mathcal{G}}(\boldsymbol{w})^\rho\psi'(r_{\mathcal{Q}_q}(\boldsymbol{w}))r_{\mathcal{G}_q}(\boldsymbol{w})z][1 + o(1)] \\
&\to \exp(-z),
\end{aligned}$$

as $q \to 1$ whenever $r_{\mathcal{G}_q}(\boldsymbol{w}) \sim \{\psi'(r_{\mathcal{Q}_q}(\boldsymbol{w}))g_{\mathcal{G}}(\boldsymbol{w})^\rho\}^{-1}$. Without loss, set $r_{\mathcal{G}_q}(\boldsymbol{w}) := \{\psi'(r_{\mathcal{Q}_q}(\boldsymbol{w}))g_{\mathcal{G}}(\boldsymbol{w})^\rho\}^{-1}$ throughout.

Suppose we have the interval $[a, b]$, $0 < a < b < \infty$, and arbitrary set $S \subset \mathbb{S}^{d-1}$. Using the normalising equations $r_{\mathcal{Q}}$ and $r_{\mathcal{G}}$, the point process of recentred and rescaled radii and of directions P_n has mean measure

$$n \Pr\left[\frac{R - r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{W})}{r_{\mathcal{G}}(\boldsymbol{W})} \in [a, b], R > r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{W}), \boldsymbol{W} \in S\right]$$

$$= n \int_S \int_a^b \Pr\left[\frac{R - r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{W})}{r_{\mathcal{G}}(\boldsymbol{W})} = z, R > r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{W}), \boldsymbol{W} = \boldsymbol{w}\right] dz d\boldsymbol{w}$$

Bringing n inside the integral, the intensity measure of P_n defined in equation (4.2) is therefore,

$$\begin{aligned}
&n \Pr\left[\frac{R - r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{W})}{r_{\mathcal{G}}(\boldsymbol{W})} = z, \boldsymbol{W} = \boldsymbol{w}, R > r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{W})\right] dz d\boldsymbol{w} \\
&= n \Pr\left[\frac{R - r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{w})}{r_{\mathcal{G}}(\boldsymbol{w})} = z \middle| \boldsymbol{W} = \boldsymbol{w}, R > r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{W})\right] \Pr\left[\boldsymbol{W} = \boldsymbol{w} \middle| R > r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{W})\right] \\
&\quad \times \Pr\left[R > r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{W})\right] dz d\boldsymbol{w} \\
&= \Pr\left[\frac{R - r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{w})}{r_{\mathcal{G}}(\boldsymbol{w})} = z \middle| \boldsymbol{W} = \boldsymbol{w}, R > r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{W})\right] \\
&\quad \times \Pr\left[\boldsymbol{W} = \boldsymbol{w} \middle| R > r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{W})\right] dz d\boldsymbol{w},
\end{aligned}$$

where the last equality holds because $r_{\mathcal{Q}_{1-n^{-1}}}$ is defined such that $\Pr\left[R > r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{W})\right] = n^{-1}$. A consequence of Proposition 5 of Papastathopoulos et al. (2025) is that $\Pr[\boldsymbol{W} = \boldsymbol{w} \mid R > r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{W})] = \Pr[\boldsymbol{W} = \boldsymbol{w}]$. With this in mind, the intensity measure becomes

$$\begin{aligned}
&\Pr\left[\frac{R - r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{w})}{r_{\mathcal{G}}(\boldsymbol{w})} = z \middle| \boldsymbol{W} = \boldsymbol{w}, R > r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{W})\right] \Pr\left[\boldsymbol{W} = \boldsymbol{w}\right] dz d\boldsymbol{w} \\
&= \Pr\left[\frac{R - r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{w})}{r_{\mathcal{G}}(\boldsymbol{w})} = z \middle| \boldsymbol{W} = \boldsymbol{w}, R > r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{W})\right] dz \mathbb{P}_{\boldsymbol{W}}(d\boldsymbol{w}) \\
&= \frac{\partial}{\partial z} \Pr\left[\frac{R - r_{\mathcal{Q}_{1-n^{-1}}}(\boldsymbol{w})}{r_{\mathcal{G}}(\boldsymbol{w})} \leq z \middle| \boldsymbol{W} = \boldsymbol{w}\right] dz \mathbb{P}_{\boldsymbol{W}}(d\boldsymbol{w}) \\
&\longrightarrow e^{-z} dz \mathbb{P}_{\boldsymbol{W}}(d\boldsymbol{w})
\end{aligned}$$

as $n \to \infty$. □

4.7.2 Marginal density of directional variable under homothetic density

Proof of Proposition (4.2). Suppose that $f(\boldsymbol{x}) = f_0(g_{\mathcal{G}}(\boldsymbol{x}))$, $\boldsymbol{x} \in \mathbb{R}^d$. Then,

$$f_{\boldsymbol{W}}(\boldsymbol{w}) = \int_0^\infty f_{R,\boldsymbol{W}}(r, \boldsymbol{w})\,\mathrm{d}r = \frac{1}{g_{\mathcal{G}}(\boldsymbol{w})^d}\int_0^\infty s^{d-1} f_0(s)\,\mathrm{d}s = \frac{1}{d\,|\mathcal{G}|\,g_{\mathcal{G}}(\boldsymbol{w})^d}, \quad (4.26)$$

where the last equality follows from Balkema and Nolde (2010, see Section 3.1). □

4.7.3 Stability of the radial Exponential distribution

Proof of stability equation (4.6). Let the random vector $\boldsymbol{Z} \in \mathbb{R}^d$ follow an rExp distribution with location $\mathcal{M}$, scale Σ, and angular component $\mathcal{W}$. Then,

$$\begin{aligned}
&\mathbb{P}\left[\boldsymbol{Z} \in \{\mathcal{M} + B_{r_1+r_2}(0)\cdot\Sigma\}' \mid \boldsymbol{Z} \in \{\mathcal{M} + B_{r_1}(0)\cdot\Sigma\}'\right] \\
&= \frac{\mathbb{P}\left[\boldsymbol{Z} \in \{\mathcal{M} + B_{r_1+r_2}(0)\cdot\Sigma\}'\right]}{\mathbb{P}\left[\boldsymbol{Z} \in \{\mathcal{M} + B_{r_1}(0)\cdot\Sigma\}'\right]} \\
&= \frac{\mathbb{P}\left[\{\|\boldsymbol{Z}\| - r_{\mathcal{M}}\left(\boldsymbol{Z}/\|\boldsymbol{Z}\|\right)\}/r_\Sigma\left(\boldsymbol{Z}/\|\boldsymbol{Z}\|\right) > r_1 + r_2\right]}{\mathbb{P}\left[\{\|\boldsymbol{Z}\| - r_{\mathcal{M}}\left(\boldsymbol{Z}/\|\boldsymbol{Z}\|\right)\}/r_\Sigma\left(\boldsymbol{Z}/\|\boldsymbol{Z}\|\right) > r_1\right]} \\
&= \frac{\exp\{-(r_1+r_2)\}}{\exp\{-r_1\}} \\
&= \exp\{-r_2\} = \mathbb{P}\left[\boldsymbol{Z} \in \{\mathcal{M} + B_{r_2}(0)\}'\cdot\Sigma\right].
\end{aligned}$$

□

4.7.4 Convergence to gauge functions for d-dimensional copulas

Proposition 4.1 is the central theoretical result of this work. In it, it is stated that, when rescaled appropriately, the negative log joint density of the d-dimensional random vector whose margins are von Mises must converge to the gauge function $g_{\mathcal{G}}$ raised to some power $\rho > 0$. Furthermore, this convergence must be uniform in the domain $\mathbb{S}^{d-1}$. This is a central assumption that must be verified to motivate the use of the rExp distribution for inference on the multivariate tail. In this

section of the appendix, we do this for numerous multivariate distributions whose margins are standardised to follow the standard Laplace distribution. In addition, we acknowledge that these ideas extend beyond von Mises margins; see Proposition 6 in Papastathopoulos et al. (2025). To this end, we also show similar uniform convergence results for multivariate distributions in margins that are not von Mises. However, statistical inference in these multivariate tails are left to future work.

4.7.4.1 Multivariate Gaussian distribution, standard Laplace margins

The negative logarithm of the probability density function of the standard d-dimensional normal copula with standard Laplace margins and positive-definite precision matrix Q is standard d-dimensional normal copula with standard Laplace margins and positive-definite precision matrix Q is

$$-\log f(t\boldsymbol{x}) = -\frac{1}{2}\log|\mathsf{Q}| + d\log 2 + t\sum_{i=1}^{d}|x_{t,i}| + \frac{1}{2}\boldsymbol{H}(t,\boldsymbol{x})^{\top}(\mathsf{Q}-I)\boldsymbol{H}(t,\boldsymbol{x})$$

where $\boldsymbol{H} : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ is defined by $\boldsymbol{H}(t,\boldsymbol{x}) = (H(t,x_i) : i = 1,\ldots,d)$, with $H(t,y) := \Phi^{-1}\{F_L(ty)\}$, for $t > 0$ and $y \in \mathbb{R}$. Using Mill's ratio to obtain an asymptotic expansion of $H(t,y)$ as t grows arbitrarily large, we conclude that the convergence condition (4.3) in Proposition 4.1 are satisfied with $\psi(t) = t$, $\rho = 1$, and gauge function

$$g_{\mathcal{G}}(\boldsymbol{x}) = \left[\operatorname{sgn}(\boldsymbol{x})|\boldsymbol{x}|^{1/2}\right]^{\top}\mathsf{Q}\left[\operatorname{sgn}(\boldsymbol{x})|\boldsymbol{x}|^{1/2}\right]$$

To inspect the convergence rate of (4.3), rewrite $-\log f(t\boldsymbol{x})$ as

$$-\log f(t\boldsymbol{x}) = tg_{\mathcal{G}}(\boldsymbol{x}) + (\log t)\,u_1(\boldsymbol{x}) + u_2(\boldsymbol{x}) + o(1) \quad \text{as } t\to\infty, \tag{4.27}$$

where

$$\begin{aligned}
u_1(\boldsymbol{x}) &= -\frac{1}{2}\left[\operatorname{sgn}(\boldsymbol{x})|\boldsymbol{x}|^{1/2}\right]^{\top}(\mathsf{Q}-\mathsf{I})\left[\operatorname{sgn}(\boldsymbol{x})|\boldsymbol{x}|^{-1/2}\right]\\
u_2(\boldsymbol{x}) &= \left[\operatorname{sgn}(\boldsymbol{x})|\boldsymbol{x}|^{1/2}\right]^{\top}(\mathsf{Q}-\mathsf{I})\left[\operatorname{sgn}(\boldsymbol{x})\log 2|\boldsymbol{x}|^{-1/2}\right] -\\
&\quad -\frac{1}{2}\left[\operatorname{sgn}(\boldsymbol{x})|\boldsymbol{x}|^{1/2}\right]^{\top}(\mathsf{Q}-\mathsf{I})\left[\operatorname{sgn}(\boldsymbol{x})\log(4\pi|\boldsymbol{x}|)|\boldsymbol{x}|^{-1/2}\right]
\end{aligned}$$

$$-\frac{1}{2}\log|\mathsf{Q}| + d\log 2. \tag{4.28}$$

We note that careful attention needs to be given to establish that the convergence (4.3) holds *uniformly*. The asymptotic expansion in (4.27) is valid for large t when all $x_i \neq 0$, and loses validity when at least one x_i is equal to 0. This is a result of the asymptotic expansion of $H(t, x_i)$ and $H(t, x_i)^2$, both depending on t and x_i, resulting in non-uniform behaviour. When $x_i = 0$, the expansions of $H(t, x_i)$ and $H(t, x_i)^2$ collapse to zero, while for $x_i \neq 0$, the leading term grows as $t^{1/2}$ with logarithmic corrections. This difference in growth rates reflects the non-uniformity in x_i, as the expansion behaves differently near the origin and away from it, complicating the asymptotic analysis of $-\log f(t\boldsymbol{x})$ whenever $\boldsymbol{x}$ is near an axis. In Supplementary Material 3.1 of Papastathopoulos et al. (2025), we show that a uniform asymptotic expansion exists, by proving $-\log f(t\boldsymbol{x})/t$ converges locally uniformly to a continuous gauge function via the continuous mapping theorem (see, for example, Section 3.2.3 of Resnick (2007)). That is, we prove that $-\log f(t\boldsymbol{x})/t \to g_{\mathcal{G}}(\boldsymbol{x})$ as $t \to \infty$, uniformly on compact sets in the variable $\boldsymbol{x} \in \mathbb{R}^d$, by showing that $-\log f(t\boldsymbol{x}_t)/t$ converges to $g_{\mathcal{G}}(\boldsymbol{x})$ whenever $\boldsymbol{x}_t \to \boldsymbol{x} \in \mathbb{R}^d$. Consequently, because $\mathbb{S}^{d-1}$ is compact, the convergence is uniform on $\mathbb{S}^{d-1}$ and therefore, the condition (4.3) of Proposition 4.1 are satisfied.

4.7.4.2 Multivariate Laplace distribution, standard Laplace margins

The joint density of the d-variate Laplace distribution in standard Laplace margins with positive definite precision matrix Q is

$$f(t\boldsymbol{x}) = \frac{2^{1+\frac{d}{2}}|\mathsf{Q}|^{1/2}}{(2\pi)^{d/2}} t^{v} \left(\frac{1}{2^2}\boldsymbol{x}^\top \mathsf{Q}\boldsymbol{x}\right)^{v/2} K_v\left\{t\left(\boldsymbol{x}^\top \mathsf{Q}\boldsymbol{x}\right)^{1/2}\right\}$$

for $t > 0$, where $v = (2-d)/2$ and K_v is the modified Bessel function of the second kind (Kotz et al., 2001). Therefore,

$$\begin{aligned}
&-\log f(t\boldsymbol{x}) \\
&= -\left(1+\frac{d}{2}\right)\log 2 + \frac{d}{2}\log(2\pi) - \frac{1}{2}\log|\mathsf{Q}| + v\log(2/t) - \frac{v}{2}\log\left(\boldsymbol{x}^\top \mathsf{Q}\boldsymbol{x}\right)
\end{aligned}$$

$$-\log\left[K_v\left\{t\left(\boldsymbol{x}^\top \mathsf{Q}\boldsymbol{x}\right)^{1/2}\right\}\right]$$

Asymptotically, we have $K_v(z) \sim (\pi/2z)^{1/2} e^{-z}(1+O(z^{-1}))$ as $z \to \infty$ (Paris, 2017). Applying the negative logarithm, obtain

$$-\log K_v(z) \sim -\frac{1}{2}\log\left(\frac{\pi}{2}\right) + \frac{1}{2}\log z + z + O(z^{-1}).$$

Substituting this in the expression for $-\log f(t\boldsymbol{x})$, obtain

$$\begin{aligned}-\log f(t\boldsymbol{x}) \sim &-\left(1+\frac{d}{2}\right)\log 2 + \frac{d}{2}\log 2\pi - \frac{1}{2}\left|\mathsf{Q}\right| - v\log t + v\log 2 - \frac{v}{2}\log\left(\boldsymbol{x}^\top \mathsf{Q}\boldsymbol{x}\right)\\ &-\frac{1}{2}\log\left(\frac{\pi}{2}\right) + \frac{1}{2}\log t + \frac{1}{4}\log\left(\boldsymbol{x}^\top \mathsf{Q}\boldsymbol{x}\right) + t\left(\boldsymbol{x}^\top \mathsf{Q}\boldsymbol{x}\right)^{1/2} + O(t^{-1}).\end{aligned}$$

Therefore, the converge condition (4.3) in Proposition 4.1 holds with $\psi(t) = t$, $\rho = 1$, and $g_{\mathcal{G}}(\boldsymbol{x}) = \left(\boldsymbol{x}^\top \mathsf{Q}\boldsymbol{x}\right)^{1/2}$. To inspect the rate of convergence of (4.3) by writing $-\log f(t\boldsymbol{x})$ as

$$-\log f(t\boldsymbol{x}) = t g_{\mathcal{G}}(\boldsymbol{x}) + (\log t)\, u_1(\boldsymbol{x}) + u_2(\boldsymbol{x}) + o(1)$$

for $t \to \infty$, where the higher order terms are given by

$$u_1(\boldsymbol{x}) = \frac{1}{2} - v$$

and

$$u_2(\boldsymbol{x}) = -\left(1+\frac{d}{2}-v\right)\log 2 + \frac{d}{2}\log 2\pi - \frac{1}{2}\log\left|\mathsf{Q}\right| - \frac{1}{2}\log\left(\frac{\pi}{2}\right) + \left(\frac{1}{4}-\frac{v}{2}\right)\log\left(\boldsymbol{x}^\top \mathsf{Q}\boldsymbol{x}\right)$$

.

4.7.4.3 Multivariate max-stable distribution, standard Fréchet margins

The joint density of max-stable random vectors when the margins of $\boldsymbol{x}$ are standard Fréchet margins is given by

$$f(\boldsymbol{x}) = \left(\sum_{\pi\in\Pi}(-1)^{|\pi|}\prod_{s\in\pi} V_s(\boldsymbol{x})\right)\exp\left\{-V(\boldsymbol{x})\right\},$$

where Π is the set of all partitions of the set of indices $\{1, \ldots, d\}$, and $V(\boldsymbol{x})$ is a -1-homogeneous exponent function. As a result, $V_s(\boldsymbol{x})$ is $-(1+|s|)$-homogeneous, and we have

$$\begin{aligned} f(t\boldsymbol{x}) &= \left(\sum_{\pi \in \Pi} (-1)^{|\pi|} \prod_{s \in \pi} t^{-(1+|s|)} V_s(\boldsymbol{x}) \right) \exp\left\{ -t^{-1} V(\boldsymbol{x}) \right\} \\ &= \left(\sum_{\pi \in \Pi} (-1)^{|\pi|} t^{-(|\pi|+d)} \prod_{s \in \pi} V_s(\boldsymbol{x}) \right) \left[1 + O(t^{-1} V(\boldsymbol{x})) \right] \\ &= \left(\min_{\pi \in \Pi} \left\{ (-1)^{|\pi|} t^{-(|\pi|+d)} \prod_{s \in \pi} V_s(\boldsymbol{x}) \right\} + \sum_{\pi \in \Pi \setminus \pi^*} (-1)^{|\pi|} t^{-(|\pi|+d)} \prod_{s \in \pi} V_s(\boldsymbol{x}) \right) \\ &\quad \times \left[1 + O(t^{-1} V(\boldsymbol{x})) \right], \end{aligned} \tag{4.29}$$

where π^* is the solution to the minimisation problem in (4.29). As $t \to \infty$, the solution of which is $\pi^* = \{\{1, \ldots, d\}\}$, and therefore $|\pi^*| = 1$. This results in

$$f(t\boldsymbol{x}) = \left(-t^{-(1+d)} V_{\{1,\ldots,d\}}(\boldsymbol{x}) + \sum_{\pi \in \Pi : |\pi| > 1} (-1)^{|\pi|} t^{-(|\pi|+d)} \prod_{s \in \pi} V_s(\boldsymbol{x}) \right) \left[1 + O(t^{-1} V(\boldsymbol{x})) \right].$$

Therefore, the following convergence holds,

$$\frac{f(t\boldsymbol{x})}{t^{-(d+1)}} \to -V_{\{1,\ldots,d\}}(\boldsymbol{x}) =: g_{\mathcal{G}}(\boldsymbol{x})^{-(d+1)}$$

as $t \to \infty$. This convergence means that $f(t\boldsymbol{x})$ has the asymptotic form

$$f(t\boldsymbol{x}) \sim t^{-(d+1)} g_{\mathcal{G}}(\boldsymbol{x})^{-(d+1)} + \sum_{k=1}^{d-1} t^{-(d+k+1)} u_k(\boldsymbol{x})$$

as $t \to \infty$, where the higher order terms are given by

$$u_k(\boldsymbol{x}) = (-1)^{k+1} \sum_{\pi \in \Pi^{(k+1)}} \prod_{s \in \pi} V_s(\boldsymbol{x})$$

for $k \in \{1, \ldots, d-1\}$. In this notation, $\Pi^{(k)} \subset \Pi$ is such that $|\pi| = k$ for all $\pi \in \Pi^{(k)}$. Note that in the context of Proposition 6 (*iii*) in Papastathopoulos et al. (2025), this convergence implies that radial exceedances of the form $(R - r_{\mathcal{Q}_q}(\boldsymbol{w}))/r_{\mathcal{Q}_q}(\boldsymbol{w})) \mid \{\boldsymbol{W} = \boldsymbol{w}, R > r_{\mathcal{Q}_q}(\boldsymbol{W})\}$ follows a generalised Pareto distribution with shape parameter (or "tail index") $\xi = 1$.

Further note that g depends on the type of dependence, i.e., the exponent function V. For example, in the logistic setting with dependence parameter $\theta \in (0,1)$, we have $V(\boldsymbol{x}) = \left(\sum_{j=1}^{d} x_j^{-1/\theta}\right)^{\theta}$. The expression of $g_{\mathcal{G}}$ is therefore given by the solution of

$$\begin{aligned} g_{\mathcal{G}}(\boldsymbol{x})^{-(d+1)} =& - V_{\{1,\ldots,d\}}(\boldsymbol{x}) \\ =& - (-\theta)^{-d} \left[\prod_{k=0}^{d-1}(\theta - k)\right] \left(\prod_{k=1}^{d} x_k\right)^{-\frac{1}{\theta}-1} \left(\sum_{j=1}^{d} x_j^{-1/\theta}\right)^{\theta-d} \\ =& \theta^{-d+1} \left[\prod_{k=1}^{d-1}(k - \theta)\right] \left(\prod_{k=1}^{d} x_k\right)^{-\frac{1}{\theta}-1} \left(\sum_{j=1}^{d} x_j^{-1/\theta}\right)^{\theta-d}. \end{aligned}$$

Furthermore, the $-(|s|+1)$-homogeneous $|s|$-order partial derivative of V with respect to inputs whose indices are in s in the higher-order terms u_k is given by

$$V_s(\boldsymbol{x}) = (-\theta)^{-|s|} \left[\prod_{k=0}^{|s|-1}(\theta - k)\right] \left(\prod_{k\in s} x_k\right)^{-\frac{1}{\theta}-1} \left(\sum_{j=1}^{d} x_j^{-1/\theta}\right)^{\theta-|s|}.$$

4.7.4.4 Multivariate inverted max-stable distribution, standard exponential margins

The class of multivariate inverted max-stable distributions is usually represented in exponential margins, and has a joint distribution function

$$F(\boldsymbol{x}) = e^{-\ell(\boldsymbol{x})}$$

where ℓ is the 1-homogeneous *stable tail dependence function*, and is defined by the -1-homogeneous exponent function V (see Appendix 4.7.4.3) through the relation $\ell(\boldsymbol{x}) = V(1/\boldsymbol{x})$. The joint density in exponential margins is given by

$$f(\boldsymbol{x}) = \left(\sum_{\pi\in\Pi}(-1)^{|\pi|}\prod_{s\in\pi}\ell_s(\boldsymbol{x})\right)\exp\{-\ell(\boldsymbol{x})\}$$

where $\ell_s(\boldsymbol{x}) = V_s(1/\boldsymbol{x})(-1)^{|s|}\prod_{j\in s} x_j^{-2}$ is $(1-|s|)$-homogeneous. Therefore, the convergence condition (4.3) in Proposition 4.1 holds with $\psi(t) = t$, $\rho = 1$, and

gauge function $g_{\mathcal{G}}(\boldsymbol{x}) = \ell(\boldsymbol{x})$. The convergence rate of (4.3) can be inspected by writing $-\log f(t\boldsymbol{x})$ as

$$-\log f(t\boldsymbol{x}) = t g_{\mathcal{G}}(\boldsymbol{x}) + u(t\boldsymbol{x})$$

where the and higher order term

$$\begin{aligned} u(t\boldsymbol{x}) &= -\log\left[(-1)^d \prod_{j=1}^{d} \ell_{\{j\}}(\boldsymbol{x}) + O(t^{-1})\right] \\ &= -\log\left[(-1)^d \prod_{j=1}^{d} \ell_{\{j\}}(\boldsymbol{x}) + o(1)\right] \end{aligned}$$

as $t \to \infty$.

4.7.4.5 Multivariate max-stable and inverted max-stable distributions with logistic dependence, Laplace margins

The joint density function in Fréchet margins is

$$f_F(\boldsymbol{z}) = \left(\sum_{\pi \in \Pi} (-1)^{|\pi|} \prod_{s \in \pi} V_s(\boldsymbol{z})\right) \exp\{-V(\boldsymbol{z})\}$$

where $V(\boldsymbol{z}) = \left(\sum_{j=1}^{d} z_j^{-1/\theta}\right)^{\theta}$ is a -1-homogeneous exponent function with dependence parameter $\theta \in (0,1)$, and V_s is the $|s|$-order partial derivative of V with respect to inputs whose indices are in s. Let Π be the set of all partitions of the set of indices $\{1,\ldots,d\}$, and let π be the set of all partitions of an arbitrary element in Π. To obtain the joint density in Laplace margins, change of variables in implemented.

Suppose $z_t(x_j) = z(tx_j)$ for $t > 0$ and $j \in \{1,\ldots,d\}$. If $x_j < 0$ (or $z_t(x_j) < (\log 2)^{-1}$), then we perform the change of variables from Fréchet to Laplace margins

$$z_t(x_j) = \left(-\log\left(\frac{1}{2}e^{tx_j}\right)\right)^{-1} = (-tx_j)^{-1}\left(1 - \log 2\,(-tx_j)^{-1} + O(t^{-2})\right)$$

with derivative given by

$$\frac{d}{d(tx_j)} z_t(x_j) = (-tx_j)^{-2}\left(1 - 2\log 2\,(-tx_j)^{-1} + O(t^{-2})\right).$$

If $x_j > 0$ (or $z_t(x_j) > (\log 2)^{-1}$), then

$$z_t(x_j) = \left(-\log\left(1 - \frac{1}{2}e^{-tx_j}\right)\right)^{-1} = 2e^{tx_j} - \frac{1}{2} + O\left(e^{-tx_j}\right)$$

with derivative given by

$$\frac{d}{d(tx_j)} z_t(x_j) = 2e^{tx_j} + O\left(e^{-tx_j}\right).$$

Lastly, if $x_j = 0$, then $z_t(x_j) = (\log 2)^{-1}$. By the inverse function and chain rules,

$$\begin{aligned}\frac{d}{d(tx_j)} z_t(x_j) = \frac{f_L(tx_j)}{f_F\left(F_F^{-1}\left(F_L(tx_j)\right)\right)} &= \begin{cases} \frac{e^{-t|x_j|}}{e^{tx_j}\left(\log\left(\frac{1}{2}e^{tx_j}\right)\right)^2} & ; \ x_j < 0 \\ \frac{\frac{1}{2}e^{-t|x_j|}}{\left(1-\frac{1}{2}e^{-tx_j}\right)\left(\log\left(1-\frac{1}{2}e^{-tx_j}\right)\right)^2} & ; \ x_j > 0 \end{cases} \\ &\xrightarrow{x_j \to 0} (\log 2)^{-2},\end{aligned}$$

where f_L and F_L are the univariate density and distribution functions of the standard Laplace distribution, respectively. For a vector $\boldsymbol{x} = (x_1, \ldots, x_d)^\top$, let $A, B, C \subset \{1, \ldots, d\}$ be the set of indices such that x_j is positive, negative, and zero for $j \in A, B, C$, respectively such that $|A| + |B| + |C| = d$. By change of variables, the joint density for the max-stable distribution with logistic dependence in Laplace margins is

$$\begin{aligned} f(t\boldsymbol{x}) =& \left| \prod_{j=1}^{d} \frac{d}{d(tx_j)} z_t(x_j) \right| f_F\left(z(tx_1), \ldots, z(tx_d)\right) \\ =& (-1)^{d+1} \left\{ \prod_{\ell=0}^{d-1} \left(1 - \frac{\ell}{\theta}\right) \right\} 2^{-\frac{|A|}{\theta}} t^{\left(\frac{1}{\theta}-1\right)|B|+1-\frac{d}{\theta}} (\log 2)^{-2|C|} \left(\prod_{k \in B} (-x_k)^{\frac{1}{\theta}-1} \right) \\ & \times \left(\sum_{k \in B} (-x_k)^{1/\theta} \right)^{\theta - d} \\ & \times \exp\left\{ -t \left[\frac{1}{\theta} \sum_{j \in A} x_j + \left(\sum_{k \in B} (-x_k)^{1/\theta} \right)^{\theta} \left(1 + O\left(e^{-\frac{t}{\theta} \min_{j \in A} x_j} \right) + O(t^{-1}) \right) \right] \right\} \\ & \times (1 + o(1)) \end{aligned}$$

as $t \to \infty$. Applying the negative logarithm, obtain the following expression

$$
\begin{aligned}
-\log f(t\boldsymbol{x}) = &-\log\left[(-1)^{d+1}\left\{\prod_{\ell=0}^{d-1}\left(1-\frac{\ell}{\theta}\right)\right\}2^{-\frac{|A|}{\theta}}(\log 2)^{-2|C|}\right]-\left(\frac{1}{\theta}-1\right)\sum_{k\in B}\log(-x_k)\\
&-(\theta-d)\log\left(\sum_{k\in B}(-x_k)^{1/\theta}\right)+\left\{-\left(\frac{1}{\theta}-1\right)|B|-1+\frac{d}{\theta}\right\}\log t\\
&+t\left[\frac{1}{\theta}\sum_{j\in A}x_j+\left\{\sum_{k\in B}(-x_k)^{1/\theta}\right\}^{\theta}\left\{1+O\left(e^{-\frac{t}{\theta}\min_{j\in A}x_j}\right)+O(t^{-1})\right\}\right]+o(1)
\end{aligned}
$$

From this, we see that the convergence assumption (4.3) in Proposition 4.1 is satisfied with $\psi(t)=t$, $\rho=1$, and

$$
g_{\mathcal{G}}(\boldsymbol{x})=\frac{1}{\theta}\sum_{j\in A}x_j+\left\{\sum_{k\in B}(-x_k)^{1/\theta}\right\}^{\theta}.
$$

The rate of convergence of (4.3) can be inspected by writing $-\log f(t\boldsymbol{x})$ as

$$
-\log f(t\boldsymbol{x})=tg_{\mathcal{G}}(\boldsymbol{x})+(\log t)u_1(\boldsymbol{x})+u_2(\boldsymbol{x}),
$$

with higher order terms are given by

$$
u_1(\boldsymbol{x})=-\left(\frac{1}{\theta}-1\right)|B|-1+\frac{d}{\theta}
$$

and

$$
\begin{aligned}
u_2(\boldsymbol{x}) = &-\log\left[\frac{(-1)^{d+1}2^{-\frac{|A|}{\theta}}}{(\log 2)^{-2|C|}}\left\{\prod_{\ell=0}^{d-1}\left(1-\frac{\ell}{\theta}\right)\right\}\right]-\left(\frac{1}{\theta}-1\right)\sum_{k\in B}\log(-x_k)\\
&-(\theta-d)\log\left\{\sum_{k\in B}(-x_k)^{1/\theta}\right\}+o(1)
\end{aligned}
$$

There are 2 special cases to consider:

- **special case 1:** Suppose $x_j>0\ \forall\ j\in\{1,\ldots,d\}$ and let $x_{(d)}=\min_{j=1,\ldots,d}x_j$. Here, the joint log-density is

$$
\begin{aligned}
-\log f(t\boldsymbol{x}) = &-\log\left[2^{-1}(-1)^d\left\{\prod_{\ell=1}^{d-1}\left(1-\frac{\ell}{\theta}\right)\right\}\right]+t\left\{\frac{1}{\theta}\sum_{j=1}^{d}x_j+\left(1-\frac{d}{\theta}\right)x_{(d)}\right\}+\\
&+2^{-1}e^{-tx_{(d)}}\left(1+o(1)\right)+o(1)=tg_{\mathcal{G}}(\boldsymbol{x})+u_1(\boldsymbol{x})+o(1)
\end{aligned}
$$

as $t \to \infty$, where the gauge function is

$$g_{\mathcal{G}}(\boldsymbol{x}) = \frac{1}{\theta}\sum_{j=1}^{d} x_j + \left(1 - \frac{d}{\theta}\right) \min_{k=1,\ldots,d} x_k$$

and the higher order term is

$$u_1(\boldsymbol{x}) = -\log\left[2^{-1}(-1)^{d+1}\left\{\prod_{\ell=1}^{d-1}\left(1 - \frac{\ell}{\theta}\right)\right\}\right].$$

- **special case 2:** Suppose $x_j < 0 \; \forall \; j \in \{1, \ldots, d\}$. Here, we have

$$-\log f(t\boldsymbol{x}) = t g_{\mathcal{G}}(\boldsymbol{x}) + (\log t) u_1(\boldsymbol{x}) + u_2(\boldsymbol{x}) + o(1)$$

as $t \to \infty$, where the gauge function is

$$g_{\mathcal{G}}(\boldsymbol{x}) = \left\{\sum_{j=1}^{d} (-x_j)^{1/\theta}\right\}^{\theta}$$

the higher order terms are given by

$$u_1(\boldsymbol{x}) = d - 1$$

and

$$u_2(\boldsymbol{x}) = -\log\left[(-1)^{d+1}\left\{\prod_{\ell=0}^{d-1}\left(1 - \frac{\ell}{\theta}\right)\right\}\left\{\prod_{j=1}^{d}(-x_j)\right\}^{\frac{1}{\theta}-1}\left\{\sum_{j=1}^{d}(-x_j)^{1/\theta}\right\}^{\theta-d}\right].$$

The case of **inverted logistic dependence**, defined by the joint distribution function in Fréchet margins

$$F_F(\boldsymbol{z}) = \exp\{-V(\boldsymbol{z})\}$$

is similar, amounting to a translation $\boldsymbol{x} \mapsto -\boldsymbol{x}$. For more detail on these derivations, the reader is referred to Supplementary Material 3.5 of Papastathopoulos et al. (2025).

4.7.4.6 Multivariate Student's t_ν distribution, Student-t_ν margins, $\nu > 0$

Suppose Q positive definite and $\nu > 0$. The joint density can be expressed as $f(\boldsymbol{x}) = f_0(g_{\mathcal{G}}(\boldsymbol{x}))$, where the homothetic function f_0 and the $g_{\mathcal{G}}$ function is given by

$$\begin{aligned} f_0(s) &= k_{\nu,\mathsf{Q}}(1+\nu^{-1}s^2)^{-\frac{1}{2}(\nu+d)}, \\ g_{\mathcal{G}}(\boldsymbol{x}) &= \left(\boldsymbol{x}^\top \mathsf{Q}\boldsymbol{x}\right)^{1/2}, \end{aligned}$$

where $k_{\nu,\mathsf{Q}} = \Gamma\left(\frac{\nu+d}{2}\right) / \left\{\Gamma\left(\frac{\nu}{2}\right)\nu^{d/2}\pi^{d/2}\left|\mathsf{Q}\right|^{-1/2}\right\}$. Note that in the context of Proposition 6 (*iii*) in Papastathopoulos et al. (2025), we have

$$\frac{f(t\boldsymbol{x})}{\psi(t)} \to g_{\mathcal{G}}(\boldsymbol{x})^{-(d+\xi^{-1})} \tag{4.30}$$

for $\psi(t) = k_{\nu,\mathsf{Q}}\nu^{\frac{1}{2}(\nu+d)}t^{-(\nu+d)}$ as $t \to \infty$. This convergence implies that radial exceedances of the form $(R - r_{\mathcal{Q}_q}(\boldsymbol{w}))/\nu^{-1}r_{\mathcal{Q}_q}(\boldsymbol{w})) \mid \{\boldsymbol{W} = \boldsymbol{w}, R > r_{\mathcal{Q}_q}(\boldsymbol{W})\}$ follows a generalised Pareto distribution with $\xi = \nu^{-1}$.

To study the rate of convergence in (4.30), note that this convergence result means that $f(t\boldsymbol{x})$ can be written as

$$f(t\boldsymbol{x}) = \psi(t)g_{\mathcal{G}}(\boldsymbol{x})^{-(d+\xi^{-1})} + u(\boldsymbol{x})$$

, where the higher-order term is

$$\begin{aligned} u(\boldsymbol{x}) =& \frac{\frac{f(t\boldsymbol{x})}{\psi(t)} - g_{\mathcal{G}}(\boldsymbol{x})^{-(\nu+d)}}{t^{-2}} \\ =& t^2\left[\psi(t)^{-1}k_{\nu,\mathsf{Q}}\nu^{\frac{1}{2}(\nu+d)}t^{-(\nu+d)}g_{\mathcal{G}}(\boldsymbol{x})^{-(\nu+d)}\left\{1+\nu t^{-2}g_{\mathcal{G}}(\boldsymbol{x})^{-2}\right\}^{-\frac{1}{2}(\nu+d)} - g_{\mathcal{G}}(\boldsymbol{x})^{-(\nu+d)}\right] \\ =& t^2\left[g_{\mathcal{G}}(\boldsymbol{x})^{-(\nu+d)}\left\{1 - \frac{1}{2}(\nu+d)\nu t^{-2}g_{\mathcal{G}}(\boldsymbol{x})^{-2} + O(t^{-4})\right\} - g_{\mathcal{G}}(\boldsymbol{x})^{-(\nu+d)}\right] \\ =& -\frac{1}{2}(\nu+d)\nu g_{\mathcal{G}}(\boldsymbol{x})^{-2-(\nu+d)} + O(t^{-2}) \\ =& -\frac{1}{2}(\nu+d)\nu g_{\mathcal{G}}(\boldsymbol{x})^{-2-(\nu+d)} + o(1) \end{aligned}$$

as $t \to \infty$.

4.7.4.7 Multivariate Student's t_ν distribution, $\nu > 0$, standard Laplace margins

The multivariate t-distribution with positive definite precision matrix $\mathsf{Q} = (q_{ij})_{i,j=1}^d$ and with univariate t-distribution margins with ν degrees of freedom is given has joint density

$$f_{\boldsymbol{t}_\nu}(\boldsymbol{z}) = k_{\nu,\mathsf{Q}} \left(1 + \frac{1}{\nu}\sum_{j=1}^{d} q_{jj} z_j^2 + \frac{2}{\nu}\sum_{1\le j<k\le d} q_{jk} z_j z_k\right)^{-\frac{1}{2}(\nu+d)}$$

where $k_{\nu,\mathsf{Q}} = \Gamma\left(\frac{\nu+d}{2}\right) / \left\{\Gamma\left(\frac{\nu}{2}\right)\nu^{d/2}\pi^{d/2}\left|\mathsf{Q}\right|^{-1/2}\right\}$. Perform the change of variables to the standard Laplace distribution, where we take advantage of a univariate t-distribution analogue of Mill's ratio (Soms, 1976). Let F_{t_ν} be the distribution function of the univariate t-distribution with ν degrees of freedom, respectively. Suppose $z = z(tx) > 0$ (or $x > 0$), then for $t > 0$ large, the change of variables from the t-distribution to the standard Laplace distribution

$$\begin{aligned} tx &= -\log\left[2\left\{1 - F_{t_\nu}(z(tx))\right\}\right] \\ &= -\log\left[\frac{2\Gamma\left(\frac{\nu+1}{2}\right)\nu^{\left(\frac{1-\nu}{2}\right)}}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}}\right] + \nu\log z(tx) + O\left(z(tx)^{-2}\right). \end{aligned}$$

Inverting this transformation, obtain

$$z(tx) = c_\nu e^{\frac{t}{\nu}x}\left(1 + O\left(e^{\frac{-2t}{\nu}}\right)\right),$$

with partial derivative with respect to tx given by

$$\frac{d}{dtx}z(tx) = \frac{c_\nu}{\nu}e^{\frac{t}{\nu}x}\left(1 + O\left(e^{\frac{-2t}{\nu}}\right)\right),$$

where $c_\nu = \left\{2\Gamma\left(\frac{\nu+1}{2}\right)\nu^{\left(\frac{1-\nu}{2}\right)}\Gamma\left(\frac{\nu}{2}\right)^{-1}(\nu\pi)^{-1/2}\right\}^{1/\nu}$. When $z(tx) < 0$ (or $x < 0$), the transformation to Laplace margins is

$$\begin{aligned} tx &= \log\left[2F_{t_\nu}(z(tx))\right] \\ &= \log 2 + \log\left[1 - F_{t_\nu}(-z(tx))\right]. \end{aligned}$$

Negating,

$$\begin{aligned}-tx =& -\log 2 - \log\left(1 - F_{t_\nu}(-z(tx))\right)\\ =& -\log\left[\frac{2\Gamma\left(\frac{\nu+1}{2}\right)\nu^{\left(\frac{1-\nu}{2}\right)}}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}}\right] + \nu\log(-z(tx)) + O\left(z(tx)^{-2}\right).\end{aligned}$$

Inverting this transformation,

$$z(tx) = -c_\nu e^{\frac{t}{\nu}|x|}\left(1 + O\left(e^{\frac{-2t}{\nu}}\right)\right),$$

with partial derivative

$$\frac{d}{dtx}z(tx) = \frac{c_\nu}{\nu}e^{\frac{t}{\nu}|x|}\left(1 + O\left(e^{\frac{-2t}{\nu}}\right)\right).$$

Therefore, by change of variables, obtain the joint density in Laplace margins

$$\begin{aligned}f(t\boldsymbol{x}) =& \left|\prod_{j=1}^{d}\frac{d}{d(tx_j)}z(tx_j)\right| f_{\boldsymbol{t}_\nu}(z(tx_1),\ldots,z(tx_d))\\ =& \frac{c_\nu^d k_{\nu,\mathtt{Q}}}{\nu^d}\exp\left\{\frac{t}{\nu}\sum_{j=1}^{d}|x_j|\right\}\frac{q_{j^\star j^\star}c_\nu^2}{\nu}\exp\left\{-t\left(1+\frac{d}{\nu}\right)\max_{j=1,\ldots,d}|x_j|\right\}(1+o(1)),\end{aligned}$$

as $t \to \infty$. where $j^\star$ is the index such that $|x_{j^\star}| = \max_{j=1,\ldots,d}|x_j|$. Taking the negative logarithm, observe that the convergence condition (4.3) holds with $\psi(t) = t$, $\rho = 1$, and gauge function

$$g_{\mathcal{G}}(\boldsymbol{x}) = -\frac{1}{\nu}\sum_{j=1}^{d}|x_j| + \left(1+\frac{d}{\nu}\right)\max_{j=1,\ldots,d}|x_j|,$$

To inspect the rate of convergence of (4.3), write $-\log f(t\boldsymbol{x})$ as

$$-\log f(t\boldsymbol{x}) = tg_{\mathcal{G}}(\boldsymbol{x}) + u(\boldsymbol{x}),$$

and note the higher order term is

$$u(\boldsymbol{x}) = -\log\left(\frac{c_\nu^{d+2}k_{\nu,\mathtt{Q}}q_{j^\star j^\star}}{\nu^{d+1}}\right) + o(1)$$

as $t \to \infty$.

4.7.4.8 Wishart distribution

For a $d \times d$ positive definite matrix $\mathbf{X}$, the Wishart distribution with ν degrees of freedom and positive definite scale matrix $\mathbf{V}$ has density function

$$f_{\mathbf{X}}(\mathbf{X}) = 2^{-(\nu d)/2} \det(\mathbf{V})^{-\nu/2} \Gamma_d(\nu/2)^{-1} \det(\mathbf{X})^{\frac{1}{2}(\nu-d-1)} e^{-\frac{1}{2}\mathrm{tr}\left(\mathbf{V}^{-1}\mathbf{X}\right)}$$

Therefore, the convergence condition (4.3) in Proposition 4.1 holds with $\psi(t) = t$, $\rho = 1$, and $g_{\mathcal{G}}(\mathbf{X}) = \frac{1}{2}\mathrm{tr}\,(\mathbf{V}^{-1}\mathbf{X})$. To inspect the rate of convergence, rewrite $-\log f_{\mathbf{X}}(t\mathbf{X})$ as $t g_{\mathcal{G}}(\mathbf{X}) + (\log t) u_1(\mathbf{X}) + u_2(\mathbf{X})$, where $u_1(\mathbf{X}) = \left(-\frac{d}{2}(\nu - d - 1)\right)$ and $u_2(\mathbf{X}) = -\frac{1}{2}(\nu - d - 1) \log\det(\mathbf{X}) + \log\left(2^{-(\nu d)2} \det(\mathbf{V})^{-\nu/2} \Gamma_d(\nu/2)^{-1}\right)$ are higher-order terms.

4.7.5 Directional densities

In Section 4.2.2, we give the exact form of the directional density $f_{\boldsymbol{W}}$ corresponding to the d-dimensional homothetic Gaussian, Laplace, Student's t distributions in their canonical margins, all of which are homothetic. In this appendix, we give the full derivations. In this appendix, we further derive the density $f_{\boldsymbol{W}}$ corresponding to the max-stable, inverted max-stable, and Wishart distributions in their canonical margins.

4.7.5.1 Multivariate Gaussian distribution, standard normal margins

The joint density can be written in the form $f(\boldsymbol{x}) = f_0(g_{\mathcal{G}}(\boldsymbol{x}))$, where $f(s) = (2\pi)^{-d/2} |\mathsf{Q}|^{1/2} \exp\{-s^2/2\}$. By (4.26),

$$\begin{aligned} f_{\boldsymbol{W}}(\boldsymbol{w}) =& g_{\mathcal{G}}(\boldsymbol{x})^{-d} \int_0^\infty s^{d-1} f_0(s) ds = (2\pi)^{-d/2} |\mathsf{Q}|^{1/2} \left(\boldsymbol{w}^\top \mathsf{Q} \boldsymbol{w}\right)^{-d/2} \int_0^\infty s^{d-1} e^{-\frac{1}{2}s^2} ds \\ =& \Gamma(d/2) 2^{\frac{d}{2}-1} (2\pi)^{-d/2} |\mathsf{Q}|^{1/2} \left(\boldsymbol{w}^\top \mathsf{Q} \boldsymbol{w}\right)^{-d/2} = \Gamma(d/2) 2^{-1} \pi^{-d/2} |\mathsf{Q}|^{1/2} \left(\boldsymbol{w}^\top \mathsf{Q} \boldsymbol{w}\right)^{-d/2} \end{aligned}$$

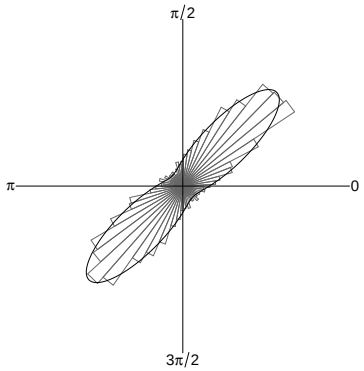

Figure 4.12: Directional density for the bivariate Gaussian distribution, standard Gaussian margins, with $\mathsf{Q}_{11}^{-1} = \mathsf{Q}_{22}^{-1} = 1$, $\mathsf{Q}_{12}^{-1} = \mathsf{Q}_{21}^{-1} = 0.8$, plotted over an empirical sample

4.7.5.2 Multivariate Laplace distribution, standard Laplace margins

We have $f_{[L]}(\boldsymbol{x}) = f_0(g_{\mathcal{G}}(\boldsymbol{x}))$, where $f_0(s) = |\mathsf{Q}|^{1/2}(2\pi)^{-d/2}s^v K_v(s)$. By (4.26), the density of angles is

$$\begin{aligned}
f_{\boldsymbol{W}}(\boldsymbol{w}) = &\int_0^\infty r^{d-1} f_0(r g_{\mathcal{G}}(\boldsymbol{w}))dr = g_{\mathcal{G}}(\boldsymbol{w})^{-d} \int_0^\infty s^{d-1} f_0(s)ds \\
=&|\mathsf{Q}|^{1/2}(2\pi)^{-d/2} \left(\boldsymbol{w}^\top \mathsf{Q}\boldsymbol{w}\right)^{-d/2} \int_0^\infty s^{d+v-1} K_v(s)ds \\
=&\Gamma(d/2)2^{\frac{d}{2}-1}(2\pi)^{-d/2} \left|\mathsf{Q}\right|^{1/2} \left(\boldsymbol{w}^\top \mathsf{Q}\boldsymbol{w}\right)^{-d/2} \\
=&\Gamma(d/2)2^{-1}\pi^{-d/2} \left|\mathsf{Q}\right|^{1/2} \left(\boldsymbol{w}^\top \mathsf{Q}\boldsymbol{w}\right)^{-d/2}
\end{aligned}$$

where the expression for the integral can be found in equation 16. of Section 6.561 of Gradshteyn and Ryzhik (2014) and equation 10.43.19 of Olver et al. (2019).

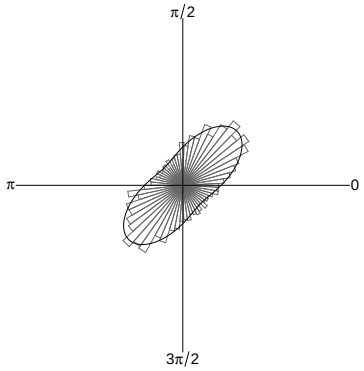

Figure 4.13: Directional density for the bivariate Laplace distribution, standard Laplace margins, with $\mathsf{Q}_{11}^{-1} = \mathsf{Q}_{22}^{-1} = 1$, $\mathsf{Q}_{12}^{-1} = \mathsf{Q}_{21}^{-1} = 0.5$, plotted over an empirical sample

4.7.5.3 Multivariate Student t_ν distribution, Student t_ν margins, $\nu > 0$

We have $f(\boldsymbol{x}) = f_0(g_{\mathcal{G}}(\boldsymbol{x}))$, where $f_0(s) = k_{\nu,\mathsf{Q}} \left(1 + \nu^{-1} s^2\right)^{-\frac{1}{2}(\nu+d)}$. By (4.26), the density of angles is

$$
\begin{aligned}
f_{\boldsymbol{W}}(\boldsymbol{w}) = & \int_0^\infty r^{d-1} f_0(r g_{\mathcal{G}}(\boldsymbol{w})) dr = g_{\mathcal{G}}(\boldsymbol{w})^{-d} \int_0^\infty s^{d-1} f_0(s) ds \\
= & k_{\nu,\mathsf{Q}} \left(\boldsymbol{w}^\top \mathsf{Q} \boldsymbol{w}\right)^{-d/2} \int_0^\infty s^{d-1} \left(1 + \nu^{-1} s^2\right)^{-\frac{1}{2}(\nu+d)} ds \\
= & \Gamma\left(\frac{\nu+d}{2}\right) \Gamma\left(\frac{\nu}{2}\right)^{-1} \nu^{-d/2} \pi^{-d/2} |\mathsf{Q}|^{1/2} \left(\boldsymbol{w}^\top \mathsf{Q} \boldsymbol{w}\right)^{-d/2} \nu^{-1+d/2} \\
& \times \Gamma\left(\frac{\nu+d}{2}\right)^{-1} \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(1 + \frac{\nu}{2}\right) \\
= & \Gamma\left(\frac{d}{2}\right) 2^{-1/2} \pi^{-d/2} |\mathsf{Q}|^{1/2} \left(\boldsymbol{w}^\top \mathsf{Q} \boldsymbol{w}\right)^{-d/2}
\end{aligned}
$$

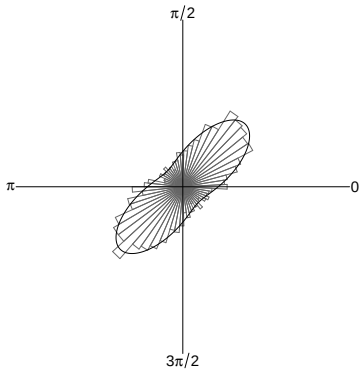

Figure 4.14: Directional density for the bivariate Student t_ν distribution, t_ν margins, with $\nu = 5$ and $\mathsf{Q}_{11}^{-1} = \mathsf{Q}_{22}^{-1} = 1$, $\mathsf{Q}_{12}^{-1} = \mathsf{Q}_{21}^{-1} = 0.6$, plotted over an empirical sample

4.7.5.4 Multivariate max-stable distribution, standard Fréchet margins

For a general -1-homogeneous exponent function,

$$\begin{aligned}
f_{\boldsymbol{W}}(\boldsymbol{w}) &= \int_0^\infty r^{d-1} f(r\boldsymbol{w}) dr \\
&= \int_0^\infty r^{d-1} \left[\sum_{\pi \in \Pi} (-1)^{|\pi|} \prod_{s \in \pi} V_s(r\boldsymbol{w}) \right] e^{-V(r\boldsymbol{w})} dr \\
&= \sum_{\pi \in \Pi} (-1)^{|\pi|} \left[\prod_{s \in \pi} V_s(\boldsymbol{w}) \right] \int_0^\infty r^{d - \sum_{s \in \pi}(1+|s|)-1} e^{-r^{-1}V(\boldsymbol{w})} dr \\
&= \sum_{\pi \in \Pi} (-1)^{|\pi|} \left[\prod_{s \in \pi} V_s(\boldsymbol{w}) \right] \Gamma \left\{ \sum_{s \in \pi} (1+|s|) - d \right\} V(\boldsymbol{w})^{\sum_{s \in \pi}(1+|s|)-d}
\end{aligned}$$

Suppose we fix $d = 2$, and we assume we have logistic dependence and Fréchet margins, then the exponent function is given by $V(x, y) = (x^{-1/\theta} + y^{-1/\theta})^\theta$, and the true directional density $f_{\boldsymbol{W}}$ can be plotted against an empirical sample of angles (see Figure 4.15).

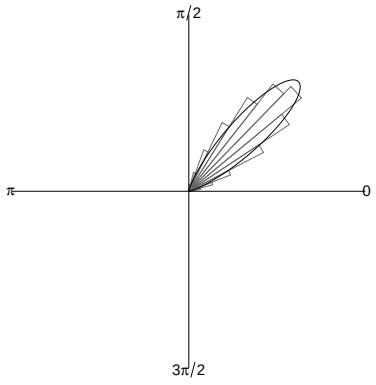

Figure 4.15: Directional density for the bivariate max-stable logistic distribution, standard Fréchet margins, with dependence parameter $\theta = 0.3$, plotted over an empirical sample

4.7.5.5 Multivariate inverted max-stable distributions, standard exponential margins

Recall, from subsubsection 4.7.4.4 that the first order gauge function for inverted max-stable distributions in exponential margins is given by the 1-homogeneous stable tail dependence function, $g_{\mathcal{G}}(\boldsymbol{x}) = \ell(\boldsymbol{x})$. With this in mind, the density of angles in this setting is given by

$$\begin{aligned}
f_{\boldsymbol{W}}(\boldsymbol{w}) &= \int_0^\infty r^{d-1} f(r\boldsymbol{w}) dr \\
&= \int_0^\infty r^{d-1} \left[\sum_{\pi \in \Pi} (-1)^{|\pi|} \prod_{s \in \pi} \ell_s(r\boldsymbol{w}) \right] e^{-\ell(r\boldsymbol{w})} dr \\
&= \sum_{\pi \in \Pi} (-1)^{|\pi|} \left[\prod_{s \in \pi} \ell_s(\boldsymbol{w}) \right] \int_0^\infty r^{d + \sum_{s \in \pi}(1-|s|) - 1} e^{-r\ell(\boldsymbol{w})} dr \\
&= \sum_{\pi \in \Pi} (-1)^{|\pi|} \left[\prod_{s \in \pi} \ell_s(\boldsymbol{w}) \right] \Gamma\left\{ d + \sum_{s \in \pi} (1 - |s|) \right\} \ell(\boldsymbol{w})^{d + \sum_{s \in \pi}(1-|s|)} \\
&= \left[\sum_{\pi \in \Pi} (-1)^{|\pi|} \left[\prod_{s \in \pi} \ell_s(\boldsymbol{w}) \right] \Gamma\left\{ d + \sum_{s \in \pi} (1 - |s|) \right\} g_{\mathcal{G}}(\boldsymbol{w})^{\sum_{s \in \pi}(1-|s|)} \right] g_{\mathcal{G}}(\boldsymbol{w})^d,
\end{aligned}$$

thus following the expression given in equation (4.5) from Section 4.2.2. Suppose we fix $d = 2$, and we assume we have logistic dependence and standard exponential margins, then the exponent function is given by $\ell(x, y) = (x^{1/\theta} + y^{1/\theta})^\theta$, and the true directional density $f_{\boldsymbol{W}}$ can be plotted against an empirical sample of angles (see Figure 4.16).

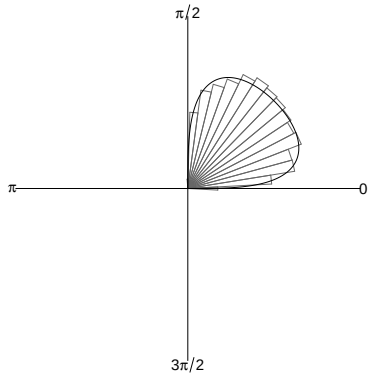

Figure 4.16: Directional density for the bivariate inverted max-stable logistic distribution, standard exponential margins, with dependence parameter $\theta = 0.7$, plotted over an empirical sample

4.7.5.6 Wishart distribution

The density of angles is computed using

$$\begin{aligned} f_{\mathbf{W}}(\mathbf{W}) = & \int_0^\infty r^{d-1} f_{\mathbf{X}}(r\mathbf{W}) dr \\ = & 2^{-(\nu d)/2} \det(\mathbf{V})^{-\nu/2} \Gamma_d(\nu/2)^{-1} \det(\mathbf{W})^{\frac{1}{2}(\nu-d-1)} \\ & \times \Gamma\left(\frac{d}{2}(1+\nu-d)\right) \left(\frac{1}{2}\mathrm{tr}\left(\mathbf{V}^{-1}\mathbf{W}\right)\right)^{-\frac{d}{2}(1+\nu-d)} \end{aligned}$$

where in this setting, radii and angles are defined using a matrix norm; $R = \|\mathbf{X}\|$, $\mathbf{W} = \mathbf{X}/\|\mathbf{X}\|$.

4.7.6 Posterior model fits on river data

On the bivariate Thames tributary data, we fit models using the three architectures M_1, M_2, and M_3 using data exceeding a high posterior quantile estimate and on all available data. Figures 4.17 and 4.18 show posterior mean limit set boundary and posterior directional densities for each of the six fitted models. In these figures, we see that the fitted model associated with the M_2 architecture and fitted on exceedance data only is best. This model's posterior limit set agrees most with $\log n$-scaled data, while the posterior mean directional density is in good agreement with an empirical sample of angles. The posterior QQ plots in Figure 4.19 corresponding to this optimal model shows good agreement with the radial exceedance model and the underlying exponential distribution, while the PP plots in 4.20 show general agreement between the posterior and the underlying directional models.

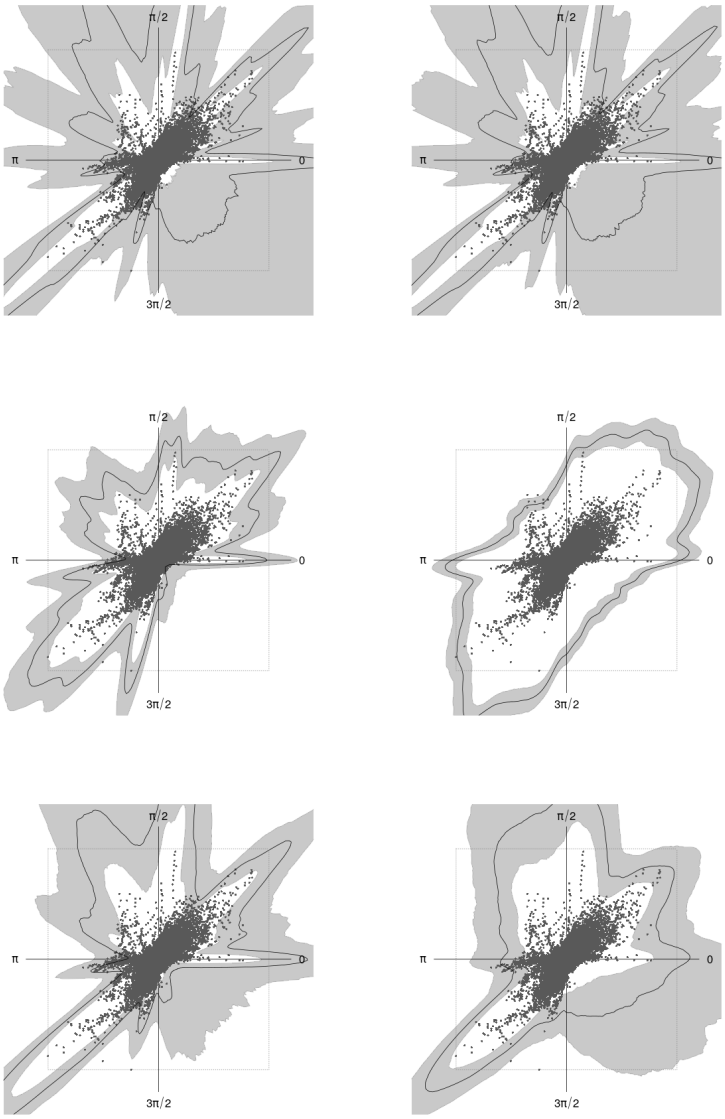

Figure 4.17: Posterior estimates of the unit level set $g_{\mathcal{G}}(\boldsymbol{x}) = 1$ for the river flow dataset. The black line corresponds to the posterior mean, with the 0.95 prediction interval shaded in grey. Black points are the original data in Laplace margins scaled by $\log(n/2)$. Dashed border line is the unit box. . Top row: model M_1. Middle row: model M_2. Bottom row: model M_3. Left: fitting using only exceedances. Right: fitting using all observations. M_1, M_2, and M_3 define the angle density kernel, as described in section 4.3.3

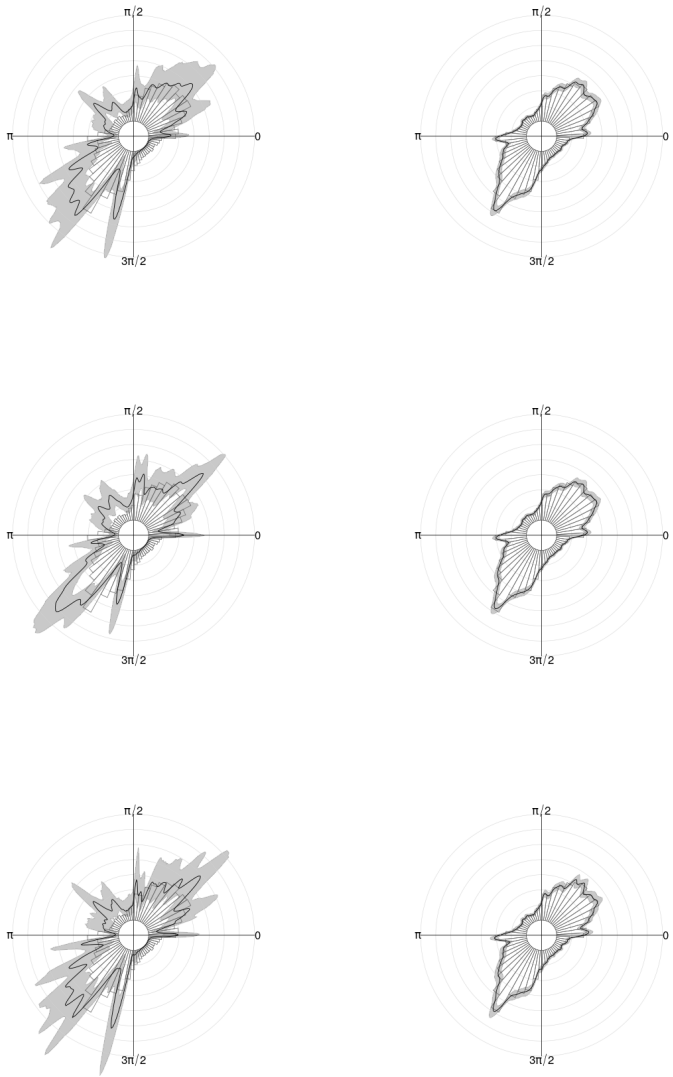

Figure 4.18: Posterior estimates of the mean angle density for the river flow dataset with 0.95 prediction intervals. The empirical density of angles is given by the underlying histogram. Top row: model M_1. Middle row: model M_2. Bottom row: model M_3. Left: fitting using only exceedances. Right: fitting using all observations. M_1, M_2, and M_3 define the angle density kernel, as described in section 4.3.3.

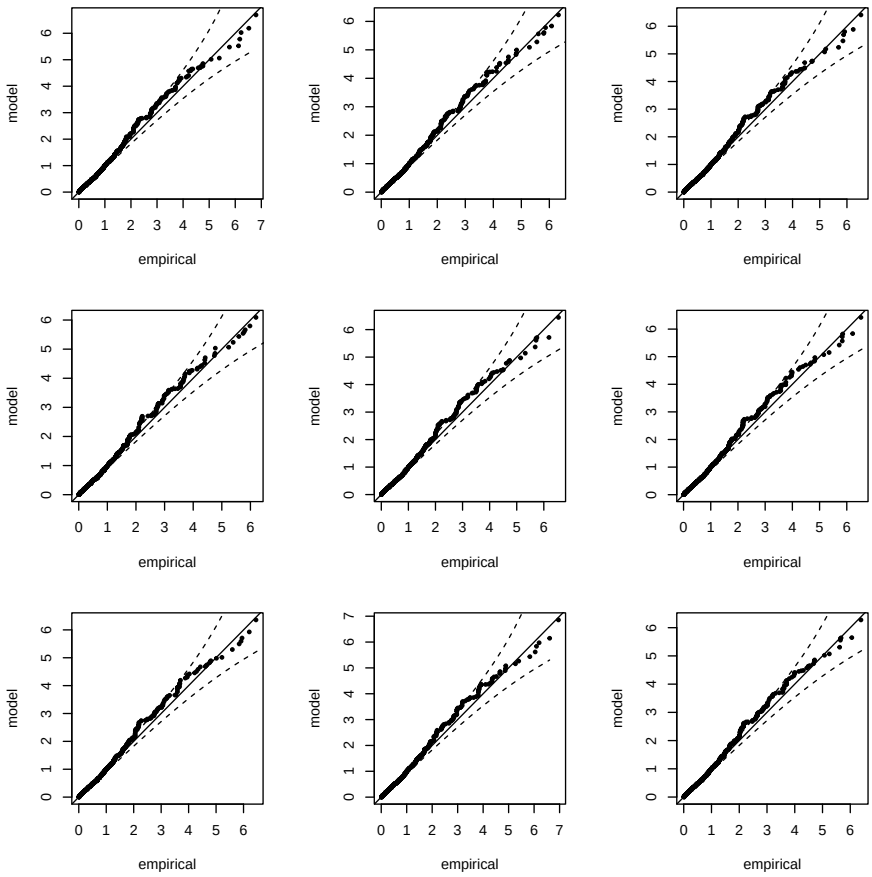

Figure 4.19: QQ plots for the Thames tributary river flow dataset exceedance model for 9 sampled posterior thresholds $r_{\mathcal{Q}_q}(\boldsymbol{w})$, with 95% confidence intervals.

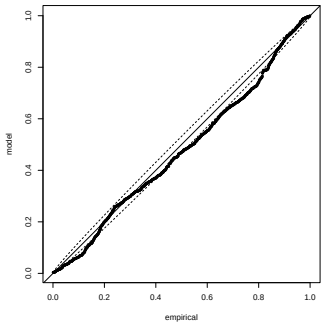

Figure 4.20: PP plots for the Thames tributary river flow dataset directional model, with 95% confidence intervals.

4.7.7 Posterior model fits on Newlyn wave height data

On the three-dimensional Newlyn wave data, we fit models using the three architectures M_1, M_2, and M_3 using data exceeding a high posterior quantile estimate and on all available data. Figure 4.21 show posterior mean limit set boundaries for each of the six fitted models. In these figures, we see that the fitted model associated with the M_2 architecture and fitted on exceedance data only agrees most with $\log n$-scaled data, though all models fit well. The posterior QQ plots in Figure 4.22 corresponding to this optimal model shows good agreement with the radial exceedance model and the underlying exponential distribution, while the PP plots in 4.23 show good agreement between the posterior and the underlying directional models. In Section 4.5 of the main body of this manuscript we see that this model is also very accurate in estimating other diagnostics. One diagnostic not mentioned in Section 4.5 of the main body is plots of $\chi_q(A)$, defined in equation (4.1) in Section 4.1 of the main body. Figure 4.10 shows posterior mean estimates of $\chi_q(A)$ for $A \in \{HP, HS, PS, HPS\}$ and for $q \in (0.9, 1)$ with posterior 95% confidence intervals. Wadsworth and Campbell (2024) establish that the variable groups HP, PS, and HPS are asymptotically independent, and that HS is asymptotically dependent. With this in mind, we conclude that our posterior joint model is able to accurately describe the extremal dependence structure presented in this dataset. Figure 4.10 correctly shows posterior estimates of $\chi_q(A)$ tending to 0 as q tends to 1 for the asymptotically independent groups, and tending to a nonzero value close to the empirical estimate for the asymptotically dependent pair HS.

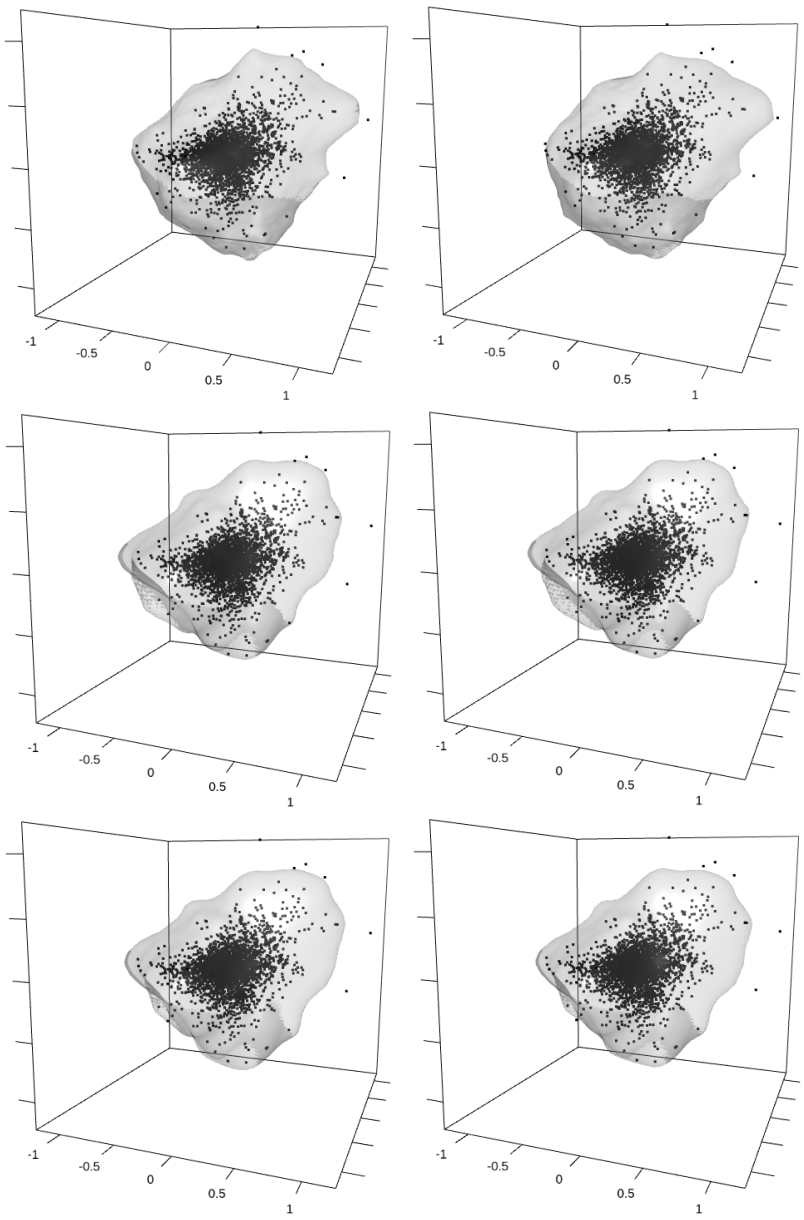

Figure 4.21: Posterior mean estimates of the unit level set $\partial\mathcal{G}$ for the wave dataset. Black points are the original data in Laplace margins scaled by $\log(n/2)$. Top row: model M_1. Middle row: model M_2. Bottom row: model M_3. Left: fitting using only exceedances. Right: fitting using all observations. M_1, M_2, and M_3 define the angle density kernel, as described in section 4.3.3.

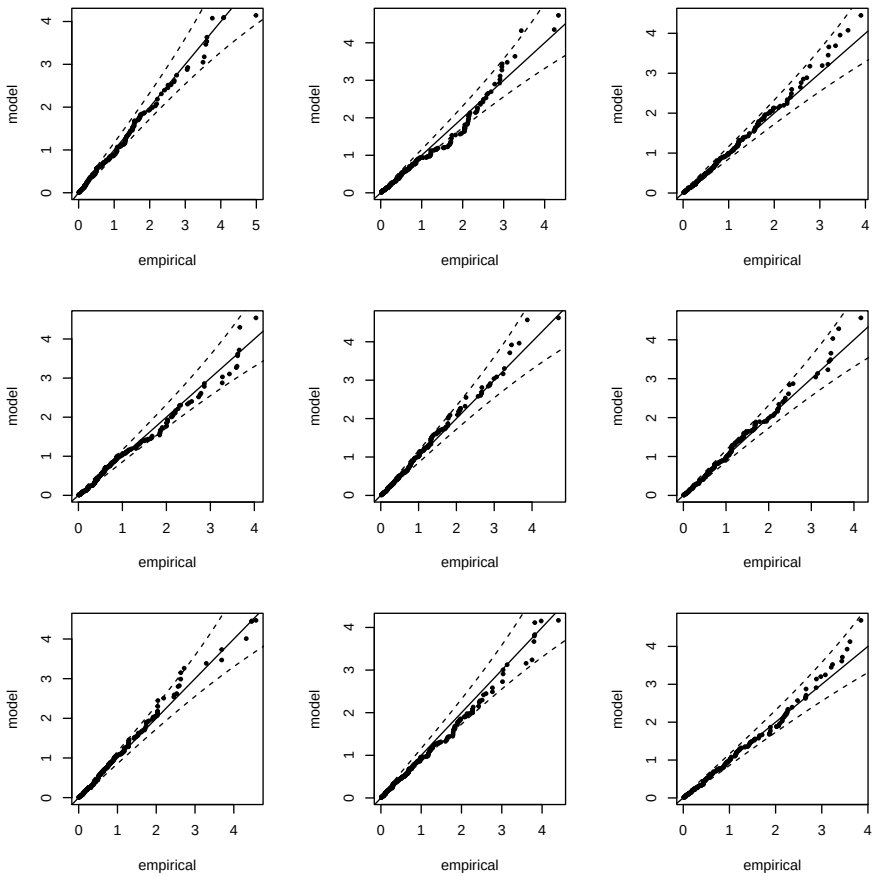

Figure 4.22: QQ plots for the Newlyn wave dataset exceedance model for 9 sampled posterior thresholds $r_{\mathcal{Q}_q}(\boldsymbol{w})$, with 95% confidence intervals.

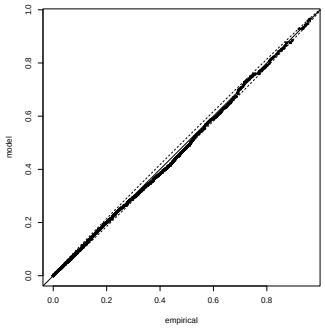

Figure 4.23: PP plots for the Newlyn wave dataset directional model, with 95% confidence intervals.

Chapter 5

Piecewise-linear modelling of multivariate geometric extremes

Abstract

A recent development in extreme value modelling uses the geometry of the dataset to perform inference on the multivariate tail. A key quantity in this inference is the gauge function, whose values define this geometry. Methodology proposed to date for capturing the gauge function either lacks flexibility due to parametric specifications, or relies on complex neural network specifications in dimensions greater than three. We propose a semiparametric gauge function that is *piecewise-linear*, making it simple to interpret and provides a good approximation for the true underlying gauge function. This linearity also makes optimization tasks computationally inexpensive. The piecewise-linear gauge function can be used to define both a radial and an angular model, allowing for the joint fitting of extremal pseudo-polar coordinates, a key aspect of this geometric framework. We further expand the toolkit for geometric extremal modelling through the estimation of high radial quantiles at given angular values via kernel density estimation. We apply the new methodology to air pollution data, which exhibits a complex extremal dependence structure.

5.1 Introduction

5.1.1 Multivariate geometric extremes

In multivariate extreme value analysis, interest lies in characterizing the extremal dependence structure of random vectors. Let $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ be a d-dimensional random vector with components X_j representing measurements of a simultaneous process. For example, $\boldsymbol{X}$ may comprise measurements of d different air pollutants at a single site, contemporaneous river flows at d locations, or values of d different stock returns. Such multivariate vectors can exhibit complex dependence structures, with some variables experiencing simultaneous extremes while others are smaller. Recently, the framework of geometric extremes has emerged as a tool for modelling extremes of potentially complex dependence structures (Wadsworth and Campbell, 2024; Papastathopoulos et al., 2025). When $\boldsymbol{X}$ has common light-tailed margins (i.e., satisfy a von Mises condition), often achieved via a transformation, a useful geometric interpretation of multivariate extremes arises.

For a wide variety of distributions, the scaled sample cloud of independent copies of light-tailed random vectors, $\{\boldsymbol{X}_1/r_n, \ldots, \boldsymbol{X}_n/r_n\}$, converges onto a *limit set* G (Davis et al., 1988; Kinoshita and Resnick, 1991; Balkema and Nolde, 2010). The limit set can be characterized by the gauge function, g, where $g : \mathbb{R}^d \to \mathbb{R}$ is 1-homogeneous, through the relation $G = \left\{\boldsymbol{x} \in \mathbb{R}^d : g(\boldsymbol{x}) \leq 1\right\}$. The scaling sequence r_n depends on the margins. In standard exponential or Laplace margins, for example, a suitable scaling factor is $r_n = \log n$, in which case the coordinatewise supremum of G is given by $(1, \ldots, 1)^\top$. The coordinatewise infimum is $(0, \ldots, 0)^\top$ for exponential margins or $(-1, \ldots, -1)^\top$ for Laplace margins. This means that $g(\boldsymbol{x}) \geq \|\boldsymbol{x}\|_\infty$, where $\|\cdot\|_\infty$ is the max-norm. When $\boldsymbol{X}$ has exponential or Laplace margins, density $f_{\boldsymbol{X}}(\boldsymbol{x})$, and g is continuous, the gauge function can be obtained via

$$g(\boldsymbol{x}) = \lim_{t \to \infty} -\log f_{\boldsymbol{X}}(t\boldsymbol{x})/t, \tag{5.1}$$

(Balkema and Nolde, 2010; Nolde and Wadsworth, 2022). The boundary ∂G of the limit set is given by the unit level set of the gauge function $g(\boldsymbol{x}) = 1$. Nolde

(2014) and Nolde and Wadsworth (2022) show how g can be used to describe the extremal dependence structure of known distributions, while Wadsworth and Campbell (2024) introduced methodology to perform inference with g. In contrast to alternative statistical methods for multivariate extremes, inference based on this new geometric framework can capture highly complex extremal dependence structures and permits extrapolation in regions where only some variables are large simultaneously. Therefore, estimating g, or equivalently G, is crucial to multivariate extremal inference.

The limit set, and therefore the gauge function, provide us with a useful description of the extremal dependence structure of the random vector $\boldsymbol{X}$ by telling us which groups of variables exhibit simultaneous extremes while the remaining variables are of smaller order. Let $D = \{1, \ldots, d\}$ and $C \subseteq D$. We say that the variables in group C can be simultaneously extreme while the others are smaller if there exists $\boldsymbol{z}^C$ such that $g(\boldsymbol{z}^C) = 1$, where $z_j^C = 1$ for all $j \in C$ and $z_j^C = \gamma_j$ for all $j \in D \setminus C$, for some $\gamma_j \in [0, 1)$ in exponential margins or $\gamma_j \in [-1, 1)$ in Laplace margins. Note these are points of intersection of the limit set boundary with the boundary box $[0, 1]^d$ or $[-1, 1]^d$. The collection of sets $C \subseteq D$ with $g(\boldsymbol{z}^C) = 1$ is denoted by $\mathcal{C}$. Each variable must be represented at least once in $\mathcal{C}$, since the coordinatewise supremum is $(1, \ldots, 1)^\top$. Figure 5.1 displays bivariate examples with $\mathcal{C} = \{\{1\}, \{2\}\}$ and $\mathcal{C} = \{\{1, 2\}\}$. Examples for $d = 3$ are also displayed with $\mathcal{C} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$ and $\mathcal{C} = \{\{1, 2, 3\}\}$. Note that our definition of simultaneous extremes based on g is slightly different to definitions that arise in the framework of multivariate regular variation (e.g., Goix et al. (2017)), although the two will often overlap theoretically and are essentially indistinguishable at a practical level. We therefore prefer this simple definition when working in the geometric framework.

To perform statistical inference using the gauge function, it is useful to consider the radial-angular decomposition $\boldsymbol{X} \longmapsto (R, \boldsymbol{W}) = (\|\boldsymbol{X}\|, \boldsymbol{X}/\|\boldsymbol{X}\|) \in \mathbb{R}_+ \times \mathcal{S}_{d-1}$, where $\|\cdot\|$ is a norm and $\mathcal{S}_{d-1} = \left\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| = 1\right\}$. We emphasize here that

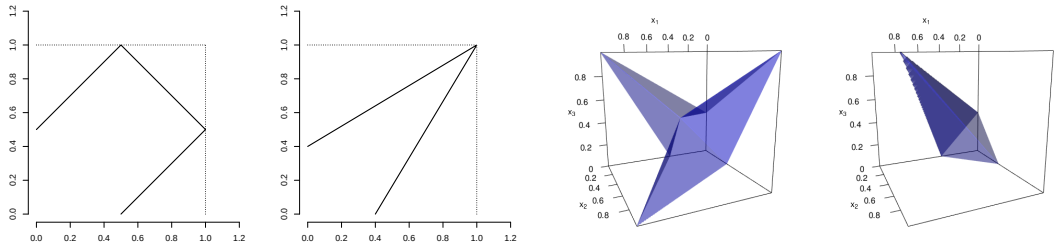

Figure 5.1: Illustration of limit set boundaries and their interpretation in terms of simultaneous extremes. Left: Example where extreme events occur separately. Center-left: Example where extreme events occur together. Center-right: Example where only pairs of variables grow large together. Right: Example where all three variables grow large simultaneously.

this decomposition is made in light-tailed margins, and therefore is fundamentally different to radial-angular decompositions in heavy-tailed margins, which are a mainstay of "classical" multivariate extremes, in which multivariate regular variation plays a key role (see, for example, Beirlant et al. (2006), Chapter 8). When appropriate, the the L_1 norm is preferred for its simplicity, defined by $\|\boldsymbol{x}\|_1 = \sum_{j=1}^{d} |x_j|$. In the radial-angular framework, we are interested in $\boldsymbol{X} = R\boldsymbol{W}$ when the conditional variable $R \mid \boldsymbol{W}$ achieves large values. Wadsworth and Campbell (2024) explain that, when working with exponential-tailed variables and the L_1 norm, the limiting behaviour in equation (5.1) leads to the asymptotic distribution of $R \mid \{\boldsymbol{W} = \boldsymbol{w}\}$ with density $f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1} \exp\{-rg(\boldsymbol{w})\,[1 + o(1)]\}$ as $r \to \infty$, where $g(\boldsymbol{w})$ is the gauge function corresponding to the joint distribution of $(R, \boldsymbol{W})$ evaluated at $\boldsymbol{w} \in \mathcal{S}_{d-1}$. However, they showed that in a wide variety of examples, the same asymptotic form also holds with the $[1+o(1)]$ outside of the exponent, i.e.,

$$f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1} \exp\{-rg(\boldsymbol{w})\}\,[1 + o(1)] \quad \text{as } r \to \infty. \tag{5.2}$$

This suggests that a gamma model is asymptotically appropriate for large values of $R \mid \boldsymbol{W}$. This limiting density is shown to hold very broadly, although for the multivariate Gaussian dependence structure, the shape parameter of the gamma distribution also depends on $\boldsymbol{w}$. Nonetheless, Wadsworth and Campbell (2024)

demonstrate that it is typically not problematic to assume that this parameter is constant. Given $\tau \in (0, 1)$ close to 1 and setting $r_\tau(\boldsymbol{w})$ as the τ^{th} quantile value of $R \mid \{\boldsymbol{W} = \boldsymbol{w}\}$, Wadsworth and Campbell (2024) model $R \mid \{\boldsymbol{W} = \boldsymbol{w}, R > r_\tau(\boldsymbol{w})\}$ with a truncated gamma distribution with rate parameter $g(\boldsymbol{w})$, using parametric forms of the gauge function g derived from known copulas to perform statistical inference with high accuracy, showing how to use this fitted model for inference on extremal probabilities through simulation of $\boldsymbol{X} \mid \{R > r_\tau(\boldsymbol{W})\}$.

5.1.2 Semiparametric estimation of the gauge function

For dimensions $d \geq 3$ in particular, the current suite of parametric models employed in Wadsworth and Campbell (2024) may not be sufficiently flexible for real datasets, where the dependence structure can be complicated. Therefore, the natural consideration is to develop semiparametric approaches for approximating g. Simpson and Tawn (2024a), Majumder et al. (2025), and Papastathopoulos et al. (2025) all aim to do this, the latter two in a Bayesian manner. The methods in Simpson and Tawn (2024a) and Majumder et al. (2025) approximate g in the bivariate case, and use these estimates to describe the underlying tail dependence structure. The theoretical guarantees in Papastathopoulos et al. (2025) are generalized for d-dimensions, and Bayesian inference using the integrated nested Laplace approximation (INLA) is suitable for problems of dimension 2 or 3. Simpson and Tawn (2024a) and Papastathopoulos et al. (2025) model exceedances $\{R - r_\tau(\boldsymbol{w})\} \mid \{\boldsymbol{W} = \boldsymbol{w}, R > r_\tau(\boldsymbol{w})\}$ using a generalized Pareto distribution in place of the truncated gamma distribution. When far enough into the tails, both choices should perform well, but the truncated gamma form may be more accurate at finite levels, especially for larger d. Deep learning methods to estimate the shape of the limit set have been explored for statistical inference using feed-forward neural networks and generative modelling through normalizing flows in Murphy-Barltrop et al. (2024b) and De Monte et al. (2025), respectively, indicating potential for higher-dimensional inference.

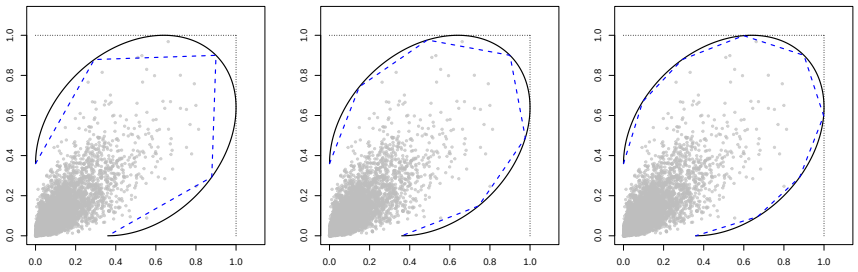

Figure 5.2: $\log n$-scaled bivariate Gaussian data in standard exponential margins, with true limit set boundary given by the solid line. The piecewise-linear limit set boundary is given by the dashed line using 5, 7, and 9 reference angles (left to right).

In this work, we present a simple and interpretable piecewise-linear representation of g. Given a choice of reference angles in $\mathcal{S}_{d-1}$, the parameters of g define the distance from the origin to the boundary of G at these reference angles. The value of g is given by linear interpolation between these points. When $d \geq 3$, a triangulation is also required to define the linear interpolation; we use a Delaunay triangulation. The simple construction of our gauge function has numerous benefits. The main benefit is a model for large radii and exceedance angles can be easily obtained with quick convergence to maximum likelihood estimates for parameters. Figure 5.2 demonstrates what our proposed limit set boundary would look like in the bivariate setting when working in exponential margins. The piecewise-linear limit set boundary represents a rough approximation of the truth, derived using the limit (5.1). The approximation is closer to the truth as the number of reference angles increases; however, this increases the number of parameters to estimate leading to a typical bias-variance trade-off. This is addressed in our work using a regularization approach. Piecewise-linear approaches have been used recently in different contexts for extreme value analysis. Barlow et al. (2023) define non-stationary shape and scale parameters of the generalized Pareto distributions in a piecewise-linear manner in the presence of covariates for univariate peaks-over-threshold modelling. Winterstein et al. (1993), Huseby et al. (2013), and Mackay and de Hauteclocque (2023) all use linearisations of the $\mathbb{R}^d$ to estimate and visualize environmental contours, which are

multivariate sets used to approximate the occurrence of extreme events.

As in Wadsworth and Campbell (2024), we model large radii using the truncated gamma distribution with rate parameter given by the gauge function value. This leads to a likelihood from which the parameters of the piecewise-linear gauge can be estimated. When simulating from the distribution of $\boldsymbol{X} \mid \{R > r_\tau(\boldsymbol{W})\}$, two things are required: the distribution of $\boldsymbol{W} \mid \{R > r_\tau(\boldsymbol{W})\}$ and the distribution of $R \mid \{\boldsymbol{W} = \boldsymbol{w}, R > r_\tau(\boldsymbol{w})\}$. The second of these is given by the truncated gamma model, while for the first, Wadsworth and Campbell (2024) used the empirical distribution of $\boldsymbol{W} \mid \{R > r_\tau(\boldsymbol{W})\}$. A natural progression, particularly for higher dimensions, is also to estimate a semiparametric form for this. We propose an angular density inspired by the homothetic case presented in Balkema and Nolde (2010), with joint density equivalent to the one used in Papastathopoulos et al. (2025). Specifically, a valid density for $\boldsymbol{W}$ over $\mathcal{S}_{d-1}$ is $f_{\boldsymbol{W}}(\boldsymbol{w}) = g(\boldsymbol{w})^{-d}/\{d\text{vol}(G)\}$, where g is a gauge function for the set G. We emphasize here that this density can be expressed and fitted independently of $R \mid \boldsymbol{W}$, so that the gauge function g in the angular density need not correspond to the gauge function in equation (5.2). This is discussed in greater detail in Section 5.4.1. The major advantage to our piecewise-linear setting is that the normalizing constant for this joint density has an explicit form, rendering its estimation simple.

When standardizing the margins of a given dataset, the choice of marginal distribution is a nuanced one. If negative dependence arises, then the limit definition of the gauge function in equation (5.1) may not hold on the axes in exponential margins, and Laplace margins are preferred for revealing greater structure. Furthermore, if the dataset has negative values in its original margins, then it may be more intuitive to use Laplace margins, as it preserves the domain to all orthants of $\mathbb{R}^d$. In many cases, real datasets are positive-valued and have no negative associations, and so exponential margins are a suitable choice. While the methodology presented is suitable for data with any choice of von Mises margins, we choose to model data in standard exponential margins for simplicity.

In Appendix 5.8.10, we explain the differences required to work in standard Laplace margins.

The outline of the paper is as follows. In Section 5.2, we detail the piecewise-linear construction of a gauge function, and consider the calculation of the angular density normalizing constant, vol(G). As we only consider radial values above a threshold, a good estimate of $r_\tau(\boldsymbol{w})$ is needed. The empirical version presented in Wadsworth and Campbell (2024) and quantile regression techniques are not currently well-suited to higher dimensions. In Section 5.3, we propose a new approach for the estimation of $r_\tau(\boldsymbol{w})$ based on kernel density estimation. Section 5.4 covers how we fit our piecewise-linear models including a regularization technique for when there are many parameters. We also consider diagnostics and probability estimation techniques using our models. In Section 5.5, simulation studies show that repeated fits of the piecewise-linear model are comparable to parametric models where knowledge of the true copula is exploited. Section 5.6 details an application to extremes of four air pollutants, demonstrating that the piecewise-linear model can be used to perform statistical inference in dimensions where other semiparametric methods struggle.

In order to select the best hyperparameters for high quantile estimation, model penalization, as well as assessing probability estimates, we assess our models on datasets generated from the following multivariate distributions, all in standard exponential margins.

(I) $d = 2$ logistic with dependence parameter $\alpha = 0.4$. $\mathcal{C} = \{\{1, 2\}\}$.

(II) $d = 2$ logistic with dependence parameter $\alpha = 0.8$. $\mathcal{C} = \{\{1, 2\}\}$, but with weaker dependence than (I).

(III) $d = 2$ Gaussian distribution, correlation $\rho = 0.8$. $\mathcal{C} = \{\{1\}, \{2\}\}$.

(IV) $d = 2$ inverted logistic with dependence parameter $\alpha = 0.7$. $\mathcal{C} = \{\{1\}, \{2\}\}$.

(V) $d = 3$ asymmetric logistic with dependence parameters $\alpha_{\{1,2\}} = \alpha_{\{1,3\}} = \alpha_{\{2,3\}} = 0.4$. $\mathcal{C} = \{\{1, 2\}, \{2, 3\}, \{1, 3\}\}$.

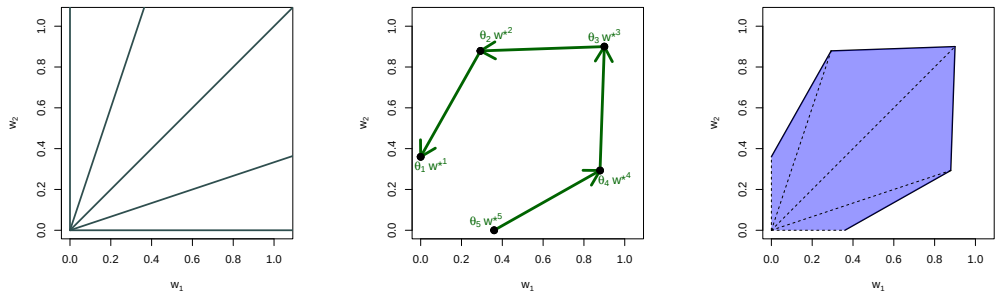

Figure 5.3: Piecewise-linear gauge function construction illustration for $d = 2$. Left: Partition of $\mathcal{S}_1$ with reference angles $w^{\star 1} = 0$, $w^{\star 2} = 0.25$, $w^{\star 3} = 0.5$, $w^{\star 4} = 0.75$, and $w^{\star 5} = 1$. Middle: Coplanar vectors $\boldsymbol{C}^{(1)}$, $\boldsymbol{C}^{(2)}$, $\boldsymbol{C}^{(3)}$, and $\boldsymbol{C}^{(4)}$ (green arrows) for pre-set parameter values $\theta_1 = 0.36$, $\theta_2 = 1.17$, $\theta_3 = 1.8$, $\theta_4 = 1.17$, and $\theta_5 = 0.36$. Right: Limit set at chosen parameter values. Solid line indicates the unit level set of the piecewise-linear gauge function at the chosen parameter and reference angle values. Dashed lines indicate the distances dictated by the parameter values.

(VI) $d = 3$ asymmetric logistic, with dependence parameters $\alpha_{\{1\}} = \alpha_{\{1,2\}} = \alpha_{\{2,3\}} = 0.4$. $\mathcal{C} = \{\{1\}, \{1,2\}, \{2,3\}\}$.

(VII) $d = 3$ equally-weighted mixture of asymmetric logistic and Gaussian. The Gaussian correlations are $\rho_{12} = \rho_{13} = \rho_{23} = 0.6$ and asymmetric logistic dependence parameters $\alpha_{\{1,2\}} = \alpha_{\{1,2,3\}} = 0.4$. $\mathcal{C} = \{\{1\}, \{2\}, \{3\}, \{1,2\}, \{1,2,3\}\}$.

A more detailed catalogue of these distributions is presented in Appendix 5.8.2.

5.2 A piecewise-linear model

5.2.1 The piecewise-linear gauge function

Construction of the piecewise-linear gauge function, denoted g_{PWL}, relies on segmenting the simplex $\mathcal{S}_{d-1}$ using N nodes called *reference angles*. A parameter value is assigned at each reference angle, defining the distance from the origin to the limit set boundary at that angle: for each reference angle $\boldsymbol{w}^\star \in \mathcal{S}_{d-1}$, the parameter

corresponds to the value $1/g_{\text{PWL}}(\boldsymbol{w}^\star)$. Because the true gauge function g satisfies $g(\boldsymbol{x}) \geq \|\boldsymbol{x}\|_\infty$, a parameter at location $\boldsymbol{w}^\star$ has an upper bound of $1/\|\boldsymbol{w}^\star\|_\infty$. Depending on the dimension, straight lines, planes, or hyperplanes are used to connect these limit set boundary values.

To provide intuition into the general approach, we begin by illustrating for $d = 2$. Taking the L_1 norm in our radial-angular decomposition, reference angles can be defined by scalar values in $[0, 1]$. Let $\left\{w^{\star 1}, w^{\star 2}, \ldots, w^{\star N}\right\}$ be an increasing sequence of reference angles such that $w^{\star 1} = 0$ and $w^{\star N} = 1$. This partition of the interval $[0, 1]$ has $N - 1$ segments with vertices $\left\{w^{\star 1}, w^{\star 2}\right\}, \left\{w^{\star 2}, w^{\star 3}\right\}, \ldots, \left\{w^{\star N-1}, w^{\star N}\right\}$. Further let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)^\top$ be N positive parameter values with $\theta_k = 1/g_{\text{PWL}}(w^{\star k}, 1 - w^{\star k})$. To define g_{PWL}, we consider coplanar vectors, $\boldsymbol{C}^{(k)} = \left(\theta_k w^{\star k} - \theta_{k+1} w^{\star k+1}, \theta_k(1 - w^{\star k}) - \theta_{k+1}(1 - w^{\star k+1})\right)^\top$, $k = 1, \ldots, N - 1$, an example of which is displayed in Figure 5.3. The equation of the line interpolating from $\theta_k \left(w^{\star k}, 1 - w^{\star k}\right)^\top$ to $\theta_{k+1} \left(w^{\star k+1}, 1 - w^{\star k+1}\right)^\top$ is $\left(C_2^{(k)} x - C_1^{(k)} y\right) / \left(C_2^{(k)} \theta_k w^{\star k} - C_1^{(k)} \theta_k (1 - w^{\star k})\right)$; therefore, $g_{\text{PWL}}(x, y; \boldsymbol{\theta})$ is given by

$$\sum_{k=1}^{N-1} \mathbf{1}_{(w^{\star k}, w^{\star k+1})} \left(\frac{x}{x+y}\right) \frac{\left[\theta_k(1 - w^{\star k}) - \theta_{k+1}(1 - w^{\star k+1})\right] x - \left[\theta_k w^{\star k} - \theta_{k+1} w^{\star k+1}\right] y}{\left[\theta_k(1 - w^{\star k}) - \theta_{k+1}(1 - w^{\star k+1})\right] \theta_k w^{\star k} - \left[\theta_k w^{\star k} - \theta_{k+1} w^{\star k+1}\right] \theta_k (1 - w^{\star k})} \tag{5.3}$$

for $(x, y) \in \mathbb{R}_+^2$, where $\mathbf{1}_A(x)$ is an indicator function with value 1 if $x \in A$ and 0 otherwise. Figure 5.3 displays an example in which a piecewise-linear gauge function is used to approximate the gauge function corresponding to the bivariate Gaussian distribution, $g_{\text{N}}(x, y; \rho) = (1 - \rho^2)^{-1} \left(x + y - 2\rho(xy)^{1/2}\right)$. The $N = 5$ reference angles correspond to an equally-spaced mesh, and parameters are set to $\theta_k = 1/g_{\text{N}}(w^{\star k}, 1 - w^{\star k}; \rho)$. The resulting limit set in Figure 5.3 does not satisfy the coordinatewise supremum property because of the absence of a reference angle $w^\star$ such that $g_{\text{N}}(w^\star, 1 - w^\star; \rho) = \left\|(w^\star, 1 - w^\star)^\top\right\|_\infty$. When performing inference, we will develop an algorithm to ensure that limit set estimates can have the coordinatewise supremum $(1, \ldots, 1)^\top$.

In dimensions $d \geq 3$, denote the set of $N \geq d$ reference angles $\left\{\boldsymbol{w}^{\star 1}, \ldots, \boldsymbol{w}^{\star N}\right\}$,

each lying in the simplex $\mathcal{S}_{d-1}$. We partition $\mathcal{S}_{d-1}$ using a Delaunay triangulation (Delaunay, 1934) with a point set based on the N reference angles. Given a set of N $(d-1)$-dimensional reference angles, the Delaunay triangulation creates a partition of $\mathcal{S}_{d-1}$ comprised of M regions with vertices given by these reference angles. These regions are constructed in such a way that their surface area (or volume) is maximized, leading to no insignificant segments. While there is no direct one-to-one correspondence between the number of reference angles N and the number of partitions M in the resulting Delaunay triangulation, it is known that M is between $O(N)$ and $O(N^{\lceil (d-1)/2 \rceil})$ in general d-dimensions. When $d = 3$, the stronger statement of $N-2 \leq M \leq 2N-5$ holds (Ber et al., 2008). Each region of the Delaunay triangulation, $\triangle^{(k)} \subset \mathcal{S}_{d-1}$, $k \in \{1, \ldots, M\}$, is defined by d vertices $\boldsymbol{w}^{\star(k),1}, \ldots, \boldsymbol{w}^{\star(k),d}$. Given the parameter values $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)^\top \in \mathbb{R}^N_+$, we define $\boldsymbol{\theta}^{(k)} = (\theta_1^{(k)}, \ldots, \theta_d^{(k)})^\top$ as the parameters from $\boldsymbol{\theta}$ associated with the d vertices of $\triangle^{(k)}$. Define the $(d-1) \times d$ *coplanar matrix* $\mathbf{C}^{(k)}$ for triangulation $k \in \{1, \ldots, M\}$ where the i^{th} row is given by the vector

$$\mathbf{C}^{(k)}_{i,\cdot} = \theta_1^{(k)} \left(\boldsymbol{w}^{\star(k),1}\right)^\top - \theta_{i+1}^{(k)} \left(\boldsymbol{w}^{\star(k),i+1}\right)^\top \in \mathbb{R}^d \; ; \;\; i = 1, \ldots, d-1.$$

For an arbitrary angle $\boldsymbol{w} \in \triangle^{(k)}$, $g_{\text{PWL}}(\boldsymbol{w}; \boldsymbol{\theta}) = \boldsymbol{n}^{(k)\top} \boldsymbol{w} / \boldsymbol{n}^{(k)\top} \theta_1^{(k)} \boldsymbol{w}^{\star(k),1}$, where the normal vector $\boldsymbol{n}^{(k)} \in \mathbb{R}^d$ to the plane defined by vertices $\theta_1^{(k)} \boldsymbol{w}^{\star(k),1}, \ldots, \theta_d^{(k)} \boldsymbol{w}^{\star(k),d}$ is

$$\boldsymbol{n}^{(k)} = \sum_{j=1}^{d} (-1)^{j+1} \det\left(\mathbf{C}^{(k)}_{\cdot,-j}\right) \boldsymbol{e}_j \, . \tag{5.4}$$

In equation (5.4), $\mathbf{C}^{(k)}_{\cdot,-j}$ is the matrix $\mathbf{C}^{(k)}$ with the j^{th} column removed, and $\boldsymbol{e}_j$ the j^{th} standard unit vector, a vector of length d of zeros except for a 1 in the j^{th} entry. Performing the summation over all regions in the Delaunay triangulation gives the proposed piecewise-linear gauge function

$$g_{\text{PWL}}(\boldsymbol{x}; \boldsymbol{\theta}) = \sum_{k=1}^{M} \mathbf{1}_{\triangle^{(k)}} \left(\boldsymbol{x} / \|\boldsymbol{x}\|\right) \frac{\boldsymbol{n}^{(k)\top} \boldsymbol{x}}{\boldsymbol{n}^{(k)\top} \theta_1^{(k)} \boldsymbol{w}^{\star(k),1}}, \qquad \boldsymbol{x} \in \mathbb{R}^d_+, \tag{5.5}$$

which is 1-homogeneous and continuous in $\boldsymbol{x}$. Note that the formulation of the $d = 2$ case in equation (5.3) is covered by equation (5.5), but was described separately to

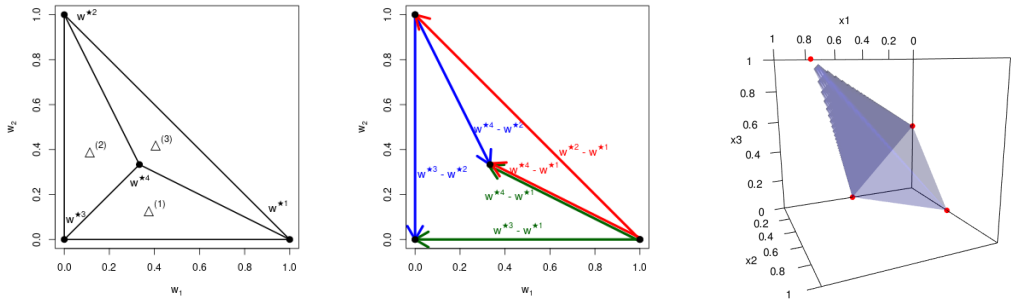

Figure 5.4: Piecewise-linear gauge function construction illustration for $d = 3$. Left to right: Delaunay triangulation based on $N = 4$ chosen reference angles; the resulting coplanar vectors; limit set boundary.

give an intuition in the bivariate setting.

As an illustration, suppose $d = 3$ and $N = 4$, with $\boldsymbol{w}^{\star 1}, \boldsymbol{w}^{\star 2}, \boldsymbol{w}^{\star 3}$ chosen to lie on each of the vertices, and $\boldsymbol{w}^{\star 4}$ in the center of the $\mathcal{S}_2$ simplex. As displayed in Figure 5.4, the resulting Delaunay triangulation gives $M = 3$ regions: $\triangle^{(1)}$ has vertices $\{\boldsymbol{w}^{\star 1}, \boldsymbol{w}^{\star 3}, \boldsymbol{w}^{\star 4}\}$, $\triangle^{(2)}$ has vertices $\{\boldsymbol{w}^{\star 2}, \boldsymbol{w}^{\star 3}, \boldsymbol{w}^{\star 4}\}$, and $\triangle^{(3)}$ has vertices $\{\boldsymbol{w}^{\star 1}, \boldsymbol{w}^{\star 2}, \boldsymbol{w}^{\star 4}\}$. Each region has two coplanar vectors that make up the following coplanar matrices of dimension 2×3,

$$\mathbf{C}^{(1)} = \begin{pmatrix} \theta_1 \left(\boldsymbol{w}^{\star 1}\right)^\top - \theta_3 \left(\boldsymbol{w}^{\star 3}\right)^\top \\ \theta_1 \left(\boldsymbol{w}^{\star 1}\right)^\top - \theta_4 \left(\boldsymbol{w}^{\star 4}\right)^\top \end{pmatrix}, \quad \mathbf{C}^{(2)} = \begin{pmatrix} \theta_2 \left(\boldsymbol{w}^{\star 2}\right)^\top - \theta_3 \left(\boldsymbol{w}^{\star 3}\right)^\top \\ \theta_2 \left(\boldsymbol{w}^{\star 2}\right)^\top - \theta_4 \left(\boldsymbol{w}^{\star 4}\right)^\top \end{pmatrix},$$

$$\mathbf{C}^{(3)} = \begin{pmatrix} \theta_1 \left(\boldsymbol{w}^{\star 1}\right)^\top - \theta_2 \left(\boldsymbol{w}^{\star 2}\right)^\top \\ \theta_1 \left(\boldsymbol{w}^{\star 1}\right)^\top - \theta_4 \left(\boldsymbol{w}^{\star 4}\right)^\top \end{pmatrix},$$

the rows of which are represented by the arrows in Figure 5.4. For $\boldsymbol{\theta} = (0.5, 0.5, 0.5, 3)^\top$, the unit level set of g_{PWL} is evaluated and is also displayed in Figure 5.4.

5.2.2 Angular model

Given that the geometric approach uses a pseudo-radial-angular decomposition, it is desirable to model the distribution of angles $\boldsymbol{W} \mid \{R > r_\tau(\boldsymbol{W})\}$ with a flexible

semiparametric model. This should reduce issues with the curse of dimensionality that can arise when using the empirical distribution of $\boldsymbol{W} \mid \{R > r_\tau(\boldsymbol{W})\}$ in higher dimensions, and even in lower dimensions may be helpful for ensuring the ability to estimate non-zero extremal probabilities. The form of $f_{\boldsymbol{W}}$ given in Section 5.1 arises as the exact angular density for a certain type of homothetic density (Balkema and Nolde, 2010). Specifically, given a gauge function g, a valid joint density is $f_{\tilde{\boldsymbol{X}}}(\boldsymbol{x}) = \exp\{-g(\boldsymbol{x})\}/\{d!\mathrm{vol}(G)\}$, as considered in some examples of Nolde and Wadsworth (2022). The margins $\tilde{\boldsymbol{X}}$ are near-exponential sub-asymptotically, and are exactly exponential asymptotically. For $R = \|\tilde{\boldsymbol{X}}\|$ and $\boldsymbol{W} = \tilde{\boldsymbol{X}}/\|\tilde{\boldsymbol{X}}\|$, we have $f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) = r^{d-1}\exp\{-rg(\boldsymbol{w})\}/\{d!\mathrm{vol}(G)\}$ and $f_{\boldsymbol{W}}(\boldsymbol{x}) = g(\boldsymbol{x})^{-d}/\{d\mathrm{vol}(G)\}$. This suggests that if the extremes of $\boldsymbol{X}$ are well-approximated by the density $f_{\tilde{\boldsymbol{X}}}$ then the angles $\boldsymbol{W}$ might be well-approximated by the density $f_{\boldsymbol{W}}$, where the gauge function is the same as the one in the gamma distribution of $R \mid \boldsymbol{W}$. However, when this approximation is poor, this still presents a way to construct a flexible model for $f_{\boldsymbol{W}}$ via a gauge function g that can be parametrized independently of that used in the truncated gamma distribution. Papastathopoulos et al. (2025) and De Monte et al. (2025) both use this form of $f_{\boldsymbol{W}}$ for probability estimation, the former in the Bayesian context where the gauge function is modelled by Matérn Gaussian random fields and the latter in the normalizing flows framework.

When fitting the model $f_{\boldsymbol{W}}$ via maximum likelihood estimation, the computation of $\mathrm{vol}(G)$ needs to be done at every likelihood evaluation, which can be computationally expensive if the form of this volume is not explicit. Numerical integration methods may be possible, but drastically slow down maximum likelihood estimation. Papastathopoulos et al. (2025) estimate the volume during model fitting via a latent variable in the likelihood, and a numerical integration procedure is performed during posterior prediction. Due to the piecewise-linear nature of g_{PWL}, the corresponding $\mathrm{vol}(G_{\mathrm{PWL}})$ is easily obtained, as the unit level set defines the union of M d-dimensional faces. In particular, computing $\mathrm{vol}(G_{\mathrm{PWL}})$ is reduced to solving M determinants of $d \times d$ matrices, whose columns correspond to the vertices of the

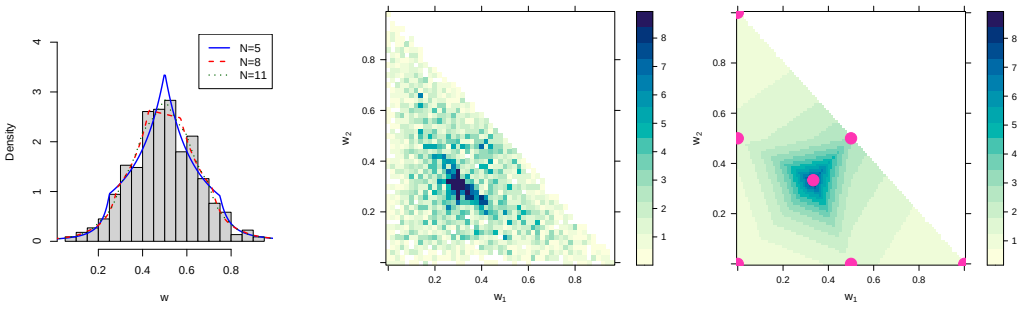

Figure 5.5: Left: Histogram of exceedance angles from bivariate Gaussian data, with fitted $f_{\boldsymbol{W}}(\boldsymbol{w})$ for $N = 5, 8, 11$ reference angles. Center: Histogram of exceedance angles from a $d = 3$ mixture model. Right: A fit of $f_{\boldsymbol{W}}(\boldsymbol{w})$ on this data with reference angles overlaid.

M piecewise-linear regions that make up the set G_{PWL}.

Proposition 5.1. *For the piecewise-linear gauge function* g_{PWL} *in equation* (5.5)*, the volume of the corresponding set* G_{PWL} *is given by*

$$\text{vol}(G_{\text{PWL}}) = \frac{1}{d!} \sum_{k=1}^{M} \left| \det \left(\theta_1^{(k)} \boldsymbol{w}^{\star(k),1} \quad \theta_2^{(k)} \boldsymbol{w}^{\star(k),2} \quad \dots \quad \theta_d^{(k)} \boldsymbol{w}^{\star(k),d} \right) \right|$$

The proof of Proposition 5.1 is given in Appendix 5.8.1.

We illustrate this construction with two examples. First, consider data from distribution (III), Gaussian dependence with exponential margins. Taking $\boldsymbol{W} = \boldsymbol{X}/\|\boldsymbol{X}\|$, we model $f_{\boldsymbol{W}}(\boldsymbol{w}) = g_{\text{PWL}}(\boldsymbol{w})^{-d}/\{d\text{vol}(G_{\text{PWL}})\}$ without any knowledge of the underlying joint distribution. After doing so, Figure 5.5 shows a good fit of the estimated density using regularly-spaced reference angles over the empirical distribution of exceedance angles, with the fit improving as N increases. Secondly, consider distribution (VII), which has a difficult angular structure to capture. Nonetheless, using the piecewise-linear method with $N = 7$ reference angles, a fitted model for $f_{\boldsymbol{W}}$ aligns reasonably well with the empirical distribution, both shown in Figure 5.5. Further diagnostics on samples obtained from the fitted density $f_{\boldsymbol{W}}$ using MCMC for these datasets are presented in Appendix 5.8.6. We note that in all density fits we implement regularization techniques to ensure smoothness

of parameter estimates across neighbouring regions. This is covered in detail in Section 5.4.2.

5.3 High quantile estimation

The truncated gamma distribution for $R \mid \boldsymbol{W}$ holds asymptotically, and therefore is fitted using datapoints which exceed a threshold. To define this threshold, denote exceedance radii as observations from the distribution corresponding to $R \mid \{\boldsymbol{W} = \boldsymbol{w}, R > \tilde{r}(\boldsymbol{w})\}$. For a probability level $\tau \in (0, 1)$ close to 1, a candidate for the radial threshold is the quantile $r_\tau(\boldsymbol{w})$ corresponding to the solution to $F_{R|\boldsymbol{W}}(r_\tau(\boldsymbol{w}) \mid \boldsymbol{w}) = \tau$. This is a natural choice, and, as Wadsworth and Campbell (2024) point out, leads to an independent estimate of the gauge function through the relation $g(\boldsymbol{w}) \approx C_\tau / r_\tau(\boldsymbol{w})$, thus can be helpful for model checking. They consider two approaches to estimate $r_\tau(\boldsymbol{w})$. For $d = 2$ they suggest additive quantile regression (AQR, Fasiolo et al. (2021)), but this could not be extended for $d > 2$ using available methodology due to a lack of basis functions defined on the simplex. A second approach using an empirical binning method was instead implemented when $d \geq 2$. This requires splitting the simplex into overlapping bins and computing the empirical τ quantile of $R \mid \{\boldsymbol{W} = \boldsymbol{w}\}$ in each bin to which the angle $\boldsymbol{w}$ belongs. The radial threshold value at new angles is computed using local means with threshold values already computed in the overlapping bins. Such an approach is not ideal as d increases, as very little data may be observed in certain bins. Furthermore, it provides a very rough approximation of $r_\tau(\boldsymbol{w})$. Papastathopoulos et al. (2025) model $\log(r_\tau(\boldsymbol{w}))$ using a Matérn Gaussian random field, with implementation currently suitable for dimension $d \leq 3$.

Given the semiparametric nature of our proposed piecewise-linear approach to modelling $R \mid \boldsymbol{W}$, a good estimate of the radial quantile across all angles in the simplex is needed to avoid regions of $\mathcal{S}_{d-1}$ about which little is known. To overcome some of the issues that persist in current methods, we develop a new approach based

on kernel density estimation (KDE). This method gives smooth results, akin to the AQR approach, and is better suited for higher dimensions than purely empirical estimation. We begin with the integral

$$F_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) = \int_0^r \frac{f_{R,\boldsymbol{W}}(\tilde{r}, \boldsymbol{w})}{f_{\boldsymbol{W}}(\boldsymbol{w})}\, d\tilde{r}\,, \tag{5.6}$$

and adopt kernel-based estimates for the densities $f_{R,\boldsymbol{W}}$ and $f_{\boldsymbol{W}}$:

$$\widehat{f}_{R,\boldsymbol{W}}(r, \boldsymbol{w}) = \frac{1}{n}\sum_{i=1}^{n} k_R\left(\frac{r - r_i}{h_R}\right) k_{\boldsymbol{W}}\left(\frac{\boldsymbol{w} - \boldsymbol{w}_i}{h_{\boldsymbol{W}}}\right) \frac{1}{h_R}\frac{1}{h_{\boldsymbol{W}}^{d-1}} \tag{5.7}$$

$$\widehat{f}_{\boldsymbol{W}}(\boldsymbol{w}) = \frac{1}{n}\sum_{i=1}^{n} k_{\boldsymbol{W}}\left(\frac{\boldsymbol{w} - \boldsymbol{w}_i}{h_{\boldsymbol{W}}}\right) \frac{1}{h_{\boldsymbol{W}}^{d-1}}\,, \tag{5.8}$$

where $h_R, h_{\boldsymbol{W}} > 0$, are bandwidths (or smoothing parameters), and $k_R, k_{\boldsymbol{W}}$ are kernels. Substituting (5.7) and (5.8) into (5.6),

$$\widehat{F}_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) = \frac{\sum_{i=1}^{n} k_{\boldsymbol{W}}\left(\frac{\boldsymbol{w}-\boldsymbol{w}_i}{h_{\boldsymbol{W}}}\right) \int_0^r k_R\left(\frac{\tilde{r}-r_i}{h_R}\right)\frac{1}{h_R} d\tilde{r}}{\sum_{i=1}^{n} k_{\boldsymbol{W}}\left(\frac{\boldsymbol{w}-\boldsymbol{w}_i}{h_{\boldsymbol{W}}}\right)} = \frac{\sum_{i=1}^{n} k_{\boldsymbol{W}}\left(\frac{\boldsymbol{w}-\boldsymbol{w}_i}{h_{\boldsymbol{W}}}\right) K_R\left(\frac{r-r_i}{h_R}\right)}{\sum_{i=1}^{n} k_{\boldsymbol{W}}\left(\frac{\boldsymbol{w}-\boldsymbol{w}_i}{h_{\boldsymbol{W}}}\right)}\,,$$

where K_R is the distribution function associated with the kernel density k_R. An estimate of $r_\tau(\boldsymbol{w})$ can be obtained by solving for r in $\widehat{F}_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) = \tau$ using numerical inversion methods presented in Brent (2013). The choice of kernel density and bandwidths will affect the smoothness of the estimate of $r_\tau(\boldsymbol{w})$. In practice, we employ the univariate Gaussian kernel for k_R and its multivariate counterpart for $k_{\boldsymbol{W}}$, with identity correlation matrix.

As a performance diagnostic for quantile estimates, we propose a metric based on K-fold cross-validation of the check function, commonly used as the objective function in quantile regression (Koenker and Bassett Jr, 1978). For $K \in \mathbb{N}$, we split the data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ into a fitting set of length $n_{fit} = n - \lfloor n/K \rfloor$ and a evaluation set of length $n_{eval} = \lfloor n/K \rfloor$. On the k^{th} set of fitting data, we obtain a radial threshold $r_\tau^{(k)}(\boldsymbol{w};\, \cdot\,)$, $k = 1, \ldots, K$. In KDE-based estimation, the obtained threshold values depend on the bandwidth $h_{\boldsymbol{W}}$. In the empirical procedure of Wadsworth and Campbell (2024), it depends on the amount of bin overlap. Once obtained, the

mean of the check function is evaluated on the evaluation data, resulting in the following score on the K cross-validation partitions,

$$S(\cdot) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{n_{eval}}\sum_{i=1}^{n_{eval}}\left[r_i - r_\tau^{(k)}(\boldsymbol{w}_i;\ \cdot\ )\right]\left[\tau - \mathbf{1}_{(-\infty,0)}\left(r_i - r_\tau^{(k)}(\boldsymbol{w}_i;\ \cdot\ )\right)\right].$$

The bandwidth or the amount of bin overlap are hyperparameters which control how smooth the radial threshold is. A smoothing parameter whose score $S(\cdot)$ is closest to zero is thought of as the optimal hyperparameter setting.

Appendix 5.8.4 provides a full comparison of the Gaussian kernel to a kernel whose support is compact and the empirical method described above. This shows little difference in the quality of quantile estimates between the methods for $d = 2, 3$, but we prefer the KDE approach for the ability to evaluate $r_\tau(\boldsymbol{w})$ for any $\boldsymbol{w} \in \mathcal{S}_{d-1}$. From a study on $d = 2, 3$ datasets in Appendix 5.8.4, we found that optimal values of $h_{\boldsymbol{W}}$ when using the Gaussian kernel were often in the neighbourhood of 0.05. We also show that varying the radial bandwidth made little difference in the performance of the quantile estimate, so we choose to fix $h_R = 0.05$.

5.4 Inference

5.4.1 Model fitting

Given a dataset comprised of n d-dimensional observations, we first transform the margins to standard exponential. This is achieved via non or semiparametric estimation of the margins. The standardized datapoints $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are then transformed to radii $r_1, \ldots, r_n$ and angles $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$, and an estimate of $r_\tau(\boldsymbol{w}_i)$ is obtained at a high quantile τ using the kernel density estimation approach from Section 5.3. Given that we have well-defined densities for radial and angular components, we can choose to model $\boldsymbol{W}$ and $R \mid \{\boldsymbol{W} = \boldsymbol{w}\}$ separately or jointly

using the following likelihood functions,

$$\begin{aligned} L_{\boldsymbol{W}}(\boldsymbol{\theta}; r_{1:n}, \boldsymbol{w}_{1:n}) &\propto \prod_{i\,:\,r_i > r_\tau(\boldsymbol{w}_i)} \text{vol}(G_{\text{PWL}}(\boldsymbol{\theta}))^{-1} g_{\text{PWL}}(\boldsymbol{w}_i; \boldsymbol{\theta})^{-d} \\ L_{R|\boldsymbol{W}}(\boldsymbol{\theta}; r_{1:n}, \boldsymbol{w}_{1:n}) &\propto \prod_{i\,:\,r_i > r_\tau(\boldsymbol{w}_i)} \frac{f_{\text{Ga}}(r_i; d, g_{\text{PWL}}(\boldsymbol{w}_i; \boldsymbol{\theta}))}{1 - F_{\text{Ga}}(r_\tau(\boldsymbol{w}_i); d, g_{\text{PWL}}(\boldsymbol{w}_i; \boldsymbol{\theta}))} \\ L_{R,\boldsymbol{W}}(\boldsymbol{\theta}; r_{1:n}, \boldsymbol{w}_{1:n}) &= L_{R|\boldsymbol{W}}(\boldsymbol{\theta}) L_{\boldsymbol{W}}(\boldsymbol{\theta}), \end{aligned}$$

where $f_{\text{Ga}}(\cdot; d, g(\boldsymbol{w}))$ and $F_{\text{Ga}}(\cdot; d, g(\boldsymbol{w}))$ are the density and distribution functions of the gamma distribution with shape parameter d and rate parameter $g_{\text{PWL}}(\boldsymbol{w}; \boldsymbol{\theta})$. In this setting, there are two ways to fit a joint model for $(R, \boldsymbol{W}) \mid \{R > r_\tau(\boldsymbol{W})\}$. One can implement a joint approach by maximizing $L_{R,\boldsymbol{W}}$. This results in lower variability of parameter estimates of g_{PWL} via the use of more data. However, potential bias can occur if g_{PWL} is taken to be the same for the radial and the angular models when the joint tail density of the random vector $\boldsymbol{X}$ is not well represented by the homothetic form discussed in Section 5.2.2. Instead a two step approach of maximizing $L_{\boldsymbol{W}}$ and $L_{R|\boldsymbol{W}}$ separately, and having models for $\boldsymbol{W} \mid \{R > r_\tau(\boldsymbol{W})\}$ and $R \mid \{\boldsymbol{W} = \boldsymbol{w}, R > r_\tau(\boldsymbol{w})\}$ can be implemented. One may also wish to model $R \mid \{\boldsymbol{W} = \boldsymbol{w}\}$ alone if the empirical distribution of $\boldsymbol{W}$ is suitable for all estimation tasks. Each of these settings is considered extensively in simulation studies.

Maximizing these likelihoods leads to parameter values that do not guarantee the marginal condition on the limit set, that $\max(G_{\text{PWL}}) = \mathbf{1}$. Algorithm 1 in Appendix 5.8.5 provides an adjustment to the parameter estimation procedure to ensure this condition holds. In it, a piecewise-linear model is first fitted via maximum likelihood using the likelihood of interest, $L_\bullet$. The parameter(s) that correspond to location where the limit set is at its largest value, but does not lie on the unit box boundary, is then divided by the fitted gauge function value at that location. These parameters are fixed and the likelihood of interest is then re-maximized with respect to the remaining parameters at the starting values given by the maximum likelihood estimates of the previous fit. This is repeated until $\max(G_{\text{PWL}}) = \mathbf{1}$. This bounding procedure is suitable when maximizing $L_{R|\boldsymbol{W}}$ and $L_{R,\boldsymbol{W}}$, as there is no such constraint in the angular model. Both unbounded and bounded gauges

are fitted in simulation studies and compared for their bias in extremal probability estimates. We additionally make note of a parameter redundancy in the angular model $f_{\boldsymbol{W}}(\boldsymbol{w}) = g_{\text{PWL}}(\boldsymbol{w})^{-d}/\{d\text{vol}(G_{\text{PWL}})\}$, notably, $f_{\boldsymbol{W}}(\boldsymbol{w}; c\boldsymbol{\theta}) = f_{\boldsymbol{W}}(\boldsymbol{w}; \boldsymbol{\theta})$ for any constant $c > 0$. To remedy this when using $L_{\boldsymbol{W}}$, we fix $\theta_1 = 1$, and maximize over the remaining $N-1$ parameters. We further note our choice to fix the gamma shape parameter at the dimension d to simplify model fitting, as the number of parameters is large in our piecewise-linear setting. In experiments, we found that estimating the shape parameter, as is done in Wadsworth and Campbell (2024), lead to little difference in terms of bias in extremal probability estimates.

5.4.2 The reference angles and penalization

5.4.2.1 Choosing the reference angles

The choice of reference angles is important for the quality of the approximation to the underlying gauge function. In essence, there are two ways of choosing the reference angles for g_{PWL}. The first is to strategically choose them where the underlying limit set boundary has a cusp or a change in direction. This is hard to do in practice without using knowledge of the true gauge function, but would lead to a model with fewer parameters. We opt for a second approach, setting a relatively fine mesh of reference angles. The result is more parameters than are perhaps needed, requiring some form of penalization during model fitting, to be discussed in Section 5.4.2.2. When $d = 2$, we take an equally spaced mesh, ensuring a reference angle is placed at $w = 1/2$, which allows for capturing whether or not the variables exhibit simultaneous extremes. In practice, we found $N = 11$ to be a good choice for approximating bivariate gauge functions. For $d = 3$, we partition at the subset of nodes $\{0, 1/6, 2/6, \ldots, 1\}^2$ which lie in $\mathcal{S}_2$, giving a triangulation of $\mathcal{S}_2$ with $N = 28$ nodes, as displayed in Figure 5.21 in Appendix 5.8.8. For $d \geq 4$, the grid-based approach leads to a very large number of angles. A sparser approach is to initially place reference angles at the edges $\boldsymbol{e}_j$, $j = 1, \ldots, d$ and the center $(1/d, \ldots, 1/d)^\top$ of $\mathcal{S}_{d-1}$, along with an angle at the center of all subfaces of $\mathcal{S}_{d-1}$, adding further angles

if diagnostics indicate the need.

5.4.2.2 Gradient-based penalization

If we maximize the likelihoods from Section 5.4.1 with a large number of reference angles N, the amount of data contributing to each parameter may be small, leading to high variability in parameter estimates. To remedy this, we propose penalizing the gradients of g_{PWL} so that they do not vary too much on either side of the reference angle locations. Linearity makes the gradient of g_{PWL} from equation (5.5) simple to calculate:

$$\nabla g_{\text{PWL}}(\boldsymbol{x};\boldsymbol{\theta}) = \sum_{k=1}^{M} \mathbf{1}_{\triangle^{(k)}}\left(\boldsymbol{x}/\|\boldsymbol{x}\|\right) \frac{\boldsymbol{n}^{(k)}}{\boldsymbol{n}^{(k)\top}\theta_1^{(k)}\boldsymbol{w}^{\star(k),1}}.$$

As the gradients do not change within a segment of the triangulated simplex, we define $\nabla g_{\boldsymbol{\theta}}^{(k)} = \boldsymbol{n}^{(k)}/\boldsymbol{n}^{(k)\top}\theta_1^{(k)}\boldsymbol{w}^{\star(k),1}$ to be the gradient of $g_{\text{PWL}}(\cdot;\boldsymbol{\theta})$ on $\triangle^{(k)}$. Further define $\mathcal{I}_\ell$ as the collection of pairs of indices of neighbouring segments in the Delaunay triangulation containing the vertex $\boldsymbol{w}^{\star\ell}$, where $\ell \in \{1,\ldots,N\}$. Neighbouring segments of the Delaunay triangulation are defined as segments that have $d-1$ matching vertices. For example, in the triangulation of $\mathcal{S}_2$ in Figure 5.4, we have $\mathcal{I}_1 = \{(1,3)\}$, $\mathcal{I}_2 = \{(2,3)\}$, $\mathcal{I}_3 = \{(1,2)\}$, and $\mathcal{I}_4 = \{(1,2),(2,3),(1,3)\}$ with $|\mathcal{I}_1| = |\mathcal{I}_2| = |\mathcal{I}_3| = 1$ and $|\mathcal{I}_4| = 3$. Given a likelihood $L_\bullet$ from Section 5.4.1, we add a penalty to give the objective function

$$-\log L_\bullet(\boldsymbol{\theta}; r_{1:n}, \boldsymbol{w}_{1:n}) + \lambda \frac{1}{N}\sum_{\ell=1}^{N} \frac{1}{|\mathcal{I}_\ell|} \sum_{(i,j)\in\mathcal{I}_\ell} \|\nabla g_{\boldsymbol{\theta}}^{(i)} - \nabla g_{\boldsymbol{\theta}}^{(j)}\|_2^2 \,. \tag{5.9}$$

This penalty term can be interpreted as the average sum of squared differences between neighbouring segment gradients at each node of the Delaunay triangulation. Figure 5.6 illustrates the effect of this penalty on two and three-dimensional data.

We propose to select the penalty value λ via K-fold cross-validation. As described in Section 5.3, the dataset is split into a fitting and an evaluation set. For each value of λ on a grid, we fit the model by minimizing (5.9), and evaluate the negative log-likelihood on the evaluation set, repeating this K times to yield $\text{CV}(\lambda) = K^{-1}\sum_{k=1}^{K} -\log L_\bullet(\widehat{\boldsymbol{\theta}}_{\lambda,k}; r_{1:n_{eval}}^{(k)}, \boldsymbol{w}_{1:n_{eval}}^{(k)})$. The parameter vector $\widehat{\boldsymbol{\theta}}_{\lambda,k}$ is the

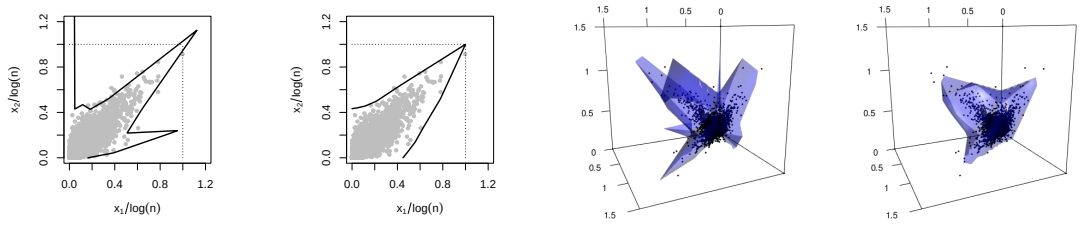

Figure 5.6: Left: $d = 2$ fitted unit level set of g_{PWL} using $N = 11$ equally-spaced reference angles on a dataset generated from distribution (I). The first panel has no penalty; the second panel uses the gradient penalty with $\lambda = 1$ and is bounded using Algorithm 1 in Appendix 5.8.5. Right: $d = 3$ fitted unit level set of g_{PWL} using $N = 28$ reference angles on a dataset generated from distribution (V). The first panel has no penalty; the second panel uses the gradient penalty with $\lambda = 1$ and is bounded using Algorithm 1 in Appendix 5.8.5.

minimizer of (5.9) evaluated at the k^{th} fitting set, and the λ value that minimizes $\text{CV}(\lambda)$ is said to be optimal. In Appendix 5.8.7, we compute the median $\text{CV}(\lambda)$ value across repeatedly-generated datasets from distributions (I)–(VII) with $K = 4$. The optimal value of λ naturally depends on the underlying distribution. Overall, selecting λ in the neighbourhood of 1 is generally acceptable when minimizing the negative log-likelihoods associated with the conditional radial model or the joint model, while a value of $\lambda = 20$ is suitable when minimizing the negative log-likelihood associated with the angular model. These penalty strength values are used henceforth.

5.4.3 Probability estimation

We perform extrapolation using a sampling-based approach in a similar manner to Wadsworth and Campbell (2024). First, n^* samples are drawn from $\boldsymbol{W} \mid \{R > r_\tau(\boldsymbol{W})\}$, with each one used to draw a conditional sample from the truncated gamma distribution $R \mid \{\boldsymbol{W} = \boldsymbol{w}, R > r_\tau(\boldsymbol{w})\}$. Once sampled, the exceedance angles $\boldsymbol{w}_1^*, \ldots, \boldsymbol{w}_{n^*}^*$ and radii $r_1^*, \ldots, r_{n^*}^*$ are multiplied, resulting in

samples $\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_{n^*}^*$ from $\boldsymbol{X} \mid \{R > r_\tau(\boldsymbol{W})\}$. Given an extremal set $B \subset \left\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| > r_\tau(\boldsymbol{x}/\|\boldsymbol{x}\|)\right\}$, probabilities can be estimated via

$$\begin{aligned}\widehat{\Pr}(\boldsymbol{X} \in B) &= \widehat{\Pr}(\boldsymbol{X} \in B \mid R > r_\tau(\boldsymbol{W}))\widehat{\Pr}(R > r_\tau(\boldsymbol{W})) \\ &= \left[\frac{1}{n^*}\sum_{i=1}^{n^*} \mathbf{1}_B(\boldsymbol{x}_i^*)\right]\left[\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{(r_\tau(\boldsymbol{w}_i),\infty)}(r_i)\right].\end{aligned}$$

In our approach, we have the option of a semiparametric model for $\boldsymbol{W} \mid \{R > r_\tau(\boldsymbol{W})\}$ with joint density $f_{\boldsymbol{W}}$. To sample from $f_{\boldsymbol{W}}$, we use Metropolis-Hastings MCMC with a beta or Dirichlet proposal density. The performance of this MCMC is assessed in Appendix 5.8.6.

5.4.4 Model performance and diagnostics

Several measures of goodness-of-fit for multivariate extremes align well with our piecewise-linear method. Wadsworth and Campbell (2024) assess the performance of a fitted truncated gamma model for exceedance radii $R \mid \{\boldsymbol{W} = \boldsymbol{w}, R > r_\tau(\boldsymbol{w})\}$ through probability-probability (PP) and quantile-quantile (QQ) plots. Plots of the fitted limit set boundary are also useful, since the shape should broadly correspond to that of the scaled sample clouds. Simpson et al. (2020) present methodology for estimating the collection of sets $\mathcal{C}$ experiencing simultaneous extremes. The methodology depends on several tuning parameters, but provides helpful insight into possible structures, that is independent of gauge function estimation. To that end, plots of the limit set boundary can be compared to findings based on the Simpson et al. (2020) coefficients to determine weather or not we accurately capture the extremal dependence structure of a dataset. For $d \leq 3$, plotting the limit set boundary is straightforward. In higher dimensions, one needs to project the gauge functions down to $d = 3$ via minimization over the $d - 3$ components. One can plot the unit level set of the projection

$$g(\boldsymbol{x}_{\{1,\ldots,d\}\setminus J}; \boldsymbol{\theta}) = \min_{\boldsymbol{x}_J \in \mathbb{R}_+^{d-3}} g(\boldsymbol{x}; \boldsymbol{\theta}) \tag{5.10}$$

where $J \subset \{1, \ldots, d\}$ is an index set of size $|J| = d - 3$. For example, when $d = 4$, we plot four unit level sets minimizing over each of the four variables individually. In order to perform these minimizations, we opt to evaluate $g(\boldsymbol{x}; \boldsymbol{\theta})$ on a mesh of $\boldsymbol{x}_J$ values in $[0, 1]^{d-3}$, and take the minimum value, which we found this to be less computationally expensive than other optimization methods. The size of the mesh is important: too small, and the resulting minima may be incorrect; too large, and the computation time will be high. In practice, a mesh of length 50^{d-3} was used in $d > 3$. Furthermore, all fitted gauge functions are bounded using Algorithm 1 in Appendix 5.8.5, so the reduced domain $[0, 1]^{d-3}$ is sufficient rather than minimizing over the entire $\mathbb{R}_+^{d-3}$ space.

Another potential goodness-of-fit measure is in the use of return-level curves (Papastathopoulos et al., 2025), sometimes also referred to as a version of "environmental contours" (Simpson and Tawn, 2024b). Given a return period T, the return-level curve defines a lower-bound of an open set such that we expect to see proportion T^{-1} points lying beyond this curve. As we only model above the threshold $r_\tau(\boldsymbol{w})$, we consider T such that $1 - T^{-1} \geq \tau$. The return curve in the truncated gamma setting is then defined as

$$\mathcal{R}(T) = \left\{ \boldsymbol{x} \in \mathbb{R}_+^d \middle| \boldsymbol{x} = F_{\text{Ga}}^{-1}\left[1 - T^{-1}; d, g(\boldsymbol{w}, \boldsymbol{\theta})\right] \boldsymbol{w},\, \boldsymbol{w} \in \mathcal{S}_{d-1} \right\}.$$

The full derivation of this expression is given in Appendix 5.8.3. Once such a curve is obtained, comparing the proportion of exceedances of $\mathcal{R}(T)$ in our data to the expected value of T^{-1} is one way to assess the predictive performance of the piecewise-linear model.

A check of how well our model captures the extremal dependence structure of the data in the joint tail can be assessed via estimates of the extremal coefficient

$$\chi_C(u) = \left(\frac{1}{1-u}\right) \Pr\left(F_X(X_j) > u\,,\; j \in C\,,\; C \subseteq \{1, \ldots, d\}\right),$$

for sufficiently high values of $u < 1$, where F_X is the distribution function common to all margins. An empirical estimate of $\chi_C(u)$ can be compared to a model-based estimate, obtained via the methods in Section 5.4.3, where the extremal set

considered is $B_u^C = \left\{\boldsymbol{x} \in \mathbb{R}_+^d \middle| x_j > F_X^{-1}(u)\ ,\ j \in C\right\}$. The value of u used needs to be high enough such that $B_u^C \subset \left\{\boldsymbol{x} \in \mathbb{R}_+^d \middle| \|\boldsymbol{x}\| \geq r_\tau(\boldsymbol{x}/\|\boldsymbol{x}\|)\right\}$. The boundary of this region in $\mathbb{R}_+^d$ is $\boldsymbol{x} = r_\tau(\boldsymbol{w})\boldsymbol{w}$, $\boldsymbol{w} = \boldsymbol{x}/\|\boldsymbol{x}\| \in \mathcal{S}_{d-1}$. For any coordinate x_j, we want u such that $r_\tau(\boldsymbol{w})w_j > F_X^{-1}(u)$ for all $j \in C$, i.e., $\min_{j\in C} x_j = \min_{j\in C} r_\tau(\boldsymbol{w})w_j > F_X^{-1}(u)$. To find the maximum point on this boundary in each coordinate we consider $\max_{\boldsymbol{w}\in\mathcal{S}_{d-1}} r_\tau(\boldsymbol{w})w_j$. The minimum value of u at which we can start estimating $\chi_C(u)$ is therefore given by $u_0 = F_X\left(\max_{\boldsymbol{w}\in\mathcal{S}_{d-1}} \min_{j\in C}\ r_\tau(\boldsymbol{w})w_j\right)$. In practice, u_0 can be obtained by taking the maximum over a mesh of values $\boldsymbol{w} \in \mathcal{S}_{d-1}$. Lastly, the empirical distribution of exceedance angles $\boldsymbol{W} \mid \{R > r_\tau(\boldsymbol{W})\}$ can be compared to the fitted distribution through density plots ($d = 2$), or comparing marginal histograms of the sample with those obtained by simulation from the fitted density ($d > 2$).

5.5 Simulation studies

An overarching goal in multivariate extreme value inference is estimation of $\Pr(\boldsymbol{X} \in B)$, where $B \subset \mathbb{R}^d$ is an extremal set generally lying outside the range of the data. The parametric geometric approach of Wadsworth and Campbell (2024) showed greater accuracy and flexibility in estimating $\Pr(\boldsymbol{X} \in B)$ compared to competing methods. Here, we will consider the case of $d = 2, 3$ for three different extremal sets B_1, B_2, B_3 for each dimension setting. We compare probability estimates obtained using the form of the true gauge as the rate of the truncated gamma, with parametric estimation of its parameters, and with those obtained using the piecewise-linear gauge. Because the parametric approach uses knowledge of the true gauge function and the piecewise-linear approach does not, the results here are intended to compete with those of Wadsworth and Campbell (2024), not to outperform them. In the simulation studies, we consider distributions (I)–(VII), exhibiting a variety of extremal dependence structures.

For each distribution, we generate $n = 5000$ observations, and use the KDE

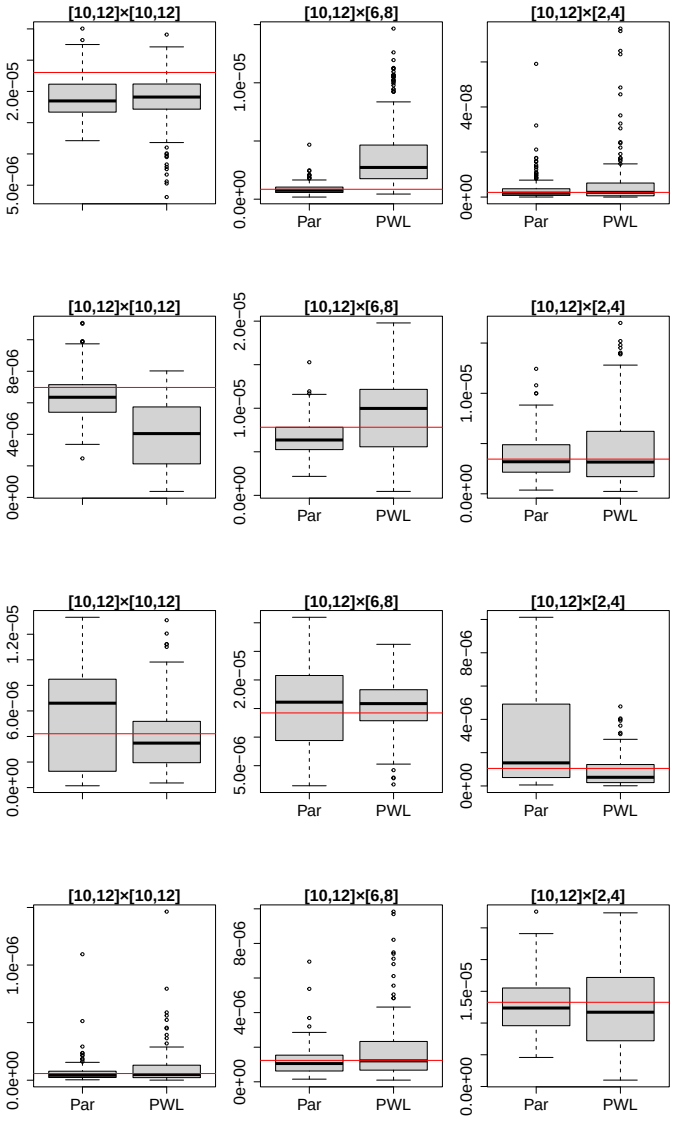

Figure 5.7: $d = 2$ simulation study probability estimates associated with distributions (I)–(IV) (top to bottom) across replicated model fits. "Par" refers to modelling with knowledge of the true parametric gauge function, "PWL" is semiparametric modelling using the piecewise-linear approach. Solid line is the true probability.

approach with Gaussian kernel to estimate the quantiles $r_\tau(\boldsymbol{w})$ for $\tau = 0.95$ and bandwidths $h_R = 0.05$, $h_{\boldsymbol{W}} = 0.05$. After obtaining $r_\tau(\boldsymbol{w})$, we first fit parametric

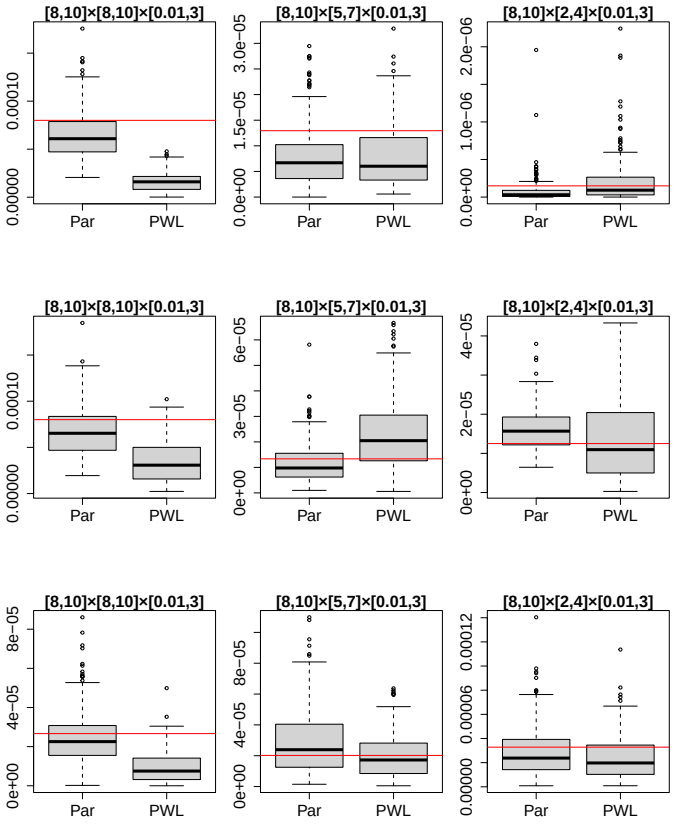

Figure 5.8: $d = 3$ extremal probability estimates associated with distributions (V)–(VII) (top to bottom) across replicated model fits. "Par" refers to modelling with knowledge of the true parametric gauge function, "PWL" is semiparametric modelling using the piecewise-linear approach. Solid line is the true probability.

models via maximization of $L_{R|\boldsymbol{W}}$ with knowledge of the true gauge function. As in Wadsworth and Campbell (2024), the empirical distribution of $\boldsymbol{W} \mid \{R > r_\tau(\boldsymbol{W})\}$ is used for probability estimation. For the piecewise-linear model, we consider six options for model fitting:

SS1: $R \mid \{\boldsymbol{W}, R > r_\tau(\boldsymbol{W})\}$ unbounded; empirical distribution for $\boldsymbol{W} \mid \{R > r_\tau(\boldsymbol{W})\}$.

SS2: $R \mid \{\boldsymbol{W}, R > r_\tau(\boldsymbol{W})\}$ bounded; empirical distribution for $\boldsymbol{W} \mid \{R > r_\tau(\boldsymbol{W})\}$.

SS3: $R \mid \{\boldsymbol{W}, R > r_\tau(\boldsymbol{W})\}$ unbounded; model for $\boldsymbol{W} \mid \{R > r_\tau(\boldsymbol{W})\}$.

SS4: $R \mid \{\boldsymbol{W}, R > r_\tau(\boldsymbol{W})\}$ bounded; model for $\boldsymbol{W} \mid \{R > r_\tau(\boldsymbol{W})\}$.

SS5: Unbounded joint model for $(R, \boldsymbol{W}) \mid \{R > r_\tau(\boldsymbol{W})\}$.

SS6: Bounded joint model for $(R, \boldsymbol{W}) \mid \{R > r_\tau(\boldsymbol{W})\}$.

Bounding is done via Algorithm 1 in Appendix 5.8.5, and likelihoods for $R \mid \boldsymbol{W}$, $\boldsymbol{W}$, and $(R, \boldsymbol{W})$ are given in Section 5.4.1. All six settings are considered when performing model fitting. In all cases we use penalized fitting on regular grid of reference angles. For each fitted model, $n^* = 50,000$ exceedance observations are generated and probabilities $\Pr(\boldsymbol{X} \in B_i)$, $i = 1, 2, 3$, are estimated and compared with the true probabilities. This procedure is repeated 200 times.

We first consider the bivariate distributions (I)–(IV), with $B_1 = [10, 12] \times [10, 12]$, $B_2 = [10, 12] \times [6, 8]$, and $B_3 = [10, 12] \times [2, 4]$. In comparing probability estimates to the true values, SS1–SS4 perform similarly, with a slight preference for SS2 and SS4. The similarity of these shows that the angular fit from maximizing L_W performs as well as its empirical counterpart. SS5 and SS6 tend to show more bias as the $\boldsymbol{W}$ distribution impacts the estimation of g_{PWL}. The probability estimates from SS4 are displayed in the boxplots in Figure 5.7, demonstrating that our semiparametric approach is comparable to the parametric method despite using no knowledge of the underlying distribution. A full summary of all possible $d = 2$ model fits is presented in Appendix 5.8.8.1.

A similar conclusion can me made from the trivariate data generated from distributions (V)–(VII). Here, the extremal regions are defined as $B_1 = [8, 10] \times [8, 10] \times [0.01, 3]$, $B_2 = [8, 10] \times [5, 7] \times [0.01, 3]$, and $B_3 = [8, 10] \times [2, 4] \times [0.01, 3]$. In comparing probability estimates to the true values, it was found that while SS4 performed slightly worse overall to the other simulation study setups, this approach is still largely comparable to the parametric method, but without knowledge of the underlying distribution (see Figure 5.8). A full summary of all possible $d = 3$ model fits is presented in Appendix 5.8.8.2.

Figure 5.9 displays fitted unit level sets of the piecewise-linear gauge function

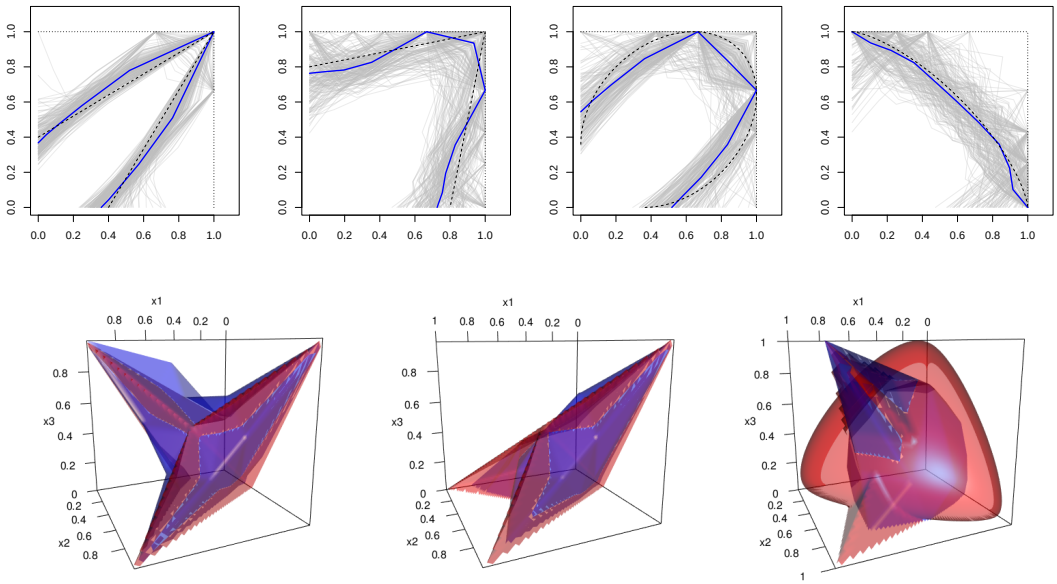

Figure 5.9: Top row: 200 estimates of the gauge function unit level sets (in grey) on data generated from distributions (I)–(IV) (left to right), with the gauge function at median parameter values in blue, and the true unit level set given by the dashed line. Bottom row: Gauge function unit level sets evaluated at median parameter values (in blue) with the true unit level set (in red) on data generated from distributions (V)–(VII) (left to right).

obtained using the bounding algorithm when performing inference on the distribution of $R \mid \{\boldsymbol{W}, R > r_\tau(\boldsymbol{W})\}$ (setting SS2 and SS4). From this, we see that parameter values across 200 fits produce limit sets that are flexible enough to capture a wide range of tail behaviours. We remark the slight difficulty in capturing asymptotic dependence in distribution (II), or in capturing one of the mixing components of distribution (VII). However, this is not a concern as probability estimates for these distributions have little overall bias. The limit set estimates for the remaining simulation study setups are displayed in Appendix 5.8.8.

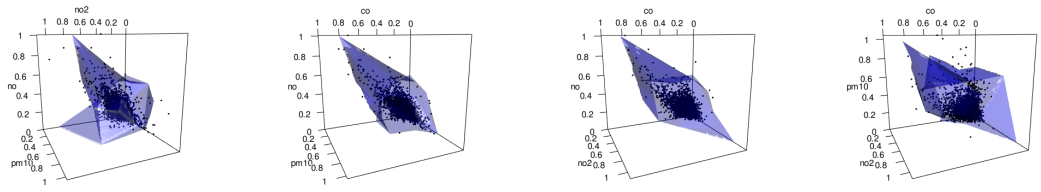

Figure 5.10: Projections of the estimated unit level set of g_{PWL} on 4-dimensional air pollution data with $\log(n)$-scaled data. Gauge functions are projected to 3-dimensions using equation (5.10) with $J = \{1\}, \{2\}, \{3\}, \{4\}$ (left to right).

5.6 Application to air pollution measurements

We consider air pollution measurements from the Automatic Urban and Rural Network (AURN), a UK-based air quality monitoring network. For the North Kensington site in London, we gather hourly measurements from April 1996 to June 2024 for carbon monoxide (CO, mg/m^3), nitrogen dioxide (NO_2, μg/m^3), particles with a diameter of 10 μm or less (PM10, mg/m^3), and nitric oxide (NO, μg/m^3). These are labelled 1, 2, 3, and 4 for brevity. We take the daily maxima over the 247,296 hourly measurements to avoid daily trends, and only consider measurements from October to April, inclusive, to reduce seasonal trends. Any measurements with missing data are excluded. The final dataset has $n = 5,584$ observations. There is no negative association between these measurements, so the margins are standardized to exponential using the empirical distribution function below the 0.95 quantile marginal threshold and a generalized Pareto distribution function fitted above this threshold, as outlined in Coles and Tawn (1991).

The radial quantile $r_\tau(\boldsymbol{w})$, $\boldsymbol{w} \in \mathcal{S}_3$, is estimated at $\tau = 0.70$, to increase the amount of exceedance data, while still focusing primarily on high values. We note that lower thresholds have been used in Simpson and Tawn (2024b) and Murphy-Barltrop et al. (2024b) without inducing large biases. The angular bandwidth was set to $h_{\boldsymbol{W}} = 0.075$. This is slightly higher than the bandwidth used in

simulation studies, but we felt that this helped eliminate excessive noise in the threshold estimate. A triangulation was obtained by first considering the vertices $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4 \in \mathbb{R}^4$, the center of the $\mathcal{S}_3$ simplex, $(1/4, 1/4, 1/4, 1/4)^\top$, and the center of the subfaces $\mathcal{S}_2, \mathcal{S}_1$. Once obtained, additional reference angles were placed in the center of each triangle from the resulting initial Delaunay triangulation. The final result is $N = 39$ reference angles. We fit the conditional radial component $R \mid \{\boldsymbol{W} = \boldsymbol{w}, R > r_\tau(\boldsymbol{w})\}$ by minimizing the associated penalized negative log-likelihood with penalty strength $\lambda = 1$ and bounding using Algorithm 1 in Appendix 5.8.5. The corresponding angular model $\boldsymbol{W} \mid \{R > r_\tau(\boldsymbol{W})\}$ was fit separately, with $\lambda = 20$.

Using the coefficients of extremal dependence from Simpson et al. (2020), we conclude that all four variables can grow large simultaneously, while PM10 can grow large when the remaining pollutants are jointly small. The projected three-dimensional unit level sets displayed in Figure 5.10 appear able to capture the joint tail dependence, with perhaps slight difficulty capturing the behaviour of PM10 when the remaining variables are jointly small (i.e., we do not have exactly $g_{\text{PWL}}(\gamma_1, \gamma_2, 1, \gamma_3) = 1$ for $\gamma_1, \gamma_2, \gamma_3 < 1$). However, the estimate is not far off, and we are able to capture this behaviour in a $d = 3$ fit, as is shown in Appendix 5.8.9.2. Also by our findings with the Simpson et al. (2020) coefficients, values of $\chi_C(u)$ are expected to be positive for all values of $u \in [0, 1]$ for any collection of variables $C \subseteq \{1, 2, 3, 4\}$. This is demonstrated by our estimated values, displayed in Figure 5.11 and in Figure 5.39 in Appendix 5.8.9.1, showing a general agreement with the corresponding empirical values. Note that probability estimates were obtained from extremal samples using our fitted angular model; therefore, good probability estimates indicate a well-fitted angular model. Figure 5.40 in Appendix 5.8.9.1 shows good agreement between estimated and true return levels $T \in [10, 1000]$, while the PP and QQ plots in Figure 5.41 generally show good agreement between the fitted model and the truncated gamma distribution. In Appendix 5.8.9.2 we present a three-dimensional fit to a subset of the pollutants for comparison, showing

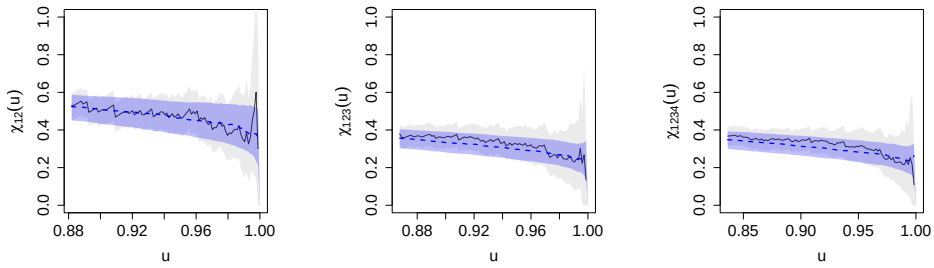

Figure 5.11: Model-based and empirical $\chi_C(u)$ plots with $C = \{1, 2\}$, $\{1, 2, 3\}$, and $\{1, 2, 3, 4\}$ for the pollution dataset. Solid lines are empirical values, and dashed lines are estimated using the piecewise-linear model. Shaded regions represent 7-day 95% block bootstrap confidence intervals.

agreement with the four-dimensional findings in detecting extremal tail behaviour.

5.7 Concluding remarks

In this work, we aimed to bypass the current difficulties in semiparametric modelling of multivariate extreme values through the geometric approach by proposing a simple piecewise-linear construction. Furthermore, the piecewise-linear construction allows for easy computation of the volume of its limit set, which in turn allows for efficient fitting of an angular model. Its calculation relies on standard operations of linear algebra, rendering it quick to evaluate and perform estimation on, including the use of gradient-based penalties. Simulation studies show that our semiparametric method is comparable to parametric methods using the true model forms, but without knowledge of the underlying distribution of the dataset. A difficulty in our proposed method is selecting the reference angles at which parameters are to be estimated. To avoid choosing these angles, we selected a regular grid on the simplex $\mathcal{S}_{d-1}$ for $d = 2, 3$, while opting for a sparser approach with $d \geq 4$. In unreported results, we were able to fit our model in dimension $d = 5$, but faced difficulties with choice of reference angles. Future avenues of work include the development of

a method for selecting only important reference angles with a view to performing modelling in higher dimensions, or a way of eliminating unimportant parameters through a different type of penalization, leading to models with fewer parameters but potentially higher predictive performance.

Acknowledgments, data availability, and code.

AURN data was imported using the `importAURN` function within the `openair` package in `R`, but is publicly available at `https://uk-air.defra.gov.uk/networks/network-info?view=aurn`. Code associated with this article can be found at `https://github.com/ryancampbell514/PWLExtremes`.

5.8 Appendix

5.8.1 Volume of G_{PWL}

Proof of Proposition 5.1. The limit set G_{PWL} with a piecewise-linear boundary is a union of subregions defined by parameters and reference angles. Therefore, its volume can be decomposed in the following manner:

$$\text{vol}(G_{\text{PWL}}) = \sum_{k=1}^{M} \text{vol}\left(\boldsymbol{\theta}^{(k)} \triangle^{(k)}\right),$$

where $\boldsymbol{\theta}^{(k)} \triangle^{(k)} \subset \mathbb{R}^d$ is the region with vertices at $\theta_1^{(k)} \boldsymbol{w}^{\star(k),1}, \ldots, \theta_d^{(k)} \boldsymbol{w}^{\star(k),d}$, and the origin if working in the positive orthant. Any point $\boldsymbol{x} \in \boldsymbol{\theta}^{(k)} \triangle^{(k)}$ can be written as

$$\boldsymbol{x} = \sum_{j=1}^{d} \theta_j^{(k)} a_j \boldsymbol{w}^{\star(k),j} \; ; \; a_j \geq 0 \, \forall \, j, \; \sum_{j=1}^{d} a_j = 1$$

where

$$\theta_j^{(k)} \boldsymbol{w}^{\star(k),j} = \begin{pmatrix} \theta_1^{(k)} \boldsymbol{w}^{\star(k),1} & \theta_2^{(k)} \boldsymbol{w}^{\star(k),2} & \ldots & \theta_d^{(k)} \boldsymbol{w}^{\star(k),d} \end{pmatrix} \boldsymbol{e}_j.$$

Let $\mathbf{M}^{(k)} = \begin{pmatrix} \theta_1^{(k)} \boldsymbol{w}^{\star(k),1} & \theta_2^{(k)} \boldsymbol{w}^{\star(k),2} & \ldots & \theta_d^{(k)} \boldsymbol{w}^{\star(k),d} \end{pmatrix}$ be the change of basis matrix, then we can write

$$\boldsymbol{x} = \mathbf{M}^{(k)} \sum_{j=1}^{d} a_j \boldsymbol{e}_j \; ; \; a_j \geq 0 \, \forall \, j, \; \sum_{j=1}^{d} a_j = 1.$$

Note that $\sum_{j=1}^{d} a_j \boldsymbol{e}_j$ defines any point in the simplex $\mathcal{S}_{d-1}$; therefore, it follows that

$$\begin{aligned}\operatorname{vol}\left(\boldsymbol{\theta}^{(k)} \triangle^{(k)}\right) = \int\limits_{\boldsymbol{\theta}^{(k)} \triangle^{(k)}} d\boldsymbol{x} &= \left|\det\left(\mathbf{M}^{(k)}\right)\right| \int\limits_{\mathcal{S}_{d-1}} d\boldsymbol{u} \\ &= \left|\det\left(\theta_1^{(k)} \boldsymbol{w}^{\star(k),1} \quad \theta_2^{(k)} \boldsymbol{w}^{\star(k),2} \quad \ldots \quad \theta_d^{(k)} \boldsymbol{w}^{\star(k),d}\right)\right| 1/d!,\end{aligned}$$

where $1/d!$ is the volume of the $d-1$-dimensional simplex (Stein, 1966).

□

5.8.2 Catalogue of considered multivariate distributions

5.8.2.1 Logistic

The joint distribution function in standard Fréchet margins is given by

$$F_F(\boldsymbol{x}; \alpha) = \exp\left\{-V(\boldsymbol{x}; \alpha)\right\}, \quad V(\boldsymbol{x}; \alpha) = \left(\sum_{j=1}^{d} x_j^{-1/\alpha}\right)^{\alpha},$$

where V is the -1-homogeneous exponent function, and $\alpha \in (0, 1]$ controls the strength of dependence. The corresponding parametric gauge function is obtained by differentiating F_F to obtain the joint density f_F, performing a change of variables to standard exponential margins, and taking the limit (5.1) to obtain

$$g(\boldsymbol{x}; \alpha) = \frac{1}{\alpha}\left(\sum_{j=1}^{d} x_j\right) + \left(1 - \frac{d}{\alpha}\right) \min\left\{x_1, \ldots, x_d\right\}.$$

5.8.2.2 Gaussian

Consider the matrix $\Sigma_{ij} = \rho_{ij}$, $i, j \in \{1, \ldots, d\}$ where $\rho_{ij} = \operatorname{Corr}(X_i, X_j) > 0$. The gauge function is obtained by performing a change of variables of a joint multivariate normal density with covariance matrix Σ in standard normal margins to standard exponential margins, then taking the limit (5.1). This results in the following

$$g(\boldsymbol{x}; \Sigma) = \sqrt{\boldsymbol{x}}^{\top} \Sigma^{-1} \sqrt{\boldsymbol{x}}. \tag{5.11}$$

In equation (5.11), all operations performed on vectors are done componentwise. The case with some $\rho_{ij} < 0$ is given for Laplace margins in equation (5.14).

5.8.2.3 Inverted logistic

The inverted logistic distribution is most simply presented by its joint survival function in standard exponential margins,

$$\bar{F}_E(x, y; \alpha) = \exp\left\{-V(1/\boldsymbol{x}; \alpha)\right\} \ , \quad V(1/\boldsymbol{x}; \alpha) = \left(\sum_{j=1}^{d} x_j^{1/\alpha}\right)^{\alpha} \ ,$$

where $\alpha \in (0, 1]$ controls the rate at which the d marginal variables grow large together. The corresponding parametric gauge function is obtained by differentiating $\bar{F}_E$ to obtain the joint density f_E and taking the limit (5.1) to obtain

$$g(\boldsymbol{x}; \alpha) = V(1/\boldsymbol{x}; \alpha).$$

5.8.2.4 Asymmetric logistic

The joint distribution function in Fréchet margins is given by

$$F_F(\boldsymbol{x}; \boldsymbol{\alpha}) = \exp\left\{-V(\boldsymbol{x}; \boldsymbol{\alpha})\right\},$$

where V is a prespecified -1-homogeneous exponent function, depending on the desired dependence structure. In general, V is given by

$$V(\boldsymbol{x}; \boldsymbol{\phi}, \boldsymbol{\alpha}) = \sum_{C \in P_D} \phi_C \left(\sum_{j \in C} x_j^{-1/\alpha_C}\right)^{\alpha_C} \ , \tag{5.12}$$

where P_D is the power set of indices $D = \{1, \ldots, d\}$,

$$\phi_C = \begin{cases} 1 & ; \text{ variables } C \text{ can grow large simultaneously.} \\ 0 & ; \text{ otherwise} \end{cases} \ ,$$

and $\alpha_C \in (0, 1]$ controls the dependence in group C. Additional parameters are required in (5.12) to make the margins standard Fréchet, though these do not affect the limiting gauge function and so are omitted. Here, the gauge function is obtained by differentiating F_F to obtain the joint density f_F, performing a change of variables to standard Gumbel margins. The standard Gumbel and standard exponential distribution are asymptotically equivalent in $\mathbb{R}_+^d$, and therefore we can obtain the

gauge function in the usual way (see Wadsworth and Campbell (2024)). The true gauge is given by

$$g(\boldsymbol{x};\boldsymbol{\alpha}) = \min_{\pi\in\Pi} \min_{C\in\mathcal{C}_s^+ \,:\, s\in\pi} \left[\sum_{s\in\pi} \left(\sum_{j\in s} \frac{x_j}{\alpha_C} + \left(1 - \frac{|s|}{\alpha_C}\right) \min_{j\in C} x_j \right) \right]$$

where Π is the set of all partitions of D, where $\mathcal{C}_s^+$ denotes a collection of indices corresponding to a group obtaining simultaneous extremes, i.e., when $\phi_s = 1$.

5.8.2.5 Mixture model

Consider a mixture model with exponential margins whose joint density is given by

$$f(\boldsymbol{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = pf_1(\boldsymbol{x};\boldsymbol{\theta}_1) + (1-p)f_2(\boldsymbol{x};\boldsymbol{\theta}_2),$$

where f_1 and f_2 are joint densities with respective parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ in exponential margins, and $p \in (0,1)$, with $p = 0.5$ throughout this work. The corresponding gauge function is given by

$$g(\boldsymbol{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \min\left\{g_1(\boldsymbol{x};\boldsymbol{\theta}_1), g_2(\boldsymbol{x};\boldsymbol{\theta}_2)\right\},$$

where g_1 and g_2 are the gauge functions corresponding to the joint densities of f_1 and f_2, respectively.

5.8.3 Truncated gamma return level sets

For a given return period T, set $\mathcal{R}(T) \subset \mathbb{R}^d$ to be the corresponding return level set. For $T = 1/(1-\tau)$, where τ is the level at which we do quantile regression, we have

$$\mathcal{R}(1/(1-\tau)) = \left\{\boldsymbol{x} \in \mathbb{R}_+^d \middle| \boldsymbol{x} = r_\tau(\boldsymbol{w})\boldsymbol{w},\, \boldsymbol{w} \in \mathcal{S}_{d-1}\right\}.$$

For general T, we have

$$\Pr\left(R \geq r_{1-T^{-1}}(\boldsymbol{w})\right) = T^{-1}.$$

It follows that, for $T \geq (1-\tau)^{-1}$,

$$\begin{aligned}
&\Pr\left(R \geq r_{1-T^{-1}}(\boldsymbol{w}) | \boldsymbol{W} = \boldsymbol{w}\right) \\
&= \Pr\left(R \geq r_{1-T^{-1}}(\boldsymbol{w}) | R \geq r_{\tau}(\boldsymbol{w}), \boldsymbol{W} = \boldsymbol{w}\right) \Pr\left(R \geq r_{\tau}(\boldsymbol{w}) | \boldsymbol{W} = \boldsymbol{w}\right) \\
&= \frac{\bar{F}_{\text{Ga}}\left[r_{1-T^{-1}}(\boldsymbol{w}); d, g(\boldsymbol{w})\right]}{\bar{F}_{\text{Ga}}\left[r_{\tau}(\boldsymbol{w}); d, g(\boldsymbol{w})\right]} \times \bar{F}_{\text{Ga}}\left[r_{\tau}(\boldsymbol{w}); d, g(\boldsymbol{w})\right] \\
&= T^{-1},
\end{aligned}$$

where $F_{\text{Ga}}^{-1}\left[\cdot; d, g(\boldsymbol{w}, \boldsymbol{\theta})\right]$ is the quantile function corresponding to the gamma distribution with shape parameter d and rate parameter $g(\boldsymbol{w}, \boldsymbol{\theta})$. Therefore, $r_{1-T^{-1}}(\boldsymbol{w}) = F_{\text{Ga}}^{-1}\left[1 - T^{-1}; d, g(\boldsymbol{w}, \boldsymbol{\theta})\right]$, and the return level set is therefore given by the curve

$$\mathcal{R}(T) = \left\{\boldsymbol{x} \in \mathbb{R}_{+}^{d} \middle| \boldsymbol{x} = F_{\text{Ga}}^{-1}\left[1 - T^{-1}; d, g(\boldsymbol{w}, \boldsymbol{\theta})\right] \boldsymbol{w},\ \boldsymbol{w} \in \mathcal{S}_{d-1}\right\}.$$

5.8.4 Quantile estimation

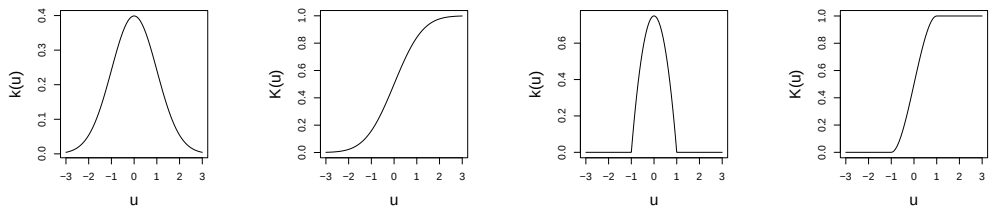

Figure 5.12: Left to right: Gaussian PDF, Gaussian CDF, Epanechnikov PDF, and Epanechnikov CDF

The kernel density estimation (KDE) approach of obtaining quantiles introduced in Section 5.3 is an important step before obtaining maximum likelihood estimates of parameters of our piecewise-linear model. Given the compact domain $\mathcal{S}_{d-1}$, the use of compactly-supported kernels may be of interest. Here, we compare the Gaussian kernel to the Epanechnikov kernel (Epanechnikov, 1969), which has a bounded support and is defined by

$$k(u) = \mathbf{1}_{(-1,1)}(u)\frac{3}{4}(1 - u^2)$$

and

$$K(u) = \begin{cases} 0 & ; u \leq -1 \\ \frac{3}{4}\left[u(1 - \frac{1}{3}u^2) + \frac{2}{3}\right] & ; -1 < u < 1 \\ 1 & ; u \geq 1 \end{cases}.$$

Both kernels' associated densities (PDFs) and distribution functions (CDFs) are shown in Figure 5.12.

Using the quantile score proposed in Section 5.3 with $K = 5$, we assess how well our KDE-based quantile estimation performs as the smoothing hyperparameter $h_{\boldsymbol{W}}$ increases when using the Gaussian and Epanechnikov kernels. Figure 5.13 shows these scores for $d = 2$ and $d = 3$ datasets (I)–(VII) at the quantile level $\tau = 0.95$, comparing to the empirical binning method of quantile estimation as the amount of overlapping increases. Note that for simplicity there is no boundary correction for either kernel in our KDE methodology. This would be important for actual density estimation, but we are simply searching for approximate high quantiles of $R \mid \boldsymbol{W}$. Because of its bounded support, the Epanechnikov kernel was found to be computationally cheaper to evaluate compared to the Gaussian kernel in the $d = 2$ case. However, quantile estimates are not visually better than using the Gaussian kernel, as is shown in Figure 5.14, and the Epanechnikov kernel is more computationally expensive to evaluate in `R` than the Gaussian kernel in the $d \geq 3$ setting, as the Gaussian kernel has an efficient multivariate evaluation using functions in the `mvtnorm` package in `R`, while the Epanechnikov setting requires taking products of the univariate kernels. For this reason, we continue using the Gaussian kernel exclusively in KDE quantile estimation for dimensions $d \geq 3$. Figures 5.14 and 5.15 display the quantile boundaries $r_\tau(\boldsymbol{w})\boldsymbol{w}$ for $\boldsymbol{w} \in \mathcal{S}_{d-1}$ and $\tau = 0.95$ at the specified adopted smoothing parameters for datasets (I)–(VII). Quantile score values at optimal levels of $h_{\boldsymbol{W}}$ are shown in Table 5.1. Results show that the KDE approach results in estimation quality similar to the empirical method in dimensions $d = 2, 3$.

The empirical binning method can lead to empty regions in higher dimensions,

Dataset	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
KDE score	0.2329	0.2296	0.2562	0.2282	0.2919	0.2656	0.3119
empirical score	0.2324	0.2291	0.2557	0.2280	0.2890	0.2643	0.3125

Table 5.1: Median scores of repeated quantile estimates for the KDE approach with Gaussian kernel and the empirical binning approach of Wadsworth and Campbell (2024) at their optimal smoothing hyperparameters.

leading to the inability to estimate quantiles in the entire $\mathcal{S}_{d-1}$ simplex. In the empirical method, given a new angle $\boldsymbol{w} \in \mathcal{S}_{d-1}$, a local average of quantile values $r_\tau(\boldsymbol{w}_i)$ already estimated are taken for angles $\boldsymbol{w}_i$ neighbouring $\boldsymbol{w}$. Therefore, we are entirely dependent on the radial quantile values of the dataset. If there is insufficient data to estimate the radial quantile on a given dataset, then one may not be able to estimate radial quantiles at new angles. This problem does not arise in the KDE approach, i.e., we can evaluate $r_\tau(\boldsymbol{w})$ for all $\boldsymbol{w} \in \mathcal{S}_{d-1}$. We note that the quantile performance score of distributions (II) and (IV) are near-independent of the amount of smoothing applied to the quantile estimation procedure. For these two distributions, $r_\tau(\boldsymbol{w})$ does not depend strongly on $\boldsymbol{w}$ for this value of τ. This threshold is therefore easy to estimate regardless of the amount of smoothing applied. Different dependence structures have different optimal bandwidths, but $h_{\boldsymbol{W}} = 0.05$ is close to optimal in all cases. In accompanying code, users can instead allow for automatic selection of $h_{\boldsymbol{W}}$ using K-fold cross-validation scoring on the check function $S(h_{\boldsymbol{W}})$ defined in Section 5.3. Figure 5.16 demonstrates that, for a fixed value of $h_{\boldsymbol{W}}$, varying the radial bandwidth h_R has no effect on the quality of radial quantile estimates.

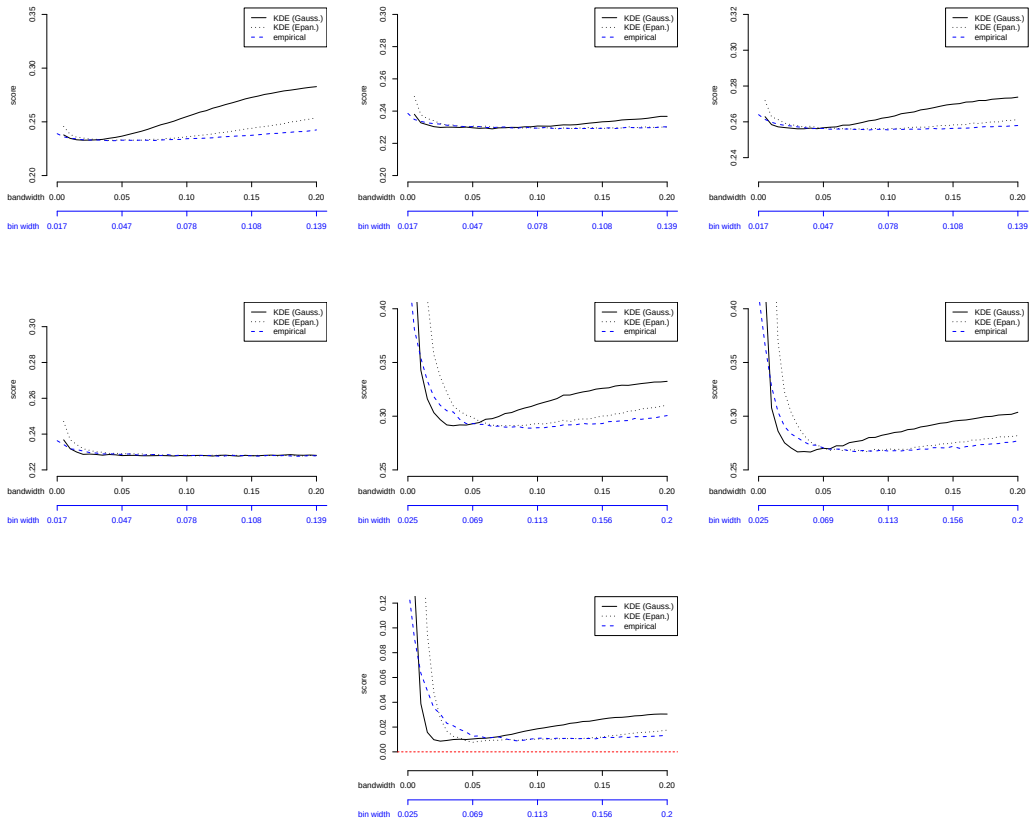

Figure 5.13: Median quantile estimation scores at $\tau = 0.95$ for data generated from distributions (I)–(III) (top row), (VII)– (VI) (middle row), and (VII) (bottom row). Quantiles are estimated using KDE with the Gaussian and the Epanechnikov kernels, and empirically.

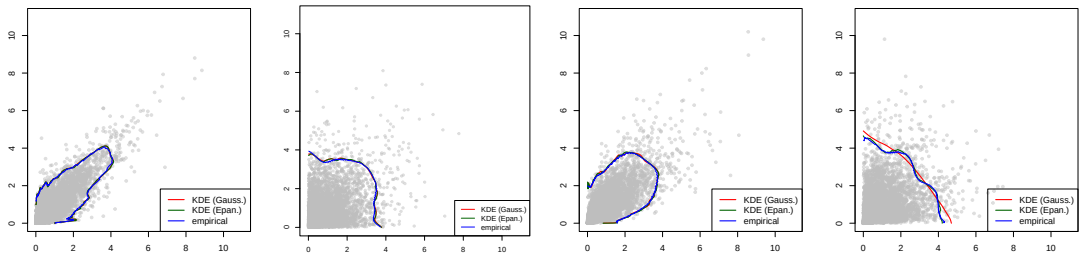

Figure 5.14: Left to right: empirical and KDE-based threshold estimates for $d = 2$ on datasets (I)–(IV) with $\tau = 0.95$. The Gaussian and Epanechnikov kernels use bandwidth values $h_R = 0.05$ and optimal values for h_W and amount of bin overlap governed by the scores in Figure 5.13.

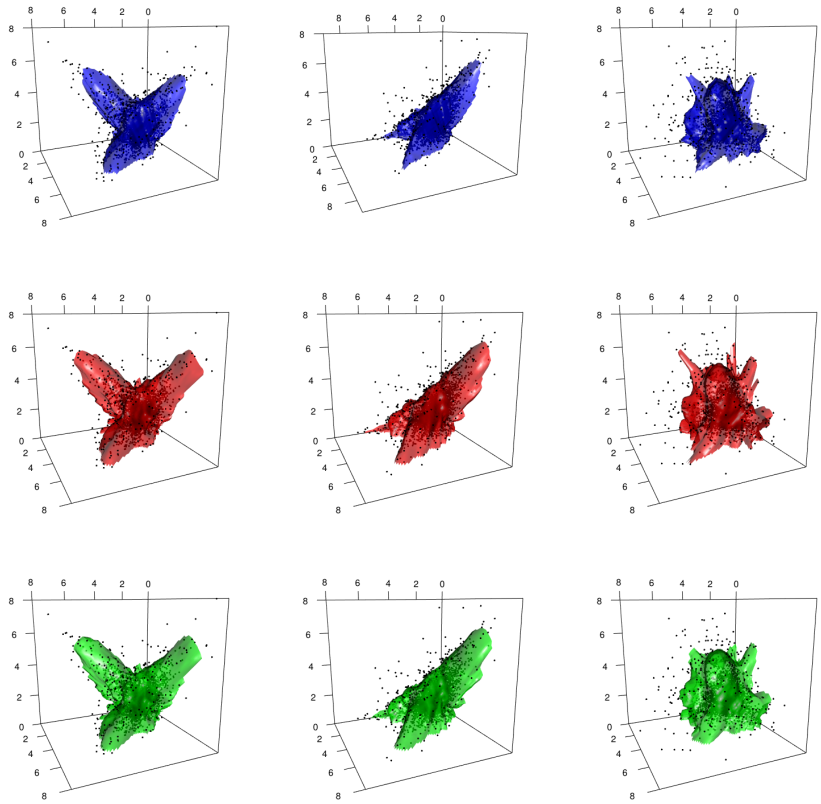

Figure 5.15: Left to right: empirical (top row), Gaussian KDE (middle row), and Epanechnikov KDE (bottom row) threshold estimates for $d = 3$ on datasets (V)–(VII) (left to right) with $\tau = 0.95$. The Gaussian and Epanechnikov kernels use bandwidth values $h_R = 0.05$ and optimal values for h_W and bin overlap governed by the scores in Figure 5.13.

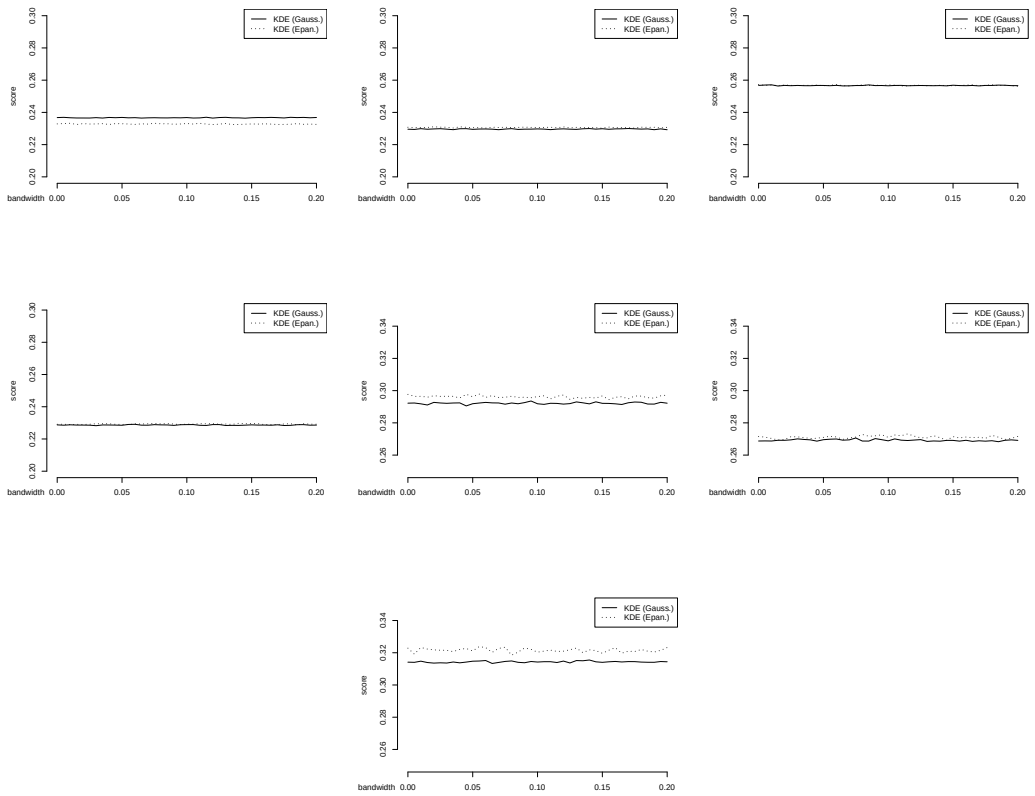

Figure 5.16: Median KDE quantile estimation scores at $\tau = 0.95$ as h_R varies with fixed $h_{\boldsymbol{W}} = 0.05$ for data generated from distributions (I)–(III) (top row), (VII)–(VI) (middle row), and (VII) (bottom row).

5.8.5 Algorithm: gauge function unit level set bounding

The following describes the bounded model fitting algorithm, introduced in Section 5.4.1 and implemented throughout the simulation studies and real data applications in Sections 5.5 and 5.6, respectively.

Algorithm 1: Bounding the piecewise-linear limit set during model fitting

Input: $r_i, \boldsymbol{w}_i$, threshold estimates $r_\tau(\boldsymbol{w}_i)$, likelihood $L_{R,\boldsymbol{W}}$ or $L_{R|\boldsymbol{W}}$, reference angle set $\left\{\boldsymbol{w}^{\star 1}, \ldots, \boldsymbol{w}^{\star N}\right\}$

$\widehat{\boldsymbol{\theta}} \leftarrow \text{argmax}_{\boldsymbol{\theta}} L_\bullet(\boldsymbol{\theta}; r_{1:n}, \boldsymbol{w}_{1:n})$;

$\mathcal{F} \leftarrow \emptyset$;

while $\max\left(w_j^{\star i}/g(\boldsymbol{w}^{\star i}; \widehat{\boldsymbol{\theta}})\right) \neq 1 \ \forall \ i = 1, \ldots, N \ \textit{and} \ j = 1, \ldots, d$ **do**

$\mathcal{F} \leftarrow$

$\mathcal{F} \bigcup \left\{ i \in \{1, \ldots, N\} : w_j^{\star i}/g(\boldsymbol{w}^{\star i}; \widehat{\boldsymbol{\theta}}) > w_j^{\star\{1,\ldots,N\}\setminus i}/g(\boldsymbol{w}^{\star\{1,\ldots,N\}\setminus i}; \widehat{\boldsymbol{\theta}}), j \in \{1, \ldots, d\} \right\}$;

$\widehat{\boldsymbol{\theta}}_{\mathcal{F}} \leftarrow \widehat{\boldsymbol{\theta}}_{\mathcal{F}}/g(\boldsymbol{w}^{\star\mathcal{F}}; \widehat{\boldsymbol{\theta}})$;

$\widehat{\boldsymbol{\theta}}_{-\mathcal{F}} \leftarrow \text{argmax}_{\boldsymbol{\theta}_{-\mathcal{F}}} L_\bullet(\boldsymbol{\theta}_{-\mathcal{F}}; r_{1:n}, \boldsymbol{w}_{1:n}, \boldsymbol{\theta}_{\mathcal{F}})$

end

Return: Scaled parameter estimates $\widehat{\boldsymbol{\theta}} \in \mathbb{R}_+^N$

5.8.6 Angular fit examples

5.8.6.1 Gaussian distribution (III), $d = 2$

We model the angles of data generated from a bivariate Gaussian distribution with correlation $\rho = 0.8$, using the density $f_{\boldsymbol{W}}(\boldsymbol{w}) = g_{\text{PWL}}(\boldsymbol{w})^{-d}/\{d\text{vol}(G_{\text{PWL}})\}$ and g_{PWL} specified piecewise-linearly using equation (5.5). Included in Figure 5.5 are estimates of the density $f_{\boldsymbol{W}}$ for increasing number of parameters N. Figure 5.17 shows samples from this fitted density using MCMC with a uniform proposal density and a beta density whose parameters were fitted using the exceedance angles of the dataset. The beta proposal is preferred, as the MCMC acceptance rate is much higher, leading to a more efficient sampling algorithm.

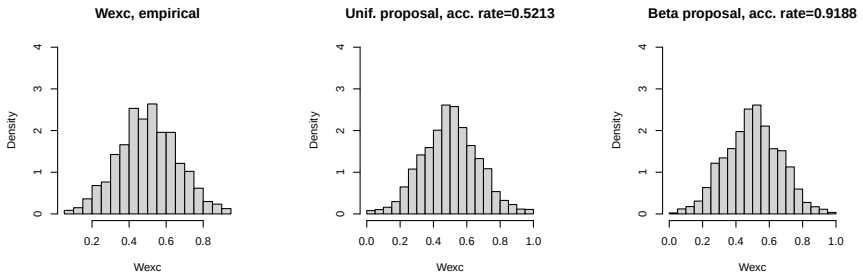

Figure 5.17: Left: empirical distribution of $W \mid \{R > r_\tau(W)\}$; center/right: samples drawn via MCMC on fitted model, using uniform (center) and beta (right) proposals.

5.8.6.2 Mixture model (VII), $d = 3$

Using data sampled from the mixture distribution (VII), we fit the exceedance angle model $f_{\boldsymbol{W}}$. Figure 5.18 includes the estimated density plotted on the $\mathcal{S}_2$ simplex. Samples obtained using the uniform and a fitted Dirichlet proposal show reasonable agreement with the underlying true angular distribution, as suggested by agreement with the empirical density. Marginal samples as seen in Figure 5.19 show that both the uniform and Dirichlet proposals reasonably capture the behaviour of the underlying angular distribution.

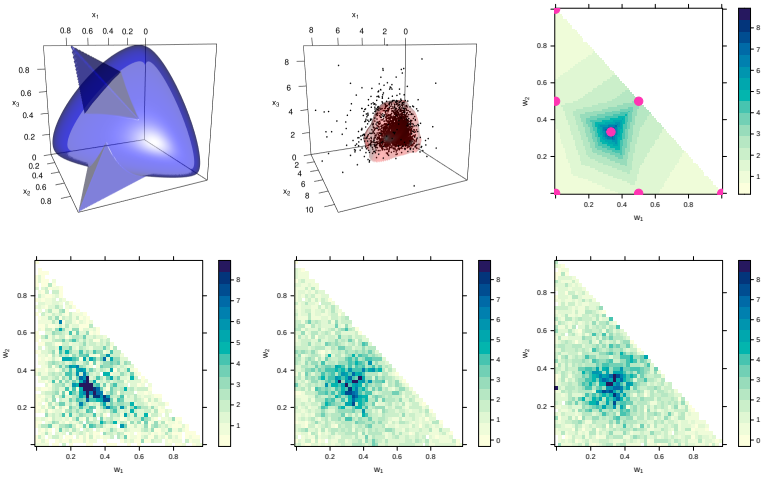

Figure 5.18: $d = 3$ mixture model example: Top-left: Unit level set of the gauge function, g. Top-middle: Dataset and estimated high quantiles. Top-right: Fit of $f_{\boldsymbol{W}}(\boldsymbol{w})$ with reference angles overlaid. Bottom-left: histogram of $\boldsymbol{W} \mid \{R > r_\tau(\boldsymbol{W})\}$. Bottom-middle and bottom-right: Histogram densities of an MCMC sample of exceedance angles using the fitted $f_{\boldsymbol{W}}$ using a uniform and a Dirichlet proposal, respectively.

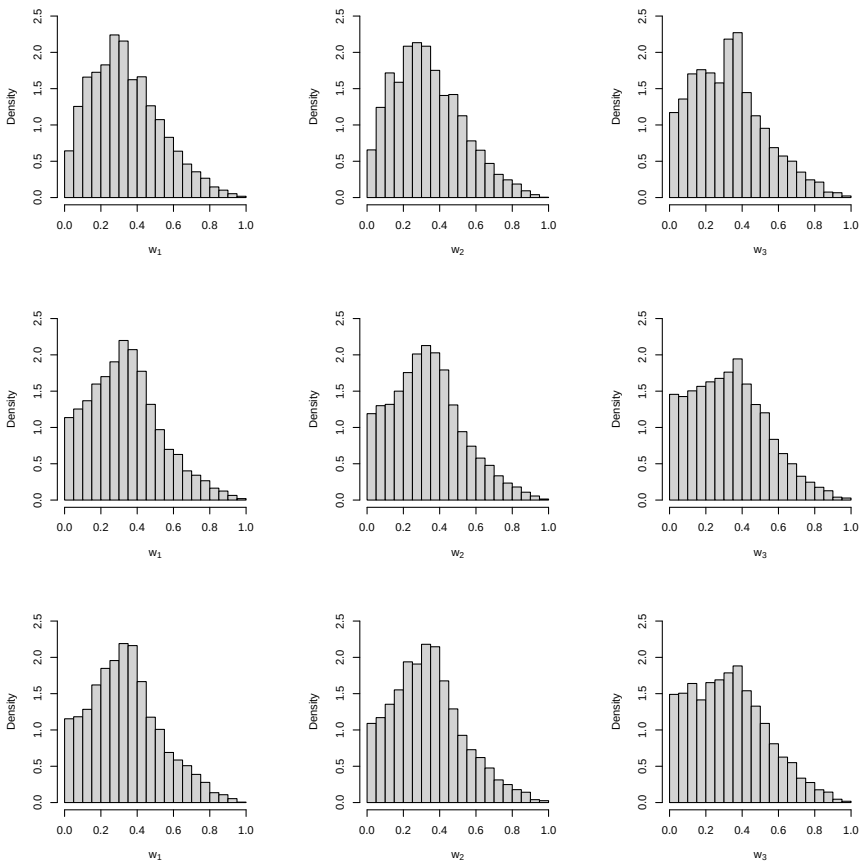

Figure 5.19: $d = 3$ mixture model example: Marginal sample of exceedance angles using the empirical distribution of exceedance angles (top row), MCMC on the fitted density $f_{\boldsymbol{W}}$ with uniform proposal (middle row) and Dirichlet proposal (bottom row).

5.8.7 Choosing the gradient penalty strength, λ

In Section 5.4.2, a gradient penalty is introduced. To select the best penalty strength $\lambda \geq 0$, a K-fold cross-validation score on negative log-likelihood (NLL) is implemented. This score is based on negative log-likelihood values at parameters obtained from a fitting set and evaluated on a hold-out set, and is explicitly given by

$$\mathrm{CV}(\lambda) = \frac{1}{K} \sum_{k=1}^{K} -\log L_{\bullet} \left(\widehat{\boldsymbol{\theta}}_{\lambda,k}; r_{1:n_k}, \boldsymbol{w}_{1:n_k} \right) ,$$

where $-\log L_{\bullet}$ is the negative log-likelihood associated with one of the three likelihood functions introduced in Section 5.4.1, $\widehat{\boldsymbol{\theta}}_{\lambda,k}$ is the vector of parameter values obtained from the k^{th} fitting set of datapoints with gradient penalty with strength $\lambda \geq 0$, and $(r_{1:n_k}, \boldsymbol{w}_{1:n_k})$ are held-out evaluation radii and angle data. Figure 5.20 shows these scores with $K = 4$ and $n = 5000$ on a mesh of λ values. From top to bottom, the score is a median value obtained across 20 datasets from distributions (I)–(VII). Columns 1–2 correspond to fitting the radial model conditioned on angles by maximizing $L_{R|\boldsymbol{W}}$ unbounded and bounded using Algorithm 1 in Appendix 5.8.5, respectively. Columns 3–4 correspond to jointly fitting the radial-angular model by maximizing $L_{R,\boldsymbol{W}}$ unbounded and bounded using Algorithm 1 in Appendix 5.8.5, respectively. Column 5 corresponds to fitting the angular model by maximizing $L_{\boldsymbol{W}}$.

As expected, the optimal degree of smoothing λ depends on the underlying dependence structure of the data. In bivariate data (rows 1–4 in Figure 5.20), Gaussian data from distribution (III) requires the least smoothing as the true limit set is curved and therefore changes in the gradient of each segment are desirable. For distributions such as (II) and (IV) (rows 2 and 4 in Figure 5.20), a flatter limit set is desired, meaning a higher λ is preferred. To account for this change in optimal smoothing hyperparameter, the code in our GitHub repository allows the user to not specify λ, and a K-fold cross-validation scoring procedure on $-\log L_{\bullet}$ is performed instead. In three-dimensions (rows 5–7 in Figure 5.20), a good middle-ground would be $\lambda = 1$ for the radial and joint models, while a λ value of around

20 seems appropriate when fitting the angular model on its own.

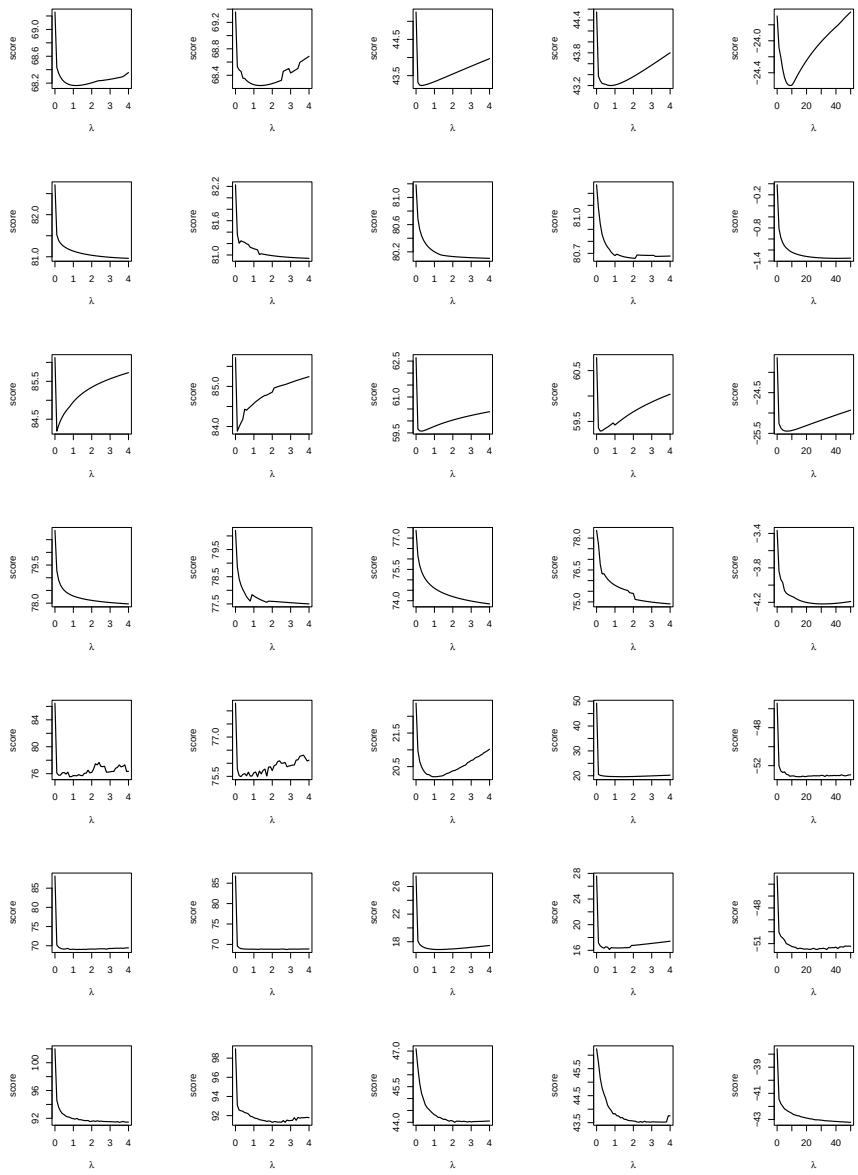

Figure 5.20: Cross-validation scores for models (left to right, see text) fitted using data generated from distributions (I)–(VII) (top to bottom) for varying λ.

.

5.8.8 Simulation studies

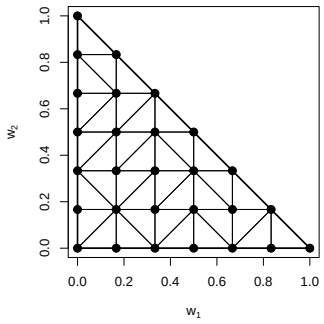

Figure 5.21: A mesh of $N = 28$ reference angles (points) and the resulting Delaunay triangulation (solid lines) on the $\mathcal{S}_2$ simplex. This setting is used throughout all $d = 3$ model fits in simulation studies and in real data examples.

As described in Section 5.5, there are six settings for fitting the piecewise-linear model for inference on multivariate extremes, labelled SS1–SS6. Here, we outline results for all possible settings for distributions (I)–(VII) in dimensions $d = 2, 3$, and compare probability estimates and estimates of the unit level set boundary when using $N = 11$ equally-spaced scalar-valued reference angles from 0 to 1, inclusive. In each case, 200 datasets of length $n = 5000$ are generated. In each replication, $r_\tau(\boldsymbol{w})$ is first estimated at the $\tau = 0.95$ threshold, using the KDE approach with Gaussian kernel, and a probability estimate is obtained for each extremal region. Furthermore, in the $d = 2$ setting we plot all 200 limit set boundaries in gray, the limit set boundary obtained at the median parameter values in blue, and the true limit set boundary using a black-dashed line. For $d = 3$, the piecewise-linear limit set boundary is plotted in blue at median parameter values across the 200 replications, and the true limit set boundary is plotted in red. Here, we use $N = 28$ reference angles, and the resulting triangulation of $\mathcal{S}_2$ is presented in Figure 5.21. Based on the penalty strength findings in Appendix 5.8.7, we fix the penalty strength $\lambda = 1$ for the radial and joint fits, while setting $\lambda = 20$ for the angular model fit.

In summary, for $d = 2$, settings SS1,3 and SS2,4 produce similar results, showing no particular advantage or disadvantage of modelling the angles separately compared to the corresponding empirical distribution. The joint fitting procedures of SS5,6 lead to good estimates in the case of distribution (I), but an increased bias in the probability estimates when $f_{\boldsymbol{W}}$ is not well-approximated by the same g in used in the radial model (distributions (II)–(IV)). Similarly, SS5 and SS6 induce more bias in the $d = 3$ distributions, while SS1,3 and SS2,4 perform similarly well overall. These findings can be seen in Tables 5.2 and 5.3, which present root mean squared error (RMSE) of the log-probability estimates. In it, we see that the piecewise-linear model fitting method competes well across dimensions two and three and across different distributions exhibiting a variety of extremal dependence properties. We note that SS5 and SS6 often have the lowest RMSEs due to reduced variance, but we place high value on unbiasedness. Therefore, for as an overall well-performing model, we report the bounded model fit presented in the setup of SS4 in the main body of this paper.

5.8.8.1 Simulation studies, $d = 2$

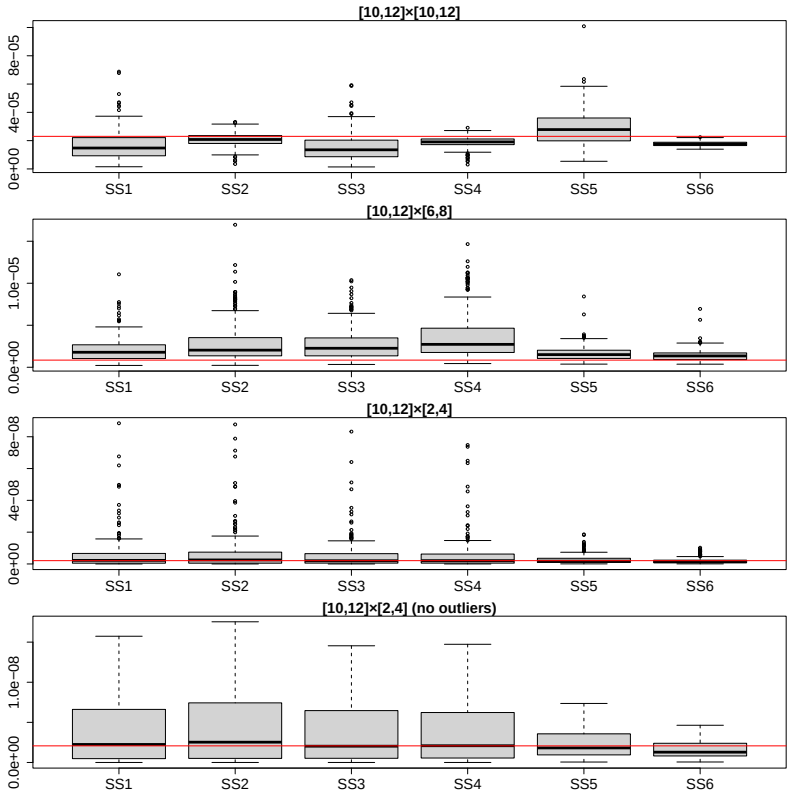

Figure 5.22: Probability estimates for distribution (I). True values shown by the solid line. Regions of interest B_i, $i = 1, 2, 3$, is given in the title, with results for B_3 given with and without outliers for clarity.

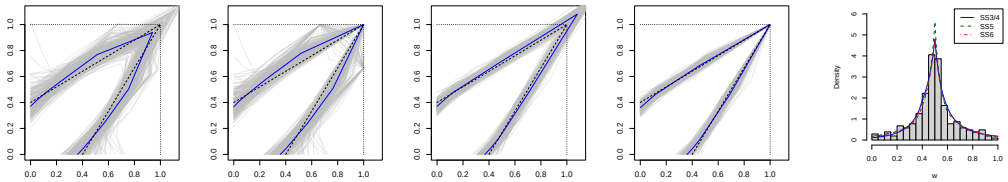

Figure 5.23: (a)–(d): Estimates of the unit level set of g for distribution (I). (e) Estimated angular density f_W, with a sample histogram from one sample.

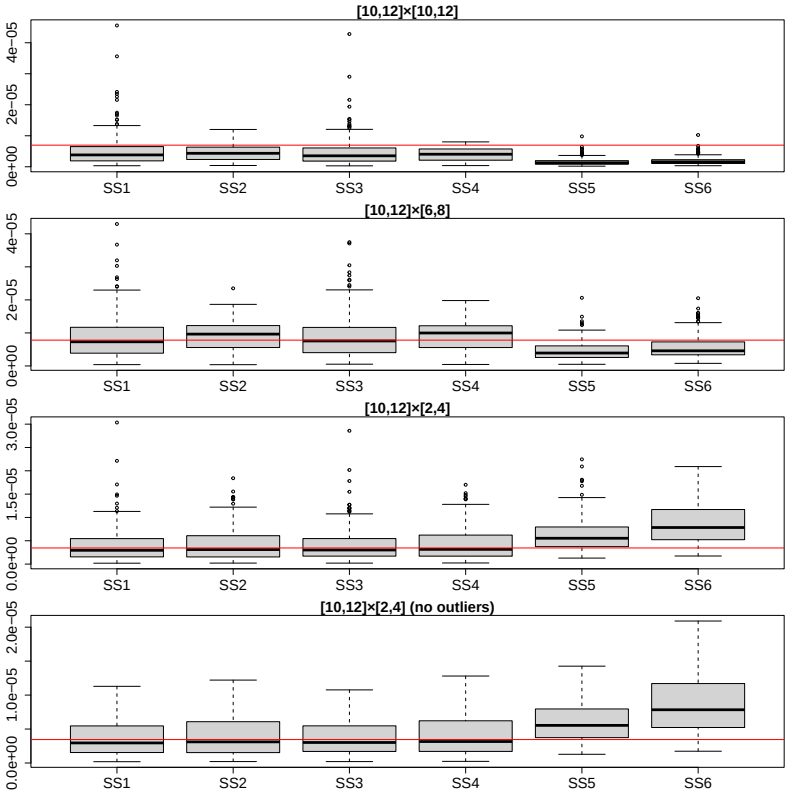

Figure 5.24: Probability estimates for distribution (II). True values shown by the solid line. Regions of interest B_i, $i = 1, 2, 3$, is given in the title, with results for B_3 given with and without outliers for clarity.

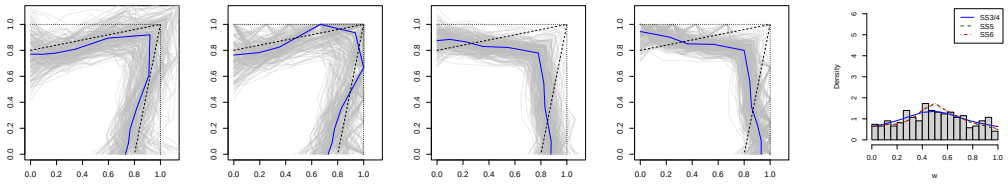

Figure 5.25: Estimates of the unit level set of g for distribution (II). Right: Estimated angular density f_W, with a sample histogram from one sample.

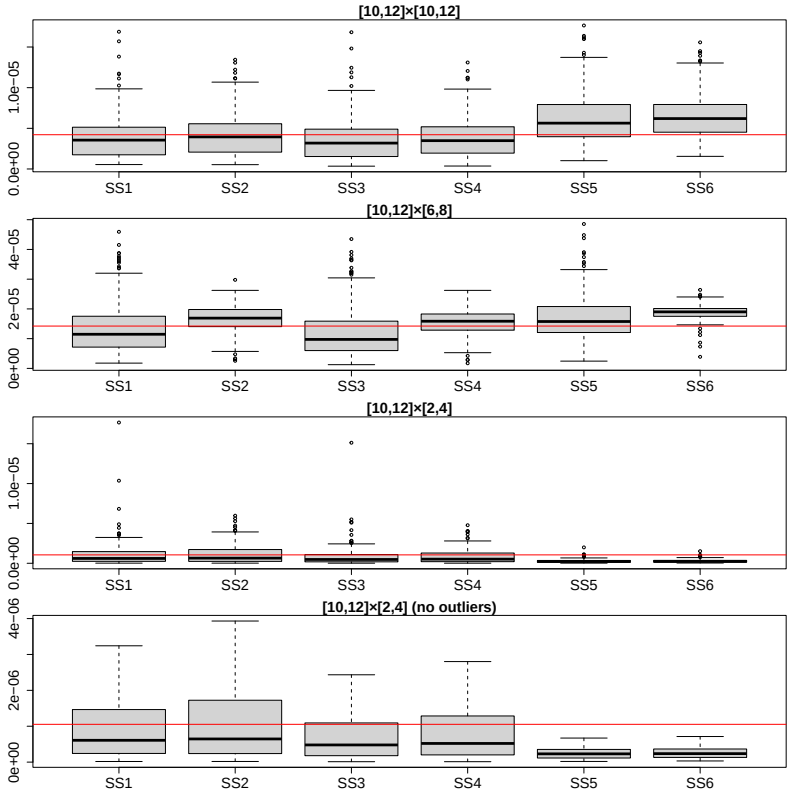

Figure 5.26: Probability estimates for distribution (III). True values shown by the solid line. Regions of interest B_i, $i = 1, 2, 3$, is given in the title, with results for B_3 given with and without outliers for clarity.

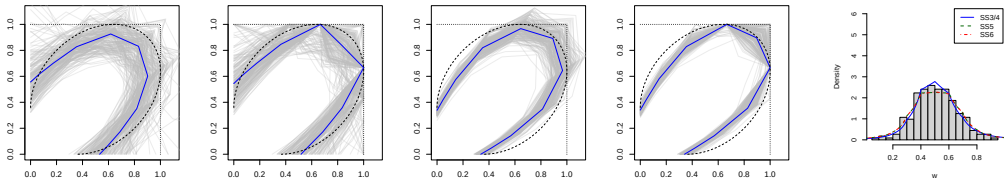

Figure 5.27: Estimates of the unit level set of g for distribution (III). Right: Estimated angular density f_W, with a sample histogram from one sample.

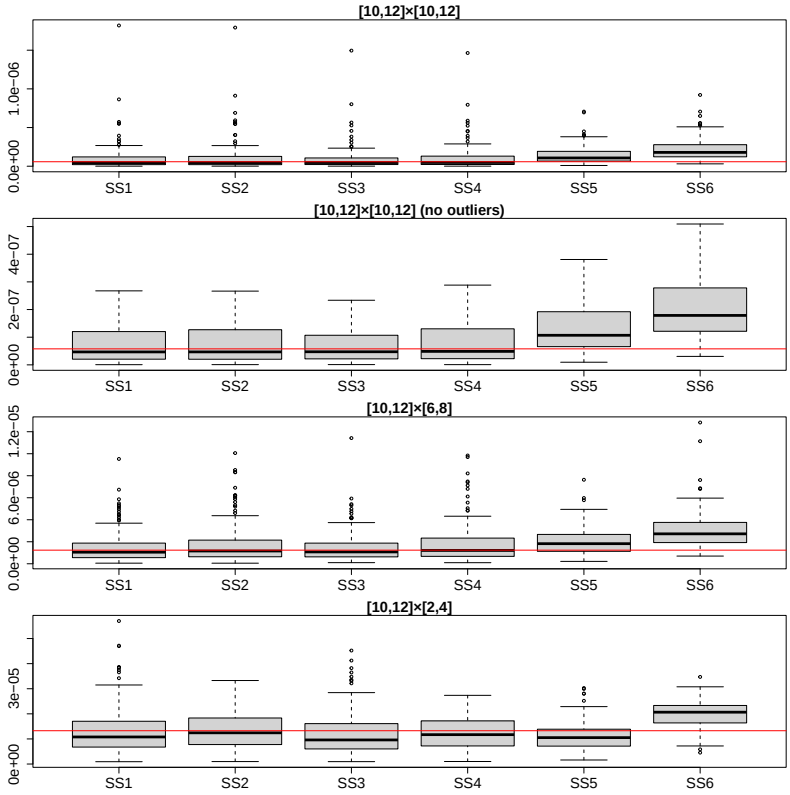

Figure 5.28: Probability estimates for distribution (IV). True values shown by the solid line. Regions of interest B_i, $i = 1, 2, 3$, is given in the title, with results for B_1 given with and without outliers for clarity.

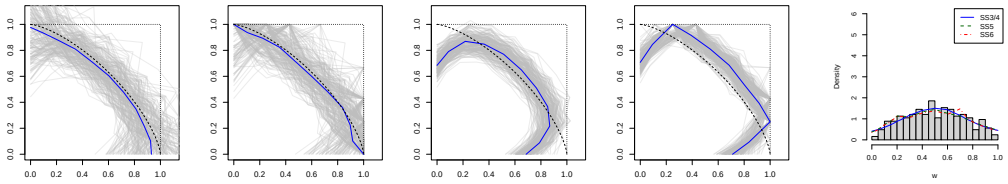

Figure 5.29: Estimates of the unit level set of g for distribution (IV). Right: Estimated angular density f_W, with a sample histogram from one sample.

Region	Setup	Dataset			
		(I)	(II)	(III)	(IV)
B1: $[10, 12] \times [10, 12]$	par	0.2670	0.2524	1.1132	0.9905
	SS1	0.7950	1.0998	0.7816	1.2671
	SS2	0.3167	0.9566	0.7066	1.2783
	SS3	0.8459	1.1404	0.8426	1.2419
	SS4	0.3649	1.0080	0.7416	1.2394
	SS5	0.4612	1.8284	0.5763	0.9888
	SS6	0.2854	1.5669	0.5549	1.2728
B2: $[10, 12] \times [6, 8]$	par	0.4907	0.3959	0.5417	0.7207
	SS1	1.0007	0.8023	0.7236	0.9285
	SS2	1.2271	0.6999	0.3941	0.9454
	SS3	1.1980	0.7814	0.7913	0.8845
	SS4	1.4300	0.6766	0.4049	0.9165
	SS5	0.7527	0.9328	0.4590	0.6808
	SS6	0.6201	0.7472	0.3237	0.9139
B3: $[10, 12] \times [2, 4]$	par	1.4067	0.6098	1.3380	0.3382
	SS1	1.7740	0.9179	1.4381	0.7816
	SS2	1.8884	0.9132	1.4235	0.6507
	SS3	1.7109	0.8893	1.5709	0.8128
	SS4	1.7456	0.8883	1.4856	0.6604
	SS5	1.1383	0.7200	1.8508	0.6130
	SS6	1.1220	0.9341	1.7152	0.4824

Table 5.2: RMSE across the 200 log-probability estimates for $d = 2$ simulation studies.

5.8.8.2 Simulation studies, $d = 3$

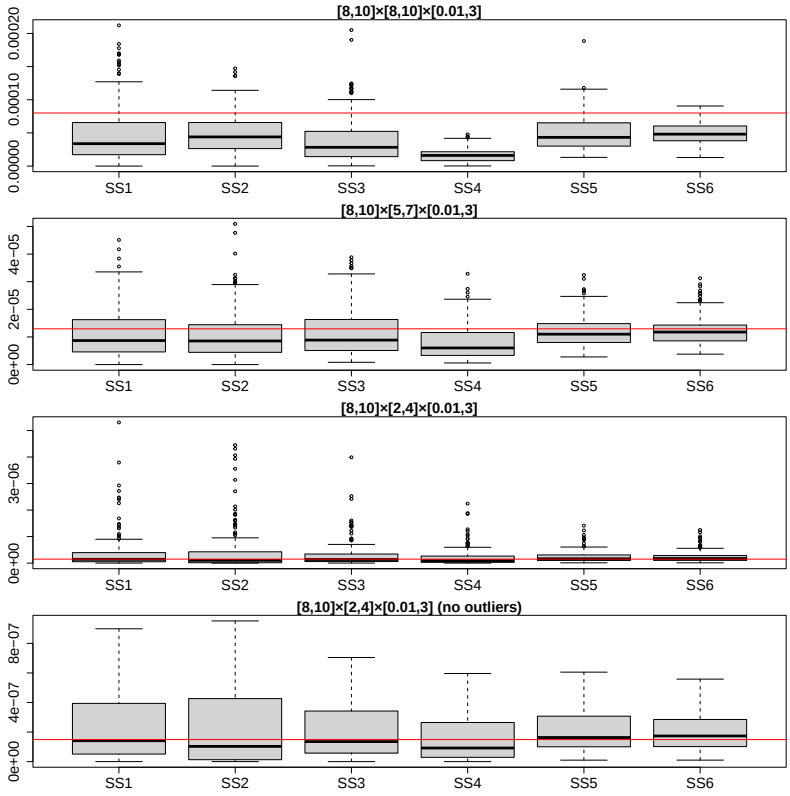

Figure 5.30: Probability estimates for distribution (V). True values shown by the solid line. Regions of interest B_i, $i = 1, 2, 3$, is given in the title, with results for B_3 given with and without outliers for clarity.

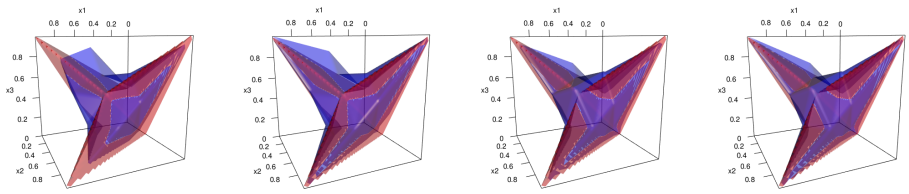

Figure 5.31: Blue: median estimates of the unit level set of g for distribution (V). Red: true unit level set.

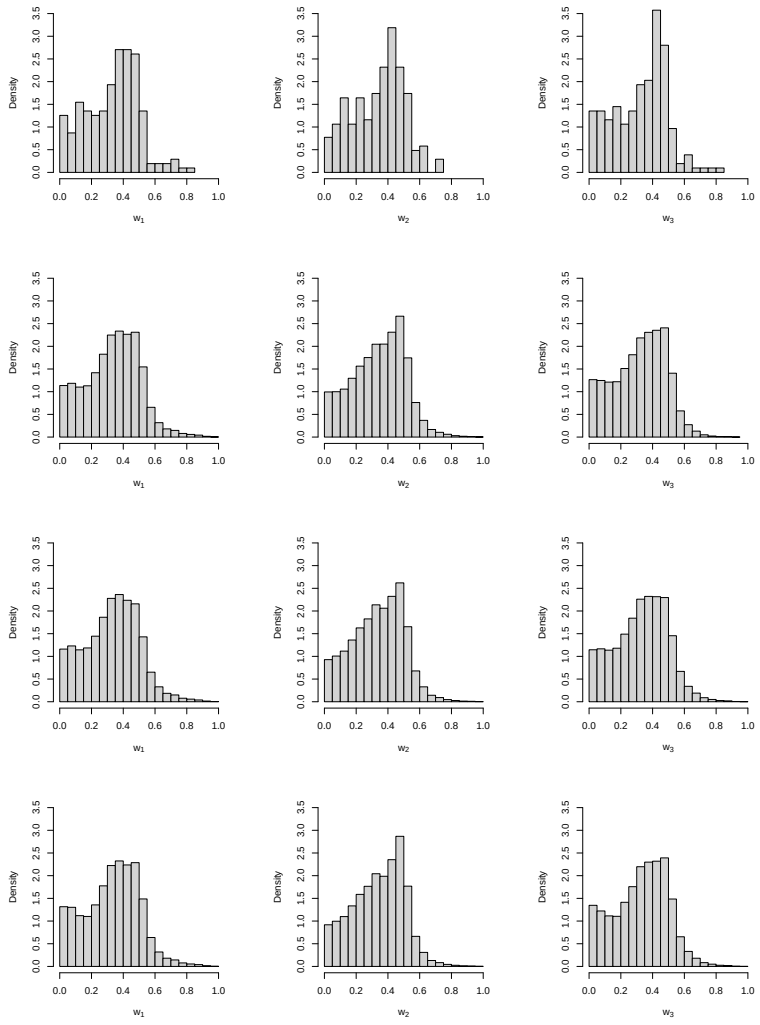

Figure 5.32: Top row: Marginal sample of exceedance angles generated of a sample from distribution (V). Rows 2–4: Marginal MCMC samples of exceedance angles from SS3/4, SS5, and SS6, respectively.

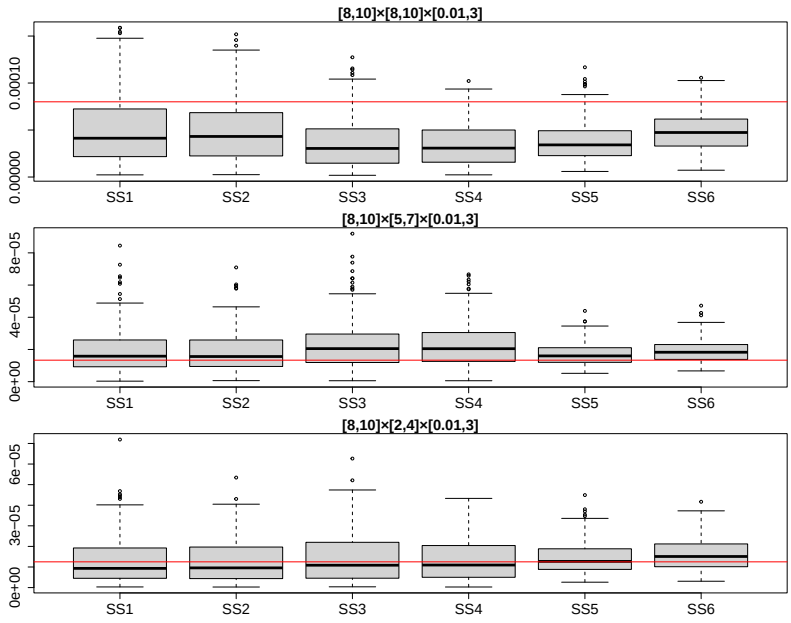

Figure 5.33: Probability estimates for distribution (VI). True values shown by the solid line. Regions of interest B_i, $i = 1, 2, 3$, is given in the title.

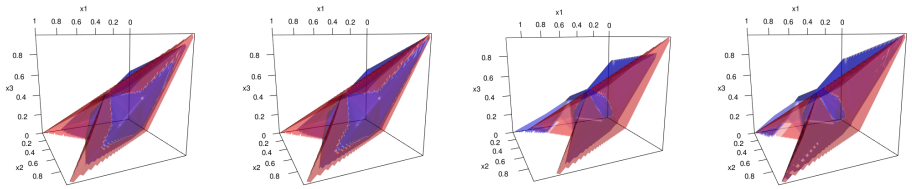

Figure 5.34: Blue: median estimates of the unit level set of g for distribution (VI). Red: true unit level set.

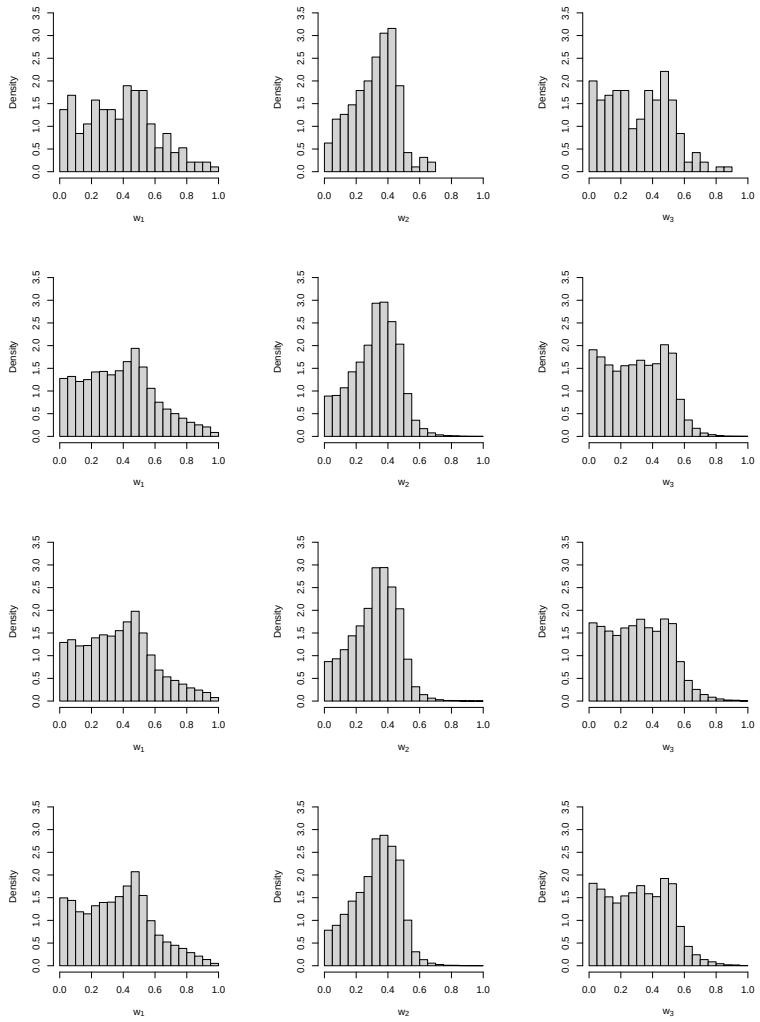

Figure 5.35: Top row: Marginal sample of exceedance angles generated of a sample from distribution (VI). Rows 2–4: Marginal MCMC samples of exceedance angles from SS3/4, SS5, and SS6, respectively.

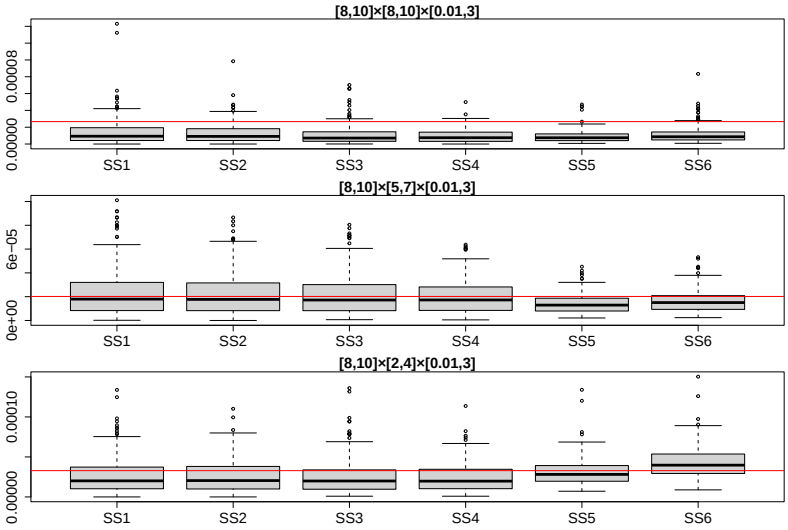

Figure 5.36: Probability estimates for distribution (VII). True values shown by the solid line.Regions of interest B_i, $i = 1, 2, 3$, is given in the title.

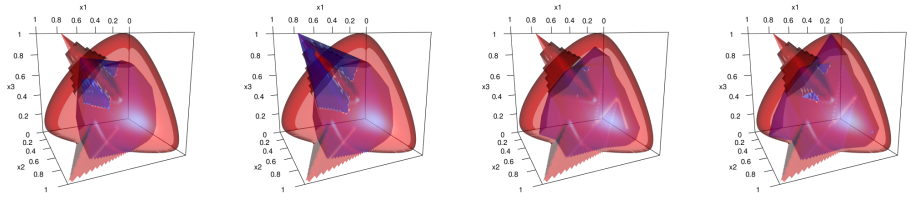

Figure 5.37: Blue: median estimates of the unit level set of g for distribution (VII). Red: true unit level set.

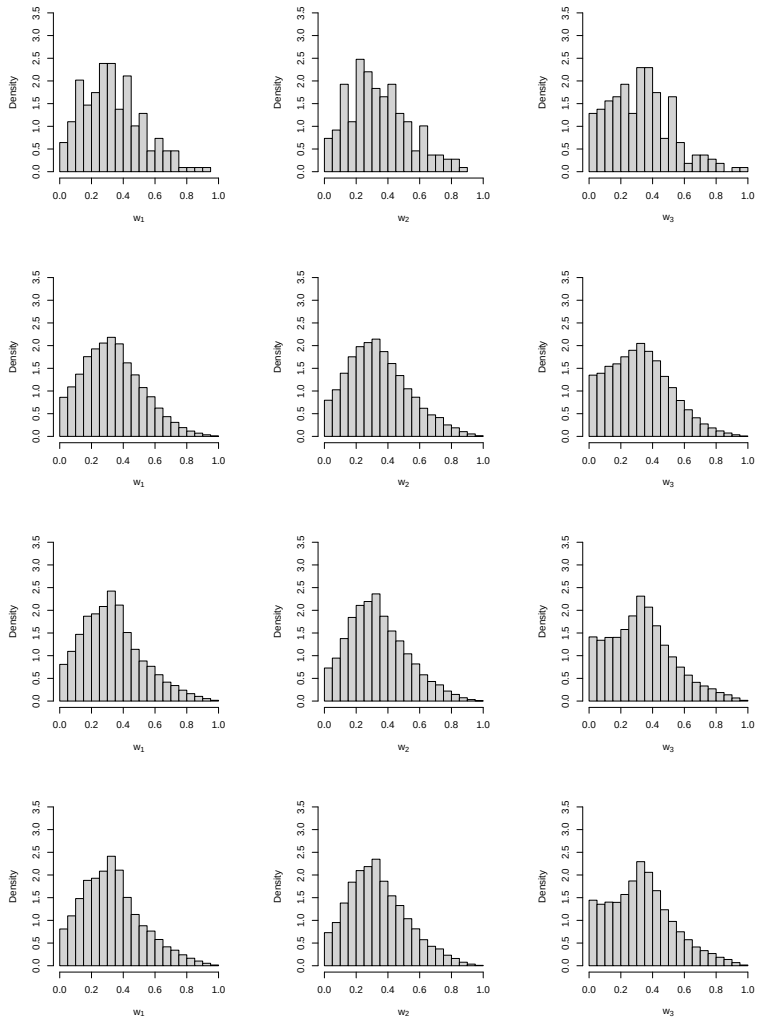

Figure 5.38: Top row: Marginal sample of exceedance angles generated of a sample from distribution (VII). Rows 2–4: Marginal MCMC samples of exceedance angles from SS3/4, SS5, and SS6, respectively.

Region	Setup	Dataset		
		(V)	(VI)	(VII)
B1: $[8,10] \times [8,10] \times [0.01,3]$	par	0.4718	0.4689	0.7422
	SS1	1.3973	1.1510	1.7031
	SS2	1.1761	1.1103	1.7008
	SS3	1.5858	1.3787	1.9959
	SS4	2.0826	1.3474	1.9339
	SS5	0.8018	1.0215	1.6056
	SS6	0.6658	0.7964	1.4271
B2: $[8,10] \times [5,7] \times [0.01,3]$	par	1.4549	0.7504	0.9414
	SS1	1.0077	0.8387	1.0864
	SS2	1.0606	0.8052	0.9941
	SS3	0.9270	0.8290	1.0006
	SS4	1.2132	0.7997	0.9698
	SS5	0.5057	0.4495	0.7624
	SS6	0.4467	0.4743	0.6673
B3: $[8,10] \times [2,4] \times [0.01,3]$	par	1.9727	0.3831	0.8872
	SS1	1.4190	1.0705	1.1091
	SS2	1.9127	1.0473	1.0885
	SS3	1.3386	1.0461	1.1305
	SS4	1.7715	1.0070	1.1027
	SS5	0.8836	0.5683	0.5294
	SS6	0.8710	0.5276	0.5009

Table 5.3: RMSE across the 200 log-probability estimates for $d = 3$ simulation studies.

5.8.9 Additional pollution data results

5.8.9.1 Setting 1: $d = 4$

Here, we present additional diagnostic plots for the $d = 4$ pollution data example from Section 5.6, in which we aim to estimate the tail behaviour of the pollutants CO, NO, PM10, and NO_2. In Figure 5.39, we show plots of $\chi_C(u)$ for high values of u for all combinations of indices $C \subseteq \{1, 2, 3, 4\}$. Values of $\chi_C(u)$ are estimated empirically and with a truncated gamma model (5.2) with gauge function estimated piecewise-linearly. We see strong agreement between the empirical values and those from the fitted model. Furthermore, all model estimates capture the asymptotic positive association between the variables in $\mathcal{C}$, as was suggested by tools introduced in Simpson et al. (2020). This demonstrates the ability of both the fitted radial and angular models to capture the extremal dependence of the for pollutants.

Figure 5.40 shows accurate estimated return level periods corresponding to curves $\mathcal{R}(T)$ against T on the log-scale. Estimates are formed by counting the proportion of points exceeding the computed boundary and taking the reciprocal as an estimate of the return period T, with good matching to the true values. The PP and QQ plots in Figure 5.41 show the fitted model for exceedance radii $R \mid \{\boldsymbol{W} = \boldsymbol{w}, R > r_\tau(\boldsymbol{W})\}$ is in general agreement with the theoretical truncated gamma model (5.2), further validating our proposed modelling approach.

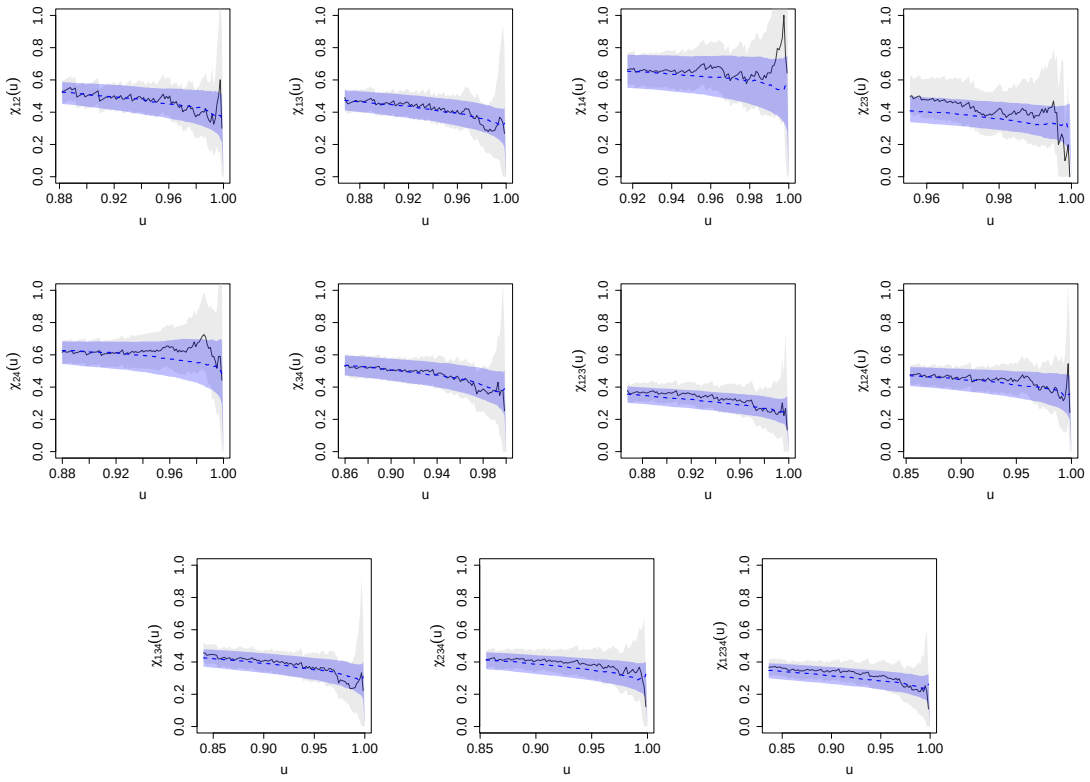

Figure 5.39: $\chi_C(u)$ plots estimated empirically (solid line) and piecewise-linearly (dashed line) on the $d = 4$ pollution dataset. Black solid lines are empirical values, and blue dashed lines are estimated using the piecewise-linear model. Shaded regions represent 7-day 95% block bootstrap confidence intervals.

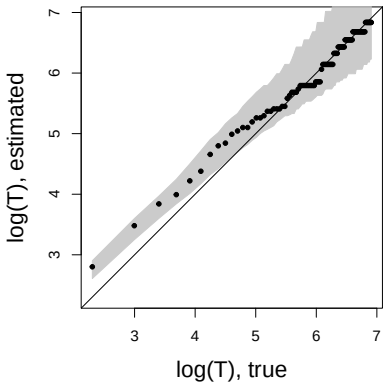

Figure 5.40: $d = 4$ pollution fitted piecewise-linear gauge estimated return periods (log-scale) compared to true values $T \in \{10, 20, 30, \ldots, 1000\}$.

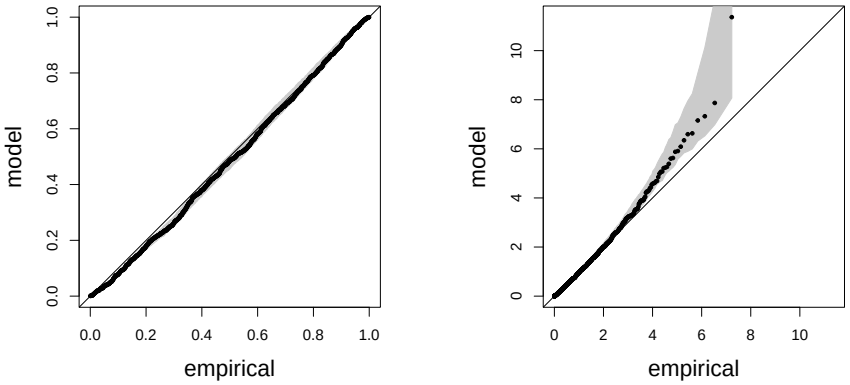

Figure 5.41: $d = 4$ pollution PP and QQ plots, with 95% 7-day block bootstrap confidence intervals in grey.

5.8.9.2 Setting 2: $d = 3$

We now consider the three-dimensional setting of modelling measurements of CO, NO_2, and PM10, encoded as in the four-dimensional setting as variables 1, 2, and 3. The first step in modelling is to obtain $r_\tau(\boldsymbol{w})$, $\boldsymbol{w} \in \mathcal{S}_2$, using the KDE approach of Section 5.3 with a Gaussian kernel with angular bandwidth $h_{\boldsymbol{W}} = 0.075$. Here, we take $\tau = 0.95$ as the quantile associated with our threshold $r_\tau(\boldsymbol{w})$. Figure 5.42 shows the resulting threshold curve $r_\tau(\boldsymbol{w})\boldsymbol{w}$ for values $\boldsymbol{w} \in \mathcal{S}_2$. Next, we obtain a triangulation using a grid of $N = 28$ reference angles at $\{0, 1/6, \ldots, 1\}^3$ within the simplex $\mathcal{S}_2$, displayed in Figure 5.21 in Appendix 5.8.8. Like the in the $d = 4$ setting, we fit the conditional radial model for $R \mid \{\boldsymbol{W} = \boldsymbol{w}, R > r_\tau(\boldsymbol{w})\}$ with penalty strength $\lambda = 1$ with bounding using Algorithm 1 in Appendix 5.8.5, along with the angular model for $\boldsymbol{W} \mid \{R > r_\tau(\boldsymbol{W})\}$ with penalty strength $\lambda = 20$.

In employing methods from Simpson et al. (2020), it was estimated that all three variables can obtain large values simultaneously, while PM10 can grow large when CO and NO_2 are both small, i.e., $\mathcal{C} = \{\{3\}, \{1, 2, 3\}\}$. The resulting limit set boundary in 5.42 is in agreement with this, since $g(1, 1, 1) = 1$ and $g(\gamma_1, \gamma_2, 1) = 1$ for $\gamma_1 = 0.673$ and $\gamma_2 = 0.3471$. The results from the methods in Simpson et al. (2020) also imply that values of $\chi_C(u)$ are expected to be positive for all values of $u \in [0, 1]$ for any collection of variables $C \subseteq \{1, 2, 3\}$. The $\chi_C(u)$ plots of Figure 5.43 also indicate this possibility, and show that our piecewise-linear model is in close agreement with the empirical estimates, demonstrating good capability of the angular and radial models in capturing the extremal behaviour of the data. Figure 5.44 shows that our model accurately estimates return periods, with three different return-level sets also displayed. Furthermore, the PP and QQ plot in Figure 5.45 show that the fitted truncated gamma model for $R \mid \{\boldsymbol{W} = \boldsymbol{w}, R > r_\tau(\boldsymbol{w})\}$ agrees with the theoretical model.

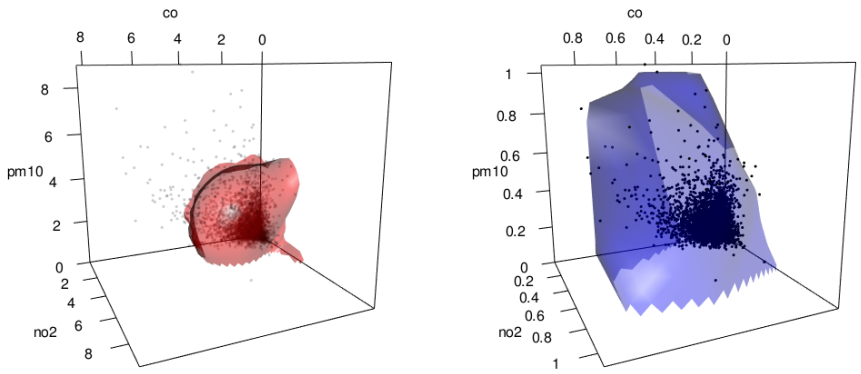

Figure 5.42: Left: $d = 3$ pollution dataset, with the radial threshold $r_{0.95}(\boldsymbol{w})$. Right: estimated gauge function unit level set with $\log(n)$-scaled data.

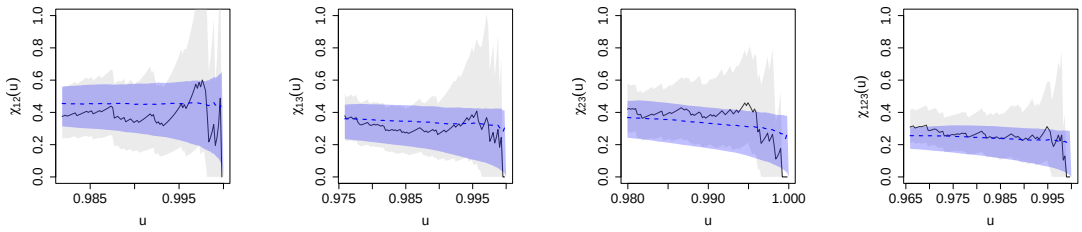

Figure 5.43: $\chi_C(u)$ plots estimated empirically (solid line) and piecewise-linearly (dashed line) on the $d = 3$ pollution dataset. From left to right: $C = \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}$. Variables 1, 2, and 3 correspond to pollutants CO, NO_2, and PM10, respectively. Shaded regions represent 7-day 95% block bootstrap confidence intervals.

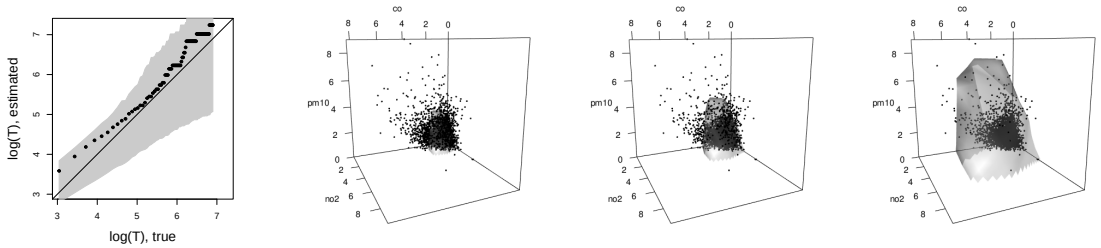

Figure 5.44: Left: $d = 3$ pollution fitted piecewise-linear gauge estimated return periods (log-scale), compared to true values $T \in \{10, 20, 30, \ldots, 1000\}$. 95% 7-day block bootstrap confidence intervals are shown in grey. Center-left–right: $T = 50, 100, 1000$ day return level set boundaries.

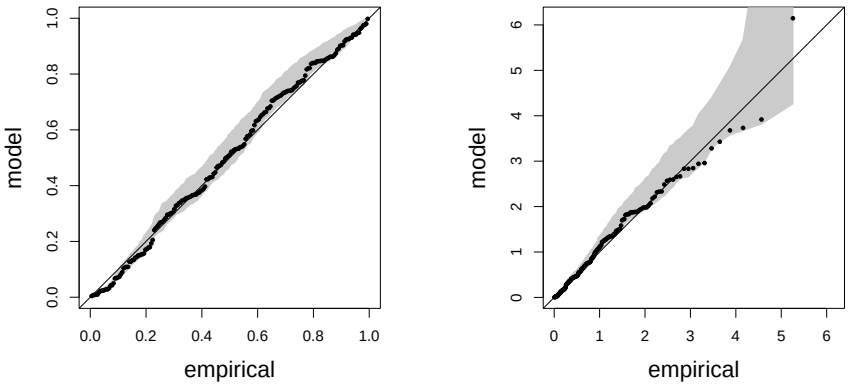

Figure 5.45: $d = 3$ pollution PP and QQ plots, with 95% 7-day block bootstrap confidence intervals in grey.

5.8.10 Piecewise-linear models for data in standard Laplace margins

In this section, we outline adaptations to the methodology for working in standard Laplace, rather than standard exponential, margins. For $d = 2$, this is straightforward, and we demonstrate the methodology. For $d > 2$, it is more complex to represent the $(d-1)$-dimensional L_1 manifold $\mathcal{S}_{d-1}^{(+,-)} = \left\{\boldsymbol{x} \in \mathbb{R}^d \middle| \|\boldsymbol{x}\|_1 = 1\right\}$ in $\mathbb{R}^{d-1}$. Figure 5.49 gives a demonstration of $\mathcal{S}_2^{(+,-)}$ in three dimensions. From this figure alone, the task of projecting to $(d-1)$-dimensions, implementing the Delaunay triangulation on this lower-dimensional space, and recovering d-dimensional vectors from their lower-dimensional L_1 representation can be seen as difficult when not restricting to the positive orthant, and is left to future work.

Mackay and Jonathan (2024) provide a useful L_1-based decomposition of bivariate copies of $\boldsymbol{X} = (X_1, X_2)^\top$ when the margins X_1, X_2 follow the standard Laplace distribution,

$$(R, W) = \left(|X_1| + |X_2|, \varepsilon \left(\frac{X_2}{|X_1| + |X_2|} \right) \left(1 - \frac{X_1}{|X_1| + |X_2|} \right) \right) \in \mathbb{R}_+ \times [-2, 2),$$

where $\varepsilon(u) = 1$ when $u \geq 0$ and $\varepsilon(u) = -1$ otherwise. In this setting, we can recover the corresponding Cartesian vectors using

$$(X_1, X_2)^\top = R \left(\frac{1 - |W|}{|1 - |W|| + |1 - |W - 1||}, \frac{1 - |W - 1|}{|1 - |W|| + |1 - |W - 1||} \right)^\top \in \mathbb{R}^2. \tag{5.13}$$

With this representation, it is possible to use the piecewise-linear framework outlined in this paper. Take, for example, the bivariate Gaussian distribution with standard Gaussian margins and correlation $\rho < 0$. When transforming to standard exponential margins, the limit set cannot be defined on the axes through a continuous gauge. However, in standard Laplace margins, the gauge function is well-defined in its $\mathbb{R}^2$ domain. In standard Laplace margins, this gauge function is given in general d-dimensions by

$$g(\boldsymbol{x}; \Sigma) = \left(\text{sign}(\boldsymbol{x}) \sqrt{|\boldsymbol{x}|} \right)^\top \Sigma^{-1} \left(\text{sign}(\boldsymbol{x}) \sqrt{|\boldsymbol{x}|} \right). \tag{5.14}$$

In equation (5.14), all operations performed on vectors are done componentwise.

We illustrate the estimation of this gauge function piecewise-linearly in a simulation study with data generated from the bivariate Gaussian distribution with standard Laplace margins and correlation $\rho = -0.5$. We generate $n = 5000$ datapoints, and perform KDE-based quantile estimation at $\tau = 0.90$. After defining a regular grid of $N = 15$ reference angles from $[-2, 2)$, including -2, we fit the models SS1–SS6 outlined in Section 5.5, and draw samples from these models to estimate the probability of lying in the regions $B_1 = [5, 9] \times [5, 9]$, $B_2 = [10, 14] \times [-2, 2]$, and $B_3 = [10, 14] \times [-14, -10]$. This is repeated 200 times. The estimated unbounded and bounded limit set boundaries from modelling $R | \{W = w, R > r_\tau(w)\}$ and $(R, W) | \{R > r_\tau(W)\}$ are displayed in Figure 5.46, where good agreement with the true gauge functions is shown. Angular models in the setting of SS3/4, SS5, and SS6 are shown in Figure 5.47. All models show good agreement with a histogram of exceedance angles. Probability estimates are displayed in Figure 5.48, where models in the settings SS2 and SS4 perform best overall. A further adjustment needs to be taken when sampling the angles W from the density $f_{\boldsymbol{W}}$ introduced in Section 5.2.2. In our MCMC algorithm, a beta proposal distribution is used. Sampled angles need to be shifted to the $[-2, 2)$ domain using the transformation $W' = 4W - 2$ before proceeding to sampling radii and obtaining extremal points using (5.13).

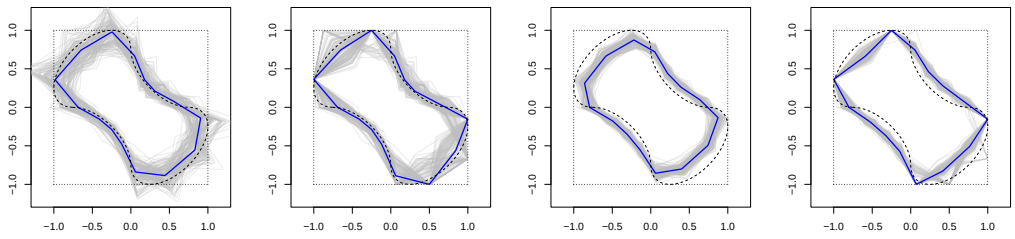

Figure 5.46: 200 estimates of the unit level set of g_{PWL} on model fits using the setup SS1, SS2, SS5, and SS6 (left to right) for the Gaussian distribution in standard Laplace margins, with median value given by the solid line. The true unit level set is given by the dashed line.

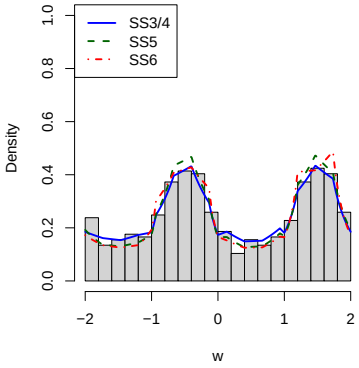

Figure 5.47: Estimate of the angular density for various simulation study setups, with an empirical angular density given by the histogram.

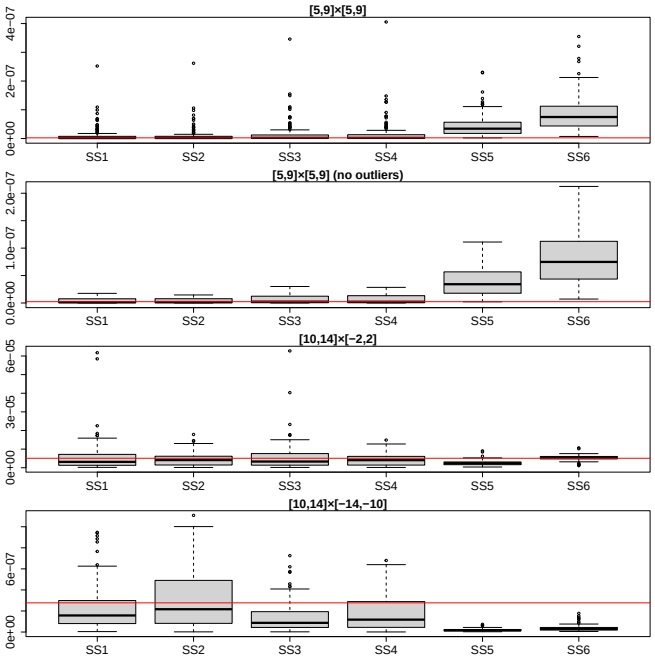

Figure 5.48: Probability estimates for bivariate Gaussian distribution in standard Laplace margins. True values given by the solid line. The region of interest is given in the figure title, and results for region B_1 are shown with and without outliers for clarity.

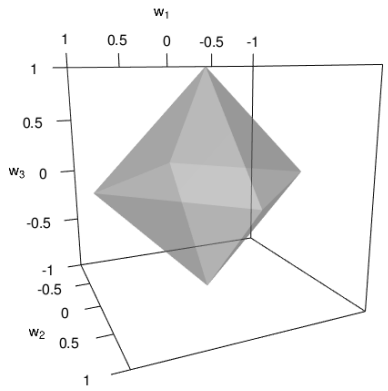

Figure 5.49: The simplex $\mathcal{S}_2^{(+,-)}$ is given by the boundary of the above surface plot.

Chapter 6

The geometric properties of generalised Pareto random vectors

Abstract

The multivariate generalised Pareto (MGP) distribution describes the limiting behaviour of random vectors whose componentwise maximum exceeds a large threshold, and is therefore important to the study of multivariate extreme value analysis. A feature of MGP random vectors is that they can be represented stochastically, allowing for their construction from any unbounded multivariate distribution. In this work, we show that the limit set of scaled sample clouds of MGP random vectors can be obtained through the geometry of one of the components of the spectral representation. When MGP random vectors do not have a non-trivial limit set, an alternative coordinate system is introduced to inspect their geometric and tail dependence properties.

6.1 Introduction

Peaks-over-threshold (POT) modelling has been a mainstay in extreme value analysis since its formalisation in the univariate setting (Pickands, 1975), and has since been successfully used to model data from applications ranging from hydrology

to finance (see, for example, Chapter 4.4 in Coles (2001)). Suppose Y is a random variable used to denote values of some process. In the POT framework, inference is performed by assuming that for large r_0,

$$Y - r \mid \{r > r_0\} \sim \mathrm{GP}(r_0, \sigma, \xi),$$

where $\mathrm{GP}(r_0, \sigma, \xi)$ is the generalised Pareto distribution with location, scale, and shape parameters $(r_0, \sigma, \xi) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$. The GP distribution is most-commonly characterised by its conditional survival function,

$$\Pr(Y > r \mid Y > r_0) = \left[1 + \xi \left(\frac{r - r_0}{\sigma}\right)\right]_+^{-1/\xi}.$$

The GP distribution arises as the limiting family for exceedances of high thresholds for a wide variety of underlying distributions. Given this fact, it is a natural modelling choice for extrapolation above high thresholds, beyond the range of any observed data. This makes it a powerful tool, and the need arises for a multivariate extension.

Suppose now that $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^\top$ is a d-dimensional random vector, with each component representing measurements of a simultaneous process. Introduced in Rootzén and Tajvidi (2006) and further formalised in Rootzén et al. (2018a) and Rootzén et al. (2018b), the multivariate generalised Pareto (MGP) arises when studying excesses of $\boldsymbol{Y}$ when at least one of its components exceeds a component of a threshold vector, $Y_j > r_{0,j}$. Rootzén et al. (2018b) focus on the theoretical properties of the MGP distributions and their associated random vectors. Namely, they introduce different parametrisations, derive the joint densities, marginal distributions, tail dependence coefficients, and establish threshold stability. Rootzén et al. (2018b) further derives a stochastic representation, showing that one could easily construct an MGP random vector from a random vector of unbounded domain. Kiriliouk et al. (2019) derive parametric MGP distributions using the stochastic representation from several known copulas, introduce statistical inference procedures, and apply the MGP model on financial and environmental data.

The multivariate POT MGP framework is inherently linked to the notion of multivariate regular variation (MRV) (Wan, 2024). Due to the connection to MRV, there are several drawbacks in working with MGP random vectors. Most prominent is the limitation in extrapolation directions. When modelling random vectors using the MGP distribution, one can only extrapolate along the direction of simultaneous growth. Therefore, this restricts the range of extremal probabilities that could be estimated. Furthermore, the MGP has an extremal tail decay rate that is only suitable for random vectors whose components are asymptotically dependent, meaning that more general scenarios cannot be accounted for. The geometric inference approach for the tail behaviour of random vectors presented in Wadsworth and Campbell (2024) bypasses these issues by modelling the tail decay across the entire domain of the random vector through use of the truncated gamma distribution. One drawback of the geometric framework however is that the popular Hüsler-Reiss MGP model (e.g., Engelke et al. (2015); Engelke and Hitz (2020)) has a degenerate limit set (Nolde and Wadsworth, 2022). In this work, we look closely at limit sets for MGP distributions. We derive the form of the limit set for a large class of MGP random vectors, and show how to handle degenerate cases like the Hüsler-Reiss with alternative coordinate systems.

We begin in Section 6.2 by covering the essential background knowledge required for linking the MGP distribution setting to the geometric framework of multivariate extremes. Section 6.3 discusses the geometry of MGP random vectors, with several examples displayed. Alternative statistical inference for MGP random vectors is illustrated in Section 6.4 using the geometric approach from Wadsworth and Campbell (2024). Section 6.5 discusses how useful geometry of MGP random vectors can be obtained when the usual limiting behaviour is degenerate. We finish with a discussion on the benefits and drawbacks of the MGP and geometric frameworks in Section 6.6.

6.2 Background

6.2.1 The multivariate generalised Pareto distribution

Before we inspect the geometric properties of MGP random vectors, it is important to characterise them in the classical sense. Suppose a random vector $\boldsymbol{Y}$ is in the domain of attraction of a generalised extreme value distribution (GEV) with joint distribution function G. In other words, there exists suitable sequences $\{\boldsymbol{a}_n\}$ and $\{\boldsymbol{b}_n\}$ such that, for i.i.d. copies $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots$ of $\boldsymbol{Y}$,

$$\Pr\left[\frac{\max_{i=1,\ldots,n} \boldsymbol{Y}_i - \boldsymbol{b}_n}{\boldsymbol{a}_n} \leq \boldsymbol{y}\right] \longrightarrow G(\boldsymbol{y}) \tag{6.1}$$

as $n \to \infty$, where $\max_{i=1,\ldots,n} \boldsymbol{Y}_i$ denotes the componentwise maximum of $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$. Here and throughout, operations on vectors are done componentwise. The margins of this GEV distribution follow a univariate GEV with parameters (μ_j, σ_j, ξ_j). If the convergence (6.1) holds, then the distribution of exceedances of $\boldsymbol{Y}$ converges to a MGP random vector, $\boldsymbol{X}$:

$$\max\left\{\frac{\boldsymbol{Y} - \boldsymbol{b}_n}{\boldsymbol{a}_n}, \boldsymbol{\eta}\right\} \leq \boldsymbol{y} \,\middle|\, \{\exists\, j \text{ s.t. } Y_j > b_{n,j}\} \longrightarrow \boldsymbol{X} \tag{6.2}$$

as $n \to \infty$, where $\boldsymbol{\eta}$ is the collection of d lower endpoints of the marginal distributions of G. The distribution function of the MGP random vector $\boldsymbol{X}$ has the form

$$F(\boldsymbol{x}) = \Pr(\boldsymbol{X} \leq \boldsymbol{x}) = \frac{\log G(\min\{\boldsymbol{x}, \boldsymbol{0}\}) - \log G(\boldsymbol{x})}{\log G(\boldsymbol{0})}$$

and the margins X_j have conditional distribution $\{X_j \mid X_j > 0\} \sim \mathrm{GP}(0, \sigma_j, \xi_j)$. To focus on dependence, suppose the components Y_j of the random vector $\boldsymbol{Y}$ follow the standard exponential distribution. Then if convergence (6.2) holds, we have

$$\boldsymbol{Y} - r \mid \{\max(\boldsymbol{Y}) > r\} \longrightarrow \boldsymbol{X}$$

as $r \to \infty$, where $\boldsymbol{X}$ is an MGP random vector with $\sigma_j = 1$ and $\xi_j = 0$, $j = 1, \ldots, d$. This standardised characterisation is important when establishing the stochastic representation of MGP random vectors. It is established in Rootzén et al. (2018b)

that a standard MGP random vector is equal in distribution to a so-called *spectral random vector* and a scalar standard exponential random variable,

$$\boldsymbol{X} \overset{d}{=} E\mathbf{1} + \boldsymbol{S}, \tag{6.3}$$

where E and $\boldsymbol{S}$ are independent, and $\max(\boldsymbol{S}) = 0$. The spectral vector $\boldsymbol{S}$ arises from $\boldsymbol{Y}$ through the limit

$$\boldsymbol{Y} - \max(\boldsymbol{Y}) \mid \{\max(\boldsymbol{Y}) > r\} \longrightarrow \boldsymbol{S}$$

as $r \to \infty$. Using the stochastic representation (6.3), one can construct MGP random vectors from *any* unbounded random vector. Suppose $\boldsymbol{T}$ lives on an unbounded domain in $\mathbb{R}^d$. Rootzén et al. (2018b) construct a valid spectral vector $\boldsymbol{S}$ by setting $\boldsymbol{S} = \boldsymbol{T} - \max(\boldsymbol{T})$, and therefore, $\boldsymbol{X} \overset{d}{=} E\mathbf{1} + \boldsymbol{T} - \max(\boldsymbol{T})$. This stochastic representation will be essential in analysing the geometric properties of MGP random vectors. After establishing some necessary background, we will show how MGP random vectors $\boldsymbol{X}$ have an inherent geometry that is solely linked to the geometry of the spectral component, $\boldsymbol{S}$.

6.2.2 Geometric extremes

The geometric approach to multivariate extremes (Nolde, 2014; Nolde and Wadsworth, 2022; Wadsworth and Campbell, 2024) induces more generality than the frameworks of MGP vectors and, by extension, MRV. In it, interest lies in characterising the multivariate tail above high quantiles of functions of random vectors. Suppose the d-dimensional random vector $\boldsymbol{Y}$ has exponential margins and joint density function f. The main assumption behind the geometric approach is that the rescaled negative log-density tends towards a non-degenerate limiting *gauge function*,

$$-r^{-1} \log f(r\boldsymbol{y}) \longrightarrow g(\boldsymbol{y}) \tag{6.4}$$

as $r \to \infty$. The gauge function g is positive, scalar-valued, and 1-homogeneous. Wadsworth and Campbell (2024) show that if (6.4) holds with g continuous, then

it is typically reasonable to assume the model

$$\|\boldsymbol{Y}\| \mid \{\|\boldsymbol{Y}\| > r_0(\boldsymbol{w}), \boldsymbol{Y}/\|\boldsymbol{Y}\| = \boldsymbol{w}\} \sim \text{truncGamma}(\alpha, g(\boldsymbol{w})) \tag{6.5}$$

as $r_0(\boldsymbol{w}) \to \infty$, where $r_0(\boldsymbol{w})$ is a high threshold of $\|\boldsymbol{Y}\| \mid \{\boldsymbol{Y}/\|\boldsymbol{Y}\| = \boldsymbol{w}\}$. In (6.5), truncGamma refers to the truncated gamma distribution, a univariate distribution most easily defined through its density function,

$$f_{\text{tG}}(r; \alpha, g(\boldsymbol{w}), r_0(\boldsymbol{w})) = f_{\text{G}}(r; \alpha, g(\boldsymbol{w}))/\bar{F}_{\text{G}}(r_0(\boldsymbol{w}); \alpha, g(\boldsymbol{w}))$$

for $r > r_0(\boldsymbol{w})$, where f_{G} and $\bar{F}_{\text{G}}$ are the density and survival function of the gamma distribution, respectively.

If the convergence assumption on the log-density (6.4) holds with g continuous, it then follows that scaled sample clouds of the form $N_n = \{\boldsymbol{Y}_1/\log n, \ldots, \boldsymbol{Y}_n/\log n\}$ converge onto a limit set $\mathcal{G} = \{\boldsymbol{x} \in \mathbb{R}^d \mid g(\boldsymbol{x}) \leq 1\}$. The intrinsic link between the limit set $\mathcal{G}$ and the gauge function g means that knowledge of g can give information about the geometric properties of a random vector. These properties were first explored in Balkema et al. (2010), Balkema and Nolde (2010), Balkema and Nolde (2012), Nolde (2014), and more recently in Nolde and Wadsworth (2022), where it was shown that knowledge of g gives information as to which groups of variables in the random vector $\boldsymbol{Y}$ grow large simultaneously. Furthermore, g gives information on the degree of dependence or independence in the joint tails between groups of variables. Together with the truncated gamma approach to inference, the geometric framework allows for inference across the entire tail region of random vectors, making it a powerful tool in overcoming some of the limitations of the MRV assumption. With this in mind, studying the geometric properties of MGP random vectors may provide a useful link between the geometric framework and classical multivariate extreme value analysis.

6.3 The limit set associated with MGP random vectors

In order to inspect the geometric properties of MGP random vectors, it is first necessary to study its joint density. Theorem 15 of Rootzén et al. (2018b) characterises the joint density of MGP random vectors $\boldsymbol{X}$ through its stochastic representation, $E + \boldsymbol{S}$. The joint density of the resulting MGP is given by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbf{1}_{\mathbb{R}_+}(\max(\boldsymbol{x})) \exp\{-\max(\boldsymbol{x})\} f_{\boldsymbol{S}}(\boldsymbol{x} - \max(\boldsymbol{x})), \tag{6.6}$$

where $\max(\boldsymbol{x}) = \max_{j\in\{1,\ldots,d\}} x_j$ and $f_{\boldsymbol{S}}$ is a Lebesgue density on the $(d-1)$-dimensional constrained space $\mathcal{S} = \{\boldsymbol{s} \in \mathbb{R}^d : \max(\boldsymbol{s}) = 0\}$. Using this expression for the joint density, it is possible to obtain the corresponding gauge function under certain conditions.

Proposition 6.1 (Gauge function of MGP random vectors)**.** *Suppose* $-\log f_{\boldsymbol{S}}(t\boldsymbol{x})/t \longrightarrow g_{\boldsymbol{S}}(\boldsymbol{x})$*, where* $g_{\boldsymbol{S}}$ *is a continuous gauge function defined on the domain* $\mathcal{S}$*, then* $\{\boldsymbol{S}_k/\log n\}_{k=1}^n$ *tends to a non-degenerate limit set defined by the gauge function* $g_{\boldsymbol{S}}$*. Furthermore,* $\log n$*-scaled sample clouds of MGP random vectors* $\{\boldsymbol{X}_n/\log n\}_{k=1}^n$ *converge onto a limit set with the associated gauge function*

$$g_{\boldsymbol{X}}(\boldsymbol{x}) = \max(\boldsymbol{x}) + g_{\boldsymbol{S}}(\boldsymbol{x} - \max(\boldsymbol{x})) \tag{6.7}$$

and is defined on the space $\big\{\boldsymbol{x} \in \mathbb{R}^d : \max(\boldsymbol{x}) \geq 0\big\}$.

The proof of the first statement in Proposition 6.1 follows as in Proposition 2 of Nolde and Wadsworth (2022). The form of $g_{\boldsymbol{X}}$ in (6.7) follows directly from the density (6.6). We can also independently derive (6.7) as a consequence of the findings in Nolde and Wadsworth (2022) on the gauge functions of additive mixtures of independent random vectors, we derive the equivalent result by minimising over the domain $\mathcal{S}$ of the spectral random vector $\boldsymbol{S}$,

$$
\begin{aligned}
g_{\boldsymbol{X}}(\boldsymbol{x}) &= \min_{a \,:\, \boldsymbol{x}-a \in \mathcal{S}} a + g_{\boldsymbol{S}}(\boldsymbol{x} - a) \\
&= \max(\boldsymbol{x}) + g_{\boldsymbol{S}}(\boldsymbol{x} - \max(\boldsymbol{x})).
\end{aligned}
$$

Given some random vector $\boldsymbol{T}$ following a multivariate distribution with an unbounded domain, we can generate spectral random vectors using the relation $\boldsymbol{S} = \boldsymbol{T} - \max(\boldsymbol{T})$. For this reason, we refer to $\boldsymbol{T}$ as a *generator* of an MGP-distributed random vector. By Theorem 12 of Rootzén et al. (2018b), we obtain the density for $\boldsymbol{X}$

$$
\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x}) =& \mathbf{1}_{\mathbb{R}_+}(\max(\boldsymbol{x})) \exp\{-\max(\boldsymbol{x})\} \int_{-\infty}^{\infty} f_{\boldsymbol{T}}(\boldsymbol{x} + s) ds \\
=& \mathbf{1}_{\mathbb{R}_+}(\max(\boldsymbol{x})) \exp\{-\max(\boldsymbol{x})\} \int_{0}^{\infty} f_{\boldsymbol{T}}(\boldsymbol{x} + \log t) t^{-1} dt
\end{aligned}
$$

This implies that, for $\boldsymbol{x} \in \mathbb{R}^d$ and assuming $\boldsymbol{T}$ has density $f_{\boldsymbol{T}}$,

$$
f_{\boldsymbol{S}}(\boldsymbol{x} - \max(\boldsymbol{x})) = \int_{-\infty}^{\infty} f_{\boldsymbol{T}}(\boldsymbol{x} + s) ds = \int_{0}^{\infty} f_{\boldsymbol{T}}(\boldsymbol{x} + \log t) t^{-1} dt \tag{6.8}
$$

Using this knowledge, beginning with a known distribution $\boldsymbol{T}$, it is possible to obtain the corresponding MGP gauge function. In what follows, we derive the spectral gauge function $g_{\boldsymbol{S}}$ and the resulting MGP gauge function $g_{\boldsymbol{X}}$ from generators previously examined by Kiriliouk et al. (2019), where the density $f_{\boldsymbol{S}}$ in equation (6.8) was derived given $\boldsymbol{T}$.

Example: Generators with independent Gumbel components, equal scale parameters Suppose $T_j \sim \text{Gumbel}(\alpha, \beta_j)$. We have

$$
f_{\boldsymbol{S}}(\boldsymbol{x} - \max(\boldsymbol{x})) = \alpha^{d-1} \Gamma(d) \prod_{j=1}^{d} e^{-\alpha(x_j - \beta_j)} \left(\sum_{j=1}^{d} e^{-\alpha(x_j - \beta_j)} \right)^{-d}
$$

Therefore, spectral random vectors associated with independent Gumbel generators have gauge function

$$
g_{\boldsymbol{S}}(\boldsymbol{x} - \max(\boldsymbol{x})) = \alpha \sum_{j=1}^{d} x_j - d\alpha \min(\boldsymbol{x}).
$$

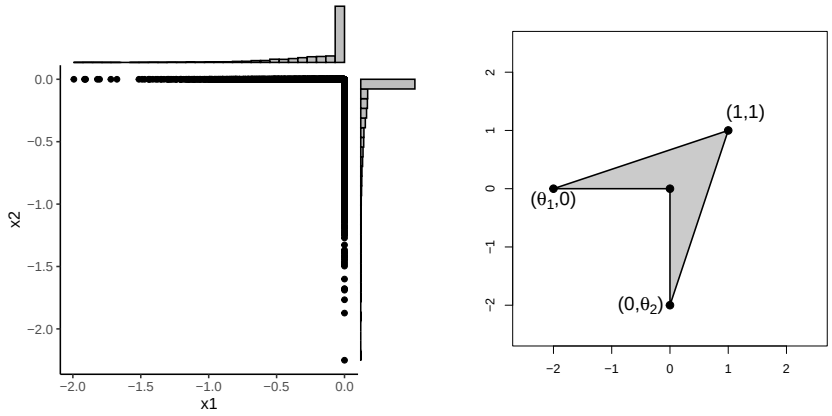

Figure 6.1: Right: Marginal histogram of a sample from the spectral random vector associated to the bivariate independent Gumbel generator with rate $\alpha = 0.5$. Left: the limit set associated to the resulting bivariate MGP random vector.

The resulting MGP distribution has gauge function

$$g_{\boldsymbol{X}}(\boldsymbol{x}) = \max(\boldsymbol{x}) + \alpha \sum_{j=1}^{d} x_j - d\alpha \min(\boldsymbol{x}).$$

Example: Generators with independent reverse exponential components Suppose T_j follows the reverse exponential distribution with rate parameter α_j and location parameter β_j. We have

$$f_{\boldsymbol{S}}(\boldsymbol{x} - \max(\boldsymbol{x})) = \left(\sum_{j=1}^{d} \alpha_j\right)^{-1} e^{-\left(\sum_{j=1}^{d} \alpha_j\right) \max(\boldsymbol{x}+\boldsymbol{\beta})} \prod_{j=1}^{d} \alpha_j e^{\alpha_j (x_j + \beta_j)}.$$

Therefore, spectral random vectors associated with independent reverse exponential generators have gauge function

$$\begin{aligned} g_{\boldsymbol{S}}(\boldsymbol{x} - \max(\boldsymbol{x})) &= \left(\sum_{j=1}^{d} \alpha_j\right) \max(\boldsymbol{x}) - \sum_{j=1}^{d} \alpha_j x_j \\ &= -\boldsymbol{\alpha}^\top (\boldsymbol{x} - \max(\boldsymbol{x})), \end{aligned}$$

and the resulting MGP distribution has gauge function

$$g_{\boldsymbol{X}}(\boldsymbol{x}) = \max(\boldsymbol{x}) - \boldsymbol{\alpha}^\top (\boldsymbol{x} - \max(\boldsymbol{x})).$$

Example: Generators with independent log-gamma components Suppose $T_j \sim \text{logGamma}(\alpha_j, 1)$. We have

$$f_{\boldsymbol{S}}(\boldsymbol{x} - \max(\boldsymbol{x})) = \frac{\Gamma\left(\sum_{j=1}^{d} \alpha_j\right)}{\prod_{j=1}^{d} \Gamma(\alpha_j)} \exp\left\{\sum_{j=1}^{d} \alpha_j x_j\right\} \left(\sum_{j=1}^{d} e^{x_j}\right)^{-\sum_{j=1}^{d} \alpha_j}$$

Therefore, spectral random vectors associated with independent log-gamma generators have gauge function

$$\begin{aligned} g_{\boldsymbol{S}}(\boldsymbol{x} - \max(\boldsymbol{x})) &= \left(\sum_{j=1}^{d} \alpha_j\right) \max(\boldsymbol{x}) - \sum_{j=1}^{d} \alpha_j x_j \\ &= -\boldsymbol{\alpha}^\top (\boldsymbol{x} - \max(\boldsymbol{x})), \end{aligned}$$

and the resulting MGP distribution vector has gauge function

$$g_{\boldsymbol{X}}(\boldsymbol{x}) = \max(\boldsymbol{x}) - \boldsymbol{\alpha}^\top (\boldsymbol{x} - \max(\boldsymbol{x})).$$

Note that when considering the $d = 2$ setting in the above examples with $\alpha_1 = \alpha_2$, $g_{\boldsymbol{S}}$ are all equivalent and produce the same limit set. This can be seen in the independent Gumbel generator example by adding and subtracting $\alpha(x_1 + x_2)$. In fact, in the bivariate setting, all MGP random vectors have the same gauge function if the spectral component converges to a non-degenerate limit set under $\log n$ scaling. The form of the limit sets described in Proposition 6.1 are illustrated in Figure 6.1.

Proposition 6.2. *Suppose $\boldsymbol{X} = E\mathbf{1} + \boldsymbol{S}$ is a bivariate MGP random vector, and that Proposition 6.1 holds. Define the values $\theta_1, \theta_2 < 0$ such that*

$$\begin{aligned} \theta_1 \quad &: \quad g_{\boldsymbol{S}}(\theta_1, 0) = 1 \\ \theta_2 \quad &: \quad g_{\boldsymbol{S}}(0, \theta_2) = 1; \end{aligned}$$

Equivalently,

$$\begin{aligned} |\theta_1| &= 1/g_{\boldsymbol{S}}(-1, 0) \\ |\theta_2| &= 1/g_{\boldsymbol{S}}(0, -1). \end{aligned}$$

Then the limit sets of $d = 2$ MGP random vectors can be obtained by joining the vertices $(1,1)^\top$, $(\theta_1, 0)^\top$, $(0,0)^\top$, and $(0, \theta_2)$. The associated gauge function is given by

$$g_{\text{MGP}}(x_1, x_2; \theta_1, \theta_2) = \frac{1}{\theta_1} x_1 + \frac{1}{\theta_2} x_2 + \left(1 - \frac{1}{\theta_1} - \frac{1}{\theta_2}\right) \max(x_1, x_2),$$

and the limit set can alternatively be written as $\mathcal{G} = \{(x_1, x_2) \in \mathbb{R}^2 : g_{\text{MGP}}(x_1, x_2; \theta_1, \theta_2) \leq 1\} \cap \{(x_1, x_2) \in \mathbb{R}^2 : \max(x_1, x_2) \geq 0\}$.

Proof. It is sufficient to derive the gauge function associated to the spectral random vector $\boldsymbol{S}$. The gauge function $g_{\boldsymbol{S}}$ must be positive, 1-homogeneous, and have domain $\mathcal{S}$. Inputs must be of the form $(x_1, 0)$ for $x_1 < 0$ or $(0, x_2)$ for $x_2 < 0$. By homogeneity,

$$\begin{aligned} g_{\boldsymbol{S}}(x_1, x_2) &= \begin{cases} |x_1| g(-1, 0) & ; \quad x_1 < 0, x_2 = 0 \\ |x_2| g(0, -1) & ; \quad x_1 = 0, x_2 < 0 \end{cases} \\ &= \begin{cases} \frac{|x_1|}{|\theta_1|} & ; \quad x_1 < 0, x_2 = 0 \\ \frac{|x_2|}{|\theta_2|} & ; \quad x_1 = 0, x_2 < 0 \end{cases} \\ &= \frac{|x_1|}{|\theta_1|} + \frac{|x_2|}{|\theta_2|}, \quad (x_1, x_2) \in \mathcal{S}. \end{aligned}$$

To ensure the domain $\mathcal{S}$, the gauge function evaluated at a generic point $\boldsymbol{x} \in \mathbb{R}^2$ is therefore

$$\begin{aligned} g_{\boldsymbol{S}}(x_1 - \max(x_1, x_2), x_2 - \max(x_1, x_2)) =& \frac{|x_1 - \max(x_1, x_2)|}{|\theta_1|} + \frac{|x_2 - \max(x_1, x_2)|}{|\theta_2|} \\ =& \frac{x_1 - \max(x_1, x_2)}{\theta_1} + \frac{x_2 - \max(x_1, x_2)}{\theta_2}. \end{aligned}$$

By (6.7), we recover the desired MGP gauge function g_{MGP}.

For the vertices of the limit set $\mathcal{G}$, $g_{\text{MGP}}(\theta_1, 0) = g_{\text{MGP}}(0, \theta_2) = 1$ by construction, and it is clear that $g_{\text{MGP}}(1, 1) = 1$. The straight lines between $(\theta_1, 0)^\top$, $(1, 1)^\top$, and $(0, \theta_2)^\top$ on $\partial\mathcal{G}$ holds because $g_{\text{MGP}}(x_1, (1 - \theta_2)x_1 + \theta_2) = 1$ for $x_1 \in (0, 1)$ and $g_{\text{MGP}}((1 - \theta_1)x_2 + \theta_1, x_2) = 1$ for $x_2 \in (0, 1)$. The vertex of $\mathcal{G}$ at $(0, 0)^\top$ holds because $\mathcal{S}$ forms part of the boundary $\partial\mathcal{G}$. □

6.4 Statistical inference for MGP random vectors using the truncated gamma model

Having an expression for the gauge g_{MGP} means that we can use inference tools such as the one presented in Wadsworth and Campbell (2024) to allow the estimation of small probabilities when at least one component is large. We will discuss the pros and cons of this approach in Section 6.6. Suppose the random vector $\boldsymbol{X}$ follows a MGP, and is constructed using $\boldsymbol{X} = E + \boldsymbol{T} - \max(\boldsymbol{T})$. In the bivariate case, we can model

$$\max(\boldsymbol{X}) \mid \{\boldsymbol{X}/\max(\boldsymbol{X}) = \boldsymbol{w}, \max(\boldsymbol{X}) > u\} \sim \text{truncGamma}(a, g_{\text{MGP}}(\boldsymbol{x}; \theta_1, \theta_2)),$$

where the model parameters (a, θ_1, θ_2) can be estimated with the standard maximum likelihood method.

With this inference set up in mind, one can easily estimate joint tail probabilities associated with MGP random vectors. Take, for example, the generator $\boldsymbol{T}$ obtained from bivariate independent Gumbel distribution with rate parameter $\alpha = 0.5$. Its true underlying limit set is shown in Figure 6.1. Using the derived expression g_{MGP}, we can estimate this. Once a high threshold is selected, the gauge function is fitted using the truncated gamma likelihood approach, and the resulting estimate is shown in Figure 6.2. Using an extremal sample generated from the truncated gamma distribution (also shown in Figure 6.2), we estimate the coefficient

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

for values of u above $u_0 = 0.94$. This is estimated by simulating 100 datasets, obtaining an empirical and a fitted geometric model's estimate for each dataset, and considering the median estimate of $\chi(u)$ for each u value, along with the 0.025 and 0.975 quantiles to give a prediction interval. The results, shown in Figure 6.2), show a good overall approximation of $\chi(u)$ when compared to the empirical counterpart, with some slight disagreement u approaches 1. However, coverage in the prediction intervals remains good.

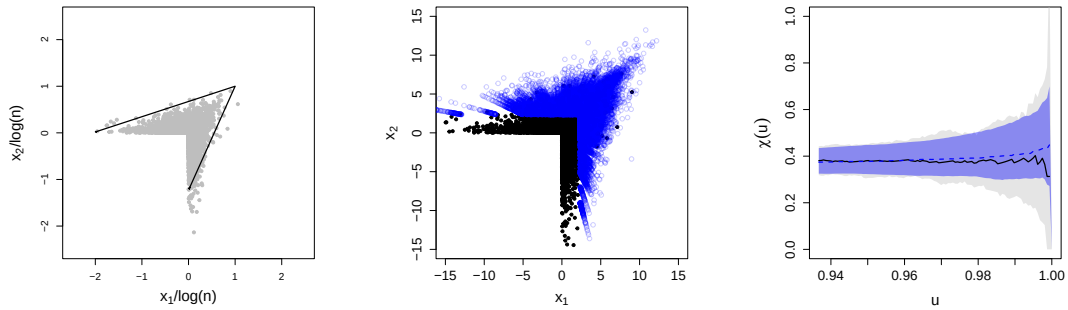

Figure 6.2: Right: Fitted g_{MGP} using the truncated gamma approach to MGP data generated from the independent Gumbel distribution with rate $\alpha = 0.5$. Centre: Extremal sample above the threshold $\max(\boldsymbol{X}) = u_0$ using these fits. Right: Empirical (in black) and estimated (in blue) median values of $\chi(u)$ with corresponding pointwise 95% prediction intervals.

6.5 Nonstandard behaviour of spectral random vectors

6.5.1 Nonstandard convergence to gauge functions

What can one do when the primary assumption of Proposition 6.1 is broken; i.e., when the spectral gauge function is degenerate under the usual scaling? This is the case, for example, when working with the MGP distribution with Hüsler Reiss dependence in standard exponential margins, as examined in Nolde and Wadsworth (2022). They show that, under the usual $\log n$ scaling, scaled sample clouds tend to a degenerate limit set given by the line $\{(x_1, x_2) \in [0,1]^2 \mid x_1 = x_2\}$. This limit set is not useful when trying to understand the extremal dependence properties of the corresponding bivariate random vector as the dependence parameter changes. It is for this reason that more general forms of convergence to the gauge function and limit set need to be considered, akin to the notion of nonstandard multivariate regular variation (Resnick, 2007, Chapter 6.5.6).

Proposition 6.3 (nonstandard convergence)**.** *Suppose* $\boldsymbol{X}$ *is* d*-dimensional with*

joint density f, $X_j \sim F_j$, and there exists functions $b_j(t) = F_j^{-1}(1 - t^{-1})$ such that $\log b_j(t)/\log t \to 0$ as $t \to \infty$ for each $j \in \{1, \ldots, d\}$. If the following holds,

$$-(\log t)^{-1} \log f(b_j(t)x_{t,j}\,,\; j = 1, \ldots, d) \longrightarrow g(\boldsymbol{x})\,, \tag{6.9}$$

for the sequence $\boldsymbol{x}_t \to \boldsymbol{x}$ as $t \to \infty$ and g continuous, then the sample cloud

$$N_n = \left\{ \left(\frac{X_{i1}}{b_1(n)}, \ldots, \frac{X_{id}}{b_d(n)} \right) \;:\; i = 1, \ldots, n \right\}$$

converges onto the limit set $\mathcal{G} = \{\boldsymbol{x} \in \mathbb{R}^d \,:\, g(\boldsymbol{x}) \leq 1\}$.

Proof. The proof follows similarly to that Proposition 2 of Nolde and Wadsworth (2022). The mean measure of N_n is given by $n\mathbb{P}\left[(X_j/b_j(n)\,,\; j = 1 \ldots, d) \in \cdot\right]$ with corresponding intensity

$$h_n(\boldsymbol{x}_n) = n \left(\prod_{j=1}^{d} b_j(n) \right) f(b_j(n)x_{n.j}\,,\; j = 1, \ldots, d)$$

Therefore,

$$-\log h_n(\boldsymbol{x}_n) \sim (g(\boldsymbol{x}_n) - 1) \log n \longrightarrow \begin{cases} -\infty & ; \quad \boldsymbol{x} \in \mathcal{G}^o \\ +\infty & ; \quad \boldsymbol{x} \in \mathcal{G}^c \end{cases}$$

as $n \to \infty$, where $\mathcal{G}^o$ is the interior of the set $\mathcal{G}$ and $\mathcal{G}^c$ is the complement of the set $\mathcal{G}$. The convergence of the mean measure of points N_n onto $\mathcal{G}$ gives convergence in probability of N_n onto $\mathcal{G}$ (Balkema and Nolde, 2010; Nolde and Wadsworth, 2022). □

To match the scaling used in the standard gauge function convergence (6.4), note that condition (6.9) is equivalent to

$$-t^{-1} \log f(b_j(e^t)x_{t,j}\,,\; j = 1, \ldots, d) \longrightarrow g(\boldsymbol{x}) \tag{6.10}$$

as $t \to \infty$. Define the function $h_j(t) = -\log(1 - F_j(t))$, it then follows that

$$\begin{aligned} h_j^{-1}(t) &= F_j^{-1}(1 - e^{-t}) \\ &= b_j(e^t) \end{aligned}$$

Putting this into (6.10), equation (6.9) is equivalent to

$$-t^{-1}\log f(h_j^{-1}(t)x_{t,j}\,,\ j=1,\ldots,d)\longrightarrow g(\boldsymbol{x}). \tag{6.11}$$

We note that the resulting g may not be 1-homogeneous. To obtain a 1-homogeneous gauge function $g^\star$ from g, suppose we restrict ourselves to cases when $h_j \in \mathrm{RV}_{\alpha_j}$, $\alpha_j > 0$. Recall the properties of regularly varying univariate functions:

1. $h(tx)/h(t) \to x^\alpha$ as $t \to \infty$.

2. $h \in \mathrm{RV}_\alpha$ implies $h^{-1} \in \mathrm{RV}_{1/\alpha}$ (Resnick, 2007, Proposition 2.6(v)).

Define the transformation $\tilde{x}_j = h_j(x_j)$ and let f^* be the joint density function of $(\tilde{X}_1,\ldots,\tilde{X}_d)^\top = (h_1(X_1),\ldots,h_d(X_d))^\top$. If (6.11) holds, then by change of variables, so does the following convergence

$$\begin{aligned}
&-t^{-1}\log f^*(th_1(x_{t,1}),\ldots,th_d(x_{t,d}))\\
&\quad= -t^{-1}\log f^*(t\tilde{x}_{t,1},\ldots,t\tilde{x}_{t,d})\\
&\quad= -t^{-1}\log f(h_1^{-1}(t\tilde{x}_{t,1}),\ldots,h_d^{-1}(t\tilde{x}_{t,d})) - t^{-1}\sum_{j=1}^{d}\log\left(\frac{d}{d\tilde{x}_{t,j}}h_j^{-1}(t\tilde{x}_{t,j})\right)\\
&\quad\sim -t^{-1}\log f(h_1^{-1}(t)\tilde{x}_{t,1}^{1/\alpha_1},\ldots,h_d^{-1}(t)\tilde{x}_{t,d}^{1/\alpha_d})\\
&\qquad\quad - t^{-1}\sum_{j=1}^{d}\log\left(\tilde{x}_{t,j}^{\alpha_j^{-1}-1}\frac{d}{dt}h_j^{-1}(t)\right) \qquad (6.12)\\
&\quad= -t^{-1}\log f(h_1^{-1}(t)\tilde{x}_{t,1}^{1/\alpha_1},\ldots,h_d^{-1}(t)\tilde{x}_{t,d}^{1/\alpha_d})\\
&\qquad\quad - t^{-1}\sum_{j=1}^{d}\log\left(\frac{d}{dt}h_j^{-1}(t)\right) - t^{-1}\sum_{j=1}^{d}(\alpha_j^{-1}-1)\log(\tilde{x}_{t,j}) \qquad (6.13)\\
&\quad\longrightarrow g(\tilde{x}_1^{1/\alpha_1},\ldots,\tilde{x}_d^{1/\alpha_d}) =: g^*(\tilde{x}_1,\ldots,\tilde{x}_d)
\end{aligned}$$

as $t \to \infty$. The asymptotic equivalence (6.12) as a result of $h_j^{-1} \in \mathrm{RV}_{\alpha_j^{-1}}$ implying $(h_j^{-1})' \in \mathrm{RV}_{\alpha_j^{-1}-1}$ (see Proposition 2.5 of Resnick (2007)). The first term in line (6.13) tends to zero as $t \to \infty$ by Proposition 2.6(i) of Resnick (2007). The result is g^*, the 1-homogeneous gauge function of the limit set associated to $N_n^* = \{h_j(X_{ij})/\log n\,:\, j=\ldots,d\}_{i=1}^n$ g and its link to g, the gauge function of the

limit set associated to $N_n = \{X_{ij}/h^{-1}(\log n) : j = \ldots, d\}_{i=1}^n$. The nonstandard convergence discussed in Proposition 6.3 can be formulated in terms of g^*.

Proposition 6.4 (Standardised nonstandard convergence)**.** *Suppose* $\boldsymbol{X}$ *is* d*-dimensional with joint density* f*,* $X_j \sim F_j$*, and there exists functions* $h_j \in RV_{\alpha_j}$ *for* $j \in \{1, \ldots, d\}$*. Let* f^* *be the joint density of* $(h_1(X_1), \ldots, h_d(X_d))^\top$*. If the following holds,*

$$-t^{-1} \log f^*(t h_j(x_{t,j}) \,,\, j = 1, \ldots, d) \longrightarrow g^*(\boldsymbol{x})$$

for the sequence $\boldsymbol{x}_t \to \boldsymbol{x}$ *and* g^* *continuous, as* $t \to \infty$*, then the sample cloud*

$$N_n^* = \left\{ \left(\frac{h(X_{i1})}{\log n}, \ldots, \frac{h(X_{id})}{\log n} \right) \;:\; i = 1, \ldots, n \right\}$$

converges onto the limit set $\mathcal{G}^* = \{\boldsymbol{x} \in \mathbb{R}^d \,:\, g^*(\boldsymbol{x}) \leq 1\}$*.*

The proof of Proposition 6.4 follows from Proposition 2 of Nolde and Wadsworth (2022).

Example: independent exponential and bivariate Gaussian components. As an illustration of convergence under different marginal scalings, suppose we consider random vectors of the form $(X, Y)^\top$ where X and Y are independent standard exponential and normal random variables. The joint density of (X, Y) is simply $f(x, y) = (2\pi)^{-1/2} e^{-x - \frac{1}{2}y^2}$. The marginal functions are $h_1(t) = t$ and $h_2(t) = -\log(1 - \Phi(t)) \sim t^2/2$, i.e., $h_2 \in \mathrm{RV}_2$. The inverse marginal functions are therefore $h_1^{-1}(t) = t$ and $h_2^{-1} \in \mathrm{RV}_{1/2}$. For the non-homogeneous gauge function corresponding to the limit set of the sample cloud $N_n = \left\{ \left(\frac{X_i}{\log n}, \frac{Y_i}{\sqrt{2 \log n}} \right) \right\}_{i=1}^n$, we have

$$\begin{aligned} -t^{-1} \log f(h_1^{-1}(t)x, h_2^{-1}(t)y) &\sim - t^{-1} \log f(tx, \sqrt{2t}y) \\ &\longrightarrow g(x, y) = x + y^2 \end{aligned}$$

where g is clearly not homogeneous. However, for the gauge function corresponding to the limit set of the sample cloud $N_n^* = \left\{ \left(\frac{X_i}{\log n}, \frac{h_2(Y_i)}{\log n} \right) \right\}_{i=1}^n$, for a sequence $y_t \to y \in$

$\mathbb{R}$ such that $|ty_t| \to \infty$ as $t \to \infty$, and letting $\tilde{y}_t = h_2(y_t)$, we have the convergence

$$\begin{aligned}
-t^{-1}\log f^*(tx, th_2(y_t)) &= -t^{-1}\log f^*(tx, t\tilde{y}_t) \\
&= -t^{-1}\log f(tx, h_2^{-1}(t\tilde{y}_t)) - t^{-1}\log\left(\frac{d}{d\tilde{y}_t}h_2^{-1}(t\tilde{y}_t)\right) \\
&\sim -t^{-1}\log f(tx, h_2^{-1}(t)\tilde{y}_t^{1/2}) - t^{-1}\log\left(\frac{d}{dt}h_2^{-1}(t)\tilde{y}_t^{1/2}\right) \\
&\sim -t^{-1}\log f(tx, \sqrt{2t}\tilde{y}_t^{1/2}) - t^{-1}\log\left(\sqrt{2t}2^{-1}\tilde{y}_t^{-1/2}\right) \\
&\longrightarrow g^*(x, \tilde{y}) = x + \tilde{y}
\end{aligned}$$

Note that $g^*(x, y) = g(x, y^{1/2})$ is 1-homogeneous, as desired. Further note that the condition $|ty_t| \to \infty$ as $t \to \infty$ is needed in order to use Mills' ratio to establish that $h_2(t) := -\log(1 - \Phi(t)) \in \mathrm{RV}_2$.

6.5.2 A coordinate transformation

In the case when $\log n$-scaled sample clouds of the spectral random vector do not tend to a non-degenerate limit set, the geometry of random vectors of the form $E + \boldsymbol{S}$ is not generally useful in describing extremal dependence between asymptotically dependent random variables. Instead, in the bivariate setting, consider the joint behaviour of $M = \frac{1}{2}(X_1 + X_2)$, and $V = X_1 - M - \frac{1}{2}(X_1 - X_2)$. Essentially, this transformation separates the exponential and spectral components of the stochastic representation into M and V, respectively. Furthermore, V lives on a linear space, and is therefore easier to handle than $\boldsymbol{S}$. By change of variables, the joint density of (M, V) is given by

$$f_{M,V}(m, v) = 2f_{X_1,X_2}(m + v, m - v)$$

Under the stochastic representation, we have

$$\begin{aligned}
f_{M,V}(m, v) \propto &\mathbf{1}_{\mathbb{R}_+}(m + \max(v, -v))e^{-m-\max(v,-v)} \\
&\times f_{\boldsymbol{S}}(v - \max(v, -v), -v - \max(v, -v))
\end{aligned}$$

The limit set associated with scaled copies of (M, V) can be obtained from the negative-logarithm of $f_{M,V}$ using the aforementioned standard or nonstandard convergence techniques.

In general d-dimensions, we generalise the transformation to

$$(M, \boldsymbol{V}) = \left(d^{-1} \sum_{j=1}^{d} X_j, \boldsymbol{X}_{-d} - d^{-1} \sum_{j=1}^{d} X_j \right),$$

where $\boldsymbol{X}_{-d}$ is the random vector $\boldsymbol{X}$ with the final component removed. The resulting in $(M, \boldsymbol{V})$ is a random vector taking values in $\{(m, \boldsymbol{v}) \in \mathbb{R}^d : m + \max(\boldsymbol{v}, -\mathbf{1}^\top \boldsymbol{v}) > 0\}$ with inverse transformation $\boldsymbol{X} = (M\mathbf{1} + \boldsymbol{V}, M - \mathbf{1}^\top \boldsymbol{V}) = (\boldsymbol{X}_{-d}, X_d)$. The density is therefore

$$f_{M,\boldsymbol{V}}(m, \boldsymbol{v}) = d\, f_{\boldsymbol{X}}(m\mathbf{1} + \boldsymbol{v}, m - \mathbf{1}^\top \boldsymbol{v})$$

Under the stochastic representation, the $(M, \boldsymbol{V})$ coordinate transformation amounts to studying the behaviour of

$$(M, \boldsymbol{V}) = \left(E + d^{-1} \sum_{j=1}^{d} S_j,\ \boldsymbol{S}_{-d} - d^{-1} \sum_{j=1}^{d} S_j \right),$$

whose joint density can be obtained from $f_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbf{1}_{\mathbb{R}_+}(\max(\boldsymbol{x})) e^{-\max(\boldsymbol{x})} f_{\boldsymbol{S}}(\boldsymbol{x} - \max(\boldsymbol{x}))$,

$$\begin{aligned} f_{M,\boldsymbol{V}}(m, \boldsymbol{v}) =& \mathbf{1}_{\mathbb{R}^+}(m + \max(\boldsymbol{v}, -\mathbf{1}^\top \boldsymbol{v})) e^{-m - \max(\boldsymbol{v}, -\mathbf{1}^\top \boldsymbol{v})} \\ & \times f_{\boldsymbol{S}}(\boldsymbol{v} - \max(\boldsymbol{v}, \mathbf{1}^\top \boldsymbol{v}), -\mathbf{1}^\top \boldsymbol{v} - \max(\boldsymbol{v}, -\mathbf{1}^\top \boldsymbol{v})) \end{aligned}$$

In this new coordinate system, interest lies in the behaviour of $-t^{-1} \log f_{M,V}(tm, h^{-1}(t)\boldsymbol{v})$ as $t \to \infty$ for some $\alpha > 0$ such that $h^{-1} \in \mathrm{RV}^{1/\alpha}$.

This coordinate system can be used to study the tail behaviour using the gauge function for multivariate distributions whose limit sets were previously shown to the degenerate under the usual $\log n$ scaling. The most notable example of this is the Hüsler-Reiss generalised Pareto distribution (Zhen Wai Olivier and Dombry, 2017), an important example in the literature due to its simple interpretation. The Hüsler-Reiss model has gained recent popularity in graphical modelling due

to its precision matrix having correspondence with the extremal graphical structure (Hentschel et al., 2024). Here, we derive the gauge function in the (M, V) coordinate system for the bivariate Hüsler-Reiss model. We show that its tail behaviour is equivalent to constructing a MGP random vector using a Gaussian generator for $\boldsymbol{T}$ by comparing the gauge functions of both models.

Example: Gaussian generator This case is of particular interest because of its similarities with the Hüsler-Reiss dependence structure (Kiriliouk et al., 2019; Wan, 2024). Suppose we have a generator $\boldsymbol{T} \sim \text{MVN}_d(\mathbf{0}, \Sigma)$ and let $\boldsymbol{S} = \boldsymbol{T} - \max(\boldsymbol{T})$. Kiriliouk et al. (2019) derived the density $f_{\boldsymbol{S}}$, given by

$$f_{\boldsymbol{S}}(\boldsymbol{x} - \max(\boldsymbol{x})) = \frac{1}{(2\pi)^{(d-1)/2} \left|\Sigma\right|^{1/2} (\mathbf{1}^\top \Sigma^{-1} \mathbf{1})^{1/2}} \exp\left\{ -\frac{1}{2} \boldsymbol{x}^\top \left[\Sigma^{-1} - \frac{\Sigma^{-1} \mathbf{1}\mathbf{1}^\top \Sigma^{-1}}{\mathbf{1}^\top \Sigma^{-1} \mathbf{1}} \right] \boldsymbol{x} \right\}$$

The quadratic form in the exponential term of $f_{\boldsymbol{S}}$ means that $\lim_{t\to\infty} -t^{-1} \log f_{\boldsymbol{S}}(t\boldsymbol{x}) = 0$ and so the scaled sample cloud $\{\boldsymbol{S}_i / \log n\}_{i=1}^n$ tends to a degenerate limit set at $\{\mathbf{0}\}$. The resulting MGP random vector $E + \boldsymbol{S}$ therefore has a limit set which constitutes the diagonal line $\{\boldsymbol{x} \in \mathbb{R}_+^d : x_1 = x_2 = \cdots = x_d\}$ coming from the common exponential component. A more useful geometric interpretation arises when inspecting $(M, \boldsymbol{V})$. Let $d = 2$ and set $\mathbb{A} = \Sigma^{-1} - \frac{\Sigma^{-1}\mathbf{1}\mathbf{1}^\top\Sigma^{-1}}{\mathbf{1}^\top\Sigma^{-1}\mathbf{1}}$. Further suppose $\Sigma_{1,1} = \Sigma_{2,2} = 1$ and $\Sigma_{1,2} = \Sigma_{2,1} = \rho$. Then $\mathbb{A} = 2^{-1}(1-\rho)^{-1}(1, -1)(1, -1)^\top$ and the MGP density of $\boldsymbol{X} = E + \boldsymbol{S}$ in the bivariate setting is given by

$$f_{X_1,X_2}(x_1, x_2) \propto \mathbf{1}_{\mathbb{R}_+}(\max(x_1, x_2)) e^{-\max(x_1,x_2)} \exp\left\{ -\frac{1}{2} \begin{pmatrix} x_1 & x_2 \end{pmatrix} \mathbb{A} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\}$$

Therefore, in the (M, V) parametrisation,

$$\begin{aligned} &f_{M,V}(m, v) \\ &\quad \propto \mathbf{1}_{\mathbb{R}_+}(m + \max(v, -v)) e^{-m - \max(v,-v)} \\ &\qquad \times \exp\left\{ -\frac{1}{2} \left(m^2 \begin{pmatrix} 1 & 1 \end{pmatrix} \mathbb{A} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 2mv \begin{pmatrix} 1 & 1 \end{pmatrix} \mathbb{A} \begin{pmatrix} 1 \\ -1 \end{pmatrix} + v^2 \begin{pmatrix} 1 & -1 \end{pmatrix} \mathbb{A} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right) \right\} \end{aligned}$$

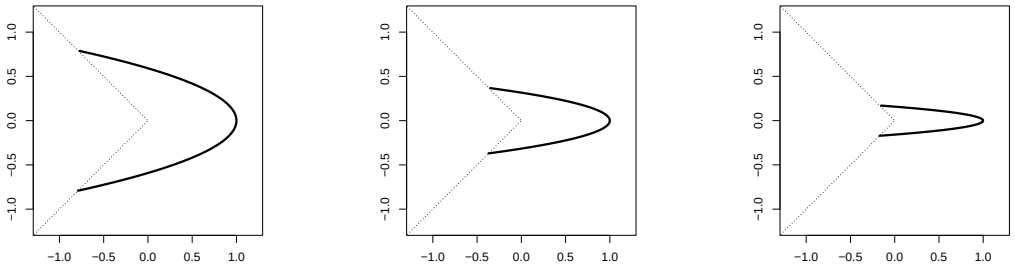

Figure 6.3: Limit set boundaries associated to the bivariate MGP random vectors with the Gaussian generator with correlation in the (M, V) representation, described by the gauge function $g(m, v) = m + (1-\rho)^{-1}v^2$. Left to right: $\rho = 0.3, 0.8, 0.95$. Dotted line is the lower-bound of the support of the random vector (M, V), given by $\{(m, v) : m + \max(v, -v) = 0\}$.

$$= \mathbf{1}_{\mathbb{R}_+}(m + \max(v,-v))e^{-m-\max(v,-v)}e^{-(m+v)} \exp\left\{ -\frac{1}{2}v^2 \begin{pmatrix} 1 & -1 \end{pmatrix} \mathbb{A} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$$

Using the marginal function $h \in \mathrm{RV}_2$ on the $\boldsymbol{V}$ component, the limit set associated to the sample cloud $N_n = \left\{ \left(\frac{M_i}{\log n}, \frac{V_i}{\sqrt{\log n}} \right) \right\}_{i=1}^{n}$ has gauge function

$$\begin{aligned} g(m, v) =& m + v^2 \begin{pmatrix} 1 & -1 \end{pmatrix} \mathbb{A} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\ =& m + (1-\rho)^{-1}v^2, \quad m + \max(v, -v) > 0. \end{aligned}$$

Using $g(m, v)$ one could inspect the geometry of bivariate GP random vectors generated from the bivariate Gaussian distribution as ρ increases. An example of this presented in Figure 6.3, where it can be seen that an increase in ρ corresponds to a narrowing of the limit set.

Example: bivariate generalised Hüsler-Reiss Pareto distribution The Hüsler-Reiss MGP distribution is parametrised by a positive-definite $d \times d$ correlation

matrix Σ and has exponent measure density in exponential margins

$$\lambda(\boldsymbol{x};\Sigma) = c\exp\left\{-\frac{1}{2}\left(\boldsymbol{x}^\top\Sigma^{-1}\boldsymbol{x} + d^{-1}\left(\Sigma^{-1}\Gamma\mathbf{1}+2\mathbf{1}\right)^\top\boldsymbol{x}\right)\right\}$$

where $\Gamma_{ij} = \Sigma_{ii}+\Sigma_{jj}-2\Sigma_{ij}$ (Hentschel et al., 2024). Nolde and Wadsworth (2022) showed a degenerate limit set occurs in the bivariate setting, considering when $\Sigma = 2(1-\rho)\mathbf{1}\mathbf{1}^\top$. This setting results in the density

$$\begin{aligned} f(x,y;\Sigma) =& \frac{\lambda(x,y;\rho)}{\iint_{\{(u,v):\max(u,v)>0\}}\lambda(u,v;\rho)du\,dv} \\ \propto& \exp\left\{-\frac{1}{2}\left(\frac{(x-y)^2}{2(1-\rho)}+x+y\right)\right\} \end{aligned}$$

Nolde and Wadsworth (2022) show that the limit set for the sample cloud $\left\{\left(\frac{X_i}{\log n},\frac{Y_i}{\log n}\right)\right\}_{i=1}^n$ is degenerate, falling on the line $x=y$.

In studying the degree of asymptotic dependence between X and Y, a better approach is to consider non-common scalings (as in Propositions 6.3 and 6.4) on the vector (M,V), where $M=\frac{1}{2}(X+Y)$, and $V = X-M$. By change of variables,

$$\begin{aligned} f_{M,V}(m,v) =& 2f_{X_1,X_2}(m+v,m-v) \\ \propto& e^{-m}e^{-v^2/(1-\rho)} \end{aligned}$$

with independent marginal distributions $M\sim\text{Exp}(1)$ and $V\sim\mathcal{N}(0,(1-\rho)/2)$. This establishes that M behaves like an exponential random variable in the right tail, and that V is univariate Gaussian.

Using $h(t)\in\text{RV}_2$, the limit set associated to the sample cloud $N_n = \left\{\left(\frac{M_i}{\log n},\frac{V_i}{\sqrt{\log n}}\right)\right\}_{i=1}^n$ has gauge function $g(m,v) = m+(1-\rho)^{-1}v^2$, while limit set associated to the sample cloud $N_n^* = \left\{\left(\frac{M_i}{\log n},\frac{h_(V_i)/(1-\rho)}{\log n}\right)\right\}_{i=1}^n$ has gauge function $g^*(m,v) = m+v$. When focusing on $N_n = \left\{\left(\frac{M_i}{\log n},\frac{V_i}{\sqrt{\log n}}\right)\right\}_{i=1}^n$, which has gauge function $g(m,v;\rho) = m+(1-\rho)^{-1}v^2$, note that it is the same as in the Gaussian generator example and has unit level set displayed in Figure 6.3,. This highlights their equivalent limiting behaviour, as desired.

6.6 Discussion

In this work, we brought together multivariate regular variation and the geometric approach by considering the limit sets of MGP random vectors. Through studying the $(d-1)$-dimensional limit sets of the spectral components of the stochastic representation of MGP random vectors, we began by showing that all bivariate MGP random vectors whose spectral component have a non-degenerate limit set under $\log n$ scaling have the same limit set. By studying the corresponding gauge function, we illustrated how any problem involving inference on such MGP distributions can be approached as a geometric multivariate extremal inference problem. When the spectral component does not have a non-degenerate limit set under $\log n$ scaling, we showed that alternative representations of the limit set can be used to effectively separate the components of the stochastic representation to study the degree of asymptotic dependence in MGP random vectors. We note that, while many results are given for general d-dimensional random vectors, we mostly illustrate our findings in the $d = 2$ setting, leaving further exploration for future work. In the MGP setting where, copies of the spectral random vector converge to a valid limit set under $\log n$ scaling, we illustrated that the truncated gamma inference method for geometric extremes (Wadsworth and Campbell, 2024) can be used when considering the L_∞-norm. While not covered here, this could potentially be extended to the nonstandard convergence setting by performing a change of variables and inferring the α_j values for the marginal regularly varying functions $h_j \in \mathrm{RV}_{\alpha_j}$.

Through studying the MGP distribution, this work highlights some connections between the classical MRV framework and the more recent geometric framework for inference on the multivariate tail. By showing that one could derive the limit set of MGP random vectors, it can then be used to study the strength of tail dependence between the components as in the geometric framework. However, when studying both methods, their differences also become apparent. While the geometric framework aims to perform inference across the *entire* multivariate tail, the MRV approach only focuses on the region where all d components grow large together. In

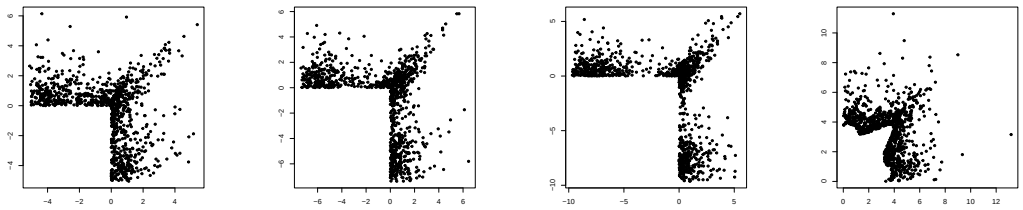

Figure 6.4: Left, centre-left, centre-right: Realisations of $\boldsymbol{Y}_i - r \mid \{\max(\boldsymbol{Y}_i) > r\}$, $i = 1, \ldots, 1000$ for increasing r. The bivariate random vectors $\boldsymbol{Y}_i$ have exponential margins and are generated from a mixture model made up of equally-weighted Gaussian and max-stable distribution with logistic dependence. Right: Realisations of $\boldsymbol{Y}_i \mid \{\|\boldsymbol{Y}_i\|_1 > r_\tau(\boldsymbol{Y}_i/\|\boldsymbol{Y}_i\|_1)\}$.

fact, in studying the limiting behaviour of $\boldsymbol{Y}_i - r \mid \{\max(\boldsymbol{Y}_i) > r\}$ as $r \to \infty$ for copies $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots$ from a random vector $\boldsymbol{Y}$, a key assumption in MRV is that any values of of $\boldsymbol{Y}_i$ that are not simultaneously extreme will be normalised to $-\infty$. Take, for example, $\boldsymbol{Y}$ following an equally-weighted mixture of Gaussian and max-stable distribution with logistic dependence. MRV assumes the Gaussian contribution of the mixture model, inducing asymptotic independence between the two components of $\boldsymbol{Y}$, should disappear when studying $\lim_{r\to\infty} \boldsymbol{Y}_i - r \mid \{\max(\boldsymbol{Y}_i) > r\}$ on the interior of the support $\{\boldsymbol{x} \in \mathbb{R}^d : \max(\boldsymbol{x}) \geq 0\}$. Figure 6.4 shows that, even with a very large threshold r, this is not satisfied in practice. A much more realistic approach is presented in the geometric setting. By not restricting to the L_∞ threshold, a much richer representation of the multivariate tail is used, as shown in Figure 6.4, right. The framework is able to treat both components of the mixture model in a non-trivial fashion. An immediate drawback of the geometric approach is that, unlike in the MRV setting, one needs to estimate this radial threshold. In low-dimensions, this is easily achieved through quantile regression (Fasiolo et al., 2021; Wadsworth and Campbell, 2024) or Bayesian semi-parametric methods (Papastathopoulos et al., 2025). In higher dimensions, empirical (Wadsworth and Campbell, 2024) and kernel density estimation (Campbell and Wadsworth, 2024) approaches have been

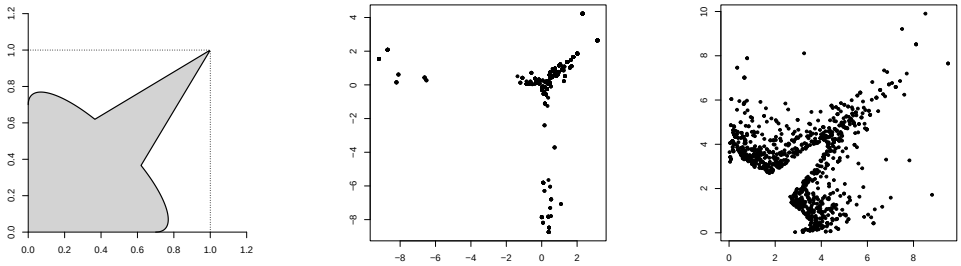

Figure 6.5: Left: Limit set of the modified Gaussian and max-stable logistic mixture model. Centre: Realisations of $\boldsymbol{Y}_i - r \mid \{\max(\boldsymbol{Y}_i) > r\}$, $i = 1, \ldots, 1000$ for increasing $r = 9.49$. Right: Realisations of $\boldsymbol{Y}_i \mid \{\|\boldsymbol{Y}_i\|_1 > r_\tau(\boldsymbol{Y}_i/\|\boldsymbol{Y}_i\|_1)\}$.

proposed. One could simply use the L_∞ threshold in the geometric approach; however, it may not be descriptive enough to capture extremal behaviour across the entire domain, particularly for semi-parametric methods.

In many instances, the convergence assumption in MRV makes it impossible to study regions of the tail away from the joint exceedances. Take, for example, a bivariate mixture model of equally-weighted max-stable logistic, and a Gaussian with modified covariance matrix, $\theta^{-1}\boldsymbol{\Sigma}$, $\theta > 1$. The resulting distribution, whose limit set is shown in Figure 6.5, has "scaled back" Gaussian components. In studying $\boldsymbol{Y}_i - r \mid \{\max(\boldsymbol{Y}_i) > r\}$, we see that the Gaussian behaviour is essentially removed, leaving only the points lying on the joint extremal region shown. In essence, the MRV framework only allows for studying extremes in this region. It may, however, be desirable to study the tail behaviour away from the axes, when one variable has extreme values and the other does not. From the Figure 6.5, right, we see that this is still possible in the geometric setting for this example.

On the other hand, when interest lies in the behaviour of asymptotically dependent random vectors and all components large, fitting the spectral component in the MRV setting leads to more accurate inference compared to the geometric approach. This is exemplified by the simulation studies in Chapter 3. When considering bivariate data drawn from a distribution with asymptotically dependent

components (Figure 3.7), probability estimates for lying in an a region when both components are large are more accurate and have less variability when using the MRV approach compared to any other competing framework. Behind this observation is the fact that the limit set of MGP random vectors only depends on the lower tail behaviour of the spectral vector $\boldsymbol{S}$, as exemplified in Proposition 6.2. The MGP framework allows one to capture the full distribution of $\boldsymbol{S}$, which offers refined estimates when extrapolating into regions where all variables are large. However, MRV was outperformed by the other methods for estimating the probability of lying in off-diagonal extremal regions, and when using non-asymptotically dependent data.

Chapter 7

Concluding Remarks

Throughout this work, focus lay on the geometric approach to multivariate extreme value analysis. The idea of limiting geometry of scaled sample clouds is nothing new (Davis et al., 1988; Kinoshita and Resnick, 1991). It was only in recent work (Nolde and Wadsworth, 2022) that its full potential to characterise the *entire* tail behaviour for random vectors with potentially complicated extremal dependence behaviour came to light. However, many gaps in the literature still existed. Namely, given data, how does one go about estimating the gauge function that is inherently linked to the geometry of interest? This was the central focus for much of this work. Chapters 3, 4, and 5 did this in different ways. Chapters 3 proposed a likelihood-based approach using parametric assumptions on the gauge function. Chapter 4 introduced model flexibility by proposing a Bayesian semiparametric framework, but inference only existed in two and three dimensions. Chapter 5 aims to increase the dimensionality by proposing a simple piecewise-linear construction for the gauge function. Apart from estimating the gauge function, Chapters 3, 4, and 5 all introduce methodologies to estimate the radial threshold, above which we perform inference of the gauge function. Each have their merit. Furthermore, Chapters 4 and 5 introduce models for the angular distribution, and show that they can potentially be linked to the gauge function as well.

In addition to the methodological work done with geometric extremes, theoretical

considerations were taken. A key feature of the geometric framework is performing inference in a radial-angular decomposition, rather than in Cartesian coordinates. Chapter 3 derived the limiting behaviour of the radial component when the original random vector has exponential margins, showing that its limiting behaviour depends on the gauge function. Rates of convergence were also provided. Chapter 4 does this as well in Laplace margins, and also angular densities are derived. In Chapter 6 we derived the limit sets for a wide array of multivariate generalised Pareto distributions, a family of distributions commonly studied and used in practical applications.

Any approach for multivariate extremal inference is prone to shortcomings. While the geometric approach addresses many of these shortcomings, its recent novelty naturally leads to open-ended areas of exploration. In a radial-angular decomposition, the geometric approach relies on modelling for exceedances above a high radial threshold at a given angle. As the underlying distribution is often intractable, an accurate estimate of this threshold is needed. This is particularly the case in the semiparametric approaches presented in Chapters 4 and 5. The proposed methods either don't scale well to high dimensions due to computational difficulties or have inaccuracies in regions of the domain with little density. With this in mind, work needs to be done to bypass these difficulties by improving this high quantile estimation, or develop methodology to bypass the estimation of this high quantiles in high-dimensions. Another area that needs further exploration is the sensitivity of the truncated gamma model fit to the set quantile level for threshold estimation. Throughout this work, we tend to set the threshold to be the 0.90 or the 0.95 quantile of the radii conditioned on angles. This was simply set based on intuition; we need a quantile high enough for the truncated gamma assumption be approximately hold, but we need it to be low enough to have sufficient data for maximum likelihood estimation. In the future, a numerical study of the pre-set quantile's affect on the quality of the resulting maximum likelihood estimates is needed. It is clear that a bias-variance trade-off would occur, but it would be

interesting to see its extent.

Issues also persist when fitting the gauge function to data as the dimension of the data increases. The parametric models of Chapter 3 are often too rigid for the complex extremal dependence structure present in real-world applications, the Bayesian semiparametric method in Chapter 4 rely on basis functions on a mesh of the angular domain that are only defined in dimensions two and three, and the piecewise linear approach in Chapter 5 struggles when data is missing in large regions of the domain, a common feature of high-dimensional data. Machine learning-based approaches have been introduced (Murphy-Barltrop et al., 2024b; De Monte et al., 2025), but it is unclear whether or not these are able to capture the tail behaviour of very complex dependence structures.

The theoretical properties of the geometric approach remain relatively unexamined. Specifically, the truncated gamma model proposed in Chapter 3 for large radial components of the radial-angular decomposition lacks any of the classic statistical convergence guarantees. Maximum likelihood estimation is used, and while it does inherently provide consistency and efficiency properties of the parameter estimates, it is unclear how they depend on the threshold. The truncated Gamma holds exactly as the threshold tends to infinity, but how do consistency results for maximum likelihood estimates depend on the threshold at high but finite levels? This is of interest, as we only consider these finite thresholds in practical applications. As a start, this needs to be examined for the truncated gamma likelihood with parametric gauges defined in Chapter 3.

Despite the current setbacks, there is no denying the potential for the geometric approach to accurately characterise the entire multivariate tail of a random process. Its simplicity in inference and probability estimation make it a powerful tool for researchers and practitioners. A key test in the coming years is how well it performs in real-world applications that occur in industry. However, based on preliminary studies, it can be confidently said that the geometric framework should find its place among the favoured approaches for extreme value analysis.

References

F. J. Acero, J. A. García, and M. C. Gallego. Peaks-over-threshold study of trends in extreme rainfall over the Iberian Peninsula. *Journal of Climate*, 24(4):1089–1105, 2011.

V. Agilan, N. Umamahesh, and P. Mujumdar. Influence of threshold selection in modeling peaks over threshold based nonstationary extreme rainfall series. *Journal of Hydrology*, 593:125625, 2021.

R. Arce Guillen, F. Lindgren, S. Muff, T. W. Glass, G. A. Breed, and U. E. Schlägel. Accounting for unobserved spatial variation in step selection analyses of animal movement via spatial random effects. *Methods in Ecology and Evolution*, 14(10): 2639–2653, 2023.

S. G. Baker. The multinomial-Poisson transformation. *Journal of the Royal Statistical Society, Series D (The Statistician)*, 43(4):495–504, 1994.

A. Balkema and N. Nolde. Asymptotic dependence for homothetic light-tailed densities. *Advances in Applied Probability*, 44:506–527, 2012.

A. Balkema, P. Embrechts, and N. Nolde. Meta densities and the shape of their sample clouds. *JMVA*, 101:1738–1754, 2010.

A. A. Balkema and L. de Haan. Residual life time at great age. *Annnals of Probability*, 2(5):792–804, 1974.

A. A. Balkema and N. Nolde. Asymptotic independence for unimodal densities. *Advances in Applied Probability*, 42(2):411–432, 2010.

A. M. Barlow, E. Mackay, E. Eastoe, and P. Jonathan. A penalised piecewise-linear model for non-stationary extreme value analysis of peaks over threshold. *Ocean Engineering*, 267:113265, 2023.

V. Barnett. The ordering of multivariate data (with discussion). *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 139(3):318–355, 1976.

S. Barthelmé and N. Chopin. The Poisson transform for unnormalised statistical models. *Statistics and Computing*, 25(4):767–780, 2015.

J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels. *Statistics of extremes: theory and applications*. John Wiley & Sons, 2006.

M. Ber, O. Cheong, M. Kreveld, and M. Overmars. *Delaunay Triangulations*, pages 191–218. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

D. Bolin and F. Lindgren. Excursion and contour uncertainty regions for latent Gaussian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):85–106, 2015.

D. Bolin and F. Lindgren. Quantifying the uncertainty of contour maps. *Journal of Computational and Graphical Statistics*, 26(3):513–524, 2017.

D. Bolin and F. Lindgren. Calculating probabilistic excursion sets and related quantities using excursions. *Journal of Statistical Software*, 86(5):1–20, 2018. doi: 10.18637/jss.v086.i05.

P. Bortot, S. Coles, and J. Tawn. The multivariate Gaussian tail model: An application to oceanographic data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 49(1):31–049, 2000.

R. P. Brent. *Algorithms for minimization without derivatives*. Courier Corporation, 2013.

A. Bücher and J. Segers. On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes*, 20(4):839–872, 2017.

T. Buishand. Statistics of extremes in climatology. *Statistica Neerlandica*, 43(1): 1–30, 1989.

J.-J. Cai, J. H. J. Einmahl, and L. de Haan. Estimation of extreme risk regions under multivariate regular variation. *The Annals of Statistics*, 39(3):1803 – 1826, 2011.

R. Campbell and J. Wadsworth. Piecewise-linear modeling of multivariate geometric extremes. *arXiv:2412.05195*, 2024.

D. Carter and P. Challenor. Estimating return values of environmental parameters. *Quarterly Journal of the Royal Meteorological Society*, 107(451):259–266, 1981.

R. E. Chandler and S. Bate. Inference for clustered data using the independence loglikelihood. *Biometrika*, 94(1):167–183, 2007.

E. Chautru. Dimension reduction in multivariate extreme value analysis. *Electronic Journal of Statistics*, 9(1):383–418, 2015.

S. Coles. *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer-Verlag, 2001.

S. Coles and F. Pauli. Models and inference for uncertainty in extremal dependence. *Biometrika*, 89(1):183–196, 2002.

S. G. Coles and J. A. Tawn. Modelling extreme multivariate events. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 53(2):377–392, 1991.

S. G. Coles and J. A. Tawn. Statistical methods for multivariate extremes: an application to structural design. *The Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 43(1):1–31, 1994.

B. Das and V. Fasen-Hartmann. Risk contagion under regular variation and asymptotic tail independence. *Journal of Multivariate Analysis*, 165:194–215, 2018.

R. A. Davis, E. Mulrow, and S. I. Resnick. Almost sure limit sets of random samples in $\mathbb{R}^d$. *Adv. Appl. Probab.*, 20(3):573–599, 1988. ISSN 00018678.

A. C. Davison. Modelling excesses over high thresholds, with an application. In *Statistical extremes and applications*, pages 461–482. 1984.

A. C. Davison and R. L. Smith. Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 52(3):393–442, 1990.

L. de Haan. *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes, by L. de Haan.* Mathematical Centre tracts. 1970.

L. de Haan. A spectral representation for max-stable processes. *Annals of Probability*, 12(3):1194–1204, 1984.

L. de Haan. Fighting the arch–enemy with mathematics. *Statistica neerlandica*, 44 (2):45–68, 1990.

L. De Haan and J. De Ronde. Sea and wind: multivariate extremes at work. *Extremes*, 1(1):7–45, 1998.

L. de Haan and S. I. Resnick. Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 40(4):317–337, 1977.

L. De Monte, R. Huser, I. Papastathopoulos, and J. Richards. Generative modelling of multivariate geometric extremes using normalising flows. *arXiv preprint arXiv:2505.02957*, 2025.

B. Delaunay. Sur la sphère vide. à la mémoire de Georges Voronoï. *News of the Russian Academy of Sciences, Mathematical series.*, (6):793–800, 1934.

C. Dombry, S. Engelke, and M. Oesting. Exact simulation of max-stable processes. *Biometrika*, 103(2):303–317, 2016.

M. Émile and J. Gumbel. Distributions des valeurs extrêmes en plusieurs dimensions. In *Annales de l'ISUP*, volume 9, pages 171–173, 1960.

S. Engelke and A. S. Hitz. Graphical models for extremes. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 82(4):871–932, 2020.

S. Engelke, A. Malinowski, Z. Kabluchko, and M. Schlather. Estimation of Hüsler-Reiss distributions and Brown-Resnick processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 77(1):239–265, 2015.

S. Engelke, M. Lalancette, and S. Volgushev. Learning extremal graphical models in high dimensions. *arXiv preprint arXiv:2111.00840*, 2022.

V. A. Epanechnikov. Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, 14(1):153–158, 1969.

K. Ewans and P. Jonathan. Evaluating environmental joint extremes for the offshore industry using the conditional extremes model. *Journal of Marine Systems*, 130: 124–130, 2014.

M. Fasiolo, S. N. Wood, M. Zaffran, R. Nedellec, and Y. Goude. Fast calibrated additive quantile regression. *Journal of the American Statistical Association*, 116 (535):1402–1412, 2021.

R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical proceedings of the Cambridge philosophical society*, volume 24, pages 180–190, 1928.

D. Fitzgerald. Single station and regional analysis of daily rainfall extremes. *Stochastic Hydrology and Hydraulics*, 3:281–292, 1989.

J. Galambos. Order statistics of samples from multivariate distributions. *JASA*, 70 (351a):674–680, 1975.

B. Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, 44(3):423–453, 1943.

N. Goix, A. Sabourin, and S. Clémençon. Sparse representation of multivariate extremes with applications to anomaly detection. *Journal of Multivariate Analysis*, 161:12–31, 2017.

I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. Academic Press, 2014.

A. Grady. Modeling daily minimum temperatures: an application of the threshold method. In *Proc. 5th Int. Meeting on Statistical Climatology. Toronto*, pages 319–324, 1992.

M. Hallin, E. del Barrio, J. Cuesta-Albertos, and C. Matrán. Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *Annnals of Statistics*, 49(2):1139 – 1165, 2021.

J. Heffernan and S. Resnick. Hidden regular variation and the rank transform. *Advances in Applied Probability*, 37(2):393–414, 2005.

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

M. Hentschel, S. Engelke, and J. Segers. Statistical inference for Hüsler–Reiss graphical models through matrix completions. *Journal of the American Statistical Association*, pages 1–13, 2024.

S. Hilal, S.-H. Poon, and J. Tawn. Portfolio risk assessment using multivariate extreme value methods. *Extremes*, 17:531–556, 2014.

S. T. Holgate. 'Every breath we take: the lifelong impact of air pollution'–a call for action. *Clinical Medicine*, 17(1):8–12, 2017.

A. B. Huseby, E. Vanem, and B. Natvig. A new approach to environmental contours for ocean engineering applications based on direct Monte Carlo simulations. *Ocean Engineering*, 60:124–135, 2013.

P. Jacob and B. Massé. Asymptotic behavior of samples from general multivariate distributions. *Journal of Statistical Planning and Inference*, 52(2):183–196, 1996.

P. Jonathan and K. Ewans. Statistical modelling of extreme ocean environments for marine design: a review. *Ocean Engineering*, 62:91–109, 2013.

C. Keef, I. Papastathopoulos, and J. A. Tawn. 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, 2013a.

C. Keef, J. A. Tawn, and R. Lamb. Estimating the probability of widespread flood events. *Environmetrics*, 24(1):13–21, 2013b.

K. Kinoshita and S. I. Resnick. Convergence of scaled random samples in $\mathbb{R}^d$. *The Annals of Probability*, pages 1640–1663, 1991.

A. Kiriliouk, H. Rootzén, J. Segers, and J. L. Wadsworth. Peaks over thresholds modeling with multivariate generalized Pareto distributions. *Technometrics*, 61 (1):123–135, 2019.

D. A. Klain. Invariant valuations on star-shaped sets. *Advances in Mathematics*, 125(1):95–113, 1997.

R. Koenker. *Quantile Regression*, volume 38. Cambridge University Press, 2005.

R. Koenker and G. Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

S. Kotz, T. J. Kozubowski, and K. Podgorski. *The Laplace distribution and generalizations*. Birkhäuser Boston, 2001.

M. Ledbetter, G. Lindgren, and H. Rootzén. Extremes and related properties of random sequences and series, 1983.

A. W. Ledford and J. A. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187, 1996.

A. W. Ledford and J. A. Tawn. Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(2): 475–499, 1997.

F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion). *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 73(4):423–498, 2011.

E. Mackay and G. de Hauteclocque. Model-free environmental contours in higher dimensions. *Ocean Engineering*, 273:113959, 2023.

E. Mackay and P. Jonathan. Modelling multivariate extremes through angular-radial decomposition of the density function. *arXiv:2310.12711*, 2024.

E. Mackay, C. Murphy-Barltrop, J. Richards, and P. Jonathan. Deep learning joint extremes of metocean variables using the SPAR model. *arXiv:2412.15808*, 2024.

R. Majumder, B. Shaby, B. Reich, and D. Cooley. Semiparametric estimation of the shape of the limiting bivariate point cloud. *Bayesian Analysis*, pages 1–27, 01 2025.

C. J. R. Murphy-Barltrop, E. Mackay, and P. Jonathan. Inference for multivariate extremes via a semi-parametric angular-radial model, 2024a.

C. J. R. Murphy-Barltrop, R. Majumder, and J. Richards. Deep learning of multivariate extremes via a geometric representation. *arXiv:2406.19936*, 2024b.

N. Nolde. Geometric interpretation of the residual dependence coefficient. *Journal of Multivariate Analysis*, 123:85–95, 2014.

N. Nolde and J. L. Wadsworth. Linking representations for multivariate extremes via a limit set. *Advances in Applied Probability*, 54(3):688–717, 2022.

N. Nolde and J. Zhang. Conditional extremes in asymmetric financial markets. *Journal of Business & Economic Statistics*, 38(1):201–213, 2020.

P. J. Northrop and P. Jonathan. Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*, 22(7):799–809, 2011.

F. W. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain. Digital library of mathematical functions. 2019.

S. Outten and S. Sobolowski. Extreme wind projections over europe from the euro-cordex regional climate models. *Weather and Climate Extremes*, 33:100363, 2021.

C. Pacifici, S. A. Padoan, and J. Mysiak. Predicting hazards of climate extremes: a statistical perspective. *arXiv preprint arXiv:2505.17622*, 2025.

S. A. Padoan and S. Rizzelli. Empirical Bayes inference for the block maxima method. *Bernoulli*, 30(3):2154–2184, 2024.

I. Papastathopoulos, G. Auld, F. Lindgren, K. Z. Gerlei, and M. F. Nolan. Bayesian inference of grid cell firing patterns using Poisson point process models with latent oscillatory Gaussian random fields. *arXiv:2303.17217*, 2023.

I. Papastathopoulos, L. de Monte, R. Campbell, and H. Rue. Statistical inference for radial generalized Pareto distributions and return sets in geometric extremes. *arXiv:2310.06130*, 2025.

R. Paris. A note on the asymptotics of the modified Bessel functions on the stokes lines. *arXiv preprint arXiv:1708.09656*, 2017.

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

A. Ramos and A. Ledford. A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(1):219–241, 2009.

S. Resnick. On the foundations of multivariate heavy-tail analysis. *Journal of Applied Probability*, 41(A):191–212, 2004.

S. I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer–Verlag, New York, 1987.

S. I. Resnick. *Heavy-tail phenomena: probabilistic and statistical modeling*, volume 10. Springer Science & Business Media, 2007.

J. Richards, J. A. Tawn, and S. Brown. Modelling extremes of spatial aggregates of precipitation using conditional methods. *The Annals of Applied Statistics*, 16(4): 2693–2713, 2022.

M. E. Robinson and J. A. Tawn. Statistics for extreme sea currents. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 46(2):183–205, 1997.

H. Rootzén and Tajvidi. The multivariate generalized Pareto distribution. *Bernoulli*, 12:917–930, 2006.

H. Rootzén, J. Segers, and J. L. Wadsworth. Multivariate peaks over thresholds models. *Extremes*, 21(1):115–145, 2018a.

H. Rootzén, J. Segers, and J. L. Wadsworth. Multivariate generalized Pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis*, 165:117–131, 2018b.

B. Seals and A. Krasner. Gas stoves: Health and air quality impacts and solutions. *Rocky Mountain Institute*, 2020.

E. Seneta. *Regularly Varying Functions*, volume 508 of *Lecture Notes in Mathematics*. Springer Berlin, Heidelberg, 1976.

E. Simiu and N. A. Heckert. Extreme wind distribution tails: a "peaks over threshold" approach. *Journal of structural engineering*, 122(5):539–547, 1996.

D. Simpson, J. B. Illian, F. Lindgren, S. H. Sørbye, and H. Rue. Going off grid: computationally efficient inference for log-Gaussian Cox processes. *Biometrika*, 103(1):49–70, 2016.

E. S. Simpson and J. A. Tawn. Estimating the limiting shape of bivariate scaled sample clouds for self-consistent inference of extremal dependence properties. *arXiv:2207.02626*, 2022.

E. S. Simpson and J. A. Tawn. Estimating the limiting shape of bivariate scaled sample clouds: with additional benefits of self-consistent inference for existing extremal dependence properties. *Electronic Journal of Statistics*, 18(2):4582 – 4611, 2024a.

E. S. Simpson and J. A. Tawn. Inference for new environmental contours using extreme value analysis. *Journal of Agricultural, Biological and Environmental Statistics*, pages 1–25, 2024b.

E. S. Simpson, J. L. Wadsworth, and J. A. Tawn. Determining the dependence structure of multivariate extremes. *Biometrika*, 107(3):513–532, 2020.

R. L. Smith. Threshold methods for sample extremes. In *Statistical extremes and applications*, pages 621–638. 1984.

R. L. Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90, 1985.

R. L. Smith. Extreme value theory. *Handbook of applicable mathematics*, 7(437-471): 18, 1990.

A. P. Soms. An asymptotic expansion for the tail area of the t-distribution. *Journal of the American Statistical Association*, 71(355):728–730, 1976.

P. Stein. A note on the volume of a simplex. *The American Mathematical Monthly*, 73(3):299–301, 1966.

H. Tabari. Extreme value analysis dilemma for climate change impact assessment on global flood and extreme precipitation. *Journal of Hydrology*, 593:125932, 2021.

N. Tajvidi. *Multivariate generalised Pareto distributions*. Citeseer, 1995.

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

J. A. Tawn. Estimating probabilities of extreme sea-levels. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 41(1):77–93, 1992.

S. Tendijck, E. Eastoe, J. Tawn, D. Randell, and P. Jonathan. Modeling the extremes of bivariate mixture distributions with application to oceanographic data. *Journal of the American Statistical Association*, 118(542):1373–1384, 2023.

S. Vettori, R. Huser, and M. G. Genton. Bayesian modeling of air pollution extremes using nested multivariate max-stable processes. *Biometrics*, 75(3):831–841, 2019.

J. Wadsworth and J. Tawn. A new representation for multivariate tail probabilities. *Bernoulli*, 19(5B):2689–2714, 2013.

J. Wadsworth and J. Tawn. Higher-dimensional spatial extremes via single-site conditioning. *Spatial Statistics*, 51:100677, 2022.

J. L. Wadsworth and R. Campbell. Statistical inference for multivariate extremes via a geometric approach. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 86(5):1243–1265, 03 2024.

E. Waldmann, T. Kneib, Y. R. Yue, S. Lang, and C. Flexeder. Bayesian semiparametric additive quantile regression. *Statistical Modelling*, 13(3):223–252, 2013.

D. Walshaw. Getting the most from your extreme wind data: a step by step guide. *Journal of Research of the National Institute of Standards and Technology*, 99(4): 399, 1994.

P. Wan. Characterizing extremal dependence on a hyperplane. *arXiv:2411.00573*, 2024.

S. R. Winterstein, T. C. Ude, C. A. Cornell, P. Bjerager, and S. Haver. Environmental parameters for extreme response: Inverse form with omission factors. *Proceedings of the ICOSSAR-93, Innsbruck, Austria*, pages 551–557, 1993.

S. N. Wood. *Generalized additive models: an introduction with R*. Chapman and Hall/CRC, 2017.

L. Wu and R. Wang. Carbon monoxide: endogenous production, physiological functions, and pharmacological applications. *Pharmacological reviews*, 57(4):585–630, 2005.

K. Yu and R. A. Moyeed. Bayesian quantile regression. *Statistics & Probability Letters*, 54(4):437–447, 2001.

Y. Yuan, F. E. Bachl, F. Lindgren, D. L. Borchers, J. B. Illian, S. T. Buckland, H. Rue, and T. Gerrodette. Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. *Annals of Applied Statistics*, 11(4):2270–2297, 2017.

Z. Zhang and J. Huang. Extremal financial risk models and portfolio evaluation. *Computational Statistics & Data Analysis*, 51(4):2313–2338, 2006.

Z. Zhang, R. Huser, T. Opitz, and J. Wadsworth. Modeling spatial extremes using normal mean-variance mixtures. *Extremes*, 25(2):175–197, 2022.

H. Zhen Wai Olivier and C. Dombry. Simple models for multivariate regular variations and the Hüsler-Reiss Pareto distribution. *Journal of Multivariate Analysis*, 173, 12 2017.