A Review of Multivariate Extremes

Matthew Speers

May 4, 2022

1 Introduction

The modelling of the extreme values is naturally motivated by a range of applications. Often, we are most interested in the tail behaviour of a distribution when studying real world phenomena; take, for example, rainfall. The most commonly occurring levels of rainfall do not pose any threat to people or infrastructure, however, an unusually large observation can result in flooding. It is thus desirable to be able to accurately predict the chance of an unusually large, or extreme, value occurring. Standard methods for estimation of distributions are often not viable for this purpose, and can lead to inconsistent estimation. For this reason, extreme value specific models are required.

Univariate study was first undertaken by the likes of Gumbel (1958), and remains an active field to this day. The focus of this report, however, is of multivariate methodology. Extension to the multivariate setting is greatly motivated by applications where there is belief of some interaction between variables in extreme cases; for example, we could expect there to be an association between observing extreme levels of rainfall and high wind speeds. The estimation of the joint distribution of environmental variables such as these plays a large role in the design of ocean structures, see Ross et al. (2020). Another possibility is extremal dependence within a single phenomenon across time, such as that considered by Winter and Tawn (2016, 2017) in the modelling of extreme temperature to assess the risk of heatwaves. The study of the dependence structure between variables in their extremes is thus a central question in this field.

Early work on multivariate extremes was conducted by the likes of Pickands (1981), this building on ideas for the bivariate case given by those such as Tiago de Oliveira (1959) and Sibuya et al. (1960). Parametric models, alongside a deeper analysis of dependence structures, were provided by Tawn (1988) and Coles and Tawn (1991), both of which are discussed in §3. Theirs and comparable work prompted a more detailed consideration of possible categories of dependence, addressed in §4 via the work of Ledford and Tawn (1996) and Coles et al. (1999). These analyses, particularly that of Ledford and Tawn (1996), motivated the development of methodology that can capture all levels of extremal dependence, e.g. Ledford and Tawn (1997) and Bortot et al. (2000). A particularly influential response was developed Heffernan and Tawn (2004), which is discussed in §5.1.

We also look to more recent takes on assessing dependence structures, such as Simpson et al. (2020) which we discuss in §5.2. Theirs provides a methodology capable of reducing higher dimensional problems to more simple cases. We conclude with a discussion of novel additions to the extremes literature such

Tendijck et al. (2019) and Engelke and Hitz (2020), and suggest possible directions in which these ideas could be taken further.

2 Univariate Extremes

We begin with an overview of the main points of univariate extreme value theory. While the focus of this report is on the multivariate literature, it is beneficial to 'set the scene' in this way since much of the multivariate theory is motivated by univariate approaches. The distributions described in §2.1 and §2.2 are particularly crucial. The entirety of this section references Coles (2013).

2.1 Distribution of Maxima

In most settings, it is more important to consider the behaviour of the largest values of a process than the smallest; flood defences are constructed to protect against the highest sea levels, or structures are designed to resist the strongest winds. There are, of course, applications for which modelling unusually small values is useful, such as when looking at the lifetime of components in a system, see Coles (2013). However, most extreme value theory focuses on modelling unusually large values and so this is the approach taken for the rest of this report.

Consider a random variable X with unknown probability distribution F_X. We wish to gain information about the largest values X can take and with what frequency they occur. A natural choice is to attempt to find the distribution of $M_n = \max\{X_i, \dots, X_n\}$, the maximum value of a block of n realisations of X. Simple approaches such as taking $\mathbb{P}(M_n \leq z) = \{F_X(z)\}^n$ are not useful, since small changes to the estimation of F_X results in large changes to F_X^n (Coles, 2013). To avoid this, F_X itself is not estimated and we instead focus on families of models for the distribution of M_n. Another issue described by Coles (2013), is that the distribution of M_n degenerates to a point mass on the upper end point of F_X as $n \to \infty$. The transformation $M_n^* = (M_n - b_n)/a_n$ solves this problem, as for appropriate choices of $\{a_n > 0\}$ and $\{b_n\}$ the distribution of M_n^* is non-degenerate. The following statements, both taken from Coles (2013), describe the possible limit distributions for M_n^*.

Definition 2.1 (Generalised Extreme Value (GEV) Family of Distributions)**.** A probability distribution whose cumulative density function has the following general form

$$G(z) = \mathbb{P}(Z \leq z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\},$$

for $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$, is said to belong to the GEV family of probability distributions. If $\xi > 0$, we have the Fréchet family, if $\xi < 0$ we have the Gumbel family, and in the limiting case of $\xi \to 0$ we have the Weibull family.

Theorem 2.1 (Extremal Types Theorem)**.** *If there exist sequences of constants* $\{a_n > 0\}$ *and* $\{b_n\}$ *such that*

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq z\right) \to G(z) \quad \text{as } n \to \infty,$$

where G *is a non-degenerate distribution function, then* G *is a member of the GEV family. Here* F_X*, the distribution of* X*, is said to be in the domain of attraction of* G*.*

Definition 2.1 defines the Generalised Extreme Value (GEV) family of distributions. It encompasses the Fréchet, Gumbel and Weibull families. Theorem 2.1 states that, after transforming M_n into M_n^*, it follows a GEV distribution in the limit. It may seem that this is not useful in practice, since the normalising constants a_n, b_n will not usually be known. However, as explained by Coles (2013), if the distribution of M_n^* can be approximated by a member of the GEV family, so can the distribution of M_n. Using this, time series data of length nm can be split into sequential block maxima $\{M_{n,1}, M_{n,2}, \ldots, M_{n,m}\}$ which can be assumed to come from a GEV distribution with parameter $\Theta = (\mu, \sigma, \xi)$. Standard inference methods such as maximum likelihood can be used to obtain an estimate $\hat{\Theta}$ (provided that $\xi > -0.5$, which is usually the case in practice). The fitted model can then be used to estimate useful quantities, such as z_p for which $\mathbb{P}(M_{n,j} > z_p) = p$. As discussed by Coles (2013), this is known as the return level associated with the return period $1/p$, i.e. z_p is expected to be exceeded once every $1/p$ block periods.

Clearly, there is a question as to how to split the nm observations into m blocks. The answer is sometimes given by the context; it may seem natural to split the data by year for storm data going back decades. However, the trade off between small or large block sizes also needs to be considered. For smaller block sizes n, more observations $\{M_{n,j} : j = 1, \ldots, m\}$ will be available, leading to smaller variance in $\hat{\Theta}$. The downside is this may lead to violations in the assumptions that support theorem 2.1, for example due to seasonality in X which has period longer than the block length. These ideas are discussed in more detail by Coles (2013).

2.2 Threshold Models

A problem with the method described in §2.1 is one of wasted data. Say that we have a time series of seasonal weather conditions, such as rainfall, which we segment into monthly blocks. A particularly wet month may contain many extreme values, and so by only considering the maximum of these values we are losing information about the others. Conversely, a dry month may not contain any extreme values but we would be fitting our model to its maximum anyway, potentially distorting the model. We avoid this problem by not segmenting the data into blocks, but by instead considering the behaviour of X when it exceeds a threshold u. In other words, we seek to find the distribution of $X - u$ given $X > u$. The next theorem from Coles (2013) provides an approximation for this.

Theorem 2.2. *Let* $X_1, X_2, \ldots$ *be a sequence of i.i.d. random variables who's block maxima have limit distribution in the GEV family with parameter* $\Theta = (\mu, \sigma, \xi) \in \mathbb{R} \times (0, \infty) \times \mathbb{R}$*. Then, for large enough* u*,*

the distribution function of (X-u), conditional on $X > u$, is approximately

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi} \tag{1}$$

defined on $\{y : y > 0$ *and* $(1 + \xi y/\tilde{\sigma} > 0\}$*, where* $\tilde{\sigma} = \sigma + \xi(u - \mu)$.

If a random variable has probability distribution $H(y)$ as in equation 1, it is said to belong to the Generalised Pareto family of distributions and have parameter $\tilde{\Theta} = (\tilde{\sigma}, \xi)$. A clear question is how to select a threshold u for the model to be valid, whilst allowing enough data to fit the model. There is a trade off analogous to that for selecting block size in the block maxima case. Too low a threshold could invalidate the assumptions of theorem 2.2, or one too high could leave not enough data to provide estimates with reasonable certainty. Various methods to select an appropriate threshold exist, such as re-parameterising $\tilde{\sigma}$ to be independent of u and then selecting the lowest threshold for which the parameter estimate appears constant. This approach, along with some others, is detailed in Coles (2013). Once the threshold has been selected using the chosen method, maximum likelihood techniques can be used to find an estimate for the parameter $\tilde{\Theta}$. Quantiles and return levels can then be estimated using $\mathbb{P}(X > u + y) = \{1 - \hat{F}_X(u)\}H(y)$, where $\hat{F}_X$ is some estimate for the distribution of X below u. Further discussion of these ideas, along with outline justifications of the theorems in this section, are contained in Coles (2013).

This section contains an admittedly brief introduction to univariate extreme value theory. We reiterate that its purpose is to serve as an introduction to the area, in preparation for a more in-depth consideration of multivariate methods. There are many interesting challenges and considerations that can be made when implementing these models in practice, such as the inclusion of covariates to better model real-world systems, or the consideration that the behaviour of a phenomenon will change over time. Some of these ideas are discussed in the multivariate setting in §6.

3 Fundamentals of Multivariate Extremes

For the remainder of the report we will consider cases where we have two or more variables, i.e. the multivariate setting. This section highlights some of the things we must consider when extending univariate concepts to the multivariate setting, and introduces some fundamental multivariate methodology.

3.1 Multivariate Ordering

We begin by asking an important question: how do we decide what an extreme value looks like in a multi-dimensional situation? Whilst this is trivial in the univariate setting, here there is no clear natural solution. Four ways of ordering multivariate vectors were discussed by Barnett (1976). Perhaps the simplest idea is to take an extreme of a set to be the vector consisting of the individual maxima of the

variables, indicated by the first plot in figure 1. Or, we could take the extremes as the points for which each variable is most extreme, as is shown in the third image. Other options include considering a convex hull, and an approach based on structure variables of the specific system being considered, illustrated in the second and fourth plots respectively. These are outlined in more detail by Barnett (1976). The most popular interpretation is the first, where the maximum of a set is taken as the component-wise maxima. This is despite it usually being the case that the jointly-extreme vector will not have actually occurred in the set. Also, all of these orderings can possess the same downside as the univariate situation shown in §2.1, where useful data can be discarded, leading to less precise results than those that could be obtained if more data was used.

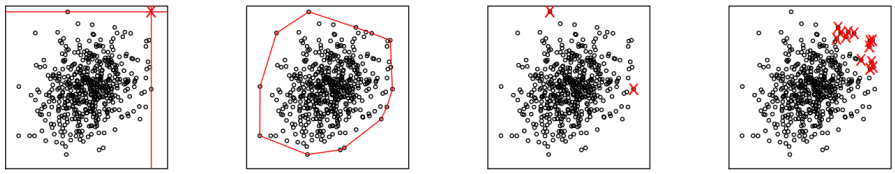

Figure 1: Illustration of various ways to order extremes (Presented by Tawn, 2022)

Models based on the first ordering are developed in detail by Tawn (1988), where data are divided into blocks and the maxima of each block taken, as with §2.1. We give an overview of some of their findings in §3.2, then move to a more detailed discussion of the work of Coles and Tawn (1991) in §3.3. The latter includes a method of estimating model parameters that utilises the strategy discussed in §2.2, and so avoids the problem of ordering altogether.

3.2 Distributions for Component-wise Maxima

We begin our review of the multivariate literature with a brief look at the work of Tawn (1988). Here, the modelling of the dependence between variables in their extremes is the primary focus. In any multivariate setting it is not enough to only determine how each variable behaves; we must also consider how the variables interact with each other. In the case of extreme value theory, we focus on the interaction between variables in extreme cases. To be specific: how does one variable being large affect other variables' chances to be large? Many applications motivate the modelling of these dependencies. Sea walls may be able to withstand an unusually high sea level or wave height individually, but not simultaneously. Assuming independence between these two phenomena would lead to an underestimation of the probability of them being extreme together, and thus an underestimation of the risk of wall failure and subsequent flooding. Studying the interaction between them from a statistical perspective allows for a better risk assessment.

The general form of bivariate extreme value distributions were originally developed by the likes of Sibuya et al. (1960), being generalised to multivariate cases later, see Pickands (1981). The extension of the methods in §2.1 to a multivariate setting, as described by Tawn (1988), proceeds as follows. For a block of n realisations of the p dimensional vector $(X_1, \ldots, X_p)$ with unknown margins, define the maximum to be the vector $(M_{1n}, \ldots, M_{pn})$, where $M_{jn} = \max_{i=1,\ldots,n} X_{ji}$ for $j = 1, \ldots, p$. Analogous to the univariate

case discussed in §2.1, normalising constants $a_{jn} > 0, b_{jn}$ for $j = 1, \ldots, p$ can be selected such that

$$\mathbb{P}\left(\frac{M_{1n} - b_{1n}}{a_{1n}} \leq z_1, \ldots, \frac{M_{pn} - b_{pn}}{a_{pn}} \leq z_p\right) \to G(z_1, \ldots, z_p),$$

as $n \to \infty$. Here, G is known as a multivariate extreme value distribution and has GEV margins Tawn (1988). A detailed analysis of the general form of the dependence function of G is provided by Tawn (1988), including methods for both parametric and non-parametric estimation. We do not cover this analysis here for the sake of brevity. Instead, we detail a particular dependence structure they define, the multivariate logistic. The distribution function G for this model, as given in Ledford and Tawn (1996), is

$$G(z_1, \ldots, z_p) = \exp\left\{-\left(z_1^{-1/\alpha} + \ldots + z_p^{-1/\alpha}\right)^{\alpha}\right\},$$

for unit Fréchet margins and dependence parameter $0 \leq \alpha \leq 1$. When $\alpha = 1$, we have exact independence of the variables. For the limiting case of $\alpha \to 0$, we have perfect dependence. We include this model in particular as it is widely applicable model both in practice and in developing theoretical results, see Ledford and Tawn (1996). Early applications of models such as the multivariate logistic, for example the work of Smith et al. (1990), involved estimation of first the marginal parameters followed by the dependence parameters. The following section details an approach which possesses the benefit of being able to estimate both the marginal and dependence structure parameters simultaneously.

3.3 Point Process Model

We look at the methodology developed by Coles and Tawn (1991), based on the work of Haan (1985). The approach taken by Joe et al. (1992) in modelling bivariate extremes also stems from this work, however, we do not focus on it here. Rather than considering only block maxima, they utilise all data which are extreme in at least one margin by approximating their distribution with that of a Poisson point process. Their method bypasses the processes of segmenting into blocks and ordering values to select maxima, and so avoids their related issues. They consider a sequence of i.i.d. vectors $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$ on $\mathbb{R}_+^p$, with unit Fréchet margins. They note that the choice of marginal distribution is inconsequential, as a suitable transformation can be applied to give any other distribution. Also, the distribution function $F_{\boldsymbol{X}}$ of $\boldsymbol{X}$ is assumed to be in the domain of attraction of some multivariate extreme value distribution G, see Tawn (1988) or §3.2 for some examples of these in the bivariate case. First, they introduce pseudo-radial and angular coordinates r and w, defined as

$$r_i = \sum_{j=1}^{p} \frac{X_{ij}}{n} \quad \text{and} \quad w_{ij} = \frac{X_{ij}}{nr_i} = \frac{X_{ij}}{\sum_{j=1}^{p} X_{ij}}, \tag{2}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, p$. Here, r_i can be interpreted as the overall 'strength' of the i^{th} observation $\boldsymbol{X}_i$, and w_{ij} as how much of that strength is placed on the j^{th} component of $\boldsymbol{X}_i$. For example, if $w_{i1} = 1$ and $w_{ij} = 0$ for $j \neq 1$, then $\boldsymbol{X}_i$ will be a sparse vector with $X_{ij} = 0$ for $j \neq 1$. Now take the point process

$P_n = \{n^{-1}\boldsymbol{X}_i; i = 1, \dots n\}$. Then, as stated by Coles and Tawn (1991), P_n converges in distribution to a non-homogeneous Poisson process P on $\mathbb{R}_+^p \setminus \{0\}$ with intensity measure μ that satisfies

$$\mu(\mathrm{d}r \times \mathrm{d}\mathbf{w}) = \frac{\mathrm{d}r}{r^2} H(\mathbf{w}), \tag{3}$$

where H is a positive finite measure. That is, for large enough n, the points in $P_n = \{n^{-1}\boldsymbol{X}_i; i = 1, \dots n\}$ that lie in a region $A \subset \mathbb{R}_+^p \setminus \{0\}$ behave approximately as a Poisson process P, where the intensity of P is given by equation 3. Importantly, the region A must be suitably far away from $\mathbf{0}$, with the necessary distance being dependent on the rate of convergence of P_n to P (Coles and Tawn, 1991). In addition to providing the intensity of the approximating Poisson process, equation 3 also shows that the measure μ can be decomposed into a known function of r and a measure H on $\mathbf{w}$, although this only holds for a suitably large r. As stated by Coles and Tawn (1991) the measure H thus contains all information about the extremal dependence of the components of $\boldsymbol{X}_i$, meaning the dependence structure can be studied by determining where H places mass on

$$S_p = \{(w_1, \dots, w_p) : \sum_{j=1}^{p} w_j = 1,\ w_j \geq 0\ j = 1, \dots, p\},$$

where S_p (sometimes denoted S_{p-1}) is known as the $(p-1)$-dimensional unit-simplex. This fact has proven crucial to the study of dependence in multivariate extremes, even being the foundation of very recent developments by Simpson et al. (2020), which will be discussed in §5.2. As such, it is desirable be able to model the measure H. This is not easy, as the only constraints placed on H are that it is a positive finite measure which satisfies

$$\int_{S_p} w_j \, \mathrm{d}H(\mathbf{w}) = 1$$

for $j = 1, \dots, p$. The following process, developed by Coles and Tawn (1991), facilitates finding a parametric measure that satisfies the constraints, provided we already have a multivariate extreme value distribution for (normalised) component-wise maxima (again, see Tawn (1988) or §3.1 for examples).

Take any limit distribution G of the normalised vector of component-wise maxima $(n^{-1}M_{n,1}, \dots, n^{-1}M_{n,p})$, where $M_{n,j}$ is the maximum over n realisations of the j^{th} component of a random vector with unknown distribution. Let $A = \mathbb{R}_+^p \setminus \{(0, x_1) \times \dots \times (0, x_p)\}$ and $\mathbf{x} = (x_1, \dots, x_p)$. Then, we have $G(\mathbf{x}) = \mathbb{P}\left(n^{-1}M_{n,j} \leq x_j, j = 1, \dots, p\right) = \mathbb{P}\left(n^{-1}\boldsymbol{X}_i \notin A, i = 1, \dots n\right)$. By the convergence to a Poisson process, $\mathbb{P}\left(n^{-1}\boldsymbol{X}_i \notin A, i = 1, \dots n\right) \to \exp\left(-\mu(A)\right)$ as $n \to \infty$, giving $G(\mathbf{x}) = \exp\left(-\mu(A)\right)$, approximately. The authors show that $\mu(A) = \int_{S_p} \max_{1 \leq j \leq p}(w_j/x_j) \, \mathrm{d}H(\mathbf{w})$ in this case, and so we can write

$$G(\mathbf{x}) = \exp\left(-V(\mathbf{x})\right) \quad \text{where} \quad V(\mathbf{x}) = \int_{S_p} \max_{1 \leq j \leq p} \frac{w_j}{x_j} \, \mathrm{d}H(\mathbf{w}). \tag{4}$$

Using this result, Coles and Tawn (1991) prove the following theorem. Before stating the theorem, we must first define some notation. Let c be a subset of the set $\{w_1, \dots, w_p\}$ and S_c be the subspace of S_p such that only the elements of c are extreme, i.e. $S_c = \{\mathbf{w} \in S_p : w_j = 0, w_j \notin c, j = 1, \dots, p\}$. We denote

the density that H places on S_c as h_c.

Theorem 3.1. *Let V and H be the measures related by equation 4. Let h_c be the density defined above on S_c. Then for $c_x = \{x_{c,1}, \ldots, x_{c,m}\}$, the set of x_j corresponding to the elements of c, we have*

$$\frac{\partial V}{\partial x_{c,1} \ldots \partial x_{c,m}} = \left(\sum_{j=1}^{m} x_{c,j} \right)^{-(m+1)} h_c \left(\frac{x_{c,1}}{\sum_j x_{c,j}}, \ldots, \frac{x_{c,m}}{\sum_j x_{c,j}} \right).$$

Using theorem 3.1, we can go from the distribution function $G(\mathbf{x}) = \exp(-V(\mathbf{x}))$ of the component-wise maxima of $\boldsymbol{X}$, to the density placed on an a subspace of S_c of S_p by the measure H associated with V. This is beneficial because, as stated previously, the extremal dependence structure of $\boldsymbol{X}$ is described entirely by where H places mass on the unit simplex. For example, if $h_c = 0$ for all c such that $|c| > 1$, we know that no components of $\boldsymbol{X}$ can be extreme together (i.e. complete independence) and that H places all mass at the vertices of S_p. Alternatively, we could have $h_c > 0$ for $|c| = p$ and $h_c = 0$ otherwise, meaning all mass it places in the interior of S_p and all the components of $\boldsymbol{X}$ must be extreme together (all components dependent). Figures 2 and 3 give simple illustrations of these two scenarios in the trivariate case of $\boldsymbol{X} = (X_1, X_2, X_3)$. Of course, in the latter scenario, the density need not be circular and uniform.

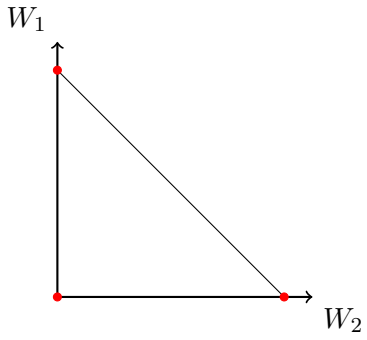

Figure 2: Mass placed on vertices

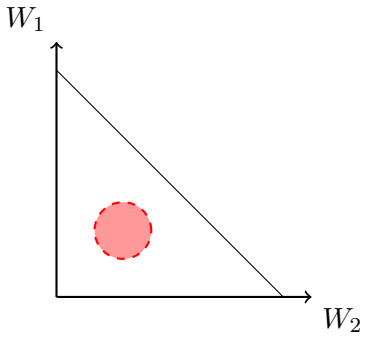

Figure 3: Mass placed on interior

There is, of course, a broad spectrum of possibilities between these two cases which reflect more nuanced dependence structures. Simply put, if H places mass on a region of S_p for which $w_{c,1}, \ldots, w_{c,m} > 0$, then there is some dependence between the components $X_{c,1}, \ldots, X_{c,m}$ of $\boldsymbol{X}$. This idea is explored more formally, and related to the dependence measures discussed in §4, by Simpson et al. (2020). This work will be covered in §5.2.

A clear issue with using theorem 3.1 to model H, is that is that it requires a multivariate extreme value distribution G, such as the logistic extreme value distribution introduced in §3.2. The authors thus propose another technique to generate parametric models for a density h on the interior of S_p, by relating the representation for G given by equation 4 to the one of Pickands (1981). The technique is summarised in the following theorem, stated in Coles and Tawn (1991). By using theorem 3.2 new valid measures and associated multivariate extreme value distributions can be obtained, the bivariate Dirichlet model given by Coles and Tawn (1991) serving as an example.

Theorem 3.2. *If h^* is any positive function on S_p, with finite first moments, then*

$$h(\boldsymbol{w}) = \langle \boldsymbol{m}, \boldsymbol{w} \rangle^{(-p+1)} \prod_{j=1}^{p} m_j h^* \left(\frac{m_1 w_1}{\langle \boldsymbol{m}, \boldsymbol{w} \rangle}, \ldots, \frac{m_p w_p}{\langle \boldsymbol{m}, \boldsymbol{w} \rangle} \right),$$

where

$$m_j = \int_{S_p} u_j h^*(\boldsymbol{u}) d\boldsymbol{u}, \quad j = 1, \ldots, p,$$

satisfies the constraints given by equation 3.3. Therefore, h is the density of a valid measure H.

Having generated the density of a valid measure H on some subsets of S_p with parameter Ψ, maximum likelihood techniques can be used to provide an estimate for Ψ. The authors do this by calculating the likelihood over the region $A = \mathbb{R}_+^p \setminus \{(0, v_1) \times \ldots \times (0, v_p)\}$ for the limiting Poisson process where the values $v_j, j = 1, \ldots, p$ are suitably chosen thresholds so that A is the region which contains all observations which are large in at least one margin. Presume that we have a realisation of the sequence of i.i.d. random vectors $\{\boldsymbol{Y}_i, i = 1, \ldots, n\}$, with unknown marginal distributions. To be able to utilise the Poisson process approximation described above, we require random vectors with unit Fréchet margins. The marginal components Y_{ij} are thus transformed to unit Fréchet random variables $X_{ij} = g(Y_{ij})$ via the probability integral transform. Of course, the distribution function of Y_{ij} is unknown and so must be estimated. For Y_{ij} values above a high threshold u_j, selected using the methods briefly introduced in §2.2, the GPD distribution from theorem 1 is used alongside an estimate for $\mathbb{P}(Y_{ij} > u_j)$. The parameter $\tilde{\Theta}$ of the GPD is then estimated alongside Ψ. For Y_{ij} values below the threshold u_j, the realisations are dense, and so the empirical distribution function is a suitable estimate for the distribution.

Also, by considering the threshold u_j, which acts as the boundary over which Y_{ij} is classified as large, we can determine the values of v_j necessary to give us the desired region A. A component Y_{ij} is considered large if $Y_{ij} > u_j$. Applying the transformation g and normalising as is required for the Poisson process approximation gives $(Y_{ij} > u_j) \implies (X_{ij} > n^{-1} g(u_j))$, hence we have $v_j = n^{-1} g(u_j)$. It also must be checked that these v_j values give a region A which contains values with large enough r to satisfy equation 3. Given all this, and the set $\{n^{-1}\mathbf{x}_{a,1}, \ldots, n^{-1}\mathbf{x}_{a,m}\}$ of transformed observations that lie in A, the likelihood over A for the Poisson process approximation with dependence measure h can be calculated. The formulation of this likelihood has been omitted here in an attempt to remain as concise as possible, so we refer to Coles and Tawn (1991) for the detail. Crucially, this method offers a way of estimating the marginal parameter $\tilde{\Theta}$ and the dependence parameter Ψ simultaneously, which as noted by Coles and Tawn (1991), allows for improved precision over sequential estimation.

4 Models for Dependence

A crucial question when modelling the extremes of a variable is the level of dependence between its components. The section motivates the need to quantify this dependence, and details approaches that

have been developed to do so.

4.1 The η parameter

Whilst a seemingly useful approach, the Poisson process representation developed by Coles and Tawn (1991) possesses problems when applied under certain conditions. To explain when these occur, we must first introduce the concept of asymptotic independence, defined here for the bivariate case. The ideas in this section can be extended to the multivariate case, see Heffernan and Tawn (2004). The following definition is based on that of Coles et al. (1999).

Definition 4.1 (Bivariate Asymptotic Independence)**.** Consider a bivariate random variable (X_1, X_2) where each element has identical marginal distributions. We say X_1 and X_2 are asymptotically independent if

$$\lim_{x \to x_u} \mathbb{P}\left(X_2 > x | X_1 > x\right) = 0, \tag{5}$$

where x_u is the upper limit of the distributions of X_1 and X_2. Otherwise, we say they are asymptotically dependent. In a sense, equation 5 is the probability of one variable being extreme, given that the other is extreme.

The model discussed in §3.3, along with the comparable method of Joe et al. (1992) referenced above, performs poorly when variables are exactly independent or asymptotically independent, i.e. satisfy equation 5. The Poisson process method will model any variables that occur sufficiently large together as asymptotically dependent (Ledford and Tawn, 1996), and so can overestimate the probability of them being large together. To circumvent this issue, a more nuanced categorisation of dependence between variables is needed. This necessity is emphasised by Ledford and Tawn (1996), who demonstrate that their own methodology falsely identifies a bivariate normal variable with $0 < \rho < 1$ as asymptotically dependent. They identify as the problem the fact that their model assumes either asymptotic dependence or exact independence, and so is incapable of coping with structures that lie between these two cases. Motivated by this, they develop the following approach that allows the classification of additional levels of dependence. Consider a bivariate random variable (X_1, X_2) with unit Fréchet margins. They show that, for large x,

$$\mathbb{P}\left(X_1 > x, X_2 > x\right) \sim \begin{cases} x^{-1} & \text{for perfect dependence,} \\ x^{-2} & \text{for exact independence.} \end{cases} \tag{6}$$

where $a_n \sim b_n$ indicates that $a_n / b_n \to 1$ as $n \to \infty$. Equation 6 represents the two possible end cases, which Ledford and Tawn (1996) link via

$$\mathbb{P}\left(X_1 > x, X_2 > x\right) \sim \mathcal{L}(x) x^{-\frac{1}{\eta}} \tag{7}$$

as $x \to \infty$, where $1/2 \leq \eta \leq 1$ is a constant, and $\mathcal{L}(x)$ is a slowly varying function. Here, η is known as the coefficient of tail dependence. Equation 7 offers a smooth link between the two bounding cases in

equation 6, with $\eta = 1, \mathcal{L}(x) = 1$ giving perfect dependence, and $\eta = 1/2, \mathcal{L}(x) = 1$ exact independence. Additionally, the interval $0 \leq \eta < 1/2$ corresponds to negative dependence, though this case is rarely applicable to extremes and so we do not consider this here. Values of η that correspond to asymptotic dependence or independence can also be obtained, by noting that the conditional probability in equation 5 can be written as approximately $\mathbb{P}(X_2 > x | X_1 > x) \sim \mathcal{L}(x) x^{1-\frac{1}{\eta}}$, given that x is large. The variables X_1 and X_2 are asymptotically independent if the right hand side approaches 0 as x gets larger, which corresponds to $\eta < 1$. Similarly, they are asymptotically dependent if the right hand side approaches a non-zero constant as x gets larger, which occurs when $\eta = 1$ and $\mathcal{L}(x) \nrightarrow 0$ as $x \to \infty$. The authors break down the case of positive tail dependence into three, more descriptive, sub-cases. These cases and their corresponding values of η are summarised as follows.

Two variables exhibit *asymptotic dependence* if $\eta = 1$ and $\mathcal{L}(x) \nrightarrow 0$ as $x \to \infty$; they exhibit *positive association* if $1/2 < \eta < 1$, and they exhibit *near independence* if $\eta = 1/2$ and $\mathcal{L}(x) \geq 1$. Under the assumptions of previous models, had two variables shown positive dependence, they would have been assumed to be asymptotically dependent and thus classified under the first case. As Ledford and Tawn (1996) point out, however, this can lead to variables which are positively, but not asymptotically, dependent being incorrectly classified into the first case. The inclusion of the second and third categories solves this problem, as they can now be classified into either of them. For example, the bivariate normal distribution with $0 < \rho < 1$ that they use as an illustration is now classified into the second case, with $\eta = (1 + \rho)/2$. Of course, in order to determine which of these categories the dependence between two variables falls into, we require a way of estimating η.

This is achieved by Ledford and Tawn (1996) through considering $T = \min(X_1, X_2)$, the so-called structure variable. We have $\mathbb{P}(T > x) = \mathbb{P}(X_1 > x, X_2 > x) \sim \mathcal{L}(x) x^{-1/\eta}$ as the survivor function of T. It can be shown that, for a high threshold u, we have $\mathbb{P}(T > u + x | T > u) \sim (1 + x/u)^{-1/\eta}$ for variables with this survivor function. Comparing this to the GPD approximation for exceedances above a high threshold in theorem 2.2, we obtain $\xi = \eta$ and $\tilde{\sigma} = \eta u$. Therefore, η for a bivariate random variable (Y_1, Y_2) can be estimated from a sample $\{(Y_{1i}, Y_{2i}); i = 1, \ldots, n\}$ in the following manner. First transform the marginal distributions of each variable to unit Fréchet, using probability integral techniques akin to those discussed in §3.3, to obtain a sample $\{(X_{1i}, X_{2i}); i = 1, \ldots, n\}$ with unit Fréchet margins. Next, construct a sample of the structure variable T as $\{T_i = \min(X_{1i}, X_{2i}); i = 1, \ldots, n\}$. A suitable threshold u can then be determined and a maximum likelihood estimate found for $\tilde{\Theta} = (\tilde{\sigma}, \xi)$, as described in §2.2. The parameter η is then estimated via $\hat{\eta} = \hat{\xi}$, where $\hat{\xi}$ is the MLE for ξ.

The ability to provide an estimate of η and thus categorise the dependence of two (or more) extremes is useful as a way to determine if models such as those of Coles and Tawn (1991) and Joe et al. (1992) can be applied. As mentioned above, these are only suitable when $\eta = 1$. The existence of the classes for which $\eta \leq 1$ thus motivates the development of extreme value models which allow for asymptotic independence. Examples of these include the bivariate joint tail model developed by Ledford and Tawn (1997), and the multivariate Gaussian tail model subsequently proposed by Bortot et al. (2000).

4.2 The χ and $\bar{\chi}$ Measures

Motivated by similar reasons to Ledford and Tawn (1996), Coles et al. (1999) developed alternative, but comparable, measures of extremal dependence within the class of asymptotically independent variables. They consider equation 5 for a bivariate (X_1, X_2) which has been transformed to (U_1, U_2) with Uniform margins. Letting $\chi = \lim_{u \to 1} \mathbb{P}\left(U_2 > u | U_1 > u\right)$, they then define

$$\chi(u) = 2 - \frac{\log \mathbb{P}\left(U_1 < u, U_2 < u\right)}{\log \mathbb{P}\left(U_1 < u\right)} \quad \text{for} \quad 0 \leq u \leq 1,$$

and show that $\chi = \lim_{u \to 1} \chi(u)$. This is a useful measure as is it constant in u when the variables exhibit asymptotic or complete independence, and so can be used to verify if models which only account for these cases can be used. This can be done by simply calculating $\chi(u)$ for various thresholds u and determining graphically if the values appear constant. The authors also develop formal inference procedures for χ, however, we omit the detail of these here. Unfortunately, the convergence of $\chi(u)$ to 0 as $u \to \infty$ can be very slow for intermediate values of u, meaning $\chi(u)$ can misleadingly appear constant when estimating from asymptotically independent data. This can lead to the incorrect classification of asymptotically independent variables as asymptotically dependent, presenting the same issues discussed in relation to previous methods. Also, χ provides no detail of the dependence of asymptotically independent variables, merely serving to determine that they fall into this class. For these reasons, Coles et al. (1999) introduce the next measure which provides easier interpretation and, similarly to η, a measure of dependence within the asymptotically independent class. They define

$$\bar{\chi}(u) = \frac{2 \log \mathbb{P}\left(U_1 > u\right)}{\log \mathbb{P}\left(U_1 > u, U_2 > 2\right)} - 1 \quad \text{for} \quad 0 \leq u \leq 1,$$

and

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

which give $-1 \leq \bar{\chi} \leq 1$. For asymptotically dependent variables, $\bar{\chi} = 1$, and conversely $\bar{\chi} \leq 1$ for asymptotically independent variables. In the latter case, $\bar{\chi}$ provides a useful measure of the strength of dependence, in a sense determining how close to asymptotic dependence the variables are. For example, for the bivariate normal case previously discussed in relation to η, we have $\bar{\chi} = \rho$. This aligns with the findings of Ledford and Tawn (1996) which yielded $\eta = (1 + \rho)/2$ in this case, showing that we only have asymptotic dependence ($\eta = 1$) in the case of perfect dependence ($\rho = 1$). In fact, Coles et al. (1999) show that, under certain conditions, $\bar{\chi} = 2\eta - 1$. The bivariate normal case illustrates this nicely with $\bar{\chi} = 2\eta - 1 = 2((1 + \rho)/2) - 1 = \rho$.

We can examine the behaviour of $\bar{\chi}$ via the same process as χ, by plotting $\bar{\chi}(u)$ for varying values of u. The convergence of $\bar{\chi}$, in particular its bounding below 1 in the asymptotically independent case, is much easier to observe graphically than that of χ (Coles et al., 1999). Therefore, $\bar{\chi}$ lends itself to more reliable interpretation of the dependence structure between U_1 and U_2 (and thus X_1 and X_2).

5 Improving Extrapolation

Given this more nuanced idea of categorising dependence between variables, methodology is needed which incorporates it to improve modelling. This section details some examples which aim to provide this.

5.1 A Conditional Approach

The η parameter introduced in §4.1 provides an additional use to that of categorising dependence; it can also be used in the estimation of extremal probabilities. As stated by Heffernan and Tawn (2004), all of the models previously discussed operate under the assumption of multivariate regular variation, see Basrak et al. (2002) for a detailed discussion of this concept. Define $t + A$ to be the set consisting of the elements of A after a component-wise translation by $t > 0$. Then, under the assumption of multivariate regular variation, we have

$$\mathbb{P}\left(\boldsymbol{Y} \in t + A\right) = \exp\left(-t/\eta\right)\mathbb{P}\left(\boldsymbol{Y} \in A\right), \tag{8}$$

for a p-dimensional variable $\boldsymbol{Y}$ with Gumbel margins (Heffernan and Tawn, 2004). Here, η is the multivariate extension of the coefficient of tail dependence for $\boldsymbol{Y}$. For $0 < \eta < 1$, we require A to contain only observations which are extreme in all margins. In the case of $\eta = 1$, however, this requirement can be relaxed to observations which are large in at least one margin, this being the choice of set in §3.3 when we were operating under the assumption of asymptotic dependence. Equation 8 provides the basis on which many multivariate extreme value methods function, including those of Ledford and Tawn (1997) and Bortot et al. (2000) developed in response to the need for asymptotically independent models highlighted by Ledford and Tawn (1996). Unfortunately, due to the restrictions imposed on A when $\eta \neq 1$, we can only use equation 8 and its associated methods to estimate the probability of sets for which all variables are large at the same time. Figure 4 gives an illustration of why this is the case. In essence, translation of a set of values which are equally extreme in each component will yield a set of values with the same property. Thus, we cannot select a set A and a scalar $t > 0$ so that $A + t = B$, where B is a set of values which are extreme in only one margin.

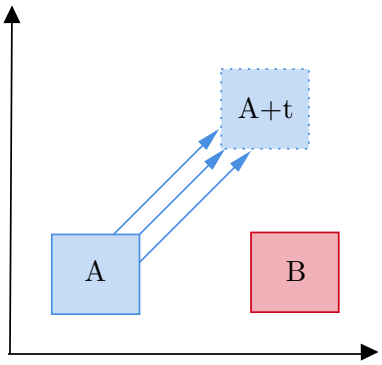

Figure 4: A cannot be diagonally translated onto B

Having identified this issue, Heffernan and Tawn (2004) proposed an alternative method that allows for

estimation of the probability of observing values which are extreme in only one dimension, for any value of η. Consider a p-dimensional random vector $\boldsymbol{X}$ with unknown margins. We seek to find the probability $\mathbb{P}(\boldsymbol{X} \in C)$ where C is a set for which all elements are large in at least one margin. The authors' approach calculating $\mathbb{P}(\boldsymbol{X} \in C)$ by partitioning $C = \cup_{i=1}^{p} C_i$, where C_i is the part of C for which element X_i is the largest. Formally, $C_j = C \cap \{\mathbf{x} \in \mathbb{R}^p : F_i(x_i) > F_j(x_j); j = 1, \ldots, p; j \neq i\}$, for $i = 1, \ldots, p$, where F_j is the marginal distribution function of X_j. Of course, estimating these summed probabilities is essentially the same problem as estimating $\mathbb{P}(\boldsymbol{X} \in C)$, so it is required to decompose this probability further. This is done by Heffernan and Tawn (2004) in the following manner. Take, for non-empty C_i, $v_i = \inf_{\mathbf{x} \in C_i}(x_i)$ to be the smallest value of X_i over all elements in C_i. The probability that $\boldsymbol{X}$ lies in C can then be written as

$$\mathbb{P}(\boldsymbol{X} \in C) = \sum_{i=1}^{p} \mathbb{P}(\boldsymbol{X} \in C_i) = \sum_{i=1}^{p} \mathbb{P}(\boldsymbol{X} \in C_i | X_i > v_i)\, \mathbb{P}(X_i > v_i). \tag{9}$$

Estimation of $\mathbb{P}(\boldsymbol{X} \in C)$ thus becomes equivalent to estimating the two summed probabilities in equation 9. The value v_i must be extreme, as X_i is the largest element of $\boldsymbol{X}$ in this case and so otherwise there would be no extreme elements of $\boldsymbol{X}$ and thus it would not lie in the extreme set C. For this reason, $\mathbb{P}(X_i > v_i)$ cannot be reliably estimated via the empirical distribution function. The authors thus adopt the approach taken by Coles and Tawn (1991) discussed in §3.3, where instead a GPD with parameter $\tilde{\Theta}_i = (\tilde{\sigma}_i, \xi_i)$ is used to model F_i above a high threshold u_i. We denote the estimated distribution function for X_i as $\hat{F}_i$. The estimation of $\mathbb{P}(\boldsymbol{X} \in C_i | X_i > v_i)$ is more involved, since the dependence structures between variables must be modelled in some way. The authors conduct this estimation via the decomposition

$$\mathbb{P}(\boldsymbol{X} \in C_i | X_i > v_i) = \int_{v_i}^{x_{u(i)}} \mathbb{P}(\boldsymbol{X} \in C | X_i = x)\, \mathrm{d}\hat{F}_i(x) / \{1 - \hat{F}_i(v_i)\},$$

where $x_{u(i)}$ is the upper end point of $\hat{F}_i$. The main undertaking here becomes the modelling of the conditional probability $\mathbb{P}(\boldsymbol{X} \in C | X_i = x)$, which they examine in the limiting case as x becomes large and thus develop a model for the asymptotic form of the conditional distribution. They consider the case of a p-dimensional random vector $\boldsymbol{Y}$ with Gumbel margins. Let h be the transformation to Gumbel margins via the probability integral transform, using the estimate for the distribution of $\boldsymbol{X}$ given by Coles and Tawn (1991). Also, let $\boldsymbol{Y}_{-i}$ be the random vector $\boldsymbol{Y}$ with component Y_i removed. Then, Heffernan and Tawn (2004) assume that, for a high threshold u_i', there are vector normalising functions $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i)$ on $\mathbb{R} \to \mathbb{R}^{p-1}$ such that

$$\mathbb{P}\left(\boldsymbol{Y}_{-i} < \mathbf{a}_{|i}(y_i) + \mathbf{b}_{|i}(y_i)\mathbf{z}_{|i} | Y_i = y_i\right) = G_{|i}(\mathbf{z}_{|i}),$$

for $y_i > u_i'$. Here, $\mathbf{z}_{|i}$ is the standardised vector resulting from the component-wise translation

$$\mathbf{z}_{|i} = \frac{\mathbf{y}_{-i} - \mathbf{a}_{|i}(y_i)}{\mathbf{b}_{|i}(y_i)}.$$

It can be shown that the transformed random vector $\mathbf{Z}_{|i}$ is independent of Y_i. This independence motivates

the sampling procedure devised by Heffernan and Tawn (2004), described in algorithm 1. However, before we can perform the sampling algorithm, we require estimates for marginal parameters $\tilde{\Theta}_i$, the normalising values $\mathbf{a}_{|i}$, $\mathbf{b}_{|i}$, and the distribution function $G_{|i}$ of $\mathbf{Z}_{|i}$. The approach taken by Heffernan and Tawn (2004) to find these estimates is as follows. Estimation of the marginal parameters $\tilde{\Theta}_i$ is done via maximum likelihood based methods, similar to that of Coles and Tawn (1991). The normalising constants $\mathbf{a}_{|i}$ and $\mathbf{b}_{|i}$ are taken to be members of the parametric family

$$\begin{aligned}\mathbf{a}_{|i}(y) &= \mathbf{a}_{|i} + I_{\{\mathbf{a}_{|i}=0,\mathbf{b}_{|i}<0\}}\{\mathbf{c}_{|i} - \mathbf{d}_{|i}\log(y)\},\\ \mathbf{b}_{|i}(y) &= y^{\mathbf{b}_{|i}},\end{aligned}$$

meaning it is also required to estimate $\mathbf{c}_{|i}, \mathbf{d}_{|i}$ on $\mathbb{R} \to \mathbb{R}^{p-1}$ in the case of $\hat{\mathbf{a}}_{|i} = 0, \hat{\mathbf{b}}_{|i} < 0$. The distribution $G_{|i}$ is assumed to have independent normal margins with means given by the vector $\boldsymbol{\mu}_{|i}$ and standard deviations by $\boldsymbol{\sigma}_{|i}$. The authors note this parametric model is likely to be misspecified, however, it is assumed here for simplicity. Due to this assumption, Gaussian estimation can be used as a basis of inference, see Hand and Crowder (2017) for details of this theory. The parameter $\boldsymbol{\theta}_{|i} = (\mathbf{a}_{|i}, \mathbf{b}_{|i}, \mathbf{c}_{|i}, \mathbf{d}_{|i}, \boldsymbol{\mu}_{|i}, \boldsymbol{\lambda}_{|i})$ is then estimated by jointly maximising over the objective function which arises from the ideas of Hand and Crowder (2017), which is not included here for the sake of brevity.

Algorithm 1 Conditional Sampling Algorithm

1: Simulate Y_i from a Gumbel distribution conditional on its exceeding $h(v_i)$.
2: Sample $\mathbf{Z}_{|i}$ from $\hat{G}_{|i}$ independently of Y_i.
3: Obtain $\boldsymbol{Y}_{-i} = \hat{\mathbf{a}}_{|i}(Y_i) + \hat{\mathbf{b}}_{|i}(Y_i)\mathbf{Z}_{|i}$.
4: Transform $\boldsymbol{Y} = (\boldsymbol{Y}_{-i}, Y - i)$ to original margins via probability integral transform using $\hat{F}_i$
The resulting vector is a sample from $\boldsymbol{X}|x_i > u_i$.

Having obtained an estimate $\hat{\boldsymbol{\theta}}_{|i}$, algorithm 1 can be followed to obtain a Monte-Carlo estimate for $\mathbb{P}(\boldsymbol{X} \in C_i | X_i > v_i)$, by finding the proportions of observations obtained from sampling that lie in C_i. With this, the probability of a realisation $\boldsymbol{X}$ occurring in C can by found via the decomposition in equation 9. This approach of Heffernan and Tawn (2004) is capable of calculating probabilities when not all margins are extreme in the asymptotically independent case, unlike other methods which are only applicable when the elements are asymptotically dependent (Coles and Tawn, 1991; Joe et al., 1992), or when they are all extreme together (Ledford and Tawn, 1997; Bortot et al., 2000).

5.2 Subset Specific Levels of Dependence

Whilst the method of Heffernan and Tawn (2004) nicely overcomes the issues present in previous approaches, there remains a strong and potentially unrealistic assumption being made. To discuss this assumption, we must first extend formally the idea of asymptotic dependence to the multivariate case. We do this via the formulation given by Simpson et al. (2020), which they present as in definition 5.1.

Definition 5.1 (Multivariate Asymptotic Independence)**.** Consider a p-dimensional random vector $\boldsymbol{X}$.

Take c to be a non-trivial subset of the components $\{X_1, \ldots, X_p\}$. Define the measure

$$\chi_c = \lim_{u \to 1} \mathbb{P}\left(F_i(X_i) > u : i \in C\right) / (1 - u).$$

We say that the components contained in c are asymptotically independent if $\chi_c = 0$. Otherwise, we say they are asymptotically dependent.

The value χ_c is essentially the probability that, given a component $X_j \in c$ is large, all other components of $\boldsymbol{X}$ that are in c are large. This quantity χ_c can be linked nicely to the measure H discussed throughout §3.3. If $\chi_c > 0$, then H places mass on the region of the unit simplex S_p given by $\{\mathbf{w} \in S_p^{\bar{c}} : \sum_{w_j \in \bar{c}'} w_j = 1\}$, where $\bar{c}'$ is the set of the angular representations of the elements of some set $\bar{c} \supset c$. That is, $S_p^{\bar{c}}$ is the region for which only the marginal components of $\boldsymbol{X}$ that are in $\bar{c}$ have weight placed on them. For a p-dimensional random vector, there are $2^p - 1$ of these subspaces, each corresponding to a vertex, edge, or the interior of S_p. Figure 5 provides a simple illustration of these subspaces in the trivariate case.

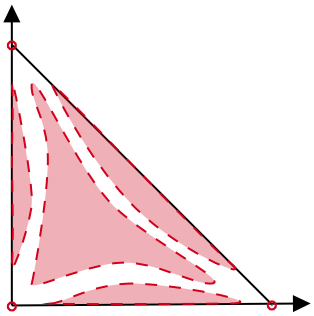

Figure 5: Illustration of Subspaces of S_3

The drawback of method discussed in §5.1 is that it can only be applied to cases for which χ_c is constant over all sets c. Therefore, if certain subsets of $\{X_1, \ldots, X_p\}$ are asymptotically dependent, but not all, we cannot utilise this approach. It is also noted by Simpson et al. (2020) that the measures χ_c do not fully describe the extremal dependence structure; for example, knowing $\chi_{\{X_1, X_2\}} > 0$ tells us mass is placed on at least one of the two subspaces for which these variables are both large simultaneously, but not which. The work of Simpson et al. (2020) aims to provide a way to obtain a more detailed estimation of extremal dependence. In short, they exploit the often satisfied assumption of multivariate regular variation to develop parameters which determine which subspaces of S_p have mass placed on them by H. Their method also possesses the desirable characteristic of being able to capture sub-asymptotic dependence structures, as well as those that are asymptotically dependent. In practice, their methodology often detects several likely dependence structures, rather than just one. Of course, this is still useful as it greatly reduces the problem from the situation where all combinations of subspaces are possible.

6 Contemporary and Open Problems

We now look to examples of work developed more recently than the majority of those discussed so far in this report. They have been selected to facilitate the discussion of open problems which are of particular interest.

6.1 Markov Chain Models

The modelling of time series of extreme events via Markov Chains has been receiving increased interest. The behaviour of heatwaves in France was modelled by Winter and Tawn (2016) using first order Markov Chains and building from the work of Heffernan and Tawn (2004). They devised joint models for the values of temperature following an observation deemed extreme by exceedance of a suitable threshold. As their methodology incorporates the ideas of Heffernan and Tawn (2004), it possesses the desirable quality of being applicable to cases of both asymptotic dependence or asymptotic independence. Comparison between the predicted and observed behaviour of clusters of extreme temperatures suggest theirs is a suitable approach. This work was then improved upon by Winter and Tawn (2017), who note the model of Winter and Tawn (2016) presents an oversimplification of the physical mechanisms behind heatwaves. Also building upon the work of Heffernan and Tawn (2004), they instead apply a kth-order Markov Chain model to values of a time series that are within k lags of a suitably high temperature. They develop novel diagnostics to determine the most appropriate choice of k, since standard tools such as the partial auto-correlation function (PACF) depend mostly on the main body of the distribution, and so may not reflect the extremal dependence structure if it differs from that of non-extreme data. The ideas of Winter and Tawn (2016, 2017), are utilised and extended by Tendijck et al. (2019) for the modelling of sea storms. They model a sea-state variable H_s, which summarises several physical features, via a 2nd-order Markov Chain, alongside a storm direction covariate, which is modelled as an autoregressive time series. Their work aims to facilitate accurate simulation of storm development given information about its peak at a single site.

The works discussed above present a rich foundation for a plethora of possible extensions to the literature. The ideas of Tendijck et al. (2019) could be extended to incorporate peak data across multiple sites, exploiting the natural spatial dependence of the phenomena to provide more powerful inference. As they themselves note, the number of covariates could be increased in an effort to capture a more realistic model of the ecosystem and thus storm evolution. Of course, these ideas would generalise suitably to be useful in other applications, such as in the modelling of heatwaves. In fact, the non-stationarity introduced through covariates could provide a nice way of accounting for trends in temperature resulting from global warming, something not achieved by Winter and Tawn (2016, 2017). Also, while Tendijck et al. (2019) identify the 2nd-order Markov Chain model as most fitting for their application, other settings may require development of higher order equivalents.

6.2 Environmental Contours and Sample Clouds

Environmental contours have been used extensively in the design of ocean structures such as ships, oil rigs and sea walls. They utilise multivariate extreme value techniques to capture regions of most likely observations of environmental variables, such as wind speed and wave height. Figure 6 provides a simple illustration of this idea in the bivariate case. Here, the majority of values fall into the solid contour which could, for example, represent the 100th return level of the variables. That is, we would expect a combination of variables outside the inner contour, coloured in red, to occur every 100th observation. The points marked in yellow thus represent the most extreme events that would be taken into design consideration, were we to only consider events that occur, on average, every 100 observations or less. The dashed contours encompass more combinations of variables, representing higher order return levels.

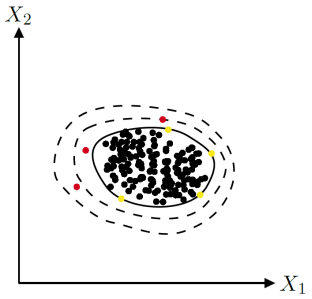

Figure 6: Illustration of Environmental Contours

There is contemporary interest in developing methodology to model environmental contours. A popular approach to contour estimation is the Inverse First Order Reliability Method (IFORM), first developed by Winterstein et al. (1993). Recently, alternatives have been suggested such as the Monte-Carlo based techniques of Bang Huseby et al. (2013) and Huseby et al. (2021), or the adaption of Qiao and Myers (2021) to the IFORM framework. A review of various other contour estimation methods is provided by Ross et al. (2020). We highlight in particular the work of Vanem et al. (2020), who describe a method for incorporating dependence on covariates into contour modelling via division of the sample into covariate 'bins'. This approach is applied to a biviariate case, leaving the extension to the multivariate case as an open problem. This would be a desirable direction to work in, as it could result in a framework for estimating contours in the multivariate setting, whilst also considering covariates.

6.3 Graphical Structure Models

Given the multivariate regular variation assumption mentioned in §5.2, we can study the distribution of the angular variables defined in equation 2 to determine the extremal dependence structure of a random vector. Interest in this approach has been shown by likes of Simpson et al. (2020) and Engelke and Hitz (2020). The latter of these employs a graphical model of threshold exceedances by exploiting the notion of conditional dependence, something not considered in other literature. This work was applied, with success, to river flow data. However, the framework developed by Engelke and Hitz (2020) does not allow

for the possibility of asymptotic independence, which was identified by Heffernan and Tawn (2004) as an important feature of extreme value models. A natural next step is thus the linking of the graphical ideas introduced by Engelke and Hitz (2020) to the concept of sub-asymptotic dependence structures that often arise when modelling extremes. Both of these approaches also perform poorly in higher dimensions due to the curse of dimensionality, and so require adapting to circumvent this issue. Alternatively, an altogether new approach that has been suggested is to combine the ideas of Engelke and Hitz (2020) with that of Haff et al. (2016). They present a learning algorithm for the structure of Bayesian networks, which have been used extensively to represent the dependence between variables.

7 Conclusion

We have attempted to provide an overview of the multivariate extremes literature via discussion of key works and their motivation. The works discussed are largely applicable, both theoretically and practically, motivating further development of related ideas and methodology. For this reason, multivariate extreme value theory remains an open and interesting area of research.

References

Bang Huseby, A., Vanem, E., and Natvig, B. (2013). A new approach to environmental contours for ocean engineering applications based on direct monte carlo simulations. *Ocean Engineering*, 60:124–135.

Barnett, V. (1976). The ordering of multivariate data. *Journal of the Royal Statistical Society: Series A (General)*, 139(3):318–344.

Basrak, B., Davis, R. A., and Mikosch, T. (2002). A characterization of multivariate regular variation. *The Annals of Applied Probability*, 12(3):908–920.

Bortot, P., Coles, S., and Tawn, J. (2000). The multivariate gaussian tail model: An application to oceanographic data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(1):31–049.

Coles, S. (2013). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer London, Limited, London.

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

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(2):377–392.

Engelke, S. and Hitz, A. S. (2020). Graphical models for extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(4):871–932.

Gumbel, E. J. (1958). Statistics of extremes. In *Statistics of Extremes*. Columbia university press.

Haan, L. F. M. (1985). *Extremes in higher dimensions: the model and some statistics*. Erasmus University.

Haff, I. H., Aas, K., Frigessi, A., and Lacal, V. (2016). Structure learning in bayesian networks using regular vines. *Computational Statistics & Data Analysis*, 101:186–208.

Hand, D. and Crowder, M. (2017). *Practical longitudinal data analysis*. Routledge.

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

Huseby, A. B., Vanem, E., Agrell, C., and Hafver, A. (2021). Convex environmental contours. *Ocean Engineering*, 235:109366.

Joe, H., Smith, R. L., and Weissman, I. (1992). Bivariate threshold methods for extremes. *Journal of the royal statistical society: series B (methodological)*, 54(1):171–183.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(2):475–499.

Pickands, J. (1981). Multivariate extreme value distribution. *Proceedings 43th, Session of International Statistical Institution, 1981*.

Qiao, C. and Myers, A. (2021). A new iform-rosenblatt framework for calculation of environmental contours. *Ocean Engineering*, 238:109622.

Ross, E., Astrup, O. C., Bitner-Gregersen, E., Bunn, N., Feld, G., Gouldby, B., Huseby, A., Liu, Y., Randell, D., Vanem, E., and Jonathan, P. (2020). On environmental contours for marine and coastal design. *Ocean Engineering*, 195:106194.

Sibuya, M. et al. (1960). Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, 11(2):195–210.

Simpson, E. S., Wadsworth, J. L., and Tawn, J. A. (2020). Determining the dependence structure of multivariate extremes. *Biometrika*, 107(3):513–532.

Smith, R., Tawn, J., and Yuen, H. (1990). Statistics of multivariate extremes. *International Statistical Review/Revue Internationale de Statistique*, pages 47–58.

Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika*, 75(3):397–415.

Tendijck, S., Ross, E., Randell, D., and Jonathan, P. (2019). A model for the directional evolution of severe ocean storms. *Environmetrics*, 30(1):e2541.

Tiago de Oliveira, J. (1959). Extremal distributions. *Univ. Lisboa Revista Fac. Ci. A.*, 7:219–227.

Vanem, E., Guo, B., Ross, E., and Jonathan, P. (2020). Comparing different contour methods with response-based methods for extreme ship response analysis. *Marine Structures*, 69:102680.

Winter, H. C. and Tawn, J. A. (2016). Modelling heatwaves in central france: a case-study in extremal dependence. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65(3):345–365.

Winter, H. C. and Tawn, J. A. (2017). kth-order markov extremal models for assessing heatwave risks. *Extremes*, 20(2):393–415.

Winterstein, S. R., Ude, T. C., Cornell, C. A., Bjerager, P., and Haver, S. (1993). Environmental parameters for extreme response: Inverse form with omission factors. In *Proc. 6th Int. Conf. on Structural Safety and Reliability, Innsbruck, Austria*.