4 Multivariate extremes

4.1 Introduction

In this section we consider the problems we face if we wish to model the extremal behaviour of two or more (dependent) processes simultaneously. There are several reasons why we may wish to do this:

- to model the extreme behaviour of a particular variable over several nearby locations (e.g. rainfall over a network of sites);
- to model the joint extremes of two or more different variables at a particular location (e.g. wind and rain at a site);
- to model the joint behaviour of extremes which occur as consecutive observations in a time–series (e.g. consecutive hourly maximum wind gusts during a storm).

All of these problems suggest fitting an appropriate limiting multivariate distribution to the relevant data. However, as we shall see, the derivation of such a multivariate distribution is not as easy as we might hope. The analogy with the Normal distribution as a model for means breaks down as we move into n dimensions! It is not even clear what the 'relevant data' should be! Most of the increased complexity is apparent in the move from 1 to 2 dimensions, so we will focus largely on bivariate problems.

4.2 Componentwise maxima models

4.2.1 Example: network of rainfall measurements

Suppose we want to study the joint extremes of daily rainfall accumulations at the network of 8 sites shown in Figure 14.

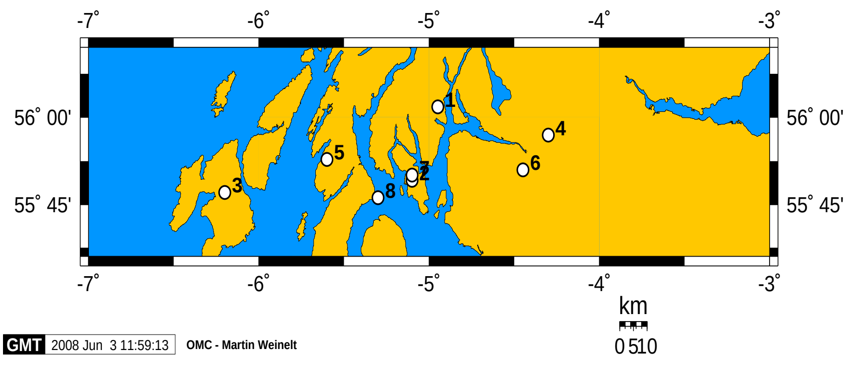

Figure 14: Eight rainfall recording stations in southern Scotland

Such issues are of great interest, especially currently, e.g. given the severe flooding experienced in the UK recently. Suppose we have sequences of daily total rainfall at each location. There is liable to be strong inter–site dependence in extremes, in the sense that days with heavy rain are liable to occur simultaneously across locations. The raw multivariate observations are 8–dimensional vectors of the daily rainfall over the eight sites.

Now suppose we wish to take a block–maxima approach, with 'blocks' being years. For any given year, the 8–dimensional vector of annual maxima is unlikely to be one of the raw multivariate observations. Let's simplify to the bivariate case. Let $(X_1, Y_1), (X_2, Y_2), \ldots$ be i.i.d. vectors with distribution function $F(x, y)$. Now consider the componentwise block maxima

$$M_{x,n} = \max_{i=1,\ldots,n} \{X_i\} \qquad \text{and} \qquad M_{y,n} = \max_{i=1,\ldots,n} \{Y_i\}.$$

We define the *vector of componentwise maxima* to be

$$\mathbf{M_n} = (M_{x,n}, M_{y,n}).$$

$\mathbf{M_n}$ is not necessarily one of the original observations (X_i, Y_i). Nevertheless, we are interested in the limiting behaviour of $\mathbf{M_n}$ as $n \to \infty$. The first point to note is that standard univariate extreme value results apply in each margin. When considering the dependence, this allows us to make a simplifying assumption.

We assume that the X_i and Y_i variables have a known marginal distribution. It is convenient to assume this is the GEV(0,1,1) distribution, also known as the unit Fréchet distribution, which has c.d.f.

$$F(z) = \exp(-1/z), \qquad z > 0.$$

This gives rise to a very simple normalization of maxima:

$$\Pr(X_i < x) = \Pr(M_{x,n}/n < x) = \exp(-1/x), \qquad x > 0,$$

(and similarly for Y_i). So if we consider the re–scaled vector

$$\mathbf{M_n^*} = \left(\max_{i=1,\ldots,n} \{X_i\}/n, \max_{i=1,\ldots,n} \{Y_i\}/n \right),$$

the margins are unit Fréchet for all n, and hence we can characterize the limiting joint behaviour of $\mathbf{M_n^*}$ without having to worry about the margins. Unfortunately no limiting parametric family exists! (for bivariate extremes, or multivariate extremes in general).

4.2.2 Theorem: limiting distributions for bivariate extremes

Let $\mathbf{M_n^*} = (M^*_{x,n}, M^*_{y,n})$ be the normalized maxima as above, where the (X_i, Y_i) are i.i.d. with standard Fréchet marginal distributions. Then if

$$\Pr(M^*_{x,n}, M^*_{y,n}) \to G(x, y),$$

where G is non–degenerate, then G has the form

$$G\left(x, y\right) = \exp\left\{-V\left(x, y\right)\right\}; \quad x > 0, \;\; y > 0 \tag{13}$$

where:

$$V\left(x, y\right) = 2 \int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) dH\left(\omega\right) \tag{14}$$

and H is a distribution function on $[0, 1]$ satisfying the mean constraint:

$$\int_0^1 \omega \; dH\left(\omega\right) = 0.5. \tag{15}$$

Hence the class of bivariate extreme value distributions is in one-to-one correspondence with distribution functions H satisfying the constraint (15). If H is differentiable with density h, then (14) becomes

$$V(x,y) = 2\int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) h(\omega)d\omega.$$

However some simple models arise when H is not differentiable. E.g. if H places mass 0.5 on each of $\omega = 0$ and $\omega = 1$, then we get

$$G(x,y) = \exp\{-(x^{-1} + y^{-1})\}, \qquad x > 0, y > 0,$$

corresponding to independent x and y.

Since the GEV provides the complete class of marginal limit distributions, then the complete class of bivariate extreme value distributions is obtained as follows. If we suppose X and Y are GEV with parameters (μ_x, σ_x, ξ_x) and (μ_y, σ_y, ξ_y) respectively, then the transformations

$$\tilde{x} = \left[1 + \xi_x\left(\frac{x-\mu_x}{\sigma_x}\right)\right]^{1/\xi_x} \quad \text{and} \quad \tilde{y} = \left[1 + \xi_y\left(\frac{y-\mu_y}{\sigma_y}\right)\right]^{1/\xi_y}$$

obtain unit Fréchet margins. Hence

$$G(x,y) = \exp\{-V(\tilde{x}, \tilde{y})\}$$

is a bivariate extreme value distribution with the appropriate margins for valid $V(.)$, and provided $[1 + \xi_x(x-\mu_x)/\sigma_x] > 0]$ and $[1 + \xi_y(x-\mu_y)/\sigma_y] > 0]$.

4.2.3 Modelling bivariate extremes in practice

In practice, modelling usually involves identifying a parametric sub–family with appropriate flexibility to handle the structure inherent in the data. Models can be fitted, e.g. by maximum–likelihood estimation, either in two steps (marginal components followed by dependence function), or in a single sweep. All of these procedures, including the choice of models, are handled in a very similar way when dealing with threshold exceedances. We consider the details in the next section.

4.3 Threshold excess models

We want to define our bivariate extremes in those observations which exceed a threshold in one or other margin. For our bivariate observation (X, Y), let's focus on X. We have already seen that the distribution function for the exceedances of a threshold u by a variable X, conditional on $X > u$ for large enough u, is given by:

$$G(x) = 1 - \lambda\left\{1 + \frac{\xi(x-u)}{\sigma}\right\}^{-1/\xi}$$

defined on $\{x - u : x - u > 0 \text{ and } (1 + \xi(x-u)/\sigma) > 0\}$, where $\xi \neq 0$, $\sigma > 0$, and $\lambda = Pr(X > u)$. Now we can obtain a unit Fréchet margin with the transformation:

$$\tilde{X} = -\left(\log\left\{1 - \lambda_x\left[1 + \frac{\xi_x(X-u_x)}{\sigma_x}\right]^{-1/\xi_x}\right\}\right)^{-1}.$$

If we apply the analogous transformation to in the Y margin, we obtain

$$\tilde{F}(\tilde{x}, \tilde{y}) = \exp\{-V(\tilde{x}, \tilde{y})\}; \quad x > u_x, \quad y > u_y,$$

where:

$$V(x, y) = 2\int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) dH(\omega)$$

and H is a distribution function on $[0, 1]$ satisfying the mean constraint:

$$\int_0^1 \omega \, dH(\omega) = 0.5.$$

4.3.1 Example: wave–surge data

Here we choose a different type of example of dependence to the rainfall problem considered in Section 4.2. Here we consider two variables recorded concurrently at the same site. A series of 3-hourly measurements on sea–surge were obtained from Newlyn, southwest England. For suitably high thresholds, we can identify which observations are extreme.

4.3.2 Threshold representation

Bivariate threshold models are complicated by the possibility that a bivariate pair (x, y) may be an 'exceedance' and yet exceed the specified threshold in only one of the two components.

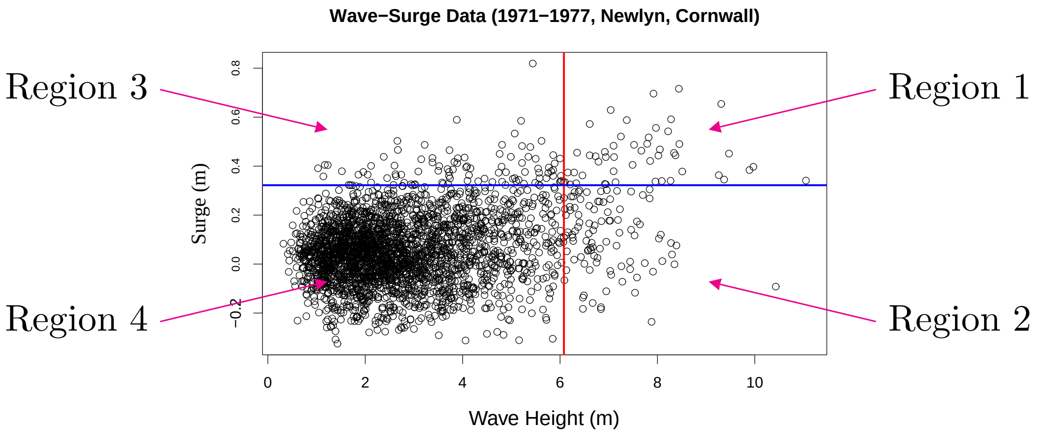

Figure 15: Threshold classification of bivariate data

4.3.3 Modelling the dependence structure

The class of bivariate extreme value models contains many families of distributions which can be used to model the dependence structure in the data. The dependence structure must satisfy the conditions on $H(\omega)$. Possible choices are:

- Logistic Model — symmetric
- Negative Logistic Model
- Bilogistic Model — asymmetric
- Dirichlet Model

Here we will focus on the logistic model and the bilogistic model as two commonly used but contrasting choices.

4.3.4 The Logistic model

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

where $x > 0$, $y > 0$ and $\alpha \in (0,1)$.

- $\alpha \to 1$ corresponds to independent variables.
- $\alpha \to 0$ corresponds to perfectly dependent variables.
- This model is symmetric — the variables are exchangeable.

4.3.5 The Bilogistic model

$$G(x,y) = \exp\left\{x\gamma^{1-\alpha} + y(1-\gamma)^{1-\beta}\right\}$$

where $0 < \alpha < 1$, $0 < \beta < 1$ and $\gamma = \gamma(x, y; \alpha, \beta)$ is the solution of:

$$(1-\alpha)\,x\,(1-\gamma)^{\beta} = (1-\beta)\,y\gamma^{\alpha}$$

- Independence is obtained when $\alpha = \beta \to 1$ and when one of α or β is fixed and the other approaches 1.
- When $\alpha = \beta$ the model reduces to the logistic model.
- The value of $\alpha - \beta$ determines the extent of asymmetry in the dependence structure.

4.3.6 Likelihood calculations

- For points in Region 1, the bivariate model structure shown applies, and the density of $\tilde{F}(\tilde{x}, \tilde{y})$ gives the appropriate likelihood component.
- In other regions, the likelihood component for the points must be censored.

4.3.7 The likelihood function

The likelihood function can be written as:

$$L(\theta; (x_1, y_1), \ldots, (x_n, y_n)) = \prod_{i=1}^{n} \psi(\theta; (x_i, y_i))$$

where θ gives the parameters of F and

$$\psi(\theta; (x,y)) = \begin{cases} \left.\frac{\partial^2 F}{\partial x \partial y}\right|_{(x,y)} & \text{if } (x,y) \in \text{Region 1} \\ \left.\frac{\partial F}{\partial x}\right|_{(x,u_y)} & \text{if } (x,y) \in \text{Region 2} \\ \left.\frac{\partial F}{\partial y}\right|_{(u_x,y)} & \text{if } (x,y) \in \text{Region 3} \\ F(u_x, u_y) & \text{if } (x,y) \in \text{Region 4} \end{cases}$$

The various models can be fitted to data by maximum likelihood estimation using routines available in the R package `evd`. We will explore this in the second R practical.

4.4 Point process representation

It helps our understanding of bivariate (and hence multivariate) extremes to think in terms of a point process model as follows. Let $(x_1, y_1), (x_2, y_2), \ldots$ be a sequence of independent bivariate observations form a distribution with standard Fréchet margins such that

$$\Pr\{M^*_{x,n} \le x, M^*_{y,n} \le y\} \to G(x, y).$$

Let N_n be a sequence of point processes defined by

$$N_n = \{(n^{-1}x_1, n^{-1}y_1), \ldots, (n^{-1}x_n, n^{-1}y_n)\}.$$

Then

$$N_n \to N$$

on regions bounded away from $(0, 0)$, where N is a non–homogeneous Poisson process on $(0, \infty) \times (0, \infty)$. Moreover, if we change our coordinates to an angular-radial form ('pseudo-polar') by setting

$$r = x \quad \text{and} \quad \omega = \frac{x}{x + y},$$

then the intensity function of N is

$$\lambda(r, \omega) = 2\frac{dH(\omega)}{r^2},$$

where H is related to G in the usual way (Equations (13) — (15)). This is helpful because r and ω are measures of distance (from the origin) and angle (from the x-axis) respectively, and the dependence function H determines the angular spread of points of N, *and is independent of radial distance*. If H is differentiable, then since ω measures the relative size of x to y in the pair (x, y), then $h(.)$ determines the density of events of different relative size. It is fairly easy now to picture what different densities $h(.)$ will look like it terms of the scatter of points in the limiting point process N.

4.4.1 The point process representation in practice

We assume the Poisson limit to be a reasonable approximation to N_n on an appropriate region. Convergence is guaranteed on any region bounded from the origin, and things are especially simple if we choose a region of from $A = \{(x, y) : x/n + y/n > r_0\}$ for suitably large r_0, since then

$$\Lambda(A) = 2\int_A \frac{dr}{r^2} dH(\omega) = 2\int_{r=r_0}^{\infty} \frac{dr}{r^2} \int_{\omega=0}^{1} dH(\omega) = 2/r_0,$$

which is constant with respect to the parameters of H. If we assume H has density h, then the likelihood is given by

$$\begin{aligned} L(\theta; (x_1, y_1), \ldots, (x_n, y_n)) &= \exp\{\Lambda(A)\} \prod_{i=1}^{N_A} \lambda(x_{(i)}/n, y_{(i)}/n) \\ &\propto \prod_{i=1}^{N_A} h(\omega_i), \end{aligned}$$

where $\omega_i = x_{(i)}/(x_{(i)} + y_{(i)})$ for the N_A points $(x_{(i)}, y_{(i)})$ which are in A. [This is based on assuming that we have already transformed the margins so that $(x_1, y_1), \ldots, (x_n, y_n)$ have standard Fréchet distributions.] Now we can fit the model using maximum–likelihood estimation.

4.4.2 Point process model for wave–surge data

A point process model was fitted to the wave–surge data after transformation to unit Fréchet margins, and using a threshold of the form $X + Y = r_0$, where r_0 was chosen so that the marginal thresholds are both at the 95th percentile. Fitting the two dependence models (logistic and bilogistic) to the wave–surge data we obtain the following results:

Model	log–lik.	α	β
Logistic	227.2	0.659 (0.013)	
Bilogistic	230.2	0.704 (0.024)	0.603 (0.032)

These results suggest a fairly weak, while clearly significant, dependence. The logistic and bilogistic models can be compared using a likelihood ratio test, and significant asymmetry is suggested. It is also possible to produce graphs of the fitted $h(\omega)$ functions, with the histograms of the empirical ω values super–imposed. Here we just show some dependence functions for the logistic model.

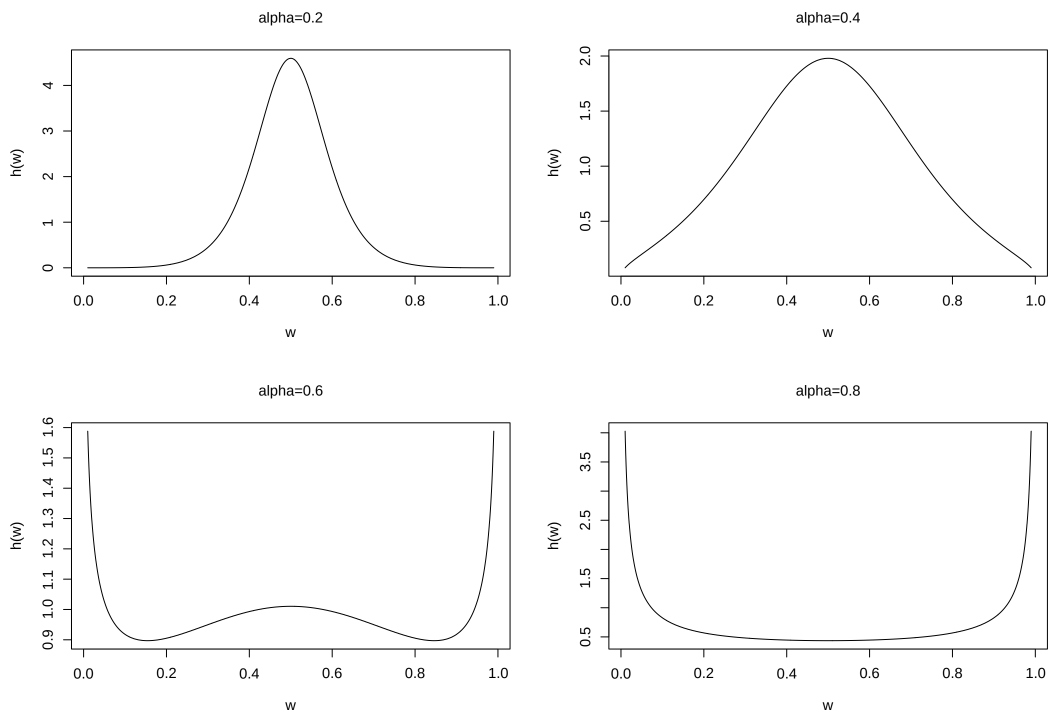

Figure 16: Some dependence functions for the logisitc model

4.5 Asymptotic dependence and independence

One key problem with using limit distributions for multivariate extremes is that they force one of two possibilities:

1. extremes occur independently in the different margins;
2. extremes occur with a dependence structure which conforms to an asymptotic extreme value distribution.

In practice this imposition is not helpful ... it is often the case that asymptotic independence is suggested by the data, and yet quite strong dependence is present, even at high levels. Data that seem to be dependent at ordinary levels may not necessarily be dependent in the limiting distribution. Consider the region $A = \left\{\frac{X}{n} > u, \frac{Y}{n} > v\right\}$. Then:

$$\Pr\left[\left(\frac{X}{n}, \frac{Y}{n}\right) \in A\right] = \begin{cases} C/n, & \textit{Asymptotic Dependence} \\ C/n^2, & \textit{Exact Independence} \end{cases}$$

where C is a constant term that does not depend on n.

4.5.1 The coefficient of tail dependence

Consider the variable:

$$T = \min\left(X, Y\right).$$

The distribution function of T is given by:

$$Pr\left(T \leq t\right) = 1 - \frac{K}{t^{1/\boldsymbol{\delta}}}, \qquad t > u,$$

where u is a threshold above which the data are regarded as extreme and K is a (almost) constant term with respect to t. $\boldsymbol{\delta}$ gives a measure of extremal dependence between X and Y and is known as the "**coefficient of tail dependence**".

4.5.2 Inference for $\boldsymbol{\delta}$

The likelihood function for T is:

$$L\left(K, \boldsymbol{\delta}; t\right) = \left(1 - \frac{K}{u^{1/\boldsymbol{\delta}}}\right)^{n-n_u} \left(\frac{K}{\boldsymbol{\delta}}\right)^{n_u} \prod_{i=1}^{n_u} t_i^{-(1+1/\boldsymbol{\delta})}$$

where n_u is the number of observations that satisfy $T > u$. Maximum likelihood estimation gives the estimate:

$$\hat{\boldsymbol{\delta}} = \frac{1}{n_u} \sum_{i=1}^{n_u} \log\left(\frac{t_i}{u}\right)$$

evaluated for the n_u points in the data set above u. $\boldsymbol{\delta}$ describes the limiting dependence structure:

- $\boldsymbol{\delta} = 1$ implies asymptotic dependence.
- $\frac{1}{2} < \boldsymbol{\delta} < 1$ implies positive association.
- $\boldsymbol{\delta} = \frac{1}{2}$ implies near independence.
- $0 < \boldsymbol{\delta} < \frac{1}{2}$ implies negative association.

4.5.3 Wave–surge data

Plots of $\hat{\delta}$ against increasing u give an indication of the level of dependence present between two processes in the limiting distribution.

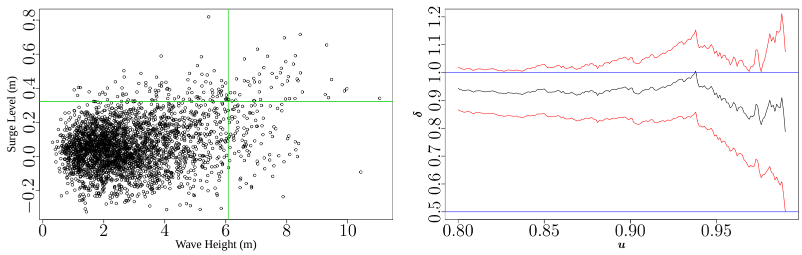

Figure 17: Wave-Surge data with 95% quantiles; $\boldsymbol{\delta}$–plot with 95% confidence bounds.

$\boldsymbol{\delta} = 1$ is within the 95% confidence bounds for all u as u increases, suggesting that wave–height and surge are **asymptotically independent**.

Research into modelling data in such instances, i.e. where there is still dependence within the 'extremes' in the data set, but yet asymptotic independence is suggested, is all fairly recent. The most prominent work is the article by Heffernan and Tawn (JRSS B, 2004). Here they develop semiparametric models based on assuming observations are extreme in at least one component, and then conditioning on this. This approach can be quite messy in implementation, combining as it does a range of different estimation procedures, and some *ad hoc* assumptions concerning the parametric forms of the key normalising constants. Here we briefly consider another approach, suggested by (Bortot *et al.*, 2000), and currently the subject of ongoing work by Atyeo and Walshaw.

4.5.4 The multivariate Gaussian tail model

The multivariate Gaussian tail model for the multivariate distribution function F is defined on the **joint tail region** (Bortot *et al.*, 2000):

$$R(\mathbf{u}) = (u_1, \infty) \times \ldots \times (u_p, \infty)$$

where $\mathbf{u} = (u_1, \ldots, u_p)$. (e.g. Region 1 in Figure 15). For each observation in the joint tail region $R(\mathbf{u})$ we transform each marginal observation to have a standard Normal marginal distribution, and then apply the p–dimensional standard Normal distribution function. We then transform *back* to extreme value margins. This provides a more realistic representation of the dependence, while retaining the asymptotic arguments for the marginal extremes.

We have been able to fit such models to the 8-dimensional rainfall problem associated with Figure 14, however inference for this problem was much simplified by adopting a Bayesian approach.