# Chapter 6

# Multivariate extremes

## 6.1 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 – simultaneous flooding at several locations could cause devastation);
- to model the joint extremes of two or more different variables at a particular location (e.g. wind and rain at a site – the combined effects of wind and rain during a hurricane can result in extreme storm surge);
- 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.

## 6.2 Componentwise maxima models

### 6.2.1 Example: Rainfall at 8 locations in Scotland

Suppose we want to study the joint extremes of daily rainfall accumulations at the network of 8 sites is southwest Scotland shown in Figure 6.1.

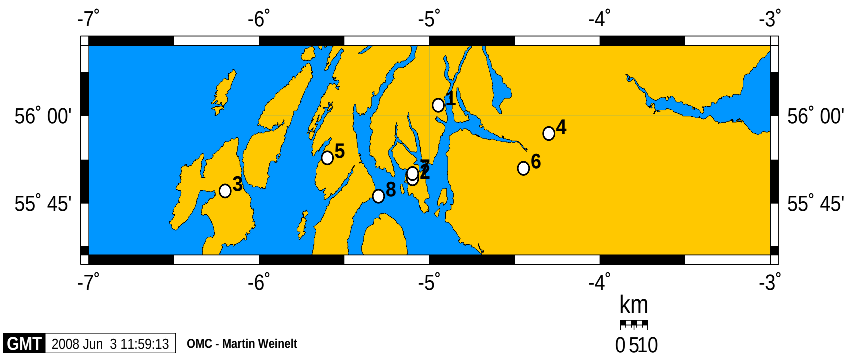


Figure 6.1: Eight rainfall recording stations in southwest Scotland.

Such issues are of great interest, especially currently – given the severe flooding experienced in the U.K. Suppose we have sequences of daily total rainfall at each location. There is likely 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 a unit Fréchet distribution (see Chapter 2), which has CDF

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

This gives rise to a very simple normalisation 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 characterise 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)!

### 6.2.2 Theorem: limiting distributions for bivariate extremes

Let $\mathbf{M_n^*} = (M_{x,n}^*, M_{y,n}^*)$ be the normalised 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

$$G\left(x,y\right) = \exp\left\{-V\left(x,y\right)\right\}; \quad x>0, \ \ y>0 \tag{6.1}$$

where:

$$V\left(x,y\right) = 2\int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) dH\left(\omega\right) \tag{6.2}$$

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

$$\int_0^1 \omega \; dH\left(\omega\right) = 0.5. \tag{6.3}$$

Hence the class of bivariate extreme value distributions is in one-to-one correspondence with distribution functions $H$ satisfying the constraint (6.3). If $H$ is differentiable with density $h$, then (6.1) becomes

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

However some simple models arise when $H$ is not differentiable. For example, 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 $(\mu_x, \sigma_x, \xi_x)$ and $(\mu_y, \sigma_y, \xi_y)$ 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 $V(.)$, provided $[1+\xi_x(x-\mu_x)/\sigma_x] > 0]$ and $[1+\xi_y(x-\mu_y)/\sigma_y] > 0]$.

## 6.3 Bivariate threshold excess models

### 6.3.1 Example: wave–surge data at Newlyn, Cornwall

Here, we choose a different type of example of dependence to the rainfall problem considered in Section 6.2.1. Specifically, 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, giving – at each time point – measurements of the wave height (in metres) and the surge (also in metres). Figure 6.2 below shows these two variables plotted against each other; this plot suggests a tendency for extremes of one variable to coincide with extremes of the other. This dependence could be important – the impact of an event that is simultaneously extreme in both variables is likely to be much greater than if extremes of either component occurred in isolation.

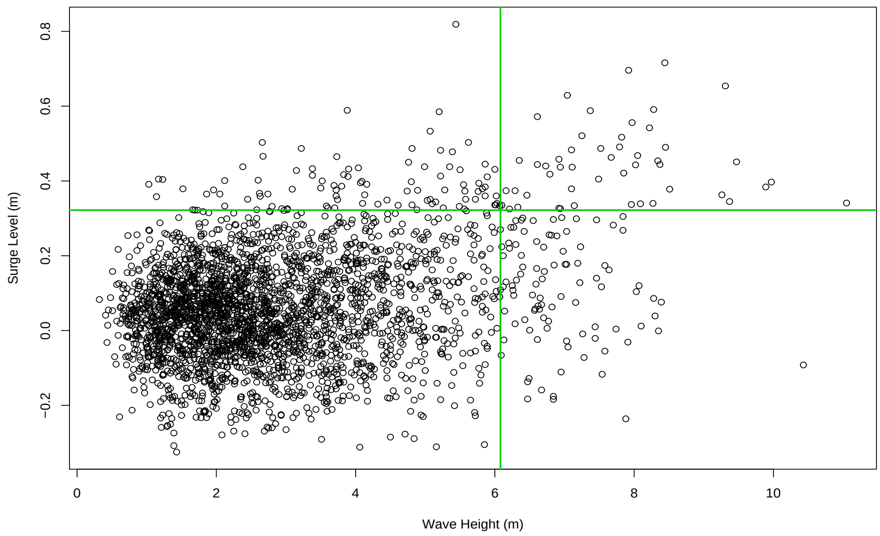


Figure 6.2: Concurrent wave and surge heights at Newlyn, Cornwall.

### 6.3.2 From componentwise maxima to threshold excesses

Now we want to define our bivariate extremes as 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:

$$H(x) = 1 - \lambda_{u_x} \left\{ 1 + \frac{\xi_x \left( x - u_x \right)}{\sigma_x} \right\}^{-1/\xi_x}$$

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

$$\tilde{x} = -\left(\log\left\{1 - \lambda_{u_x}\left[1 + \frac{\xi_x\,(x - u_x)}{\sigma_x}\right]^{-1/\xi_x}\right\}\right)^{-1}. \tag{6.4}$$

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

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

where $V(x, y)$ is as defined in Equation (6.1), again satisfying the mean constraint in Equation (6.3).

## 6.4 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 either bivariate componentwise maxima or bivariate threshold exceedances.

### 6.4.1 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.

**1. The Logistic model**

Here, for $V(x, y)$ in Equation (6.1), we have

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

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*.

**2. The Bilogistic model**

Now we have the following form for $V(x, y)$:

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

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

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

- Independence is obtained when $\alpha = \beta \to 1$ and when one of $\alpha$ or $\beta$ 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.

### 6.4.2 Likelihood calculations

After transformation to unit Fréchet margins, we can obtain the probability density function of the chosen dependence model by differentiation of Equation (6.1) to give $g(x, y)$; from this, the likelihood can be formed (and maximised) in the usual way. However, inference for the bivariate threshold excess setup is complicated by the fact that a bivariate pair may exceed a specified threshold in just one of its components. For example, consider Figure 6.3 below, an updated version of Figure 6.2 for the wave–surge data now including marginal thresholds (obtained by using mean residual life plots for both components). We obtain contributions to the likelihood function for a pair of observations in the following way, where $\theta$ represents the parameter(s) in our dependence model (e.g. $\theta = \alpha$ if we use the logistic model; $\theta = (\alpha, \beta)$ in the bilogistic model):

$$g(x, y; \theta) = \begin{cases} \left.\dfrac{\partial^2 G}{\partial x \partial y}\right|_{(\tilde{x},\tilde{y})} & \text{if } (x, y) \in \text{Region 1} \\ \left.\dfrac{\partial G}{\partial x}\right|_{(\tilde{x},\tilde{u}_y)} & \text{if } (x, y) \in \text{Region 2} \\ \left.\dfrac{\partial G}{\partial y}\right|_{(\tilde{u}_x,\tilde{y})} & \text{if } (x, y) \in \text{Region 3} \\ G(\tilde{u}_x, \tilde{u}_y) & \text{if } (x, y) \in \text{Region 4} \end{cases}$$

Then, we have

$$L(\theta; x, y) = \prod_{i=1}^{n} g(\tilde{x}_i, \tilde{y}_i).$$

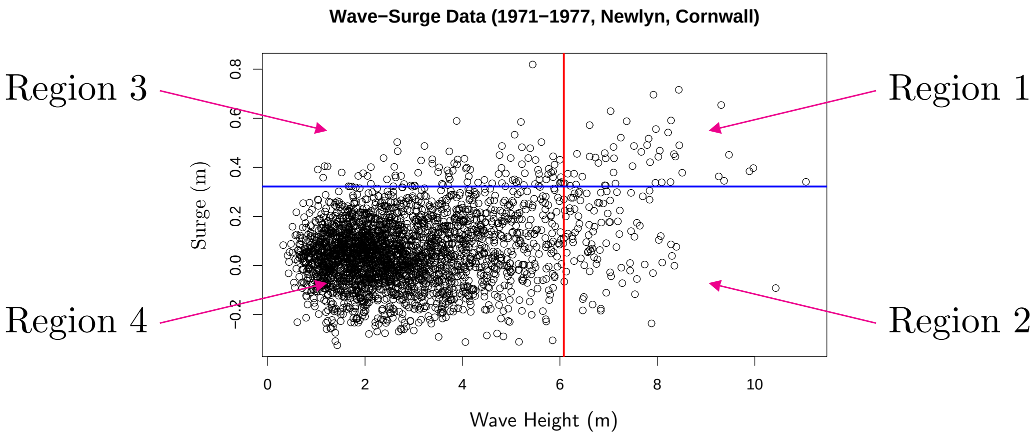


Figure 6.3: Threshold classification of bivariate data.

## Example 6.1: Wave–surge analysis at Newlyn

Consider the wave–surge data from Newlyn, Cornwall, shown in Figures 6.2 and 6.3. Flood defences in Newlyn have been designed to withstand a sea swell resulting from, at most, a wave height of $x = 9$ metres combined with a surge height of $y = 0.7$ metres.

A threshold–based approach to modelling is to be used for the wave–surge data shown in Figures 6.2 and 6.3. Mean residual life plots suggest marginal thresholds of $u_x = 6.1$ metres and $u_y = 0.32$ metres (as shown in Figure 6.2) for identifying wave height and surge height as extreme.

(a) Assuming extreme wave heights occur independently of extreme surge heights, find the probability that the flood defence system in Newlyn will be overwhelmed.

(b) Now assume there is extremal dependence between wave height and surge height.

 (i) Assuming the logistic model for this dependence, obtain the likelihood contributions to $L(\alpha; x_i, y_i)$ if (1) both $x > u_x$ and $y > u_y$; (2) only $x > u_x$; (3) only $y > u_y$; (4) neither $x$ nor $y$ exceed their marginal thresholds.

 (ii) Use the evd function fbvpot to fit the logistic model with likelihood contributions you identified in part (i). Show your estimated dependence parameter $\alpha$, with it's standard error.

 (iii) Using your fitted model in (ii), find the probability that the flood defence system in Newlyn will be overwhelmed. Compare your answer to that in part (a).

(c) Check for the presence of asymmetry in the dependence structure.

## Example 6.1: Solution

The Newlyn wavesurge data are available to download from the ismev package:

```
> library(ismev)
> data(wavesurge)
> wavesurge
   wave  surge
1  1.50 -0.009
2  1.83 -0.053
3  2.44 -0.024
4  1.68  0.000
5  1.49  0.079
6  1.20  0.068
7  1.35 -0.009
.     .      .
.     .      .
.     .      .
```

For convenience, we can “pull out” the wave and surge components:

```
> wave=wavesurge[,1]
> surge=wavesurge[,2]
```

(a) Thus, we can estimate $(\hat{\sigma}_x, \hat{\xi}_x)$ and $(\hat{\sigma}_y, \hat{\xi}_y)$ using `gpd.fit`. For example:

```
> gpd.fit(wave,6.1)
$threshold
[1] 6.1

$nexc
[1] 141

$conv
[1] 0

$nllh
[1] 155.1482

$mle
[1]  1.3335727 -0.1875704

$rate
[1] 0.04872149

$se
[1] 0.14084399 0.06565004
```

Using the same command on `surge`, i.e. `gpd.fit(surge, 0.32)`, we have

$$\hat{\lambda}_{u_x} = 0.049(0.004) \quad \hat{\sigma}_x = 1.334(0.141) \quad \hat{\xi}_x = -0.188(0.066)$$

and

$$\hat{\lambda}_{u_y} = 0.051(0.004) \quad \hat{\sigma}_y = 0.093(0.011) \quad \hat{\xi}_y = -0.041(0.080),$$

with standard errors in parentheses.

(b)

(i) For the logistic model, we have

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

**Region 1:**

If $x > u_x$ and $y > u_y$, then we are in Region 1 of Figure 6.3, and so

$$g(x,y;\alpha) = \frac{\partial^2 G}{\partial x \partial y}.$$

Differentiating w.r.t. $x$ gives

$$\begin{aligned}
\frac{\partial G}{\partial x} &= \exp\left\{-\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}\right\} \\
&\quad \times -\alpha\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha-1} \times (-1/\alpha)x^{-(1/\alpha+1)} \\
&= \exp\left\{-\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}\right\}\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha-1} x^{-(1/\alpha-1)}.
\end{aligned}$$

Differentiating the result w.r.t. $y$ then gives (after some algebra)

$$\begin{aligned}
\frac{\partial^2 G}{\partial x \partial y} &= (xy)^{-(1/\alpha+1)}\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha-2}\left[\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha} - (1-1/\alpha)\right] \\
&\quad \exp\left\{-\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}\right\} \\
&= (\tilde{x}\tilde{y})^{-(1/\alpha+1)}\left(\tilde{x}^{-1/\alpha} + \tilde{y}^{-1/\alpha}\right)^{\alpha-2}\left[\left(\tilde{x}^{-1/\alpha} + \tilde{y}^{-1/\alpha}\right)^{\alpha} - (1-1/\alpha)\right] \\
&\quad \exp\left\{-\left(\tilde{x}^{-1/\alpha} + \tilde{y}^{-1/\alpha}\right)^{\alpha}\right\},
\end{aligned}$$

evaluated at $x = \tilde{x}$ and $y = \tilde{y}$.

**Region 2:**

If only $x > u_x$, then we are in Region 2 of Figure 6.3, and so

$$g(x,y;\alpha) = \frac{\partial G}{\partial x}.$$

Thus, we have:

**Region 3:**

If only $y > u_y$, then we are in Region 3 of Figure 6.3, and so

$$g(x, y; \alpha) = \frac{\partial G}{\partial y}.$$

Thus, we have:

**Region 4:**

If both $x$ and $y$ are sub-threshold, then we are in Region 4 of Figure 6.3. Thus,

$$g(x, y; \alpha) = G(\tilde{u}_x, \tilde{u}_y) = \exp\left\{-\left(\tilde{u}_x^{-1/\alpha} + \tilde{u}_y\right)^{\alpha}\right\}.$$

(ii) For each pair of wave height/surge height measurements, we would now use the marginal GPD fits from part (a) to transform to unit Fréchet margins (via Equation 6.4); then the full likelihood

$$L(\alpha; x, y) = \prod_{i=1}^{2894} g(\tilde{x}_i, \tilde{y}_i)$$

can be maximised in the usual way to estimate the dependence parameter $\alpha$. However, we can make use of the `fbvpot` command in `evd`:

```
>  fbvpot(wavesurge, threshold=c(6.1,0.32), model='log')

Call: fbvpot(x = wavesurge, threshold = c(6.1, 0.32), model = "log")
Deviance: 2025.254
AIC: 2035.254
Dependence: 0.3023499

Threshold: 6.1 0.32
Marginal Number Above: 141 147
Marginal Proportion Above: 0.0487 0.0508
Number Above: 48
Proportion Above: 0.0166

Estimates
   scale1     shape1     scale2     shape2        dep
 1.265706  -0.139014   0.091818   0.006741   0.763539

Standard Errors
 scale1   shape1   scale2   shape2      dep
0.13280  0.06886  0.01052  0.08404  0.02933

Optimization Information
  Convergence: successful
  Function Evaluations: 37
  Gradient Evaluations: 8
```

Thus, we have

$$\hat{\lambda}_{u_x} = 0.049(0.004) \quad \hat{\sigma}_x = 1.266(0.133) \quad \hat{\xi}_x = -0.139(0.069)$$

$$\hat{\lambda}_{u_y} = 0.051(0.004) \quad \hat{\sigma}_y = 0.092(0.011) \quad \hat{\xi}_y = 0.007(0.084)$$

and

$$\hat{\alpha} = 0.764(0.029)$$

Notice that there are slight differences in the estimates of the marginal GPD parameters compared to those we found in part (a) (although these differences are not significant). This is because the full log–likelihood $\ell(\lambda_{u_x}, \lambda_{u_y}, \sigma_x, \sigma_y, \xi_x, \xi_y, \alpha)$ has been optimised in a single sweep for all of the parameters *simultaneously*.

Although our estimated dependence parameter does not indicate the presence of strong extremal dependence (recall that as $\alpha \to 1$ we have decreasing dependence), at least we have correctly accounted for this dependence.

(c) To check for asymmetry in the dependence structure, we can fit a bilogistic model in R and then compare this to the fitted logistic (as the logistic is *nested* within the bilogistic; when $\alpha = \beta$, the bilogistic reduces to the logistic).

```
> fbvpot(wavesurge,threshold=c(6.1,0.32),model="bilog")

Call: fbvpot(x = wavesurge, threshold = c(6.1, 0.32), model = "bilog")
Deviance: 2024.823
AIC: 2036.823
Dependence: 0.3032494

Threshold: 6.1 0.32
Marginal Number Above: 141 147
Marginal Proportion Above: 0.0487 0.0508
Number Above: 48
Proportion Above: 0.0166

Estimates
  scale1    shape1    scale2    shape2     alpha      beta
 1.28033  -0.14768   0.09074   0.01283   0.79655   0.71996

Standard Errors
 scale1   shape1   scale2   shape2    alpha     beta
0.13597  0.06955  0.01047  0.08315  0.05209  0.07883

Optimization Information
  Convergence: successful
  Function Evaluations: 48
  Gradient Evaluations: 10
```

Notice the similarity in the estimates of $\alpha$ and $\beta$; in fact, constructing confidence intervals for these two parameters gives:

$$\begin{aligned} \alpha &: (0.694, 0.899) \\ \beta &: (0.565, 0.874) \end{aligned}$$

Notice that these intervals overlap, suggesting that $\alpha \approx \beta$; thus, allowing for asymmetry in the dependence structure is probably not worthwhile.