Chapter 5

Modelling Componentwise Maxima

5.0 Introduction

The class of multivariate extreme-value distributions (2.7) was shown in Chapters 2 and 3 to be an asymptotically justified model for componentwise maxima from multivariate sequences. Estimation within this class is the focus of this chapter.

Let $X = (X_1, \ldots, X_D)$ be a vector random variable with distribution function G and component distribution functions G_d, $1 \leq d \leq D$. Suppose for now that the components are standard Fréchet: $G_d(x_d) = \exp(-1/x_d)$ when $x_d > 0$. Let $x = (x_1, \ldots, x_D)$ be a realisation of X with radial component $r = x_1 + \ldots + x_D$ and angular component $w = (w_1, \ldots, w_D)$, where $w_d = x_d/r$. Recall from Theorem 2.3

that G is a multivariate extreme-value distribution function if

$$G(x) = \exp\{-V(x)\} \quad \text{for all } x \in \mathbb{R}_+^D,$$

where

$$V(x) = \frac{1}{r} \int_{S_D} \max_{1 \le d \le D} \left\{ \frac{z_d}{w_d} \right\} \mathrm{d}H(z)$$

for some positive, finite measure H on the simplex

$$S_D = \{z = (z_1, \ldots, z_D) \in \mathbb{R}_+^D : z_1 + \ldots + z_D = 1\}$$

for which

$$\int_{S_D} z_d \, \mathrm{d}H(z) = 1 \quad \text{for } 1 \le d \le D.$$

In this chapter, only the bivariate case, $D = 2$, is considered. Because H is restricted to the unit simplex,

$$V(x) = \frac{1}{r} \int_0^1 \max\left\{ \frac{z}{w}, \frac{1-z}{1-w} \right\} \mathrm{d}H(z),$$

where $w = w_1$ and $z = z_1$ lie in $[0, 1]$, and H is now a positive, finite measure on $[0, 1]$ satisfying

$$\int_0^1 z \, \mathrm{d}H(z) = \int_0^1 (1-z) \, \mathrm{d}H(z) = 1. \tag{5.1}$$

It will be convenient to write the distribution function as

$$G(x) = \exp\left\{ -\frac{A(w)}{rw(1-w)} \right\}, \tag{5.2}$$

where A is called the dependence function and satisfies the integral relationship

$$A(w) = \int_0^1 \max\{z(1-w), w(1-z)\}\, \mathrm{d}H(z). \tag{5.3}$$

Attention will also be confined to differentiable distributions, for which G has density

$$g(x) = r^{-4}\{C(w) + rA''(w)\} \exp\left\{-\frac{A(w)}{rw(1-w)}\right\}, \tag{5.4}$$

where $C(w) = \{A(w) - wA'(w)\}\{A(w) + (1-w)A'(w)\}/\{w^2(1-w)^2\}$, and the derivative of H exists on $(0,1)$. In the differentiable case, the constraints implied by (5.1) and (5.3) can be written (Tawn, 1988) explicitly in terms of A:

$$\begin{gathered} A(0) = A(1) = 1, \\ -1 \le A'(0) \le 0 \le A'(1) \le 1, \\ A''(w) \ge 0 \quad \text{for all } 0 < w < 1. \end{gathered} \tag{5.5}$$

The graphical interpretation of these constraints is that A is convex, must lie within the triangle with vertices (0,1), (1,1) and $(\frac{1}{2}, \frac{1}{2})$, and is fixed at the first two vertices. The two bounding cases, $A(w) \equiv 1$ and $A(w) = \max\{w, 1-w\}$, correspond to independence and complete dependence between the components.

Let $\{(r_i, w_i)\}_{i=1}^n$ be a sample of independent realisations of the vector random variable (R, W), where $R = X_1 + X_2$, $W = X_1/R$ and (X_1, X_2) has bivariate extreme-value distribution (5.2). Estimation for the dependence function A currently falls into two camps: parametric and non-parametric methods, which will be reviewed in Sections 5.1 and 5.2. A semi-parametric approach that attempts to overcome the deficiencies of the parametric and non-parametric estimators is described in Section 5.3. The performances of several of the estimators are compared with a simulation study in Section 5.4 and their application is illustrated with a data example in Section 5.5.

5.1 Parametric Estimators

One approach to estimating the dependence function, A, is to choose a differentiable, parametric model that satisfies the constraints (5.5) and then fit the model by maximum likelihood. This approach has a number of benefits. Foremost, the resulting estimate of G is guaranteed to be a valid extreme-value distribution function. An estimate of the density is also obtained so that the likelihood is available. Joint estimation of the dependence structure and the parameters of the component distributions is therefore possible, which can have considerable benefits, as found, for example, by Barão and Tawn (1999). This also allows uncertainty in component parameters to be incorporated into confidence statements about G; covariate information can be included in the model; and likelihood-based tests of model features (such as independence between components or equality of component shape parameters) can be performed.

The disadvantage of the parametric approach is that it restricts attention to only a subset of the possible dependence functions. Therefore, sufficiently flexible models must be used. The construction of flexible models that satisfy the constraints (5.5) is a difficult task, particularly in higher dimensions, and, while there are now quite a few parametric models available, there is still a requirement for more flexibility, as found by Joe *et al.* (1992).

One parametric model is the logistic model, introduced by Gumbel (1960) and extended by Tawn (1988). The model for the dependence function is

$$A(w) = \{(1-w)^{1/\alpha} + w^{1/\alpha}\}^{\alpha} \tag{5.6}$$

for a dependence parameter $\alpha \in (0, 1]$. The model reaches independence when $\alpha = 1$ and approaches complete dependence as $\alpha \to 0$. The logistic model is not a particularly flexible class however: the dependence function is symmetric and $A'(1) = -A'(0) = 1$.

Another model is the Dirichlet model, introduced by Coles and Tawn (1991). The model for the dependence function is

$$A(w) = (1-w)\{1 - \mathrm{Be}(\alpha_1 + 1, \alpha_2; z)\} + w\mathrm{Be}(\alpha_1, \alpha_2 + 1; z) \tag{5.7}$$

for positive dependence parameters α_1 and α_2, where $z = \alpha_1 w / \{\alpha_1 w + \alpha_2(1-w)\}$ and

$$\mathrm{Be}(a, b; z) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^z y^{a-1}(1-y)^{b-1}\,\mathrm{d}y$$

is the incomplete Beta function. The model is asymmetric unless $\alpha_1 = \alpha_2 = \alpha$, in which case the model reaches independence at $\alpha = 0$ and approaches complete dependence as $\alpha \to \infty$. Again, $A'(1) = -A'(0) = 1$.

Other parametric models include the negative logistic (Galambos, 1975), asymmetric logistic (Tawn, 1988), Gaussian (Hüsler and Reiss, 1989), negative asymmetric logistic (Joe, 1989), bilogistic (Joe *et al.*, 1992) and negative bilogistic (Coles and Tawn, 1994). All of these are listed in Kotz and Nadarajah (2000).

5.2 Non-parametric Estimators

In order to avoid the restrictions imposed by parametric models for the dependence function, non-parametric estimators have been developed. Unfortunately, this approach has a number of drawbacks. It is particularly difficult to ensure that non-parametric estimates of A satisfy the constraints (5.5), so that crude adjustments are typically required. The estimators are often not differentiable on the whole interval $[0, 1]$; consequently, G contains point masses, the likelihood has degeneracies and many of the benefits of the parametric approach are out of reach. A good non-parametric estimate of A can, nevertheless, be a valuable guide to selecting a suitable parametric model. With this motivation, two non-parametric estimators are now reviewed.

The first is introduced by Capéraà *et al.* (1997). Making the change of variables $(x_1, x_2) \to (r, w)$, the density (5.4) becomes

$$g_{R,W}(r, w) = r^{-3}\{C(w) + rA''(w)\} \exp\left\{-\frac{A(w)}{rw(1-w)}\right\}$$

and it follows that the distribution function of the angular component is

$$G_W(w) = w + w(1-w)A'(w)/A(w). \tag{5.8}$$

Capéraà *et al.* (1997) use this to derive an estimator based on the empirical distribution function, $\hat{G}_W$, of the angular components, $\{w_i\}_{i=1}^n$, specifically

$$A_c(w) = \exp\left\{(1-w)\int_0^w \frac{\hat{G}_W(z)-z}{z(1-z)}\,\mathrm{d}z - w\int_w^1 \frac{\hat{G}_W(z)-z}{z(1-z)}\,\mathrm{d}z\right\}. \tag{5.9}$$

While this estimator does satisfy some of the conditions (5.5), it is not necessarily convex, nor does it necessarily lie above $\max\{w, 1-w\}$. A valid estimator can be obtained by enforcing the constraints artificially, that is by taking the convex hull of $\max\{A_c(w), w, 1-w\}$. Note, however, that taking the convex hull pulls the estimator towards stronger dependence.

The second non-parametric estimator is a modification proposed by Hall and Tajvidi (2000) of the estimator of Pickands (1981). Pickands uses the fact that the quantity $1/\max\{(1-w)X_1, wX_2\}$ has an exponential distribution with mean $1/A(w)$ to motivate the piecewise-linear estimator

$$n\left[\sum_{i=1}^n \min\left\{\frac{1}{(1-w)x_{i1}}, \frac{1}{wx_{i2}}\right\}\right]^{-1}. \tag{5.10}$$

The asymptotic properties of this estimator, which is also the maximum-likelihood estimator for $A(w)$ when w is fixed, are investigated by Deheuvels (1991). The estimator is not guaranteed to satisfy any of the constraints (5.5) however. Hall and Tajvidi (2000) propose a normalised version,

$$A_p(w) = n\left[\sum_{i=1}^{n} \min\left\{\frac{\bar{x}_1}{(1-w)x_{i1}}, \frac{\bar{x}_2}{wx_{i2}}\right\}\right]^{-1}, \tag{5.11}$$

where $\bar{x}_d = n/\sum_{i=1}^{n} x_{id}^{-1}$. This ensures that $A(0) = A(1) = 1$, as noted by Deheuvels (1985). Again, a valid estimator is obtained by taking the convex hull of $\max\{A_p(w), w, 1-w\}$.

These estimates of the dependence function are not differentiable on the whole of $[0,1]$. Two attempts have been made at smoothing that yield differentiable estimates. The first is proposed by Hall and Tajvidi (2000), who apply spline smoothing to both A_c and A_p. By imposing constraints on the spline functions they also ensure that the resulting estimates are valid dependence functions. The second approach is proposed by Smith *et al.* (1990), who smooth Pickands's estimator (5.10) with a kernel function. The idea of kernel smoothing will appear again in the following section.

5.3 Semi-parametric Estimators

In this section, a new, semi-parametric approach to the problem of estimating the dependence function is introduced. The approach attempts to combine the flexibility of non-parametric methods with the coherency of parametric models. The estimator yields estimates of A and its first two derivatives that are self-consistent and satisfy naturally the criteria (5.5) of a dependence function. This makes the likelihood available. The estimator is also flexible and differentiable as it is constructed using kernel density estimation. Of the current non-parametric methods, only the constrained spline-smoothing approach of Hall and Tajvidi (2000) enjoys the same benefits.

5.3.1 Asymptotic motivation

Recall that, for differentiable models, the measure H has a density, h, on $(0,1)$. Point masses, $0 \leq \phi_0 \leq 1$ and $0 \leq \phi_1 \leq 1$, can also be admitted at zero and one (Smith, 1985). In this case, the integral relationship (5.3) can be rewritten as

$$A(w) = w\phi_0 + (1-w)\phi_1 + \int_{(0,1)} \max\{z(1-w), w(1-z)\} h(z)\,\mathrm{d}z. \tag{5.12}$$

Differentiating yields

$$A'(w) = \phi_0 - \phi_1 + \int_{(0,w)} (1-z)h(z)\,\mathrm{d}z - \int_{(w,1)} zh(z)\,\mathrm{d}z \tag{5.13}$$

and

$$A''(w) = h(w). \tag{5.14}$$

Note that $A'(0) = -1 + \phi_0$ and $A'(1) = 1 - \phi_1$ so that allowing non-zero point-masses at zero and one increases the flexibility of A. The moment constraints (5.1) can also be rewritten as

$$\phi_0 + \int_{(0,1)} (1-z)h(z)\,\mathrm{d}z = \phi_1 + \int_{(0,1)} zh(z)\,\mathrm{d}z = 1. \tag{5.15}$$

The plan is to estimate ϕ_0, ϕ_1 and h under constraint (5.15). Substituting the estimates into equations (5.12) – (5.14) will yield estimates of A, A' and A'' that automatically define a valid estimate of the extreme-value density (5.4). This approach contrasts with the non-parametric estimators encountered in Section 5.2, which estimate A directly, and has the advantage that it is relatively simple to ensure that an estimator for h satisfies the appropriate conditions. This is likely to be particularly useful in higher dimensions.

To see how h might be estimated, consider the distribution of the angular component, W, conditional on a large radial component, R. By Theorem 3 of de Haan and Resnick (1977),

$$\lim_{r\to\infty} P(W \leq w \mid R > r) = \frac{H\{(0,w)\}}{H\{(0,1)\}} \quad \text{for } w \in (0,1)$$

as long as H has no mass at zero or one, that is $\phi_0 = \phi_1 = 0$. The main assumption in what follows is that the limit holds for the density,

$$q(w) = \lim_{r\to\infty} \frac{\mathrm{d}}{\mathrm{d}w} P(W \leq w \mid R > r) = \frac{h(w)}{H\{(0,1)\}} \quad \text{for } w \in (0,1), \tag{5.16}$$

and when H has point masses at zero and one. If this is the case then information about h will be found in the angular component of observations with a large radial component.

5.3.2 Kernel estimator

Rewriting the limit (5.16) yields

$$h(w) = \phi q(w) \quad \text{for } w \in (0,1),$$

where $\phi = H\{(0,1)\} = (1-\phi_0)+(1-\phi_1)$ by the moment conditions (5.15). Suppose for now that the values of ϕ_0 and ϕ_1 are known. When $\phi \neq 0$, the problem reduces to estimating the limiting angular density, q, subject to the constraints

$$\int_{(0,1)} q(z)\,\mathrm{d}z = 1 \quad \text{and} \quad \int_{(0,1)} zq(z)\,\mathrm{d}z = \psi, \tag{5.17}$$

where $\psi = (1-\phi_1)/\phi$. If $\phi = 0$ then q is arbitrary.

The shape of q influences the shapes of A and its derivatives on $(0,1)$. Parametric models force a structure on this shape that can be quite restrictive, particularly if a symmetric model is chosen. Any restriction can be avoided by estimating q non-parametrically.

First choose a radial threshold above which the limit (5.16) is believed to hold approximately. For some $1 \le m \le n$, the threshold used here is r_m, the m-th largest of the ordered radial components, $r_1 > \ldots > r_n$. Re-label $w_1, \ldots, w_n$ so that w_i is the angle that occured with r_i. A kernel density estimate of q is

$$q_n(w) = \frac{1}{m\lambda} \sum_{i=1}^{m} \tilde{k}\left(\frac{w - w_i}{\lambda}\right), \tag{5.18}$$

based on the angular components of the m largest observations. Define the kernel, $\tilde{k}$, to be

$$\tilde{k}\left(\frac{w-z}{\lambda}\right) = k\left(\frac{w-z}{\lambda}\right) + k\left(\frac{w+z}{\lambda}\right) + k\left(\frac{w-2+z}{\lambda}\right),$$

where k is the standard Gaussian density and λ is a smoothing parameter. This choice of kernel compensates for the mass that would spill over the boundaries of $(0,1)$ if a simple kernel such as k were used instead.

The estimator (5.18) is not constrained to have the required mean (5.17), but this can be achieved with a simple power transformation:

$$q_n^*(w) = \frac{1}{m\lambda}\sum_{i=1}^{m} \tilde{k}\left(\frac{w-w_i^\kappa}{\lambda}\right),$$

where $\kappa \geq 0$ solves

$$\int_{(0,1)} z q_n^*(z)\,\mathrm{d}z = \psi. \tag{5.19}$$

Note that the more common approach (Hall and Presnell, 1999) of modifying the weights $1/m$ rather than the angles w_i is not applicable here since the mean ψ could be outside the range of the w_i. An alternative would be to modify both the weights and angles, an example of which is the generalisation of Theorem 2 of Coles and Tawn (1991) described in Section 3.7.1 of Coles (1991), but this option is not considered here.

An estimator has been defined for q given ϕ_0, ϕ_1, m and λ. The four parameters could be chosen to maximise the likelihood

$$L(\phi_0, \phi_1, m, \lambda) = \prod_{i=1}^{n} g_n(x_i),$$

where g_n is obtained from (5.4) by replacing A, A' and A'' with the estimators

$$\begin{aligned} A''_n(w) &= \phi q^*_n(w), \\ A'_n(w) &= \phi_0 - \phi_1 + \phi \int_{(0,w)} (1-z) q^*_n(z)\,\mathrm{d}z - \phi \int_{(w,1)} z q^*_n(z)\,\mathrm{d}z, \\ A_n(w) &= w\phi_0 + (1-w)\phi_1 + \phi \int_{(0,1)} \max\{w(1-z), z(1-w)\} q^*_n(z)\,\mathrm{d}z. \end{aligned} \tag{5.20}$$

Unfortunately, the likelihood has a singularity when λ equals zero. To see this, consider m fixed and note that, as $\lambda \to 0$, $A''_n(w^\kappa_i) \to \infty$ for each $1 \leq i \leq m$. The likelihood contains the terms $A''_n(w_i)$, $1 \leq i \leq n$. Therefore, ϕ_0 and ϕ_1 will be chosen so that $\psi = \bar{w}$, the mean of $w_1, \ldots, w_m$, since then $\kappa \to 1$ as $\lambda \to 0$. Instead of estimating all four parameters with the likelihood, only ϕ_0, ϕ_1 and m will be estimated; λ is fixed as a function of m using the over-smoothing formula (Wand and Jones, 1995, page 61) for a standard Gaussian kernel, that is

$$\lambda = s \left(\frac{243}{70 m \pi^{3/2}} \right)^{1/5}, \tag{5.21}$$

where s is an estimate of the standard deviation of q. When an automatic choice of smoothing parameter is required, s will be set equal to the median of $s_1, \ldots, s_n$, where, for $1 \leq m \leq n$,

$$s_m^2 = m^{-1} \sum_{i=1}^{m} (w_i - \bar{w}_m)^2 \quad \text{and} \quad \bar{w}_m = m^{-1} \sum_{i=1}^{m} w_i. \tag{5.22}$$

A more subjective choice is made in the data example of Section 5.5.

By fixing λ in this way, the mean of q_n^* is constrained to lie in the interval $(\psi_\infty, \psi_0]$, where

$$\psi_\infty = \frac{1}{\lambda} \int_{(0,1)} z\tilde{k}\left(\frac{z}{\lambda}\right) \mathrm{d}z \quad \text{and} \quad \psi_0 = \frac{1}{\lambda} \int_{(0,1)} z\tilde{k}\left(\frac{z-1}{\lambda}\right) \mathrm{d}z.$$

Through (5.19) and (5.17), this imposes an additional linear constraint,

$$\frac{\psi_\infty}{1-\psi_\infty} < \frac{1-\phi_1}{1-\phi_0} \leq \frac{\psi_0}{1-\psi_0},$$

on ϕ_0 and ϕ_1.

Note finally that, because the full likelihood is used, all of the data contribute to the estimates of ϕ_0, ϕ_1, m and, through these, the dependence function and its derivatives.

5.3.3 Symmetric kernel estimator

A symmetric version of the estimator can also be defined by setting $\phi_0 = \phi_1$, $\kappa = 1$ and

$$q_n^*(w) = \frac{1}{2m\lambda} \sum_{i=1}^{m} \left\{ \tilde{k}\left(\frac{w-w_i}{\lambda}\right) + \tilde{k}\left(\frac{w-1+w_i}{\lambda}\right) \right\}.$$

The only constraints now are that ϕ_0 and ϕ_1 lie in $[0,1]$. The smoothing parameter becomes

$$\lambda = s\left(\frac{243}{140m\pi^{3/2}}\right)^{1/5}$$

and $s_m^2 = m^{-1}\sum_{i=1}^{m}(w_i - 1/2)^2$ is used to guide selection of s. The performance of this symmetric estimator will be compared with the unrestricted estimator in the simulation study of Section 5.4.

5.3.4 Unknown component distributions

Although the component distributions have been assumed to be standard Fréchet, this is not necessary and the maximum likelihood estimates of the component parameters can be found simultaneously. Suppose that the d-th component, here represented by $\tilde{X}_d$, has generalised extreme-value distribution with parameters (μ_d, σ_d, ξ_d). Then, transforming via (2.6) to standard Fréchet components

$$X_d = \left\{1 + \xi_d\left(\frac{\tilde{X}_d - \mu_d}{\sigma_d}\right)\right\}^{1/\xi_d}$$

yields the likelihood required for joint estimation. The terms in the likelihood are of the form

$$g(x)\frac{x_1^{1-\xi_1}x_2^{1-\xi_2}}{\sigma_1\sigma_2},$$

where the component parameters, (μ_1, σ_1, ξ_1) and (μ_2, σ_2, ξ_2), are constrained so that x_1 and x_2 are positive.

5.3.5 Asymptotic properties

Asymptotic properties for the kernel estimators described in the previous subsections have not been found, but uniform consistency can be shown for simplified estimators in a restricted setting. Define $g_{W|R}(w \mid r)$ to be the density of $P(W \leq w \mid R > r)$ and suppose that the following conditions hold:

$$\phi_0 = \phi_1 = 0, \tag{5.23}$$

$$\lim_{w \to 0} q(w) = \lim_{w \to 1} q(w) = 0, \tag{5.24}$$

$$\lim_{r \to \infty} \sup_{w \in [0,1]} |g_{W|R}(w \mid r)| < \infty, \tag{5.25}$$

$$\lim_{r \to \infty} \sup_{w \in [0,1]} |g'_{W|R}(w \mid r)| < \infty. \tag{5.26}$$

Consider the estimators (5.20) for A, A' and A'' when the parameters ϕ_0 and ϕ_1 are known to be zero and the kernel density estimator q_n^* is replaced with q_n, the simpler version (5.18) that does not incorporate the power transformation.

For the density, $g_{W|R}(\cdot \mid r_m)$, estimated by $q_n(\cdot)$ to approach $q(\cdot)$, it is necessary that r_m approaches infinity as the sample size n increases. For consistency, it is also necessary that $m \to \infty$, $\lambda \to 0$ and $m\lambda^2 \to \infty$ as $n \to \infty$ (Parzen, 1962). Recall that $\lambda \propto m^{-1/5}$, so this is achieved if $m = m_n = o(n)$ and $m_n \to \infty$ as $n \to \infty$.

Uniform consistency of A_n'' follows from

$$
\begin{aligned}
E \sup_{w\in[0,1]} |A_n''(w) - A''(w)| \;&\le\; 2E \sup_{w\in[0,1]} |q_n(w) - g_{W|R}(w \mid r_m)| \\
&\quad + 2 \sup_{w\in[0,1]} |g_{W|R}(w \mid r_m) - q(w)|.
\end{aligned}
$$

The first term approaches zero by Theorem 3A of Parzen (1962), where conditions (5.25) and (5.26) are required instead of his uniform continuity conditions because the density being estimated changes with n. The second term tends to zero under conditions (5.23) and (5.24) by Theorem 1 of Nadarajah (2000). Uniform consistency of A_n follows from

$$
\begin{aligned}
|A_n(w) - A(w)| \;&\le\; w \int_0^w (1-z)|A_n''(z) - A''(z)|\,\mathrm{d}z \\
&\quad + (1-w)\int_w^1 z|A_n''(z) - A''(z)|\,\mathrm{d}z \\
&\le\; \sup_{z\in[0,1]} |A_n''(z) - A''(z)|,
\end{aligned}
$$

and similarly for A_n'.

The conditions under which these results hold are strong: for example, condition (5.24) holds for the logistic model (5.6) only when $\alpha < 1/2$, and for the Dirichlet model (5.7) only when $\alpha_1 > 1$ and $\alpha_2 > 1$. It will be seen in the following section, however, that the estimators perform well for finite sample sizes even when conditions (5.23) – (5.26) are not satisfied.

5.4 Simulation Study

The performances of the parametric, non-parametric and semi-parametric estimators of the three previous sections are compared here for small samples with a simulation study. To simulate data from a bivariate extreme-value distribution with arbitrary dependence function, A, and standard Fréchet margins, the following algorithm (Ghoudi *et al.*, 1998) is employed. First, generate a uniform(0,1) variate, u, and then generate the angular component from distribution (5.8) by solving $G_W(w) = u$ for w. To generate the radial component, use the fact that the conditional density, $g_{R|W}(\cdot \mid W = w)$, is a mixture of two inverse Gamma densities:

$$g_{R|W}(r \mid W = w) = \pi p(r\,;\,1, \beta) + (1 - \pi)p(r\,;\,2, \beta),$$

where $\beta = A(w)/w(1-w)$, $\pi = [1+w(1-w)C(w)/\{A(w)A''(w)\}]^{-1}$ and $p(\cdot\,;\,a, b)$ is an inverse Gamma density with parameters a and b.

Data are simulated from two bivariate extreme-value distributions: one with symmetric, logistic dependence (5.6) and another with asymmetric, Dirichlet dependence (5.7). For each model, three different sets of parameter values are considered: for the logistic model, $\alpha = 0.25, 0.5$ and 0.75; for the Dirichlet model, $(\alpha_1, \alpha_2) = (1, 9), (\frac{1}{3}, 9)$ and $(\frac{1}{9}, 1)$. These parameter values are listed in order of weakening asymptotic dependence; the shapes of the true dependence functions and their second derivatives are shown in Figures 5.1 and 5.2.

For each of the six dependence structures, samples of sizes 25, 50 and 100 are

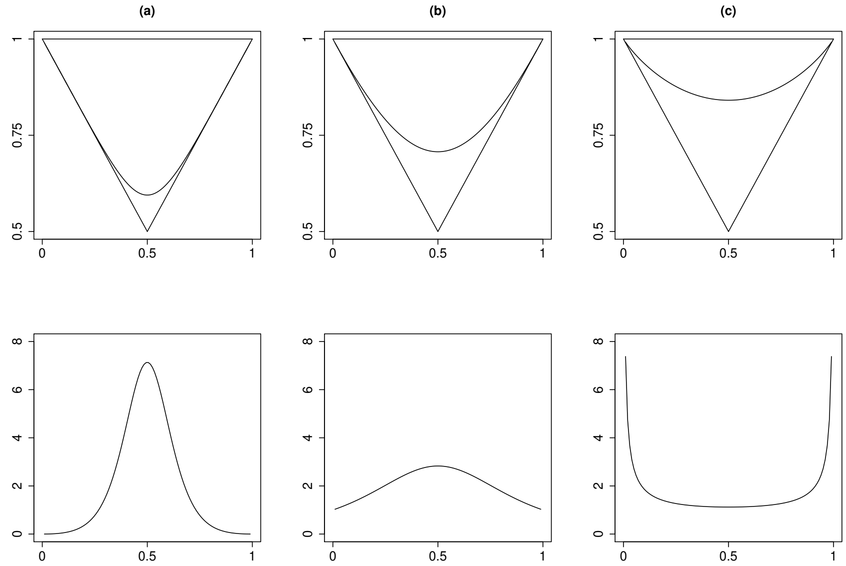

Figure 5.1: The dependence functions A (above) and A'' (below) of the three logistic models. The dependence parameters are (a) 0.25, (b) 0.5, (c) 0.75.

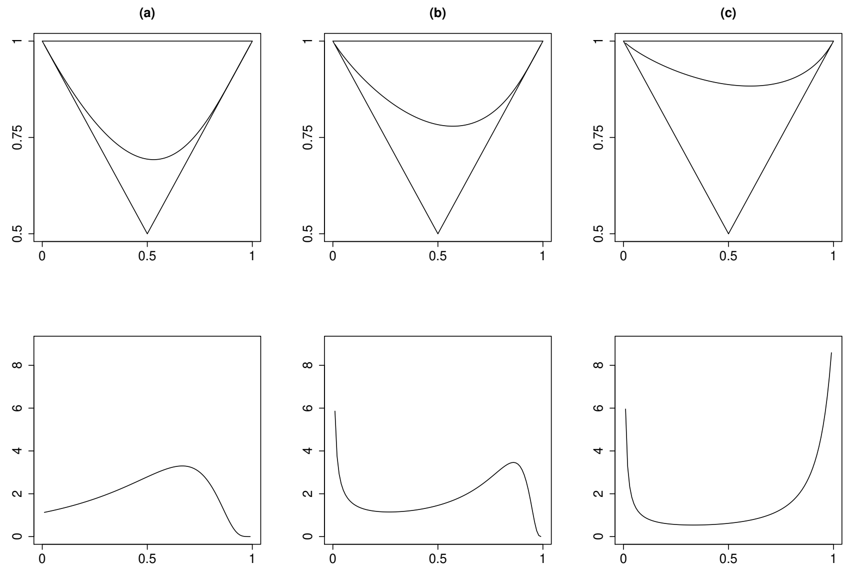

Figure 5.2: The dependence functions A (above) and A'' (below) of the three Dirichlet models. The dependence parameters are (a) $(1, 9)$, (b) $(\frac{1}{3}, 9)$, (c) $(\frac{1}{9}, 1)$.

generated. In each case the performances of four estimators for the dependence function are compared to that of the true, parametric model (labelled M) fitted by maximum likelihood. The four estimators are the kernel estimator (K) defined in (5.20), the symmetric kernel estimator (Ks), the estimator (C) of Capéraà *et al.* (1997) defined in (5.9) and the modified Pickands estimator (P) of Hall and Tajvidi (2000) defined in (5.11). Kernel estimates of A' and A'' are also compared to those of the true model: no such results for the non-differentiable, non-parametric estimators C and P are presented. The component distributions are assumed known and performance is summarised by Monte Carlo estimates of mean-integrated-square error (MISE) based on one-hundred replicated samples. The results are shown in Tables 5.1, 5.2 and 5.3.

Consider first the results in Table 5.1 for the estimates of A. Of the two non-parametric estimators, C almost always outperforms P except for the logistic model with strong dependence. Compare now the two kernel estimators. As expected, the symmetric version (Ks) is preferred only when the data are generated from the symmetric, logistic model and this superiority decreases as dependence weakens, where there is less scope for asymmetry. Impressively, the performance of Ks for the logistic data is almost equal to that of the true model fitted by maximum likelihood. The unconstrained kernel estimator (K) also performs well, typically outperforming both of the non-parametric estimators and occasionally matching the true model when the sample size is at its smallest.

The spline smoothing approach of Hall and Tajvidi (2000) has not been implemented but a tentative comparison of the results in Table 5.1 can be made with their simulation study in the case of the logistic dependence structure with $\alpha = 0.5$. Both studies produce the same ordering of the estimators M, C and P, and yield similar values of mean-integrated-square error. Hall and Tajvidi find that spline smoothing reduces the mean-integrated-square error for C by 31, 32 and 19%, and for P by 47, 30 and 21% when the sample size is 25, 50 and 100. If this improvement were replicated in the current study then estimator C would become superior to K at all sample sizes; estimator P would become superior to K at sample size 25.

Estimating the derivatives, A' and A'', is more difficult, as made clear by Tables 5.2 and 5.3. Nevertheless, the performance of the kernel estimator is encouraging, particularly for small samples of the Dirichlet data, where fitting the true model does not give vastly superior estimates. Again, the symmetric kernel estimator is significantly better than the unconstrained version for the logistic data.

The estimators are also applied to data simulated from the logistic model with $\alpha = 1$, that is independence between components. The results are in Table 5.4 and provoke comments similar to those recorded above for the other dependence structures.

It seems that the kernel estimator, K, is worthy of consideration as an estimator for the dependence function and its derivatives. Its overall performance for the small samples considered here is significantly better than that of the non-parametric estimators examined, particularly when asymmetry is present, and often compares favourably with maximum-likelihood fitting of the true model. Moreover, some

	Logistic(0.25)						Dirichlet(1,9)					
	25		50		100		25		50		100	
M	8	(1)	3	(1)	2	(0)	59	(10)	22	(4)	11	(1)
C	24	(2)	13	(2)	6	(1)	96	(11)	46	(7)	22	(2)
P	16	(2)	7	(1)	4	(1)	107	(12)	51	(8)	27	(6)
K	15	(2)	8	(2)	2	(0)	67	(9)	36	(6)	17	(2)
Ks	8	(1)	3	(1)	1	(0)	76	(10)	34	(6)	19	(2)
	Logistic(0.5)						Dirichlet($\frac{1}{3}$, 9)					
	25		50		100		25		50		100	
M	83	(12)	31	(6)	16	(2)	144	(22)	49	(7)	27	(4)
C	111	(12)	53	(8)	26	(3)	257	(37)	105	(19)	57	(7)
P	142	(14)	70	(11)	35	(7)	325	(38)	146	(24)	76	(14)
K	122	(20)	49	(8)	22	(2)	191	(27)	79	(10)	49	(6)
Ks	93	(13)	34	(6)	19	(3)	187	(29)	89	(12)	60	(5)
	Logistic(0.75)						Dirichlet($\frac{1}{9}$, 1)					
	25		50		100		25		50		100	
M	264	(37)	104	(18)	59	(8)	237	(26)	106	(14)	63	(8)
C	397	(54)	168	(28)	91	(11)	554	(88)	215	(36)	119	(14)
P	576	(69)	258	(40)	131	(21)	794	(116)	327	(50)	169	(26)
K	264	(29)	118	(16)	75	(8)	209	(21)	122	(16)	84	(10)
Ks	273	(38)	119	(20)	66	(9)	275	(29)	134	(16)	95	(11)

Table 5.1: MISE ($\times 10^5$) for five estimators for A and six models of dependence. Columns are labelled with sample size; estimated standard errors are in brackets.

features of the kernel estimator might be improved, in particular the choice of smoothing parameter which is far from optimal. A detailed comparison with the spline smoothing estimators of Hall and Tajvidi (2000) is needed to discriminate the performances of their method and the kernel estimator. Finally, notice that it is possible to apply kernel-smoothing ideas to non-parametric estimators such as C and P. These estimators for the dependence function A imply a point-mass function for A'' that could be smoothed and adjusted to ensure the appropriate moment conditions, and then integrated to obtain estimates for A and A'.

	Logistic(0.25)			Dirichlet(1,9)		
	25	50	100	25	50	100
M	23 (3)	10 (2)	5 (1)	95 (14)	39 (6)	18 (2)
K	53 (5)	27 (3)	12 (1)	130 (12)	74 (8)	36 (3)
Ks	30 (4)	14 (2)	7 (1)	129 (12)	72 (7)	49 (3)
	Logistic(0.5)			Dirichlet($\frac{1}{3}$, 9)		
	25	50	100	25	50	100
M	95 (13)	37 (7)	20 (3)	199 (26)	72 (10)	35 (5)
K	184 (23)	89 (10)	43 (4)	279 (30)	137 (13)	81 (7)
Ks	124 (14)	50 (7)	28 (4)	279 (28)	171 (12)	136 (6)
	Logistic(0.75)			Dirichlet($\frac{1}{9}$, 1)		
	25	50	100	25	50	100
M	260 (37)	102 (18)	58 (8)	315 (28)	144 (17)	87 (10)
K	358 (28)	184 (17)	117 (10)	320 (23)	199 (19)	133 (11)
Ks	292 (36)	139 (20)	78 (8)	353 (29)	206 (16)	157 (11)

Table 5.2: MISE ($\times 10^4$) for three estimators for A' and six models of dependence. Columns are labelled with sample size; estimated standard errors are in brackets.

	Logistic(0.25)			Dirichlet(1,9)		
	25	50	100	25	50	100
M	23 (3)	11 (2)	6 (1)	48 (6)	22 (3)	10 (1)
K	80 (8)	41 (3)	23 (2)	71 (6)	45 (4)	26 (2)
Ks	50 (6)	26 (3)	15 (2)	71 (5)	48 (3)	36 (2)
	Logistic(0.5)			Dirichlet($\frac{1}{3}$, 9)		
	25	50	100	25	50	100
M	34 (4)	14 (2)	7 (1)	108 (14)	40 (6)	18 (3)
K	72 (6)	44 (3)	26 (2)	154 (9)	111 (5)	89 (3)
Ks	52 (5)	27 (3)	17 (2)	136 (6)	109 (4)	98 (2)
	Logistic(0.75)			Dirichlet($\frac{1}{9}$, 1)		
	25	50	100	25	50	100
M	58 (9)	19 (4)	9 (1)	288 (25)	175 (16)	125 (15)
K	224 (7)	175 (6)	149 (5)	273 (14)	221 (12)	182 (7)
Ks	176 (8)	142 (6)	118 (5)	290 (10)	246 (8)	213 (6)

Table 5.3: MISE ($\times 10^2$) for three estimators for A'' and six models of dependence. Columns are labelled with sample size; estimated standard errors are in brackets.

	25		50		100	
M	184	(38)	94	(24)	42	(11)
C	1091	(146)	473	(66)	241	(27)
P	1660	(261)	636	(88)	324	(47)
K	312	(43)	163	(27)	75	(13)
Ks	225	(42)	113	(25)	53	(12)

	25		50		100	
M	181	(37)	93	(23)	41	(11)
K	384	(48)	214	(32)	97	(16)
Ks	224	(42)	115	(26)	54	(12)

	25		50		100	
M	77	(13)	53	(9)	30	(6)
K	141	(17)	100	(16)	46	(9)
Ks	36	(8)	22	(6)	11	(3)

Table 5.4: MISE ($\times 10^5$, $\times 10^4$ and $\times 10^2$) for estimators for A (top), A' (middle) and A'' (bottom) under independence. Columns are labelled with sample size; estimated standard errors are in brackets.

5.5 Data Example

In this section, the estimators of Sections 5.1, 5.2 and 5.3 are applied to a bivariate time series of seventy-four annual maximum sea-levels recorded in metres at two sites, Lowestoft (X_1) and Sheerness (X_2), on the east coast of England between 1899 and 1983. The data are plotted in Figure 5.3. First, the component distributions are estimated and fixed as the dependence structure is estimated. These results are compared with those obtained by estimating the component parameters and the dependence structure simultaneously.

Following Tawn (1988), linear trends are included in the location parameters of the component generalised extreme-value distributions by replacing (μ_1, μ_2) with $(\alpha_1 + \beta_1 t, \alpha_2 + \beta_2 t)$, where t is a time covariate for which an increment of 0.01 corresponds to one year and $t = 0$ corresponds to 1941. Fitting the components

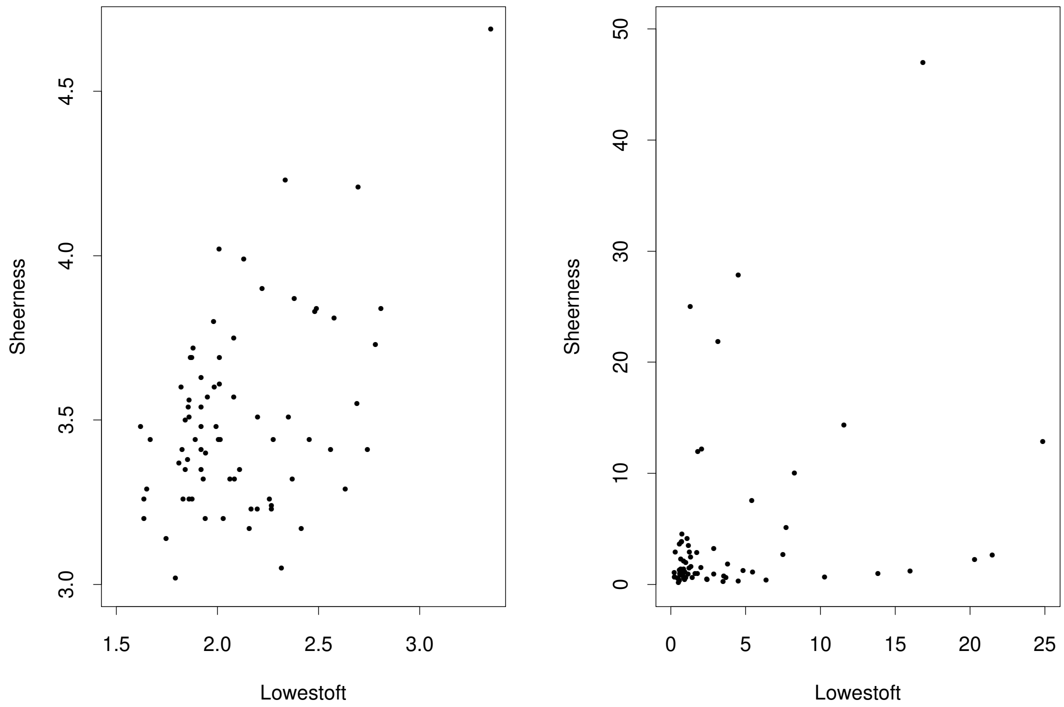

Figure 5.3: Annual maximum sea-levels (metres) recorded at Sheerness against those recorded at Lowestoft. The raw data are plotted on the left, and the data transformed to standard Fréchet components by the independence fit are plotted on the right. The outlier is not shown in the right-hand plot: it has coordinates (126, 236).

separately yields the parameter estimates displayed in the bottom half of Table 5.5. The data transformed to standard Fréchet components using these estimates are also plotted in Figure 5.3.

Having transformed to standard Fréchet components the dependence function can be estimated using the non-parametric estimators (5.9) and (5.11). These estimates are plotted in Figure 5.4; they have not been made convex for illustrative purposes. The estimates point to stronger dependence when w is less than one-half, which corresponds to relatively larger sea-levels at Sheerness than at Lowestoft.

Joint	α	β	σ	ξ
Lowestoft	1.95	0.050	0.23	0.070
Sheerness	3.39	0.379	0.19	0.019
Independence	α	β	σ	ξ
Lowestoft	1.95	0.032	0.23	0.083
Sheerness	3.39	0.369	0.19	0.072

Table 5.5: Component parameters estimated jointly (top) and under independence (bottom).

Another feature is the very strong dependence at w close to 0.94. This corresponds to the string of observations along the bottom of the right-hand plot in Figure 5.3.

For comparison, a parametric model is also fitted, the asymmetric logistic, with dependence function

$$A(w) = (1-\theta_1)(1-w) + (1-\theta_2)w + \{\theta_1^{1/\alpha}(1-w)^{1/\alpha} + \theta_2^{1/\alpha}w^{1/\alpha}\}^{\alpha},$$

where $0 \leq \theta_1 \leq 1$, $0 \leq \theta_2 \leq 1$ and $0 \leq \alpha \leq 1$. The parameter estimates are $\theta_1 = 0.33$, $\theta_2 = 1.00$ and $\alpha = 0.51$, the log-likelihood is -310.20, and the estimated dependence function is shown in Figure 5.4. The principal difference to the non-parametric estimates is near $w = 1$: the parametric model places mass $1-\theta_1 = 0.67$ at $w = 1$, interpreting the aforementioned string of points in Figure 5.3 as events at Lowestoft that are independent of the sea-level at Sheerness.

For the kernel estimator, the smoothing parameter, λ, is set using the over-smoothing formula (5.21) with the standard deviation, s, chosen by examining the plot, shown in Figure 5.5, of s_m (5.22) for $m = 1, \ldots, n$. There is the usual trade-off between bias and variability as m decreases, but the estimates appear to stabilise around $s = 0.31$. With this choice for s, the parameter estimates are

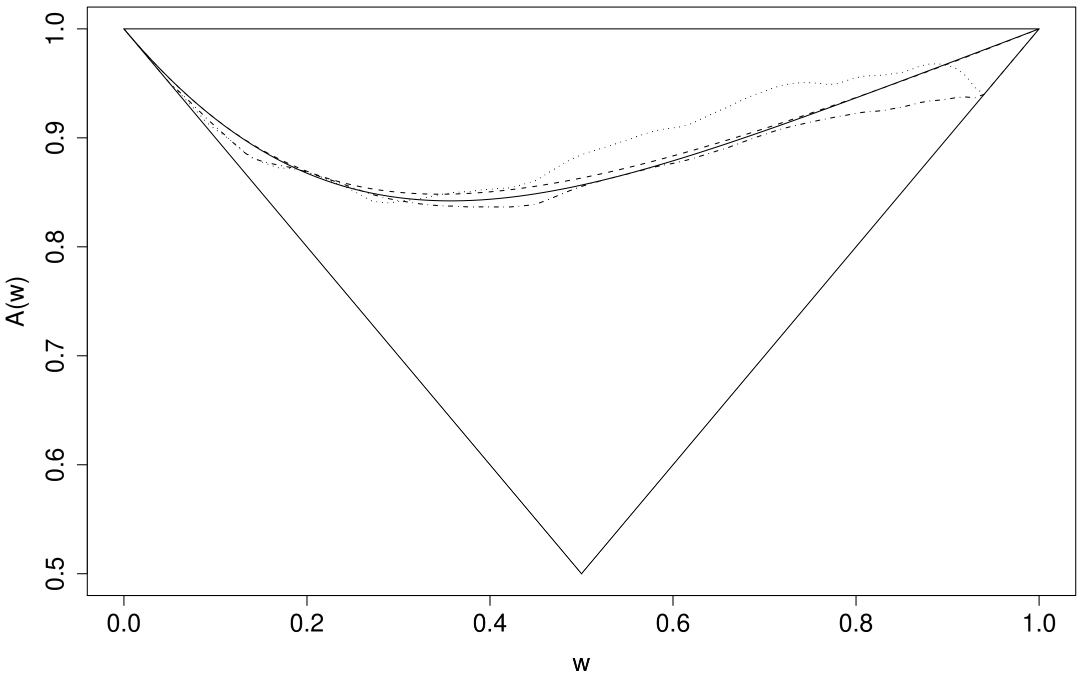

Figure 5.4: Four estimates of A for the transformed Lowestoft-Sheerness data: kernel (——), asymmetric logistic (- - -), Pickands (· · ·) and Capéraà *et al.* (-·-·-).

$m = 3$, $\phi_0 = 0.00$, $\phi_1 = 0.68$ and $\lambda = 0.22$; the log-likelihood is -310.11. The stability of parameter estimates from choosing a wider range of values for s was also investigated. Figure 5.5 indicates that the estimates for m, ϕ_0 and ϕ_1 are indeed stable when s is about 0.31. The kernel estimator places mass $\phi_1 = 0.68$ at $w = 1$, almost the same as with the asymmetric logistic model. The parametric and kernel estimates of the dependence function are also similar: see Figure 5.4.

Now the components and dependence structure are fitted simultaneously using the kernel estimator. With starting values dictated by the previous analysis, this yields the component parameter estimates in the top half of Table 5.5. Comparing these with the estimates from the separate component fits confirms the common finding

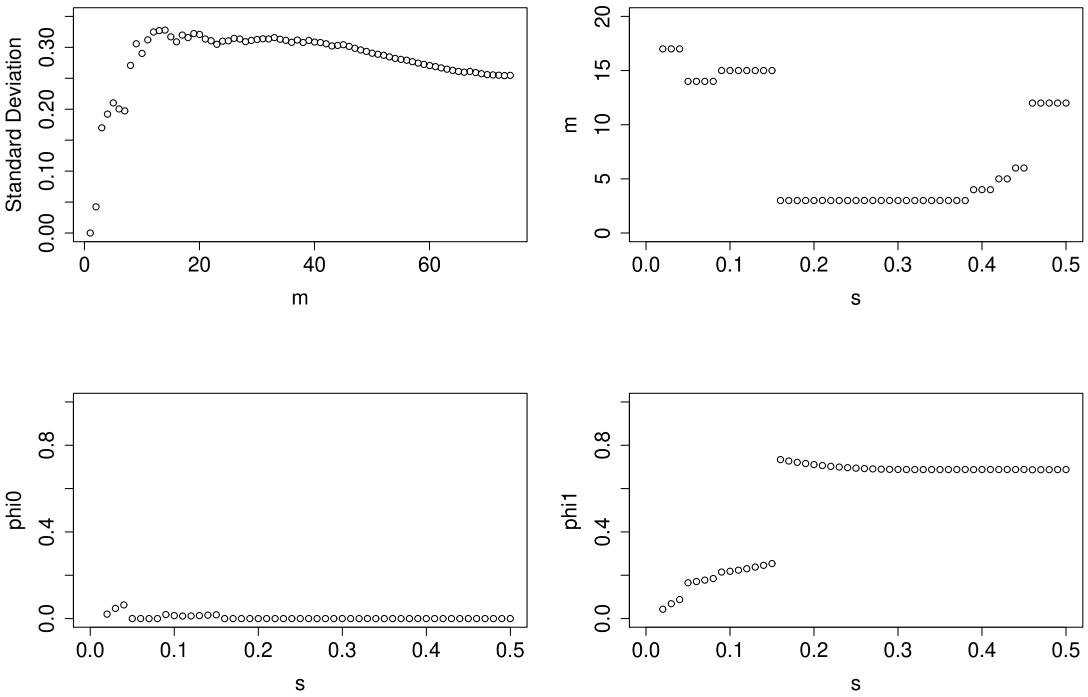

Figure 5.5: Top-left: standard deviation (s_m) of the angular component computed from the largest m observations. The remaining plots show estimates of the kernel dependence parameters with different values for s.

that joint estimation has most effect on the shape parameters. The parameter estimates of the dependence structure are $m = 14$, $\phi_0 = 0.00$, $\phi_1 = 0.72$ and $\lambda = 0.16$. Estimating the asymmetric logistic model simultaneously with the component parameters yields dependence parameters $\theta_1 = 0.28$, $\theta_2 = 1.00$ and $\alpha = 0.49$, and the two pairs of component parameter estimates are (1.96, 0.030, 0.25, 0.050) and (3.39, 0.384, 0.19, 0.023). Again, the mass placed on $w = 1$ by the two estimators is almost equal.

To conclude, the fitted models are used to estimate two failure probabilities:

$$
\begin{aligned}
p_{X_1 \cup X_2} &= P(X_1 > x_1 \text{ or } X_2 > x_2), \\
p_{X_2 | X_1} &= P(X_2 > x_2 \mid X_1 > x_1)
\end{aligned}
$$

for some threshold values x_1 and x_2. As an illustration, these are set equal to the upper p-quantiles (estimated under independence) of the two component distributions at time $t = 0.46$, corresponding to the year 1987, that is

$$
\begin{aligned}
x_1 &= \alpha_1 + \beta_1 t - \frac{\sigma_1}{\xi_1}\left[1 - \{-\log(1-p)\}^{-\xi_1}\right], \\
x_2 &= \alpha_2 + \beta_2 t - \frac{\sigma_2}{\xi_2}\left[1 - \{-\log(1-p)\}^{-\xi_2}\right].
\end{aligned}
$$

These are estimated to be $x_1 = 3.28$ and $x_2 = 4.58$ for $p = 0.01$. Of the observed data, only the outlier corresponding to the 1953 flood exceeds these thresholds: the empirical estimates of the probabilities are $\hat{p}_{X_1 \cup X_2} = 1/74 \approx 0.014$ and $\hat{p}_{X_2|X_1} = 1$. The estimates from the joint kernel fit are, with bootstrapped 90% confidence intervals, 0.012 (0.007, 0.029) and 0.20 (0.14, 0.26), while the joint asymmetric logistic fit yields 0.014 (0.004, 0.032) and 0.18 (0.04, 0.63).

5.6 Discussion

There are potentially many ways to improve the semi-parametric kernel estimator. One is the choice of smoothing parameter λ, which could be selected with a cross-validation procedure (cf. Hall and Tajvidi, 2000), or the kernel density estimator

(5.18) could be replaced with a simple histogram. Another issue is the scaling of q_n^* by ϕ to obtain the estimate (5.20) of $h = A''$. This was suggested by the limit (5.16), but when there is mass at the end points the estimate of q is likely to be contaminated by surplus observations near to zero and one: at finite levels, no angles will be observed that are precisely equal to zero or one. It may be profitable, therefore, to downweight q_n^* near the end points rather than scale uniformly over the entire interval.

Perhaps the most pressing issue concerns the likelihood. The parameter space of (ϕ_0, ϕ_1, m) is discrete, which makes optimisation expensive and invalidates standard likelihood procedures. A possible solution is to define a kernel density estimator with mass p_i on w_i, $1 \leq i \leq n$, and optimise the likelihood with respect to $(\phi_0, \phi_1, p_1, \ldots, p_n)$. The n masses could be chosen to belong to some parametric family in order to reduce the number of parameters. Another option is to formulate the estimator in a Bayesian framework and make inferences with a suitable Markov chain Monte Carlo scheme.