Bivariate Extreme Value Theory: Models and Estimation
Author(s): Jonathan A. Tawn
Source: *Biometrika*, Vol. 75, No. 3 (Sep., 1988), pp. 397-415
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: http://www.jstor.org/stable/2336591
Accessed: 15-06-2018 19:42 UTC

Biometrika (1988), **75**, 3, *pp.* 397–415
Printed in Great Britain

Bivariate extreme value theory: Models and estimation

BY JONATHAN A. TAWN
Department of Probability and Statistics, The University, Sheffield S3 7RH, U.K.

SUMMARY

Bivariate extreme value distributions arise as the limiting distributions of renormalized componentwise maxima. No natural parametric family exists for the dependence between the marginal distributions, but there are considerable restrictions on the dependence structure. We consider modelling the dependence function with parametric models, for which two new models are presented. Tests for independence, and discriminating between models, are also given. The estimation procedure, and the flexibility of the new models, are illustrated with an application to sea level data.

Some key words: Bivariate exponential distribution; Extreme value theory; Maximum likelihood; Nonregular estimation; Stable distribution; Survival data.

1. INTRODUCTION

Extreme value theory has recently been an area of much theoretical and practical work. Univariate theory is a well documented area, whereas bivariate/multivariate extreme value theory has, until recently, received surprisingly little attention. In the multivariate case, no natural parametric family exists for the dependence structure, so this must be modelled in some way. In the analysis of environmental extreme value data, there is a need for models of dependence between extremes from different sources: for example at various sea ports, or at various points of a river.

In this paper we consider bivariate extreme value distributions. We assume, without loss of generality, that we have exponential marginal distributions with unit means. The class of bivariate exponential distributions in which we are interested, satisfy a strong stability relation. Exponential variables (X, Y) satisfy the stability relation if and only if $W = \min(aX, bY)$ is also exponentially distributed for all $a, b > 0$ (Pickands, 1981). Therefore, the models we will consider have particular application in reliability and survival analysis.

One approach to modelling the dependence structure is via parametric models. This requires a flexible family of models which satisfy certain constraints. Models are of two kinds: either differentiable, or nondifferentiable. All nondifferentiable models give distributions which are singular, with nonzero probability concentrated on a certain subspace. The differentiable models have densities, but the existing models are symmetric which leads to the variables being exchangeable. Here, we present two new asymmetric differentiable models, which have increased flexibility. Properties of the differentiable models are examined.

Estimation of the parametric models has previously been by ad hoc methods, because there is a nonregular estimation problem when the margins are independent. For the cases when margin parameters are known and unknown, this nonregular problem is resolved for the existing differentiable models, and for one of the new models. As a result, maximum likelihood is proposed for estimation of parameters, and tests of independence are developed. Since many models for expressing dependence are presented, various methods are given to help decide between them. Finally, we present an example of extreme sea level data to illustrate the estimation procedure, the flexibility of the new models and the hypothesis tests developed in the paper.

2. Univariate motivation

The extreme value distributions arise in the following way: let the random variables $X_1, \ldots, X_n$ be independent and identically distributed, with distribution function $F(.)$. Define $M_n = \max(X_1, \ldots, X_n)$ and suppose there exist sequences of normalizing constants $a_n > 0$, b_n such that

$$\mathrm{pr}\{(M_n - b_n)/a_n \leqslant z\} = F^n(a_n z + b_n) \to G(z) \tag{2·1}$$

as $n \to \infty$, where G is a proper nondegenerate distribution function. It follows that G must be one of the extreme value distributions first identified by Fisher & Tippett (1928). These distributions can be summarized by the Generalized Extreme Value distribution, with distribution function,

$$G(z; \mu, \sigma, k) = \exp\left[-\left\{1 - k\left(\frac{z-\mu}{\sigma}\right)\right\}^{1/k}\right], \tag{2·2}$$

where $\sigma > 0$ and z has the range determined by $1 - k(z-\mu)/\sigma > 0$. There are three particular forms of G, corresponding to $k > 0$, that is Weibull, $k < 0$, Fréchet, and $k = 0$, Gumbel, taken as the limit as $k \to 0$. Maximum likelihood estimation of the parameters of the Generalized Extreme Value distribution is considered by Prescott & Walden (1980, 1983), and Smith (1985a) gives theoretical details for the asymptotic behaviour of the estimators.

3. Bivariate extreme value theory

We now consider $Z_i = (X_i, Y_i)$ $(1 \leqslant i \leqslant n)$ to be independent and identically distributed vector random variables with distribution function $F(.,.)$. The extension of univariate results is not entirely immediate. The obvious problem is the lack of natural order in higher dimensions (Barnett, 1976). Here, we use componentwise ordering, so that we define

$$\max_{1 \leqslant i \leqslant n} Z_i = (M_{1n}, M_{2n}),$$

where $M_{1n} = \max X_i$ and $M_{2n} = \max Y_i$.

A difficulty with this approach is that in some applications it may be impossible for (M_{1n}, M_{2n}) to occur as a vector observation. Despite this problem, this is the approach most widely used in bivariate extreme value analysis.

As in the the univariate case, we suppose that there exist sequences of normalizing constants $a_{in} > 0$, b_{in} $(i = 1, 2)$ such that

$$\mathrm{pr}\{(M_{1n} - b_{1n})/a_{1n} \leqslant x, (M_{2n} - b_{2n})/a_{2n} \leqslant y\} = F^n(a_{1n}x + b_{1n}, a_{2n}y + b_{2n}) \to G(x, y) \tag{3·1}$$

as $n \to \infty$, where G is a proper distribution function, nondegenerate in each margin. It can be easily shown that G must satisfy the max-stability relation: for all $n \geq 1$ there exist $\alpha_{in} > 0$, β_{in} for $i = 1, 2$ such that

$$G^n(\alpha_{1n}x + \beta_{1n}, \alpha_{2n}y + \beta_{2n}) = G(x, y). \tag{3·2}$$

Clearly, from (3·1) and (2·1) the marginal distributions must be Generalized Extreme Value, given by (2·2). As all three univariate classes can be transformed into each other, there is no theoretical loss of generality in assuming the marginals are identically distributed. The choice of marginal distribution is arbitrary; Tiago de Oliveira (1962/63; 1980) and Galambos (1978) assume Gumbel margins, whereas de Haan & Resnick (1977) assume Fréchet margins. We shall assume unit exponential margins, as do Pickands (1981) and Deheuvels (1983, 1985). The key reasons for this choice are the simplicity in the structure of the dependence and the simple extension to the multivariate representation.

The general structure for bivariate extreme value distributions has been known since the work of Tiago de Oliveira (1958), Geffroy (1958/59) and Sibuya (1960). Only recently have similar results been obtained for the multivariate case (de Haan & Resnick, 1977; Pickands, 1981). Via the corresponding bivariate representation theorem of Pickands, we develop models which may easily be extended to the multivariate case. We do not present the extension here.

Let (X, Y) be a random vector and let $G(x, y) = \text{pr}\,(X > x, Y > y)$. We say that (X, Y) follows an extreme value distribution with unit exponential margins if and only if

$$\text{pr}\,(X > x) = e^{-x}, \quad \text{pr}\,(Y > y) = e^{-y} \quad (x > 0, y > 0), \tag{3·3}$$

and for any $n \geq 1$

$$G^n(x, y) = G(nx, ny) \quad (x > 0, y > 0). \tag{3·4}$$

Equation (3·4) is just the exponential form of (3·2). Pickands (1981) shows that (X, Y) follows such a distribution if and only if the joint survivor function can be written in the form

$$G(x, y) = \exp\left\{-(x+y)A\left(\frac{y}{x+y}\right)\right\} \quad (x > 0, y > 0), \tag{3·5}$$

where

$$A(w) = \int_0^1 \max\{(1-w)q, w(1-q)\}\, dH(q). \tag{3·6}$$

Here H is a positive finite measure on $[0, 1]$. In order that (3·3) is satisfied we need

$$1 = \int_0^1 q\, dH(q) = \int_0^1 (1-q)\, dH(q). \tag{3·7}$$

Equations (3·7) give that $\frac{1}{2}H$ is the distribution function of a random variable with mean $\frac{1}{2}$. Following Pickands (1981) we call $A(.)$ the dependence function of (X, Y). This must not be confused with dependence functions introduced by other authors. Good accounts of the connexions between the various dependence functions are given by Deheuvels (1984) and Weissman (1985). An alternative result obtained by Pickands (1981) is that (X, Y) satisfy (3·3) and (3·4) if and only if $W = \min\,(aX, bY)$ is exponential for all $a, b > 0$.

Combining the dependence structure representation with unspecified univariate distributions gives the joint distribution expressed in its most general form

$$G(x, y) = \exp\left(\int_0^1 \log\left[\min\left\{G_1^q(x),\, G_2^{1-q}(y)\right\}\right] dH(q)\right).$$

Here G_1 and G_2 are univariate extreme value distributions, given by (2·2). By the univariate version of (3·2) we see that the location and scale parameters of each of the marginal distributions vary as the measure is integrated over $[0, 1]$.

Returning to the case of unit exponential margins, we now consider properties of the dependence function and its relationship to the measure. From (3·7) we have that $A(0) = A(1) = 1$. In addition, due to representation (3·6) we have that

$$\max(w, 1-w) \leqslant A(w) \leqslant 1 \quad (0 \leqslant w \leqslant 1)$$

and, most importantly, $A(.)$ is a convex function within this region. The structure of the dependence makes the variables (X, Y) positively associated. In fact, a stronger stochastic ordering condition (Lehmann, 1966) must also hold, namely that the conditional probability $\mathrm{pr}(Y > y \mid X > x)$ is increasing in x for fixed y.

Two important examples of the dependence function are those on the boundary of the functional space as follows.

(a) If $A(w) = 1$ $(0 \leqslant w \leqslant 1)$, then (X, Y) are independent. The corresponding measure puts mass one at each of the endpoints 0 and 1.

(b) If $A(w) = \max(w, 1-w)$, then (X, Y) are completely dependent, that is $\mathrm{pr}(X = Y) = 1$. The corresponding measure puts mass two at $\frac{1}{2}$.

The class of dependence functions is a convex set so that, if $A_1, \ldots, A_p$ are dependence functions, then, with $\alpha_i \geqslant 0$ $(i = 1, \ldots, p)$,

$$A(w) = \alpha_1 A_1(w) + \ldots + \alpha_p A_p(w)$$

is a dependence function, if $\alpha_1 + \ldots + \alpha_p = 1$. Some interesting properties of the variables can be formulated in terms of $A(.)$. For example, the variables are exchangeable if and only if $A(.)$ is symmetrical about $\frac{1}{2}$. The correlation between (X, Y) is

$$\rho = \int_0^1 \frac{dw}{\{A(w)\}^2} - 1$$

and is always nonnegative because $A(w) \leqslant 1$. An alternative dependence measure, which is independent of the marginal distributions, is $2\{1 - A(\frac{1}{2})\}$. This is zero in the independent case, and one in the completely dependent case. In the literature, dependence measures have received much attention; see, for example, Tiago de Oliveira (1980).

A notable feature of bivariate extreme value theory is that there is no finite-dimensional parametric family for the dependence function. We must therefore estimate the dependence function by some method. We have the usual option of using nonparametric or parametric methods. We concentrate here on parametric models, but for purposes of comparison we will also discuss a nonparametric estimator.

4. Nonparametric estimation

An estimator suggested by Pickands (1981) was derived in the following way. If we have independent and identically distributed pairs (X_i, Y_i) for $1 \leqslant i \leqslant n$, from a bivariate extreme value distribution with exponential margins, let

$$Z_i(w) = \min\{(1-w)^{-1} X_i,\, w^{-1} Y_i\} \quad (1 \leqslant i \leqslant n,\, 0 \leqslant w \leqslant 1).$$

Then

$$\mathrm{pr}\,\{Z_i(w) > z\} = \exp\{-zA(w)\},$$

so that the maximum likelihood estimator is

$$\hat{A}(w) = n\left\{\sum_{i=1}^{n} Z_i(w)\right\}^{-1}. \tag{4·1}$$

There is no restriction on this estimator to be convex, as is seen in the example in § 10. Various methods based on this estimator have been proposed to ensure we obtain a convex estimator; see Pickands (1981) and Smith (1985b). In forcing convexity, such methods often result in the joint distribution being singular. In addition to this weakness, the greater convenience and tractability of parametric methods leads us to restrict ourselves to parametric models. Therefore, here the role of the nonparametric estimator (4·1) is in illustrating the flexibility and suitability of the parametric models.

5. Parametric methods

Parametric models for the dependence function have been known since Gumbel (1960), and have received much attention in Tiago de Oliveira's work. The models can be separated into two classes: differentiable and those which are not everywhere differentiable. The differentiable models are those for which (X, Y) have a density, whereas the nondifferentiable models are singular. For the existing nondifferentiable models, see Tiago de Oliveira (1984). These are the biextremal, natural and Gumbel models. The only distributions which have limiting singular distributions for the renormalized maxima are themselves singular. Since, in most environmental applications, singular distributions do not occur, we concentrate on the differentiable models. The constraints on A for differentiable parametric models now become

$$A(0) = A(1) = 1, \quad -1 \leq A'(0) \leq 0, \quad 0 \leq A'(1) \leq 1, \quad A''(w) \geq 0 \quad (0 \leq w \leq 1). \tag{5·1}$$

The class of existing differentiable parametric models contains two distinct models: the mixed and logistic models, although more models can be generated via convex combinations.

(a) The mixed model has dependence function $A(w) = \theta w^2 - \theta w + 1$ $(0 \leq \theta \leq 1)$, with corresponding joint survival function

$$G(x, y) = \exp\left\{-(x+y) + \frac{\theta xy}{x+y}\right\}.$$

This model can easily be shown to be the complete class of quadratic functions satisfying conditions (5·1). Independence corresponds to $\theta = 0$, but we cannot have complete dependence. With this model the variables (X, Y) are exchangeable and have correlation

$$\rho = (1 - \tfrac{1}{4}\theta)^{-3/2}\theta^{-\frac{1}{2}}\{\sin^{-1}(\tfrac{1}{2}\theta^{\frac{1}{2}}) - \tfrac{1}{2}\theta^{\frac{1}{2}}(1 - \tfrac{1}{4}\theta)^{\frac{1}{2}}(1 - \tfrac{1}{2}\theta)\}.$$

(b) The logistic model has dependence function $A(w) = \{(1-w)^r + w^r\}^{1/r}$ for $r \geq 1$, with corresponding joint survival function

$$G(x, y) = \exp\{-(x^r + y^r)^{1/r}\}. \tag{5·2}$$

Independence and complete dependence correspond to $r=1$ and $r=+\infty$ respectively. With this model the variables are exchangeable and have correlation $\rho=r^{-1}\{\Gamma(2/r)\}^{-1}\{\Gamma(1/r)\}^2-1$.

The logistic model has appeared in the survival analysis literature: see, for example, Hougaard (1986). Alternative parameterizations for this model are possibly advisable: for example $\nu=1/r$ $(0\leq\nu\leq 1)$. It is interesting to note, and of importance later, that independence and complete dependence correspond to parameters on the boundary of the parameter space: this is a direct result of these cases lying on the boundary of the functional space for dependence functions.

In some contexts exchangeability would not be a reasonable assumption, and the degree of nonexchangeability may be of interest in itself. So, we present two new models which have additional flexibility, the asymmetric mixed and the asymmetric logistic models.

(c) The asymmetric mixed model has dependence function

$$A(w)=\phi w^3+\theta w^2-(\theta+\phi)w+1 \quad (\theta\geq 0,\ \theta+\phi\leq 1,\ \theta+2\phi\leq 1,\ \theta+3\phi\geq 0),$$

with corresponding joint survival function

$$G(x,y)=\exp\left[-(x+y)+xy\{x(\theta+\phi)+y(2\phi+\theta)\}(x+y)^{-2}\right].$$

This model is the complete class of cubic functions satisfying the conditions in (5·1). When $\phi=0$ we have the mixed model. Independence corresponds to $\theta=\phi=0$, a corner of the parameter space, but we cannot have complete dependence.

(d) The asymmetric logistic model has dependence function

$$A(w)=\{\theta^r(1-w)^r+\phi^r w^r\}^{1/r}+(\theta-\phi)w+1-\theta \quad (0\leq\theta,\ \phi\leq 1,\ r\geq 1),$$

with corresponding joint survival function

$$G(x,y)=\exp\{-(1-\theta)x-(1-\phi)y-(x^r\theta^r+y^r\phi^r)^{1/r}\}.$$

When $\theta=\phi=1$ we have the logistic model, but this model contains other existing models. If $\theta=\phi$ we get a mixture of logistic and independence. If $r\to+\infty$ we have

$$A_\infty(w)=\max\{1-\phi w, 1-\theta(1-w)\}, \tag{5·3}$$

a new nondifferentiable model with $\operatorname{pr}(Y\phi=X\theta)=\phi\theta/(\theta+\phi-\theta\phi)$.

In (5·3), when $\theta=1$ and $\phi=\alpha$ we have the biextremal (α) model, whereas when $\theta=\alpha$ and $\phi=1$ we have the dual of the biextremal (α) model, which corresponds to X and Y being exchanged. If $\theta=\phi=\alpha$ we have the Gumbel model. So this contains three of the existing models and therefore must be very flexible. Complete dependence corresponds to $\theta=\phi=1$ and $r=+\infty$, whereas independence corresponds to $\theta=0$ or $\phi=0$ or $r=1$.

The asymmetric mixed model is a crude model, but has the advantage of having a single parameter clearly identified with nonexchangeability. On the other hand, the asymmetric logistic model is flexible and simply expressible, but this model has identifiability problems. In small-sample practical applications, a possible simplification, without much loss of flexibility, could be to let either θ or ϕ equal one.

6. Estimation in the differentiable parametric models

In this section, we consider only the case when all the margin parameters are known. When the margins are unknown, see § 8.

Previously, estimation has been by various ad hoc methods (Gumbel & Mustafi, 1967; Posner et al., 1969; Tiago de Oliveira, 1975). None of these methods easily extends to multiparameter models, so maximum likelihood was considered. Maximum likelihood has previously not been used because of nonregular behaviour (Tiago de Oliveira, 1980). Since the problems are the same for each of the existing models, we illustrate the problem and solution only for the logistic model. The same problems are identified for this model by Hougaard (1986).

If $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent and identically distributed with joint survival function given by (5·2), and we let the corresponding density be $g(x, y; r)$: then

$$g(x, y; r) = \frac{\partial^2 G(x, y)}{\partial x\, \partial y}$$
$$= (xy)^{r-1}(x^r + y^r)^{-2+1/r}\{(x^r+y^r)^{1/r} + r - 1\} \exp\{-(x^r+y^r)^{1/r}\}. \qquad (6{\cdot}1)$$

Let

$$L_n(r) = \sum_{i=1}^{n} \log g(x_i, y_i; r), \quad U_n(1) = \frac{dL_n(1)}{dr} = \sum_{i=1}^{n} u(x_i, y_i),$$

where

$$u(x, y) = \log(xy) + (x + y - 2)\log(x + y) - x\log x - y\log y + (x + y)^{-1}. \qquad (6{\cdot}2)$$

Here, $U_n(1)$ is the score statistic for independence in the logistic model. It can easily be shown that $E\{u(X, Y)\} = 0$, but $E\{u(X, Y)^2\} = +\infty$.

We let $\bar{r}_n$ denote the maximum likelihood estimator of r when the marginal distributions are known. In the independence case, $r = 1$, the Cramér–Rao bound is zero and n var $(\bar{r}_n) \to 0$ as $n \to \infty$, which shows that $\bar{r}_n$ has a nonregular behaviour at $r = 1$. For $1 < r < +\infty$ the estimation problem is regular, but the expected information cannot be found in closed form. If the nonregular behaviour can be understood, maximum likelihood is a desirable estimation procedure because in the case of independence, all other procedures have zero asymptotic relative efficiency. The exact behaviour of $U_n(1)$ has not been found, but the asymptotic distribution is obtained below.

We require stable law theory because we are dealing with sums of independent random variables with infinite variance. To apply such theory, we need the tail behaviour of the density function of $u(X, Y)$. The only term in (6·2) with infinite variance is $V = (X + Y)^{-1}$, so the upper tail behaviour of the density of $u(X, Y)$ is dominated by the tail behaviour of V, whereas the lower tail yields regular behaviour. Changing variables to V and $W = X$ we have the joint density

$$f_{v,w}(v, w) = v^{-2} e^{-1/v} \quad (0 < w < v^{-1} < \infty).$$

Thus $f_v(v) \sim v^{-3}$ for large v.

To find the asymptotic distribution of $U_n(1)$ under some suitable renormalizing, we now appeal to Feller (1971, § XVII.5) or Woodroofe (1972). Defining $\mu(x) = \int v^2 f_v(v)\, dv$, where the integral is over $(-x, x)$, then $\mu(x) \sim \log x$ for large x.

Now, by Feller (1971), $\{u(X_1, Y_1) + \ldots + u(X_n, Y_n)\}/c_n$ converges in distribution to a standard normal random variable. Here $c_n \to \infty$ and $c_n^{-2} n\mu(c_n) \to 1$ as $n \to \infty$, so it follows that $c_n \sim (\frac{1}{2} n \log n)^{\frac{1}{2}}$. Therefore the score statistic converges in distribution,

$$(\tfrac{1}{2} n \log n)^{-\frac{1}{2}} U_n(1) \to Z, \qquad (6{\cdot}3)$$

where Z is a standard normal random variable. To obtain the asymptotic behaviour of the maximum likelihood estimator, we also need to know the asymptotic behaviour of

$$\sum_{i=1}^{n} \{u(X_i, Y_i)\}^2 = \sum_{i=1}^{n} (X_i + Y_i)^{-2} + o_p\left\{\sum_{i=1}^{n} (X_i + Y_i)^{-2}\right\}.$$

By Feller (1971, § VII.7, Th. 3, p. 233) there exist constants d_n such that in probability

$$\{u(X_1, Y_1)^2 + \ldots + u(X_n, Y_n)^2\}/d_n \to 1.$$

The proof of the theorem shows how to find suitable d_n: in our case $d_n \sim \frac{1}{2}n \log n$, so that in probability

$$(\tfrac{1}{2}n \log n)^{-1} \sum_{i=1}^{n} \{u(X_i, Y_i)\}^2 \to 1. \tag{6·4}$$

We now consider the asymptotic behaviour of the maximum likelihood estimator, $\bar{r}_n$, defined to be the value of the parameter r which maximizes the likelihood within the parameter space $r \geq 1$. We are interested in the special case when $r = 1$, which is the boundary of the parameter space. Such situations are considered by Moran (1971), but some of his regularity conditions fail here because the expected information is infinite.

We first expand the log likelihood in a one-sided $(r > 1)$ Taylor expansion:

$$L_n(r) = L_n(1) + (r-1)\frac{dL_n(1)}{dr} + \tfrac{1}{2}(r-1)^2 \frac{d^2L_n(r^*)}{dr^2}, \tag{6·5}$$

where r^* is in the interval $(1, r)$. It can be shown that

$$\limsup \frac{d^2L_n(r)}{dr^2}\left\{\frac{d^2L_n(1)}{dr^2}\right\}^{-1} = 1,$$

where the supremum is for $r \in [1, 1 + M(n \log n)^{-\frac{1}{2}}]$, for all fixed M.

For simplicity we consider the cases $U_n(1) \leq 0$ and $U_n(1) > 0$ separately. If $U_n(1) \leq 0$ then, from (6·5), we have $dL_n(1)/dr = U_n(1) \leq 0$ so that $\bar{r}_n = 1$.

Now suppose $U_n(1) > 0$, if $r = 1 + M(n \log n)^{-\frac{1}{2}}$, then by (6·3) and (6·4)

$$\frac{dL_n(r)}{dr} = U_n(1) + M(n \log n)^{-\frac{1}{2}} \frac{d^2L_n(r^*)}{dr^2} < 0,$$

with probability one for sufficiently large M: hence $\bar{r}_n \in (1, 1 + M(n \log n)^{-\frac{1}{2}})$. Thus

$$0 = U_n(1) + (\bar{r}_n - 1)\{d^2L_n(1)/dr^2\} W_n,$$

where $W_n \to 1$ in probability, so that

$$(\bar{r}_n - 1)(\tfrac{1}{2}n \log n)^{\frac{1}{2}} = \frac{U_n(1)(\frac{1}{2}n \log n)^{-\frac{1}{2}}}{-[\{d^2L_n(1)/dr^2\} W_n(\frac{1}{2}n \log n)^{-1}]}. \tag{6·6}$$

From (6·3), (6·4) and the Yule–Slutsky lemma, the asymptotic behaviour of (6·6) can be found for $U_n(1) > 0$. Combining the above results, we have that the asymptotic behaviour of the maximum likelihood estimator is

$$(\bar{r}_n - 1)(\tfrac{1}{2}n \log n)^{\frac{1}{2}} \to S, \tag{6·7}$$

where convergence is in distribution and S is a nonnegative random variable. Moreover, $\mathrm{pr}(S \leq s) = h(s)\Phi(s)$, where $h(.)$ is the Heaviside function and $\Phi(.)$ is the standard normal distribution function. This is precisely the limit Moran (1971) obtains, but is achieved with a different renormalizing sequence.

The development is similar for the mixed model. If we let $U_n(0)$ be the score statistic for independence, then $U_n(0) = \Sigma\, u(x_i, y_i)$, where the sum is over $i = 1, \ldots, n$, with

$$u(x, y) = xy(x+y)^{-1} + 2xy(x+y)^{-3} - (x^2+y^2)(x+y)^{-2}.$$

We obtain identical limiting distributions using normalizing sequences differing from above only in the constant, with $\frac{1}{15}$ replacing $\frac{1}{2}$. In the case of the new models the problem is more difficult. The asymmetric mixed model has similar problems at independence, and a similar method of solution was used. The score vector converges to a bivariate normal, whereas the maximum likelihood estimators converge to truncated normal distributions determined by the configuration of the parameter space. In the asymmetric logistic model, problems again arise in the independence case. This is unsolved, because there are additional identifiability aspects.

7. Tests for independence

The importance of independence as the asymptotic limit has been known since Geffroy (1958/59), Sibuya (1960) and Mardia (1964). They showed that the normalized componentwise maxima are asymptotically independent for many distributions. In particular, independence arises if

$$\lim \frac{\operatorname{pr}(X > x,\, Y > y)}{1 - \operatorname{pr}(X < x,\, Y < y)} = 0$$

where as $x \to w_x$, $y \to w_y$, the upper endpoints of X and Y respectively. The generality of the condition shows that independence should play an important role in many applications. Sibuya (1960) showed that the componentwise maxima of a bivariate normal distribution with correlation coefficient ρ for $\rho < 1$ are asymptotically independent, with Gumbel margins. Another interesting example arises for

$$\operatorname{pr}(X > x,\, Y > y) = \exp\{K^{\nu} - (x^{1/\nu} + y^{1/\nu} + K)^{\nu}\}$$

a survival function suggested by Crowder (1988). It can be shown that componentwise minima are asymptotically independent if $K \neq 0$, but are logistic, with parameter $1/\nu$, if $K = 0$, in each case with unit exponential margins.

Ad hoc tests of independence have been presented by Gumbel & Goldstein (1964), Gumbel & Mustafi (1967) and Tiago de Oliveira (1984). Here we give score and likelihood ratio tests. For simplicity, let θ denote the dependence parameter and c denote the renormalizing constant for the score, so $c = \frac{1}{2}, \frac{1}{15}$ for the logistic and mixed models respectively.

The locally most powerful test is to reject independence at the α level when

$$U_n(\theta_0) > (cn \log n)^{\frac{1}{2}} \Phi^{-1}(1-\alpha),$$

where $\theta_0 \leqslant \theta$, is the value of the parameter that corresponds to independence. If we let λ denote the likelihood ratio then we have

$$\operatorname{pr}(2 \log \lambda \leqslant x) \to h(x)\Phi(x^{\frac{1}{2}}),$$

where convergence is in distribution and $h(.)$ is the Heaviside function.

An important question must be: how applicable are these asymptotic results? We consider this only for the locally most powerful test. Because of the nonregularity, the

Berry–Esseen bound is not applicable, so the rate of convergence may be very slow. By Hall (1983, Th. 2, 3) we have

$$\sup_{-\infty<x<\infty} |\mathrm{pr}\{(cn \log n)^{-\frac{1}{2}} U_n(\theta_0) + 2(\log n)^{-1} \leqslant x\} - \Phi(x)| = O\{(\log n)^{-1}\}$$

is the optimal choice of normalizing constants. As the theoretical rate of uniform convergence is so slow, it is particularly useful, for comparison purposes, to have small-sample results. An appropriate way to measure the correctness of the asymptotic approximation is via comparison of suitable percentage points. For the two existing models, percentage points for small samples were obtained by simulation. The simulation study involved generating 100 000 replications of the normalized score $(cn \log n)^{-\frac{1}{2}} U_n(\theta_0)$; see Table 1.

Table 1. *Simulated and asymptotic results for the normalized score statistic for independence in the logistic model; standard errors are given in parentheses*

		Exceedance levels		
		10%	5%	$2\frac{1}{2}$%
Asymptotic approximation		1·282	1·645	1·960
Simulated value	$n = 50$	1·73 (0·04)	2·55 (0·06)	4·07 (0·20)
	$n = 100$	1·64 (0·03)	2·43 (0·07)	3·82 (0·12)
	$n = 200$	1·53 (0·04)	2·31 (0·06)	3·15 (0·18)
	$n = 500$	1·41 (0·04)	2·02 (0·05)	2·67 (0·05)

The asymptotic approximation is poor for the sample sizes that occur in practice. We therefore recommend using the simulated values. Over the range of sample sizes considered, there is a good linear relationship, $y = \alpha + \beta \log n$, between the empirical percentage points, y, and log sample size. By standard regression we get

$$(\alpha, \beta) = (2{\cdot}26, -0{\cdot}137), (3{\cdot}47, -0{\cdot}227), (6{\cdot}63, -0{\cdot}639)$$

for the 10%, 5% and $2\frac{1}{2}$% percentage points respectively. The problems of testing independence in the asymmetric mixed model are similar. Testing independence in the asymmetric logistic model is even less straightforward, because of the identifiability problem mentioned earlier. Davies (1977) and Berman (1986) have shown that there is no immediate easy general solution to such problems. A simple suggestion is to accept independence in this case, if we accept independence in the logistic model.

8. Estimation when the margins are unknown

In § 6 we obtained the asymptotic behaviour of the maximum likelihood estimator of the dependence parameter, when the marginal distributions were assumed to be known. Here, we consider the more complicated estimation problem of finding the joint asymptotic behaviour of the maximum likelihood estimators of the dependence and margin parameters, when the marginal distributions are known only to be Generalized Extreme Value distributions.

Smith (1985a) showed that the asymptotic behaviour of the maximum likelihood estimators of the margin parameters depends on the true value of the shape parameter, k. In particular, different behaviour arises if $k < \frac{1}{2}$, $k = \frac{1}{2}$, $\frac{1}{2} < k < 1$ or $k \geqslant 1$. In the bivariate case, we have two shape parameters, and for simplicity, we assume both are less than $\frac{1}{2}$.

This corresponds to the case when all the maximum likelihood estimators of the margin parameters behave regularly.

Now, as in § 6, if there is dependence between the variables, then the estimation problem is regular, so standard results follow. Instead, suppose that a model for dependence is fitted when the margins are actually independent. In § 6 we found that the resulting estimation problem is nonregular. We consider this problem only for the existing one-parameter differentiable dependence models.

Let θ denote the dependence parameter and θ_0 be the true value of θ, which corresponds to independence, $\theta \geqslant \theta_0$. Also let $\phi = (\phi^1, \ldots, \phi^q)$ be the vector of margin parameters, with true value ϕ_0. We define the global maximum likelihood estimators by $(\hat{\theta}_n, \hat{\phi}_n)$. This point need not satisfy the likelihood equations because of the restricted parameter space. For the special case when $\theta = \theta_0$, that is independence is known, we let $\bar{\phi}_n$ denote the maximum likelihood estimator for ϕ. The existence and consistency of $\bar{\phi}_n$ follows from Smith (1985a). Finally, for the special case when $\phi = \phi_0$, that is when the margins are known, we let $\bar{\theta}_n$ denote the maximum likelihood estimator of θ. The behaviour of $\bar{\theta}_n$ has been discussed in § 6.

Before we can state our main results, we require the following preliminaries. Let E_{θ_0,ϕ_0} denote expectation with respect to the true density. Then define

$$nm_{ij} = E_{\theta_0,\phi_0}\{-\partial^2 L_n(\theta_0, \phi_0)/\partial\phi^i\, \partial\phi^j\} \quad (i, j = 1, \ldots, q), \tag{8·1}$$

$$nm_i^0 = E_{\theta_0,\phi_0}\{-\partial^2 L_n(\theta_0, \phi_0)/\partial\phi^i\, \partial\theta\} \quad (i = 1, \ldots, q). \tag{8·2}$$

Also, we write $Y_n <_{\mathrm{p}} s_n$ for a sequence random variables $\{Y_n\}$ and positive constants $\{s_n\}$ if

$$\lim_{a\to\infty} \limsup_{n\to\infty} \mathrm{pr}\,\{|Y_n| > as_n\} = 0.$$

As discussed above, the following results are for both shape parameters less than $\frac{1}{2}$. For both Theorems 1 and 2, similar results have been obtained for both shape parameters less than one, with the only difference being the joint asymptotic behaviour of the estimators of the margin parameters.

THEOREM 1. *Suppose the estimation problem is as stated above. Also suppose M is a strictly positive-definite matrix with entries m_{ij} $(i, j = 1, \ldots, q)$ defined by* (8·1). *Then there exists a sequence of maximum likelihood estimators $(\hat{\theta}_n, \hat{\phi}_n)$ such that*

$$\hat{\theta}_n - \theta_0 <_{\mathrm{p}} (n \log n)^{-\frac{1}{2}}, \quad \hat{\phi}_n^j - \phi_0^j <_{\mathrm{p}} n^{-\frac{1}{2}} \quad (j = 1, \ldots, q).$$

Moreover in probability

$$(n \log n)^{\frac{1}{2}}(\hat{\theta}_n - \bar{\theta}_n) \to 0, \quad n^{\frac{1}{2}}(\hat{\phi}_n^j - \bar{\phi}_n^j) \to 0 \quad (j = 1, \ldots, q).$$

Theorem 1 guarantees the asymptotic existence of maximum likelihood estimators. We now give the asymptotic distribution of $(\hat{\theta}_n, \hat{\phi}_n)$.

THEOREM 2. *Under the conditions of Theorem* 1 *let $(\hat{\theta}_n, \hat{\phi}_n)$ denote a sequence of maximum likelihood estimators satisfying the conclusions of Theorem* 1. *Then*

$$\{(cn \log n)^{\frac{1}{2}}(\hat{\theta}_n - \theta_0),\ n^{\frac{1}{2}}(\hat{\phi}_n - \phi_0)\}$$

converges in distribution to a random vector $(S, Z_1, \ldots, Z_q)$, where S is as defined in (6·7), *and $Z_1, \ldots, Z_q$ are zero mean normal variables. The covariance matrix of the random vector is* diag $(1, M^{-1})$, *where M is as defined in Theorem* 1.

COROLLARY. *We have that*

$$(\bar{\theta}_n - \theta_0)\left\{-\frac{\partial^2 L_n(\bar{\theta}_n, \phi_0)}{\partial \theta^2}\right\}^{\frac{1}{2}} \to S, \tag{8·3}$$

$$(\hat{\theta}_n - \theta_0)\left\{-\frac{\partial^2 L_n(\hat{\theta}_n, \hat{\phi}_n)}{\partial \theta^2}\right\}^{\frac{1}{2}} \to S, \tag{8·4}$$

where convergence is in distribution and S is as defined in (6·7).

Proofs of these results are in the Appendix.

Three interesting points arise from Theorems 1 and 2. First, the results for the asymptotic behaviour of the maximum likelihood estimator of the dependence parameter have a strong resemblance to those for the endpoint of the Generalized Extreme Value distribution when $k = \frac{1}{2}$. Secondly, the corollary shows that the variance of the estimators may be estimated asymptotically by means of the observed information, in a case where the expected information does not exist. In regular estimation problems, Efron & Hinkley (1978) argued by second-order approximations and conditional arguments that the observed information is superior to the expected information as an estimator of variance. Here we have an important practical example in which the superiority of the observed information is easily demonstrated. Finally, the maximum likelihood estimator of the dependence parameter possesses all the desirable properties Cox & Reid (1987) obtain, in regular estimation problems, via orthogonalization. So in our case, despite the expected information not existing, we have the benefits of orthogonality. For further discussion of orthogonality in nonregular problems, see Tawn (1987).

As in regular estimation problems, the asymptotic behaviour of the score and likelihood ratio tests for independence are unchanged by having to estimate the margin parameters.

9. DISCRIMINATING BETWEEN MODELS

So far we have presented, and referenced, many parametric models for fitting bivariate extreme value data, without any method of deciding which model fits the data best.

First we deal with the four differentiable models presented here. These models fall into two separate families. The standard likelihood ratio test is appropriate for testing between models of the same family. In the logistic family, due to boundary problems, the limiting distribution of this is a suitably adjusted chi-squared distribution. All that remains is to discuss testing between the mixed and logistic families. We could use the results of Cox (1961, 1962), but here we suggest using the convexity of the set of dependence functions. Thus, if we wish to discriminate between dependence functions A_1 and A_2, each having the same number of parameters, we fit the model with dependence function

$$A(w) = \gamma A_1(w) + (1-\gamma) A_2(w) \quad (0 \leqslant \gamma \leqslant 1) \tag{9·1}$$

and accept A_1 if $\hat{\gamma} > \frac{1}{2}$, otherwise accept A_2. Problems only really arise when the data are independent, because we then have nonidentifiability. In the case of the mixed and logistic models, we already know the asymptotic behaviour of the score statistics for independence. By similar methods, the joint asymptotic behaviour can be found. Let $\Sigma\, u_M(x_i, y_i)$ and $\Sigma\, u_L(x_i, y_i)$ denote the score for independence for the mixed and logistic models respectively. Then, if we wish to accept independence with probability $1-\alpha$, and accept each of the logistic and mixed models with probability $\frac{1}{2}\alpha$, the test becomes as follows.

Accept independence if both

$$\sum_{i=1}^{n} u_M(x_i, y_i) \leq C_\alpha(\tfrac{1}{2}n \log n)^{\frac{1}{2}}, \quad \sum_{i=1}^{n} u_L(x_i, y_i) \leq C_\alpha(\tfrac{1}{15}n \log n)^{\frac{1}{2}}.$$

Accept the logistic model if both

$$\sum_{i=1}^{n} u_L(x_i, y_i) > C_\alpha(\tfrac{1}{2}n \log n)^{\frac{1}{2}}, \quad (\tfrac{2}{15})^{\frac{1}{2}} \sum_{i=1}^{n} u_L(x_i, y_i) > \sum_{i=1}^{n} u_M(x_i, y_i).$$

Accept the mixed model if both

$$\sum_{i=1}^{n} u_M(x_i, y_i) > C_\alpha(\tfrac{1}{15}n \log n)^{\frac{1}{2}}, \quad \sum_{i=1}^{n} u_M(x_i, y_i) > (\tfrac{2}{15})^{\frac{1}{2}} \sum_{i=1}^{n} u_L(x_i, y_i).$$

Examples of C_α are $C_{0\cdot 1} = 1\cdot 430$, $C_{0\cdot 05} = 1\cdot 789$ and $C_{0\cdot 025} = 2\cdot 100$.

To test between the differentiable and nondifferentiable models, the asymmetric logistic model is a natural tool, as it contains models of both kinds. Because the nondifferentiable models are asymptotically exactly identifiable, there is no need for asymptotic tests. A simple small-sample test could be to reject nondifferentiable models if $r < C$, for some suitable constant C.

As the new nondifferentiable model, given by (5·3), is just the bivariate Marshall & Olkin (1967) distribution, transformed to have unit exponential margins, we see that the asymmetric logistic model leads to a differentiable generalization of this with exponential margins. The survival function corresponding to (5·3) satisfies the Marshall–Olkin loss of memory condition only if $\theta = \phi$. Block & Basu (1974) develop an alternative generalization preserving the loss of memory feature, but, as a result, have mixtures of exponentials for the margins. Our model has the advantage that tests between it and the Marshall–Olkin distribution can easily be constructed.

10. An example

The data we will consider are annual maxima sea levels for Lowestoft and Sheerness, two ports on the east coast of Britain. The data, plotted in Fig. 1, consist of 74 pairs of annual maxima. These correspond to all the years when there is a record of the annual maxima at both ports.

The outlier corresponds to the 1953 flood. As this is the largest observation for each margin, we expect some dependence; but the dependence will be weak, because there are many cases where the values are large at one site but small at the other. The knowledge that the 1953 value corresponds to the same storm at each port shows that our analysis does not take into account all the relevant information. In this analysis, we shall not be concerned with such matters, as similar information is not available for earlier maxima.

For computational ease, we first estimated the parameters of the marginal Generalized Extreme Value distributions. For Lowestoft the estimated parameters were $(\hat{\mu}, \hat{\sigma}, \hat{k}) = (1\cdot 95, 0\cdot 24, -0\cdot 076)$, whereas for Sheerness, which has a linear trend $\mu_n = \alpha + \beta n$, the estimates were $(\hat{\alpha}, \hat{\beta}, \hat{\sigma}, \hat{k}) = (3\cdot 22, 0\cdot 043, 0\cdot 19, -0\cdot 064)$. Using these, the data were transformed to unit exponential distributions. Then the four models in the paper were fitted by maximum likelihood and the referenced ad hoc methods. Table 2 gives the relevant results for the maximum likelihood estimation of these models. In all the cases the standard errors are based on the observed information. For the asymmetric mixed model the joint estimate is on the boundary $(\theta + 3\phi = 0)$ of the parameter space. The likelihood

Fig. 1

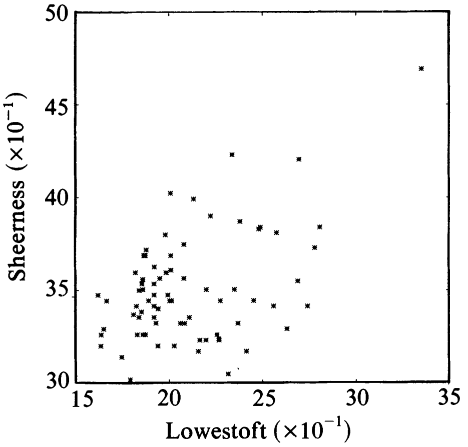

Fig. 2

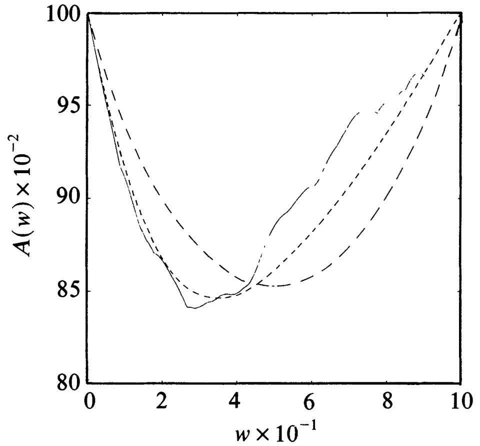

Fig. 1. Annual maxima sea levels, in metres, for Lowestoft versus Sheerness.

Fig. 2. The dependence function for Lowestoft versus Sheerness. Solid line, nonparametric estimator; dotted and dashed line, logistic model; broken line, asymmetric logistic model.

Table 2. *Estimation results for the differentiable models*

Model	L	Standardized score for independence	Estimate and standard errors
Independence	−147.80	—	—
Mixed	−141·17	11·75	$\bar{\theta} = 0{\cdot}56\ (0{\cdot}18)$
Asym. mixed	−139·40	—	$\bar{\theta} = 1{\cdot}28\ (0{\cdot}24)$, $\bar{\phi} = 0{\cdot}43\ (0{\cdot}09)$
Logistic	−140·61	10·04	$\bar{r} = 1{\cdot}30\ (0{\cdot}12)$
Asym. logistic, $\phi = 1$	−138·60	—	$\bar{r} = 1{\cdot}96\ (0{\cdot}54)$, $\bar{\theta} = 0{\cdot}34\ (0{\cdot}12)$

Asym., asymmetrical; L, maximized log likelihood.

based 95% confidence intervals for the logistic and mixed models are $(0{\cdot}23, 0{\cdot}85)$, $(1{\cdot}11, 1{\cdot}52)$ respectively. These are surprisingly symmetric considering the sample size involved. The profile likelihood based interval for the asymmetric logistic model is slightly skew, but this may be overcome by using the parameterization $\nu = 1/r$.

In the mixed, logistic and asymmetric mixed models we can test independence. For these models, each of the tests developed in the paper gives that we should definitely reject independence, even though the estimated correlation is approximately 0·24. As mentioned in § 7, testing independence in the asymmetric logistic model is more difficult. It is interesting that with this model, if we assume $\phi = 1$, then the maximum likelihood estimates are not within two standard errors of independence, $r = 1$ or $\theta = 0$.

The maximum likelihood estimators are invariant to the chosen marginal distribution, whereas nearly all the ad hoc methods are not. If ad hoc methods are to be used, experience suggests that estimation using Gumbel margins is preferable to estimation using exponential margins, because of the stability of the transformation. The ad hoc methods gave estimates which were reasonably close to the maximum likelihood estimates, but with much larger asymptotic confidence intervals.

We must now decide which model to use. The asymmetric logistic model is significant at the $2\frac{1}{2}$% level, whereas the evidence for the asymmetric mixed model is not as strong, being significant only at the 10% level. To decide between the separate families, the model in (9·1) was fitted, with A_1 logistic and A_2 mixed. We obtained $\hat{\gamma}=1$ so we accept the logistic family and hence the asymmetric logistic model. Because $\bar{r}$ is not very large in the asymmetric logistic model, we need not consider fitting the nondifferentiable models.

In Fig. 2 we compare the nonparametric estimator, given by (4·1), with the logistic and asymmetric logistic models. Pickands (1981) proposes the convex hull of (4·1) as an estimator of the dependence function. This would give a much closer fit to each of the parametric models, but has the disadvantage of giving a nondifferentiable model. The asymmetric logistic model appears to be a convex smoother of the nonparametric estimators. The quality of the fit illustrates the flexibility of the new models.

Now we have a model, we can illustrate the importance of considering dependence and nonexchangeability. To do this, we examine the influence of these factors on the design heights of sea defences. In particular, we compare the joint quantiles required to protect against certain types of events. Let p_1 denote an exceedance probability: then

$$x_{p_1}=\mu_1+\sigma_1 k_1^{-1}[1-\{-\log(1-p_1)\}^{k_1}]$$

is the level, which is exceeded by the annual maximum, at site 1, with probability p_1. Similarly, let y_{p_2} denote the level exceeded by the annual maximum at site 2 with probability p_2. In our analysis Lowestoft and Sheerness are sites 1 and 2 respectively. A minor problem is that Sheerness has a linear trend, so we consider quantiles only for the year 1987.

If the sea walls are of height (x_{p_1}, y_{p_2}), we are interested in three different probabilities:

(i) the probability, p_*, that at least one port floods,

$$p_*=1-\mathrm{pr}(X<x_{p_1}, Y<y_{p_2})=1-\{(1-p_1)(1-p_2)\}^{A(s)},$$

where $s=\log(1-p_2)/\log\{(1-p_1)(1-p_2)\}$;

(ii) the probability, p^*, of flooding at both ports, $p^*=p_1+p_2-p_*$;

(iii) the probability, $\tilde{p}_{i,j}$, that the level at site j is less extreme than that observed at site i, $\tilde{p}_{1,2}=\mathrm{pr}(Y<y_p \mid X=x_p)$.

To obtain an optimal choice of x_{p_1} and y_{p_2} requires a loss function. As such a function is not easily available, we consider the choice $p_1=p_2=p$ for examining (i) and (ii). As $1\leqslant 2A(\frac{1}{2})\leqslant 2$, then for small fixed p_* we have $p \simeq p_*/\{2A(\frac{1}{2})\}$ so $\frac{1}{2}p_*<p\leqslant p_*$. In our example, for fixed p_*, the effect of dependence on quantiles is weak, as it makes little difference if we assume independence instead of the estimated asymmetric logistic model.

For fixed p^*, the effect of dependence on quantiles is very strong, as is seen in Table 3. In practice, the sea defences are designed with $p_1 \simeq p_2$ and $0{\cdot}001<p_1<0{\cdot}01$. At present,

Table 3. *Comparison of quantiles under independence and the estimated model for various probabilities of joint flooding*

	Model				
	Asymmetric logistic		Independence		p^* if independence
p^*	X_p	Y_p	X_p	Y_p	is wrongly assumed
0·25	2·08	3·66	2·04	3·62	0·30
5×10^{-2}	2·44	3·93	2·29	3·83	$9{\cdot}4\times10^{-2}$
1×10^{-2}	2·86	4·28	2·53	4·02	$3{\cdot}4\times10^{-2}$
1×10^{-3}	3·61	4·85	2·88	4·28	$9{\cdot}5\times10^{-3}$
1×10^{-4}	4·52	5·53	3·25	4·57	$2{\cdot}9\times10^{-3}$
1×10^{-5}	5·61	6·32	3·65	4·88	$8{\cdot}9\times10^{-4}$

the design takes no account of sea levels at other ports, so present procedures effectively assume independence between joint extremes. An interesting associated probability is p^*_{ij}, the probability of flooding at site j given that flooding occurred at site i, because

$$\lim_{p\to 0} p^*_{12} = \lim_{p\to 0} p^*_{21} = 2\{1 - A(\tfrac{1}{2})\}.$$

The limit is the dependence function, introduced in § 3, in our case, the limit is 0·28, whereas for independence, it is zero. The limit is a good approximation for $p \leqslant 0{\cdot}01$.

So far we have identified only the dependence. The influence of nonexchangeability is best illustrated by examining $R = \tilde{p}_{12}/\tilde{p}_{21}$. If the data are symmetric then $R = 1$, whereas for the estimated asymmetric logistic model, $R = 0{\cdot}819$. This implies that if we observe a level at Sheerness with exceedance probability p, then the probability that the level is less extreme at Lowestoft is greater than with the ports interchanged.

Obtaining the significant asymmetric result for extreme levels at ports along the east cost of England is pleasing, because it is supported by physical reasons. These are based on the dynamics of the North Sea and the knowledge of the weather systems that induce storms. An additional advantage in having the best fitting model for dependence could be in fitting covariates to the margin parameters, thus enabling predictions to be made at ports where only a limited sea level record is available.

Acknowledgements

The work was supported at University of Surrey by the Science and Engineering Research Council in the form of a CASE Studentship with the Proudman Oceanographic Laboratory, formerly the Institute of Oceanographic Sciences (Bidston). I would particularly like to thank Professor R. L. Smith, Mr H. K. Yuen and the referees for many helpful suggestions.

Appendix

Proofs of Theorems 1 and 2 and the Corollary

Before we give the proof of Theorem 1, we require the following lemma, which is a simple extension of Aitchison & Silvey (1958, Lemma 2).

LEMMA. *Let t be a continuously differentiable real-valued function of $q+1$ variables x, where the first variable x_1 is restricted to $-\infty < e \leqslant x_1$. Let T denote the gradient vector of t. Suppose that the scalar product of x and $T(x)$ is negative whenever $|x| = 1$ and $e \leqslant x_1$. Then t has a local maximum for some x with $|x| < 1$ and $e \leqslant x_1$.*

Proof of Theorem 1. Except for the problems associated with the restricted parameter space, the proof is similar to Smith (1985a, Th. 3(ii)). Therefore, here we will give only an outline of the proof, emphasizing the boundary problems. The argument hinges on showing that $\hat{\theta}_n$, $\hat{\phi}_n$ are respectively near $\bar{\theta}_n$, $\bar{\phi}_n$. Because of the boundary problem, we do not necessarily have $\partial L_n(\bar{\theta}_n, \phi_0)/\partial\theta = 0$. Therefore

$$\begin{aligned}\frac{\partial L_n(\theta,\phi)}{\partial\theta} &= \frac{\partial L_n(\theta,\phi)}{\partial\theta} - \frac{\partial L_n(\bar{\theta}_n,\phi_0)}{\partial\theta} + \frac{\partial L_n(\bar{\theta}_n,\phi_0)}{\partial\theta}\\ &= (\theta - \bar{\theta}_n)\frac{\partial^2 L_n(\theta^+,\phi^+)}{\partial\theta^2} + \sum_{j=1}^{q}(\phi^j - \phi_0^j)\frac{\partial^2 L_n(\theta^+,\phi^+)}{\partial\phi^j\,\partial\theta} + \frac{\partial L_n(\theta_n,\phi_0)}{\partial\theta},\end{aligned}$$

where (θ^+, ϕ^+) is on the hyperline joining (θ, ϕ) and $(\bar{\theta}_n, \phi_0)$. So that we have

$$n^{-1}\frac{\partial L_n(\theta,\phi)}{\partial\theta} = (\theta - \bar{\theta}_n)n^{-1}\frac{\partial^2 L_n(\theta^+,\phi^+)}{\partial\theta} - \sum_{j=1}^{q}(\phi^j - \phi_0^j)m_j^0 + n^{-1}\frac{\partial L_n(\bar{\theta}_n,\phi_0)}{\partial\theta} + e_{1,n}(\theta,\phi), \quad \text{(A·1)}$$

with m_j^0 defined by (8·2). Similarly for $i=1,\ldots,q$

$$n^{-1}\frac{\partial L_n(\theta,\phi)}{\partial\phi^i}=-(\theta-\theta_0)m_i^0-\sum_{j=1}^{q}(\phi^j-\bar{\phi}_n^j)m_{ij}+e_{2,n}(\theta,\phi). \qquad (A\cdot2)$$

By § 6, we have that $\bar{\theta}_n-\theta_0<_{\mathrm{p}}(n\log n)^{-\frac{1}{2}}$, and from Smith (1985a) that

$$\bar{\phi}_n^j-\phi_0^j<_{\mathrm{p}} n^{-\frac{1}{2}} \quad (j=1,\ldots,q). \qquad (A\cdot3)$$

For $y\in R^q$ and $x\in R$, constrained such that $\bar{\theta}_n+xn^{-\frac{1}{2}}(\log n)^{-5/6}\geqslant\theta_0$, let

$$f_n(x,y)=(\log n)^{\beta(\bar{\theta}_n)}L_n\{\bar{\theta}_n+xn^{-\frac{1}{2}}(\log n)^{-5/6},\ \bar{\phi}_n+yn^{-\frac{1}{2}}(\log n)^{-1/3}\},$$

where

$$\beta(\bar{\theta}_n)=\begin{cases}\frac{2}{3} & (\bar{\theta}_n>\theta_0),\\ \frac{1}{3} & (\bar{\theta}_n=\theta_0,\ x>0),\\ \frac{2}{3} & (\bar{\theta}_n=\theta_0,\ x=0).\end{cases}$$

Here we will consider only the case $x>0$: the case $x=0$ is immediate. Now by (A·1)

$$\frac{\partial f_n}{\partial x}=n^{\frac{1}{2}}(\log n)^{\beta(\bar{\theta}_n)-5/6}\Bigg[xn^{-3/2}(\log n)^{-5/6}\frac{\partial^2L_n(\theta^+,\phi^+)}{\partial\theta^2}$$
$$-\sum_{j=1}^{q}\{\bar{\phi}_n^j-\phi_0^j+y^jn^{-\frac{1}{2}}(\log n)^{-1/3}\}m_j^0+n^{-1}\frac{\partial L_n(\bar{\theta}_n,\phi_0)}{\partial\theta}$$
$$+e_{1,n}\{\bar{\theta}_n+xn^{-\frac{1}{2}}(\log n)^{-5/6},\ \bar{\phi}_n+yn^{-\frac{1}{2}}(\log n)^{-1/3}\}\Bigg].$$

But if $\bar{\theta}_n>\theta_0$ then $\partial L_n(\bar{\theta}_n,\phi_0)/\partial\theta=0$, whereas if $\bar{\theta}_n=\theta_0$ then $\partial L_n(\bar{\theta}_n,\phi_0)/\partial\theta<0$. Also it can be shown in probability

$$-(n\log n)^{-1}\frac{\partial^2L_n(\theta^+,\phi^+)}{\partial\theta^2}\to m^{00}>0$$

and so

$$\frac{\partial f_n}{\partial x}\to\begin{cases}-xm^{00} & (\bar{\theta}_n>\theta_0),\\ -Z_+ & (\bar{\theta}_n=\theta_0),\end{cases} \qquad (A\cdot4)$$

where in the case $\bar{\theta}_n>\theta_0$ convergence is in probability, and in the case $\bar{\theta}_n=\theta_0$ convergence is in distribution. Here Z_+ is a positive random variable, with distribution function $2\Phi(z)-1$, for $z\geqslant0$, obtained as in § 6.

Similarly for $i=1,\ldots,q$, by (A·2) in probability

$$\frac{\partial f_n}{\partial y^i}\to\begin{cases}-\Sigma\, y^jm_{ij} & (\bar{\theta}_n>\theta_0),\\ 0 & (\bar{\theta}_n=\theta_0),\end{cases} \qquad (A\cdot5)$$

where the sum is over $j=1,\ldots,q$. Combining (A·4) and (A·5), we have

$$x\frac{\partial f_n}{\partial x}+\sum_{i=1}^{q}y^i\frac{\partial f_n}{\partial y^i}\to\begin{cases}-x^2m^{00}-yMy^{\mathrm{T}}<0 & (\bar{\theta}_n>\theta_0),\\ -xZ_+<0 & (\bar{\theta}_n=\theta_0).\end{cases}$$

In the case $\bar{\theta}_n>\theta_0$ the result is obtained by the strict positive-definite property of M and $m^{00}>0$. When $\bar{\theta}_n=\theta_0$ the result follows because $x>0$ and Z_+ is a positive random variable. By the Lemma f_n has, with probability tending to one, a local maximum within the truncated ball,

$x^2+|y|^2 \leqslant \delta$ such that $\bar{\theta}_n + xn^{-\frac{1}{2}}(\log n)^{-5/6} \geqslant \theta_0$, for all n and any fixed $\delta > 0$. Hence L_n has a local maximum at $(\hat{\theta}_n, \hat{\phi}_n)$ satisfying

$$\hat{\theta}_n - \bar{\theta}_n <_{\mathrm{p}} n^{-\frac{1}{2}}(\log n)^{-5/6}, \quad \hat{\phi}^j_n - \bar{\phi}^j_n <_{\mathrm{p}} n^{-\frac{1}{2}}(\log n)^{-1/3} \quad (j=1,\ldots,q) \tag{A·6}$$

with probability tending to one. The results follow immediately from (A·3) and (A·6). □

Proof of Theorem 2. Since by Theorem 1 in probability

$$(n \log n)^{\frac{1}{2}}(\hat{\theta}_n - \bar{\theta}_n) \to 0, \quad n^{\frac{1}{2}}(\hat{\phi}^j_n - \bar{\phi}^j_n) \to 0 \quad (j=1,\ldots,q),$$

it suffices to prove the result with $\bar{\theta}_n, \bar{\phi}_n$ in place of $\hat{\theta}_n, \hat{\phi}_n$. In § 6 we found the asymptotic distribution of $\bar{\theta}_n$, whereas the asymptotic joint distribution of $\bar{\phi}_n$ is given by Smith (1985a, Th. 3(i)). Therefore, we need only to show the asymptotic independence of $\bar{\theta}_n$ and $\bar{\phi}_n$. By previous arguments, if $\bar{\theta}_n > \theta_0$

$$(cn \log n)^{\frac{1}{2}}(\bar{\theta}_n - \theta_0) = (cn \log n)^{-\frac{1}{2}} \frac{\partial L_n(\theta_0, \phi_0)}{\partial \theta} W_n,$$

where $W_n \to 1$ in probability. So, unconditionally, the left-hand side converges to the random variable S, with distribution function given by (6·7). Similar arguments applied to $\bar{\phi}_n - \phi_0$ together with the Cramér–Wold device and an adaption of Lemma 6 of Smith (1985a) allow us to assert that S is independent of the asymptotic distribution of $n^{\frac{1}{2}}(\bar{\phi}_n - \phi_0)$ as required. □

Proof of Corollary. We can write (8·3) as

$$(cn \log n)^{\frac{1}{2}}(\bar{\theta}_n - \theta_0)\left\{-(cn \log n)^{-1} \frac{\partial^2 L_n(\bar{\theta}_n, \phi_0)}{\partial \theta^2}\right\}^{\frac{1}{2}},$$

the product of two factors, the first converging to S, defined in (6·7), and the second to one; this gives (8·3). By a similar approach we obtain (8·4). □

REFERENCES

AITCHISON, J. & SILVEY, S. D. (1958). Maximum likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.* **29**, 813–28.

BARNETT, V. (1976). The ordering of multivariate data (with discussion). *J. R. Statist. Soc.* A **139**, 318–54.

BERMAN, M. (1986). Some unusual examples of extrema associated with hypothesis tests when nuisance parameters are present only under the alternative. In *Pacific Statistical Congress*, Ed. I. S. Francis, B. F. J. Manly and F. C. Lam, pp. 141–6. Amsterdam: North-Holland.

BLOCK, H. W. & BASU, A. P. (1974). A continuous bivariate exponential extension. *J. Am. Statist. Assoc.* **69**, 1031–7.

COX, D. R. (1961). Tests of separate families of hypotheses. *Proc. 4th Berkeley Symp.* **1**, 105–23.

COX, D. R. (1962). Further results on tests of separate families of hypotheses. *J. R. Statist. Soc.* B **24**, 406–24.

COX, D. R. & REID, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc.* B **49**, 1–39.

CROWDER, M. J. (1988). A multivariate distribution with Weibull connections. *J. R. Statist. Soc.* B. To appear.

DAVIES, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64**, 247–54.

DEHEUVELS, P. (1983). Point processes and multivariate extreme values. *J. Mult. Anal.* **13**, 257–71.

DEHEUVELS, P. (1984). Probabilistic aspects of multivariate extremes. In *Statistical Extremes and Applications*, Ed. J. Tiago de Oliveira, pp. 117–30. Dordrecht: Reidel.

DEHEUVELS, P. (1985). Point processes and multivariate extreme values, II. In *Multivariate Analysis VI*, Ed. P. R. Krishnaiah, pp. 145–64. Amsterdam: North-Holland.

EFRON, B. & HINKLEY, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika* **65**, 457–87.

FELLER, W. (1971). *An Introduction to Probability Theory and its Applications*, **2**, 2nd ed. New York: Wiley.

FISHER, R. A. & TIPPETT, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.* **24**, 180–90.

GALAMBOS, J. (1978). *The Asymptotic Theory of Extreme Order Statistics.* New York: Wiley.

GEFFROY, J. (1958/59). Contributions à la théorie des valeurs extrème. *Publ. Inst. Statist. Univ. Paris*, **7/8**, 37–185.
GUMBEL, E. J. (1960). Bivariate exponential distributions. *J. Am. Statist. Assoc.* **55**, 698–707.
GUMBEL, E. J. & GOLDSTEIN, N. (1964). Analysis of empirical bivariate extremal distributions. *J. Am. Statist. Assoc.* **59**, 794–816.
GUMBEL, E. J. & MUSTAFI, C. K. (1967). Some analytical properties of bivariate extremal distributions. *J. Am. Statist. Assoc.* **62**, 269–88.
HAAN, L. DE & RESNICK, S. I. (1977). Limit theory for multivariate sample extremes. *Zeit. Wahrscheinl.-theorie* **40**, 317–37.
HALL, P. (1983). Fast rates of convergence in the central limit theorem. *Zeit. Wahrscheinl.-theorie* **62**, 491–507.
HOUGAARD, P. (1986). A class of multivariate failure time distributions. *Biometrika* **73**, 671–8.
LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37**, 1137–53.
MARDIA, K. V. (1964). Asymptotic independence of bivariate extremes. *Calcutta Statist. Assoc. Bull.* **13**, 172–8.
MARSHALL, A. W. & OLKIN, I. (1967). A multivariate exponential distribution. *J. Am. Statist. Assoc.* **62**, 30–44.
MORAN, P. A. P. (1971). Maximum likelihood estimation in non-standard conditions. *Proc. Camb. Phil. Soc.* **70**, 441–50.
PICKANDS, J. (1981). Multivariate extreme value distributions. *Bull. Int. Statist. Inst.* 859–78.
POSNER, E. C., RODERICK, E. R., ASHLOCK, J. C. & LURIE, S. (1969). Application of an estimator of high efficiency in bivariate extreme value theory. *J. Am. Statist. Assoc.* **64**, 1403–14.
PRESCOTT, P. & WALDEN, A. T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika* **67**, 723–4.
PRESCOTT, P. & WALDEN, A. T. (1983). Maximum likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution from censored samples. *J. Statist. Comput. Simul.* **16**, 241–50.
SIBUYA, M. (1960). Bivariate extreme statistics. *Ann. Inst. Statist. Math.* **11**, 195–210.
SMITH, R. L. (1985a). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72**, 67–90.
SMITH, R. L. (1985b). Statistics of extreme values. *Bull. Int. Statist. Inst.* **26.1**, 1–17.
TAWN, J. A. (1987). Discussion of paper by D. R. Cox and N. Reid. *J. R. Statist. Soc.* B **49**, 33–4.
TIAGO DE OLIVEIRA, J. (1958). Extremal distributions. *Rev. Fac. Ciências Lisboa*, 2 ser., A Mat. **7**, 215–27.
TIAGO DE OLIVEIRA, J. (1962/63). Structure theory of bivariate extremes, extensions. *Est. Mat., Estat. e. Econ.* **7**, 165–95.
TIAGO DE OLIVEIRA, J. (1975). Statistical decision for extremes. *Trab. Estad. y Inv. Oper.* **26**, 433–71.
TIAGO DE OLIVEIRA, J. (1980). Bivariate extremes; foundations and statistics. In *Multivariate Analysis V*, Ed. P. R. Krishnaiah, pp. 349–66. Amsterdam: North-Holland.
TIAGO DE OLIVEIRA, J. (1984). Bivariate models for extremes; statistical decision. In *Statistical Extremes and Applications*, Ed. J. Tiago de Oliveira, pp. 131–153. Dordrecht: Reidel.
WEISSMAN, I. (1985). Multidimensional extreme value theory (Discussion). *Bull. Int. Statist. Inst.* Book 5, 187–90.
WOODROOFE, M. (1972). Maximum likelihood estimation of a truncated distribution. *Ann. Math. Statist.* **43**, 113–22.

[*Received June* 1987. *Revised December* 1987]