Bernoulli **12**(5), 2006, 917–930

Multivariate generalized Pareto distributions

HOLGER ROOTZÉN[1] and NADER TAJVIDI[2]

[1]*Chalmers University of Technology, S-412 96 Göteborg, Sweden. E-mail: rootzen@math.chalmers.se*
[2]*Lund Institute of Technology, Box 118, S-221 00 Lund, Sweden. E-mail: tajvidi@maths.lth.se*

Statistical inference for extremes has been a subject of intensive research over the past couple of decades. One approach is based on modelling exceedances of a random variable over a high threshold with the generalized Pareto (GP) distribution. This has proved to be an important way to apply extreme value theory in practice and is widely used. We introduce a multivariate analogue of the GP distribution and show that it is characterized by each of following two properties: first, exceedances asymptotically have a multivariate GP distribution if and only if maxima asymptotically are extreme value distributed; and second, the multivariate GP distribution is the only one which is preserved under change of exceedance levels. We also discuss a bivariate example and lower-dimensional marginal distributions.

Keywords: generalized Pareto distribution; multivariate extreme value theory; multivariate Pareto distribution; non-homogeneous Poisson process; peaks-over-threshold method

1. Introduction

Statistical modeling of extreme values has developed extensively over recent decades. This is witnessed by several recent books (Coles 2001; Embrechts *et al.* 1997; Kotz and Nadarajah 2000; Kowaka 1994; Beirlant *et al.* 2004; Reiss and Thomas 2005) and a large journal literature which includes both theoretical and applied papers, the latter concerned with a wide range of important problems such as extreme wind speeds, waveheights, floods, insurance claims, price fluctuations, ... For references to some of this literature, see Kotz and Nadarajah (2000) and Beirlant *et al.* (2004).

The main emphasis has been on univariate extremes, and so far the univariate results are the most complete and the most directly usable. Two main sets of methods, the block maxima method and the peaks-over-threshold method, have been developed (Coles 2001). Here we only consider independent and identically distributed variables. However, the methods are also widely useful for dependent and non-stationary situations.

In the *block maxima* method one is supposed to have observed the maximum values of some quantities over a number of 'blocks', a typical example being that a block is a year and the observed quantities may be some environmental quantity such as the wind speed at a specific location. In this method, the block maxima are modelled by an extreme value (EV) distribution with distribution function (d.f.)

$$G(x) = \exp\left\{-\left(1+\gamma\frac{x-\mu}{\sigma}\right)_+^{-1/\gamma}\right\}.$$

This choice of distribution is motivated by the fact that the EV distributions are the only ones which can appear as the limit of linearly normalized maxima, and that they are the only ones which are 'max-stable', that is, such that a change of block size only leads to a change of location and scale parameter in the distribution.

In the *peaks-over-threshold* method, on the other hand, one is supposed to have observed all values which are larger than some suitable threshold, for example all wind speeds in excess of 20 m/s. These values are then assumed to follow the generalized Pareto (GP) distribution with d.f.

$$H(x) = 1 - \left(1 + \gamma \frac{x}{\sigma}\right)_+^{-1/\gamma}.$$

This choice of distributions is motivated by characterizations due to Balkema and de Haan (1974) and Pickands (1975). One characterization is that the distribution of a scale-normalized exceedance over a threshold asymptotically (as the threshold tends to the right-hand endpoint of the distribution) converges to a GP distribution if and only if the distribution of block maxima converges (as the block length tends to infinity) to an EV distribution. The other one is that the GP distributions are the only 'stable' ones, that is, the only ones for which the conditional distribution of an exceedance is a scale transformation of the original distribution. Pickands gives the full statement of this, although we believe there is a small gap in his proof. Balkema and de Haan only consider the infinite-endpoint case, but give a complete proof. Some basic papers on the peaks-over-threshold method are Smith (1985, 1987), Smith *et al.* (1990, 1997) and Davidson and Smith (1990). Ledford and Tawn (1996) develop threshold-based models for joint tail regions of multivariate extreme for asymptotically independent cases. Since the peaks-over-threshold method uses more of the data it can sometimes result in better estimation precision than the block maxima method. As an aside, there are several other variants of the (one-dimensional) Pareto distribution; see, for example, Arnold (1983).

Multivariate EV distributions arise in connection with extremes of a random sample from a multivariate distribution. They are extensively discussed by Resnick (1987), Kotz and Nadarajah (2000) and Beirlant *et al.* (2004) and in the review by Fougères (2004). Several recent papers (e.g. Joe *et al.* 1992; Coles and Tawn 1991; Tawn 1988, 1990; Smith *et al.* 1990) have explored their statistical application.

There are several possibilities for ordering multivariate data; see the review by Barnett (1976). For extreme values the most widely used method is the marginal or M-ordering where the maximum is defined by taking componentwise maxima. Then, for a series of vectors $\{\mathbf{X}_i,\ i \geqslant 1\} = \{(X_i^{(1)}, \ldots, X_i^{(d)}),\ i \geqslant 1\}$, the maximum, $\mathbf{M}_n$, is defined by $\mathbf{M}_n = (M_n^{(1)}, \ldots, M_n^{(d)}) = (\bigvee_{i=1}^n X_i^{(1)}, \ldots, \bigvee_{i=1}^n X_i^{(d)})$, where $\bigvee$ denotes maximum. Under rather general conditions, the distribution of the linearly normalized $\mathbf{M}_n$ converges to a multivariate EV distribution. In applications $\mathbf{M}_n$ is often the vector of annual maxima, and block maxima methods can be applied similarly to when the observations are one-dimensional. However, as in the univariate case it is also of interest to study methods which utilize more of the data and which can contribute to better estimation of parameters. For multivariate observations a further reason to study such methods is that block maxima hide the 'time structure' since they do not show if the component maxima have occurred simultaneously or not.

The aim of this paper is to define the multivariate GP distributions and to prove that this definition indeed is the right one. The multivariate GP distribution should (a) be the natural distribution for exceedances of high thresholds by multivariate random vectors, and (b) describe what happens to the other components when one or more of the components exceed their thresholds. In complete analogy with the one-dimensional case we interpret (a) to mean that the multivariate GP distribution should be characterized by each of the following two properties:

- exceedances (of suitably coordinated levels) asymptotically have a multivariate GP distribution if and only if componentwise maxima asymptotically are EV distributed;
- the multivariate GP distribution is the only one which is preserved under (a suitably coordinated) change of exceedance levels.

In the next section we prove that this indeed is the case for the definition given in this paper. The section also explains the caveat 'suitably coordinated levels'. Further, requirement (b) is taken care of by the choice of support for the GP distribution.

There is a close connection between the multivariate GP distribution and the multivariate point process methods used in Coles and Tawn (1991) and Joe *et al.* (1992); see Section 2. In Section 3 we show how an 'explicit' formula for the multivariate EV distributions directly leads to a corresponding expression for the multivariate GP distributions and also give a few concrete examples. Lower-dimensional 'marginals' of multivariate GP distributions may be thought of in different ways. This is discussed in Section 4. Proofs are given in Section 5.

This paper is a further development of the results in Tajvidi (1996). A set of related work is Falk and Reiss (2001, 2002, 2003a, 2003b, 2005) which introduced a class of distributions named bivariate GP distributions with uniform margins and suggested a canonical parameterization for the distributions. The papers also discussed estimation and asymptotic normality. To the best of our knowledge no multivariate generalization of these distributions has been discussed by these authors.

2. Multivariate generalized Pareto distributions

In this section we give the formal definition of the multivariate GP distribution and reformulate the motivating characterizations in mathematical terms. Proofs of the characterizations are given in Section 5. However, we begin with some preliminaries.

Suppose $\{\mathbf{X}_i,\ i \geqslant 1\} = \{(X_i^{(1)}, \ldots, X_i^{(d)}),\ i \geqslant 1\}$ are independent and identically distributed d-dimensional random vectors with d.f. F. As before, let $\mathbf{M}_n$ be the vector of componentwise maxima,

$$\mathbf{M}_n = (M_n^{(1)}, \ldots, M_n^{(d)}) = \left(\bigvee_{i=1}^{n} X_i^{(1)}, \ldots, \bigvee_{i=1}^{n} X_i^{(d)} \right).$$

Assume that there exist normalizing constants $\sigma_n^{(i)} > 0$, $u_n^{(i)} \in \mathbb{R}$, $1 \leqslant i \leqslant d$, $n \geqslant 1$, such that as $n \to \infty$,

$$P[(M_n^{(i)} - u_n^{(i)})/\sigma_n^{(i)} \leqslant x^{(i)},\ 1 \leqslant i \leqslant d] = F^n(\sigma_n^{(1)}x^{(1)} + u_n^{(1)}, \ldots, \sigma_n^{(d)}x^{(d)} + u_n^{(d)}) \to G(\mathbf{x}) \tag{1}$$

with the limit distribution G such that each marginal G_i, $i = 1, \ldots, d$, is non-degenerate. If (1) holds, F is said to be in the domain of attraction of G, and we write $F \in D(G)$, and G is said to be a multivariate extreme value distribution.

By setting all x except $x^{(i)}$ to $+\infty$ it is seen that each marginal G_i of G must be an EV d.f., so that

$$G_i(x) = \exp\left(-\left(1 + \gamma_i \frac{x - \mu_i}{\sigma_i}\right)_+^{-1/\gamma i}\right).$$

Here μ_i is a location parameter, $\sigma_i > 0$ is a scale parameter, γ_i is a shape parameter, and the '+' signifies that if the expression in parentheses is negative then it should be replaced by 0. For $\gamma_i = 0$ the expression for the d.f. should be interpreted to mean $\exp(-\exp(-(x - \mu_i)/\sigma_i)$.

As in the univariate case, a multivariate convergence of types argument shows that the class of limit d.f.s for (1) is the class of max-stable distributions, where a d.f. G in $\mathbb{R}^d$ is max-stable if, for $i = 1, \ldots, d$ and every $t > 0$, there exist functions $\alpha^{(i)}(t) > 0$, $\beta^{(i)}(t)$ such that

$$G^t(x) = G(\alpha^{(1)}(t)x^{(1)} + \beta^{(1)}(t), \ldots, \alpha^{(d)}(t)x^{(d)} + \beta^{(d)}(t)).$$

It is convenient to have a convention to handle vectors occurring in the same expression but not all of the same length. We use the convention that the value of the expression is a vector with the same length as that of the longest vector occurring in the expression. Shorter vectors are *recycled* as often as need be, perhaps fractionally, until they match the length of the longest vector. In particular, a single number is repeated the appropriate number of times. All operations on vectors are performed element by element. For example, if $\mathbf{x}$ and $\mathbf{y}$ are bivariate vectors and α is a scalar, then we have

$$\alpha\mathbf{x} = (\alpha x_1, \alpha x_2), \qquad \alpha + \mathbf{x} = (\alpha + x_1, \alpha + x_2)$$

and

$$\mathbf{xy} = (x_1 y_1, x_2 y_2), \qquad \mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2).$$

This convention also applies when we take the supremum or infimum of a set, so that, for example, a coordinate of the supremum of a set is the supremum of all the values this coordinate takes in the set. Thus, for $\mathbf{u}_n = (u_n^{(1)}, \ldots u_n^{(d)})$, $\boldsymbol{\sigma}_n = (\sigma_n^{(1)}, \ldots \sigma_n^{(d)})$, we can write (1) as

$$P((\mathbf{M}_n - \mathbf{u}_n)/\boldsymbol{\sigma}_n \leqslant \mathbf{x}) \to G(\mathbf{x}).$$

For the definition we also use the convention that $0/0 = 1$. The definition has also independently been noticed by Beirlant *et al.* (2004: Chapter 9).

Definition 2.1. *A distribution function H is a multivariate generalized Pareto distribution if*

$$H(\mathbf{x}) = \frac{1}{-\log G(\mathbf{0})} \log \frac{G(\mathbf{x})}{G(\mathbf{x} \wedge \mathbf{0})} \tag{2}$$

for some extreme value distribution G with non-degenerate margins and with $0 < G(\mathbf{0}) < 1$. In particular, $H(\mathbf{x}) = 0$ for $\mathbf{x} < \mathbf{0}$ and $H(\mathbf{x}) = 1 - \log G(\mathbf{x})/\log G(\mathbf{0})$ for $\mathbf{x} > \mathbf{0}$.

Perhaps more elegantly, the class of multivariate GP distributions could alternatively be taken to be all distributions of the form

$$H(\mathbf{x}) = \log \frac{G(\mathbf{x})}{G(\mathbf{x} \wedge \mathbf{0})},$$

for G an EV distribution with $G(\mathbf{0}) = \mathrm{e}^{-1}$. This is not less general than (2) since if we let $t = 1/(-\log G(\mathbf{0}))$ then the H in (2) is of the form $\log(G(\mathbf{x})^t/G(\mathbf{x} \wedge \mathbf{0})^t)$, and by max-stability G^t is again an EV distribution, with $G(\mathbf{0})^t = \exp(-\log G(\mathbf{0})/\log G(\mathbf{0})) = \mathrm{e}^{-1}$. However, for statistical applications one would want to parameterize G, and then the form (2) is more convenient.

As mentioned in the Introduction, there is a strong connection between this multivariate GP distribution and the point process approach of Coles and Tawn (1991) and Joe *et al.* (1992). The major difference is that (2) holds for all values $\mathbf{x} \nless \mathbf{0}$ whereas in the point process approach only values $\mathbf{x} > \mathbf{0}$ are modelled parametrically. This might prove to be an improvement for statistical analysis of extremes, since then negative x_is also contribute to making inference on the distribution.

Our first motivation for this definition is the following theorem. It shows that exceedances (of suitably coordinated levels) asymptotically have a multivariate GP distribution if and only if maxima are asymptotically EV distributed. To state the theorem, let $\mathbf{X}$ be a d-dimensional random vector with d.f. F and write $\bar{F} = 1 - F$ for the tail function of a distribution F. Further, let $\{\mathbf{u}(t) \mid t \in [1, \infty)\}$ be a d-dimensional curve starting at $\mathbf{u}(1) = \mathbf{0}$, let $\boldsymbol{\sigma}(\mathbf{u}) = \boldsymbol{\sigma}(\mathbf{u}(t)) > \mathbf{0}$ be a function with values in $\mathbb{R}^d$, and let

$$\mathbf{X}_\mathbf{u} = \frac{\mathbf{X} - \mathbf{u}}{\boldsymbol{\sigma}(\mathbf{u})}$$

be the vector of normalized exceedances of the levels $\mathbf{u}$. In the characterizations we consider exceedances of d levels which 'tend to infinity' (interpreted to mean that the levels move further and further out into the tails of F). However, asymptotic distributions can differ for different relations between the levels. The components of the curve $\{\mathbf{u}(t)\}$ give these levels and the curve specifies how the levels increase 'in a suitably coordinated way'.

Theorem 2.1. (i) *Suppose G is a d-dimensional EV distribution with $0 < G(\mathbf{0}) < 1$. If $F \in D(G)$ then there exists an increasing continuous curve $\mathbf{u}$ with $F(\mathbf{u}(t)) \to 1$ as $t \to \infty$, and a function $\sigma(\mathbf{u}) > \mathbf{0}$ such that*

$$P(\mathbf{X}_\mathbf{u} \leqslant \mathbf{x} | \mathbf{X}_\mathbf{u} \nleq \mathbf{0}) \to \frac{1}{-\log G(\mathbf{0})} \log \frac{G(\mathbf{x})}{G(\mathbf{x} \wedge \mathbf{0})} \tag{3}$$

as $t \to \infty$, *for all* $\mathbf{x}$.

(ii) *Suppose there exists an increasing continuous curve* $\mathbf{u}$ *with* $F(\mathbf{u}(t)) \to 1$ *as* $t \to \infty$, *and a function* $\sigma(\mathbf{u}) > \mathbf{0}$ *such that*

$$P(\mathbf{X}_u \leqslant \mathbf{x} | \mathbf{X}_u \nleqslant \mathbf{0}) \to H(\mathbf{x}), \tag{4}$$

for some function H, *as* $t \to \infty$, *for* $\mathbf{x} > \mathbf{0}$, *where the marginals of* H *on* $\mathbb{R}_+$ *are non-degenerate. Then the left-hand side of* (4) *converges to a limit* $H(\mathbf{x})$ *for all* $\mathbf{x}$ *and there is a unique multivariate extreme value distribution* G *with* $G(\mathbf{0}) = \mathrm{e}^{-1}$ *such that*

$$H(\mathbf{x}) = \log \frac{G(\mathbf{x})}{G(\mathbf{x} \wedge \mathbf{0})}. \tag{5}$$

This G *satisfies* $G(\mathbf{x}) = \mathrm{e}^{-\overline{H}(\mathbf{x})}$ *for* $\mathbf{x} > \mathbf{0}$, *and* $F \in D(G)$.

The next motivation for Definition 2.1 is that distribution (2) is the only one which is preserved under (a suitably coordinated) change of exceedance levels.

Theorem 2.2. (i) *Suppose* $\mathbf{X}$ *has a multivariate generalized Pareto distribution. Then there exists an increasing continuous curve* $\mathbf{u}$ *with* $P(\mathbf{X} \leqslant \mathbf{u}(t)) \to 1$ *as* $t \to \infty$, *and a function* $\boldsymbol{\sigma}(\mathbf{u}) > \mathbf{0}$ *such that*

$$P(\mathbf{X}_u \leqslant \mathbf{x} | \mathbf{X}_u \nleqslant 0) = P(\mathbf{X} \leqslant \mathbf{x}), \tag{6}$$

for $t \in [1, \infty)$ *and all* $\mathbf{x}$.

(ii) *If there exists an increasing continuous curve* $\mathbf{u}$ *with* $P(\mathbf{X} \leqslant \mathbf{u}(t)) \to 1$ *as* $t \to \infty$, *and a function* $\boldsymbol{\sigma}(\mathbf{u}) > 0$ *such that* (6) *holds for* $\mathbf{x} > \mathbf{0}$, *and* $\mathbf{X}$ *has non-degenerate margins, then* $\mathbf{X}$ *has a multivariate generalized Pareto distribution.*

A useful tool in extreme value theory is the convergence of the point process of large values; see Resnick (1987). The close relation between the previous results and point process convergence is the content of the next result. In it we use the rather standard notation of Resnick (1987), and let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be independent and indentically distributed with d.f. F.

Theorem 2.3. (i) *Suppose one of the conditions of Theorem 2.1 holds. Write* S *for the support of* G, *so that* $S = \{\mathbf{x} : G(\mathbf{x}) \in (0, 1)\}$ *and let* μ *be the measure on* S *which is determined by* $\mu(-\infty, \mathbf{x}]^c = -\log G(\mathbf{x})$. *Then there exist* d*-dimensional vectors of constants* $\mathbf{u}_n$ *and* $\sigma_n > \mathbf{0}$ *such that*

$$\sum_{i=1}^{n} \varepsilon_{(1/n, (\mathbf{X}_i - \mathbf{u}_n)/\boldsymbol{\sigma}_n)} \Rightarrow \mathrm{PRM}(\mathrm{d}t \times \mathrm{d}\mu) \text{ on } S. \tag{7}$$

(ii) *Suppose* (7) *holds on* $\mathbf{x} \geqslant \mathbf{0}$ *for some measure* μ *with* $\mu(-\infty, \mathbf{0}]^c \in (0, \infty)$ *and where the function* $\mu(-\infty, \mathbf{x}]^c$ *has non-degenerate marginals. Then the conditions of Theorem 2.1 hold, and hence also* (i) *above is satisfied.*

3. A representation and a bivariate example

In this section we exhibit a general expression for the multivariate GP distributions and give one specific bivariate example: the bivariate logistic distribution.

Several authors, among them Resnick (1987) and Pickands (1981), have given equivalent characterizations of multivariate EV distributions, assuming different marginal distributions. For example, according to Proposition 5.11 in Resnick (1987), all max-stable distributions with the unit Fréchet EV distribution $\Phi_1(x) = \exp(-x^{-1})$, $x > 0$, as marginal distribution can be written as

$$G_*(\mathbf{x}) = \exp\{-\mu_*[\mathbf{0}, \mathbf{x}]^c\}, \qquad \mathbf{x} \geqslant \mathbf{0}, \tag{8}$$

with

$$\mu_*[\mathbf{0}, \mathbf{x}]^c = \int_{\aleph} \bigvee_{i=1}^{d} \left(\frac{a^{(i)}}{x^{(i)}}\right) S(\mathrm{d}\mathbf{a}). \tag{9}$$

Here S is a finite measure on $\aleph = \{\mathbf{y} \in {}^d : \|\mathbf{y}\| = 1\}$, which below is assumed to satisfy $\int_{\aleph} a^{(i)} S(\mathrm{d}\mathbf{a}) = 1$, $1 \leqslant i \leqslant d$, where $\|\cdot\|$ is an arbitrary norm in $\mathbb{R}^d$, and μ_* is called the exponent measure.

This leads to the description of the multivariate EV distribution G with arbitrary marginals as all distributions of the form

$$G(\mathbf{x}) = G_*\left(\left(\mathbf{1} + \frac{\boldsymbol{\gamma}(\mathbf{x} - \boldsymbol{\mu})}{\boldsymbol{\sigma}}\right)^{1/\gamma}\right).$$

Here $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\gamma}$ are d-dimensional vectors with potentially different entries. By Definition 2.1 this in turn gives the following expression for the multivariate GP distributions.

Proposition 3.1. *$H(\mathbf{x})$ is a multivariate GP distribution if there exists a finite measure S, normalized as described above, such that for $\mathbf{x} \nleqslant \mathbf{0}$,*

$$H(\mathbf{x}) =$$

$$\frac{\int_{\aleph} \bigvee_{i=1}^{d} (a^{(i)}(1 + \gamma^{(i)}(x^{(i)} \wedge 0 - \mu^{(i)})/\sigma^{(i)})^{-1/\gamma^{(i)}}) S(\mathrm{d}\mathbf{a}) - \int_{\aleph} \bigvee_{i=1}^{d} (a^{(i)}(1 + \gamma^{(i)}(x^{(i)} - \mu^{(\mathrm{i})})/\sigma^{(i)})^{-1/\gamma^{(i)}}) \mathrm{S}(\mathrm{d}\mathbf{a})}{\int_{\aleph} \bigvee_{i=1}^{d} (a^{(i)}(1 - \gamma^{(i)}\mu^{(i)}/\sigma^{(i)})^{-1/\gamma^{(i)}}) S(\mathrm{d}\mathbf{a})}$$

$$= \frac{\mu_*([\mathbf{0}, (1 + \gamma(\mathbf{x} \wedge \mathbf{0} - \mu)/\sigma)^{1/\gamma}]^c) - \mu_*([\mathbf{0}, (1 + \gamma(\mathbf{x} - \mu)/\sigma)^{1/\gamma}]^c)}{\mu_*([\mathbf{0}, (1 - \gamma\mu/\sigma)^{1/\gamma}]^c)}.$$

The parameters $\boldsymbol{\mu}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\sigma} > 0$ have to satisfy $\mu^{(i)} < \sigma^{(i)}/\gamma^{(i)}$ if $\gamma^{(i)} > 0$ and $\mu^{(i)} > \sigma^{(i)}/\gamma^{(i)}$ for $\gamma^{(i)} < 0$, for $i = 1, \ldots, n$ (to make $0 < G(\mathbf{0}) < 1$).

Example 3.1. The *symmetric logistic model* has exponent measure

$$\mu_*([\mathbf{0}, (x, y)]^c) = (x^{-r} + y^{-r})^{1/r}, \qquad r \geqslant 1.$$

The independent case corresponds to $r = 1$ and for $r = +\infty$ we obtain complete dependence, which is the only situation without density.

By transforming marginals to an arbitrary EV distribution we obtain the following bivariate GP distribution:

$$H(x, y) = \frac{\begin{array}{l}((1+\gamma_x(x\wedge 0-\mu_x)/\sigma_x)_+^{-r/\gamma_x}+(1+\gamma_y(y\wedge 0-\mu_y)/\sigma_y)_+^{-r/\gamma_y})^{1/r}\\ \qquad -((1+\gamma_x(x-\mu_x)/\sigma_x)_+^{-r/\gamma_x}+(1+\gamma_y(y-\mu_y)/\sigma_y)_+^{-r/\gamma_y})^{1/r}\end{array}}{((1-\gamma_x\mu_x/\sigma_x)^{-r/\gamma_x}+(1-\gamma_y\mu_y/\sigma_y)^{-r/\gamma_y})^{1/r}}.$$

As above we assume that the parameterization is such that it ensures $0 < G(0, 0) < 1$.

For $(x, y) > (0, 0)$ this corresponds to

$$H(x, y) = 1 - \frac{((1+\gamma_x(x-\mu_x)/\sigma_x)_+^{-r/\gamma_x}+(1+\gamma_y(y-\mu_y)/\sigma_y)_+^{-r/\gamma_y})^{1/r}}{((1-\gamma_x\mu_x/\sigma_x)^{-r/\gamma_x}+(1-\gamma_y\mu_y/\sigma_y)^{-r/\gamma_y})^{1/r}},$$

while for $x < 0$, $y > 0$ we have

$$H(x, y) = \frac{\begin{array}{l}((1+\gamma_x(x-\mu_x)/\sigma_x)_+^{-r/\gamma_x}+(1-\gamma_y\mu_y/\sigma_y)^{-r/\gamma_y})^{1/r}\\ \qquad -((1+\gamma_x(x-\mu_x)/\sigma_x)_+^{-r/\gamma_x}+(1+\gamma_y(y-\mu_y)/\sigma_y)_+^{-r/\gamma_y})^{1/r}\end{array}}{((1-\gamma_x\mu_x/\sigma_x)^{-r/\gamma_x}+(1-\gamma_y\mu_y/\sigma_y)^{-r/\gamma_y})^{1/r}}.$$

For the independent case, $r = 1$, this simplifies to

$$H(x, y) = \frac{(1-\gamma_y\mu_y/\sigma_y)^{-1/\gamma_y}-(1+\gamma_y(y-\mu_y)/\sigma_y)_+^{-1/\gamma_x}}{(1-\gamma_x\mu_x/\sigma_x)^{-1/\gamma_x}+(1-\gamma_y\mu_y/\sigma_y)^{-1/\gamma_y}}.$$

Now, still considering the independent case, assume $\gamma_x, \gamma_y > 0$ and let X_∞ and Y_∞ be independent random variables where X_∞ has the GP d.f.

$$1 - \left(1 + \frac{\gamma_x x}{\sigma_x}\right)_+^{-1/\gamma_x} \qquad \text{for} \quad x > 0,$$

and Y_∞ is defined similarly. Further, let

$$p = \frac{(1-\gamma_x\mu_x/\sigma_x)^{-1/\gamma_x}}{(1-\gamma_x\mu_x/\sigma_x)^{-1/\gamma_x}+(1-\gamma_y\mu_y/\sigma_y)^{-1/\gamma_y}}.$$

Then H is the distribution function of a bivariate random variable which equals $(X_\infty, \mu_y - \sigma_y/\gamma_y)$ with probability p and $(\mu_x - \sigma_x/\gamma_x, Y_\infty)$ with probability $q = 1 - p$. Thus, in either case one of the components is degenerate at the lower bound of the distribution while the other is a GP random variable. The same holds also when the γs are non-positive, but then the lower bound is $-\infty$.

This completely agrees with intuition. Suppose the the distribution of the exceedances of a bivariate vector after normalization converges as the levels increase. Then, for independent components the event that one of the component exceeds its level does not influence the value of the other one. Hence asymptotically, as the levels tend to infinity the normalization will force the this component down to its lower bound. The roles of the components can of course be interchanged in this argument.

4. Lower-dimensional marginal distributions

Interpreted in the usual way, lower-dimensional marginal distributions of multivariate GP distributions are not GP distributions. For example, if $H(x, y)$ is the bivariate GP distribution from Example 3.1 and $H_1(x)$ is the marginal distribution of the first component, then

$$\begin{aligned} H_1(x) &= H(x, \infty) \\ &= \frac{((1+\gamma_x(x \wedge 0 - \mu_x)/\sigma_x)_+^{-r/\gamma_x} + (1-\gamma_y \mu_y/\sigma_y)^{-r/\gamma_y})^{1/r} - (1+\gamma_x(x-\mu_x)/\sigma_x)_+^{-1/\gamma_x}}{((1-\gamma_x\mu_x/\sigma_x)^{-r/\gamma_x} + (1-\gamma_y\mu_y/\sigma_y)^{-r/\gamma_y})^{1/r}}. \end{aligned}$$

This is not a one-dimensional GP distribution. However, if X_1 has distribution H_1 then the conditional distribution of $X_1|X_1 > 0$ is GP. This property holds for all marginal distributions regardless of the dimension of the original problem.

The reason is that H_1 is the asymptotic conditional distribution of the first component of the random vector given that either the first or the second component is large. In contrast, a one-dimensional GP distribution is the asymptotic conditional distribution of a random variable, given that it is large. In general, a (standard) lower-dimensional marginal distribution of a multivariate GP distribution is *the asymptotic conditional distribution of a subset of random variables given that at least one of a bigger set of variables is large.* Sometimes these may be the appropriate lower-dimensional marginals of multivariate GP distributions. However, the following concept may also be useful.

Let H be a d-dimensional multivariate GP distribution with representation

$$H(\mathbf{y}) = \frac{1}{-\log G(\mathbf{0})} \log \frac{G(\mathbf{y})}{G(\mathbf{y} \wedge \mathbf{0})} \tag{10}$$

in terms of a multivariate EV distribution G. For a $(d-1)$-dimensional vector $\mathbf{x} = (x_1, \ldots x_{d-1})$, let $G_{(i)}(\mathbf{x}) = G((x_1, \ldots x_{i-1}, \infty, x_i, \ldots x_{d-1}))$ be the $(d-1)$-dimensional marginal distribution of G, with the ith component removed. The $(d-1)$-dimensional *generalized Pareto marginal distribution* $H_{(i)}^{\mathrm{GP}}$ of H is defined to be

$$H_{(i)}^{\mathrm{GP}}(\mathbf{x}) = \frac{1}{-\log G_{(i)}(\mathbf{0})} \log \frac{G_{(i)}(\mathbf{x})}{G_{(i)}(\mathbf{x} \wedge \mathbf{0})}. \tag{11}$$

Since $G_{(i)}$ is an EV distribution it follows directly that $H_{(i)}^{\mathrm{GP}}$ is a GP distribution.

The interpretation is that if H is the asymptotic conditional distribution of a d-dimensional random vector $(X_1, \ldots X_d)$ given that at least one of its components is large, then $H_{(i)}^{\mathrm{GP}}$ is the asymptotic conditional distribution of $(X_1, \ldots X_{i-1}, X_{i+1}, \ldots X_d)$ given that at least one of the components of $(X_1, \ldots X_{i-1}, X_{i+1}, \ldots X_d)$ is large.

Expression (11) is implicit since it involves G. However, the GP marginal distribution can also be expressed directly in terms of the parent GP distribution H. For $\mathbf{x}$ a $(d-1)$-dimensional vector $\mathbf{x}$, write $\mathbf{x}_{(i)}$ for the d-dimensional vector which is obtained from $\mathbf{x}$ by inserting an ∞ at position i, that is, $\mathbf{x}_{(i)} = (x_1, \ldots x_{i-1}, \infty, x_i, \ldots x_{d-1})$. Then

$$H_{(i)}^{\mathrm{GP}}(\mathbf{x}) = \frac{H(\mathbf{x}, \infty) - H(\mathbf{x} \wedge 0, \infty)}{1 - H(\mathbf{x} \wedge \mathbf{0}, \infty)}. \tag{12}$$

We have formally only discussed $(d-1)$-dimensional marginal distributions of d-dimensional GP distributions. However, of course, for both definitions of marginal distributions, $(d-k)$-dimensional marginal distributions can be obtained by repeating the above procedure k times, and it is obvious that the resulting $(d-k)$-dimensional distributions do not depend on the order in which one steps down in dimension.

5. Proofs

Proof of Theorem 2.1. (i) If $F \in D(G)$ then there exist $\boldsymbol{\sigma}_n > \mathbf{0}$, $\mathbf{u}_n$ in $\mathbb{R}^d$ such that

$$F^n(\boldsymbol{\sigma}_n \mathbf{x} + \mathbf{u}_n) \to G(\mathbf{x}) \tag{13}$$

for all $\mathbf{x}$, since G is continuous. The components of the norming constants $\boldsymbol{\sigma}_n > \mathbf{0}$, $\mathbf{u}_n$ may be chosen as in the univariate case, where we may choose each component of $\mathbf{u}_n$ to be non-decreasing (see Leadbetter *et al.* 1983: 18). Further, making a suitably small pertubation of the $\mathbf{u}_n$ we may, and will in what follows, assume $\mathbf{u}_n$ to be strictly increasing.

By (13) also $F^n(\boldsymbol{\sigma}_{n+1}\mathbf{x} + \mathbf{u}_{n+1}) \to G(\mathbf{x})$ and by the convergence of types theorem (Leadbetter *et al.* 1983: 7) applied to each marginal, it follows that

$$\boldsymbol{\sigma}_{n+1}/\boldsymbol{\sigma}_n \to \mathbf{1} \quad \text{and} \quad (\mathbf{u}_{n+1} - \mathbf{u}_n)/\boldsymbol{\sigma}_n \to \mathbf{0}. \tag{14}$$

Taking logarithms, it will be seen that (13) is equivalent to

$$n\overline{F}(\boldsymbol{\sigma}_n \mathbf{x} + \mathbf{u}_n) \to -\log G(\mathbf{x}). \tag{15}$$

Now, define $\mathbf{u}$ by $\mathbf{u}(t) = \mathbf{u}_n$ for $t = n$ and by linear interpolation for $n < t < n+1$, and set $\boldsymbol{\sigma}(\mathbf{u}(t)) = \boldsymbol{\sigma}_n$ for $n \leqslant t < n+1$. It then follows from (14) and (15) that

$$t\overline{F}(\boldsymbol{\sigma}(\mathbf{u}(t))\mathbf{x} + \mathbf{u}(t)) \to -\log G(\mathbf{x}). \tag{16}$$

By straightforward argument,

$$P(\mathbf{X_u} \leqslant \mathbf{x} | \mathbf{X_u} \nleqslant \mathbf{0}) = \frac{P(\mathbf{X_u} \nleqslant \mathbf{x} \wedge \mathbf{0}) - P(\mathbf{X_u} \nleqslant \mathbf{x})}{P(\mathbf{X_u} \nleqslant \mathbf{0})}, \tag{17}$$

for $\mathbf{x} \nleqslant \mathbf{0}$. Since $P(\mathbf{X_u} \nleqslant \mathbf{x}) = \overline{F}(\boldsymbol{\sigma}_{\mathbf{u}}\mathbf{x} + \mathbf{u})$, (3) now follows from (16).

(ii) Suppose that (4) holds. Since $\mathbf{u}(t)$ is strictly increasing we may reparameterize so that $t = \inf\{s : \overline{F}(\mathbf{u}(s)) \leqslant 1/t\}$ for large t. Then, $t\overline{F}(\mathbf{u}(t-)) \geqslant 1 \geqslant t\overline{F}(\mathbf{u}(t))$. Further, for any continuity point $\varepsilon > 0$ of H, since $\boldsymbol{\sigma} > \mathbf{0}$,

$$\limsup_{t\to\infty} \frac{\overline{F}(\mathbf{u}(t-)}{\overline{F}(\mathbf{u}(t)} \leqslant \limsup_{t\to\infty} \frac{\overline{F}(\mathbf{u}(t))}{\overline{F}(\mathbf{u}(t)+\boldsymbol{\sigma}(\mathbf{u}(t))\varepsilon)} = \frac{1}{\overline{H}(\varepsilon)}.$$

Since $H(x)$ is a limit of distribution functions it is right continuous, and $\overline{H}(0) = 1$, so letting $\varepsilon \to 0$ through continuity points of H gives that $\overline{F}(\mathbf{u}(t-)/\overline{F}(\mathbf{u}(t) \to 1$, and hence

$$t\overline{F}(\mathbf{u}(t)) \to 1.$$

It follows from (4) and (17) that, for $\mathbf{x} > \mathbf{0}$,

$$\begin{aligned}\lim_{t\to\infty} t\overline{F}(\mathbf{u}(t)+\boldsymbol{\sigma}(\mathbf{u}(t))\mathbf{x}) &= \lim_{t\to\infty} \frac{\overline{F}(\mathbf{u}(t)+\boldsymbol{\sigma}(\mathbf{u}(t))\mathbf{x})}{\overline{F}(\mathbf{u}(t)} \qquad (18)\\ &= \lim_{t\to\infty} \frac{P(\mathbf{X}_{\mathbf{u}} \nleqslant \mathbf{x})}{P(\mathbf{X}_{\mathbf{u}} \nleqslant \mathbf{0})} = \overline{H}(\mathbf{x}).\end{aligned}$$

We next show that (18) also holds when $\mathbf{x}$ is not positive. From (18) it follows that $F(\mathbf{u}(t)+\boldsymbol{\sigma}(t)\mathbf{x})^t \to \mathrm{e}^{-\overline{H}(\mathbf{x})}$ for $\mathbf{x} > \mathbf{0}$. Further, $t\overline{F}(\mathbf{u}(\Delta t)+\boldsymbol{\sigma}(\mathbf{u}(\Delta t))\mathbf{x}) \to \overline{H}(\mathbf{x})/\Delta$ for $\Delta > 0$ and hence

$$F(\mathbf{u}(\Delta t)+\boldsymbol{\sigma}(\Delta t))^t \to \mathrm{e}^{-\overline{H}(\mathbf{x})/\Delta}.$$

By the extremal types theorem (Leadbetter *et al.* 1983: 7) there exist $\mathbf{c}_\Delta > \mathbf{0}$, $\mathbf{x}_\Delta$ such that

$$\frac{\boldsymbol{\sigma}(\Delta t)}{\boldsymbol{\sigma}(t)} \to \mathbf{c}_\Delta, \qquad \frac{\mathbf{u}(\Delta t)-\mathbf{u}(t)}{\boldsymbol{\sigma}(t)} \to \mathbf{x}_\Delta.$$

Thus, for any $\mathbf{x} \geqslant -\mathbf{x}_\Delta/\mathbf{c}_\Delta$,

$$\begin{aligned}\Delta t\overline{F}(\mathbf{u}(\Delta t)+\boldsymbol{\sigma}(\Delta t)\mathbf{x}) &= \Delta t\overline{F}\left(\mathbf{u}(t)+\boldsymbol{\sigma}(t)\left(\frac{\boldsymbol{\sigma}(\Delta t)}{\boldsymbol{\sigma}(t)}\mathbf{x}+\frac{\mathbf{u}(\Delta t)-\mathbf{u}(t)}{\boldsymbol{\sigma}(t)}\right)\right)\\ &\to \Delta\overline{H}(\mathbf{c}_\Delta\mathbf{x}+\mathbf{x}_\Delta).\end{aligned}$$

This may be rephrased as

$$t\overline{F}(\mathbf{u}(t)+\boldsymbol{\sigma}(t)\mathbf{x}) \to \Delta\overline{H}(\mathbf{c}_\Delta\mathbf{x}+\mathbf{x}_\Delta).$$

Hence the limit does not depend on the choice of Δ, and we may uniquely define $-\log G(\mathbf{x})$ as $\Delta\overline{H}(c\mathbf{x}+\mathbf{x}_\Delta)$ for any $\mathbf{x} \geqslant \inf_{\Delta\geqslant 1}\mathbf{x}_\Delta/\mathbf{c}_\Delta$ to obtain that, for such $\mathbf{x}$,

$$t\overline{F}(\mathbf{u}(t)+\boldsymbol{\sigma}(t)\mathbf{x}) \to -\log G(\mathbf{x}). \qquad (19)$$

Suppose that one coordinate in $\mathbf{x}$ is less than the corresponding coordinate of $\inf_{\Delta\geqslant 1}\mathbf{x}_\Delta/\mathbf{c}_\Delta$. Then if we let $\mathbf{x}^\infty$ be the vector which has all other components set to ∞, and $\mathbf{x}_0^\infty$ the vector were all other components are set to ∞ and this coordinate is set to 0, we have that

$$\liminf_{t\to\infty} t\overline{F}(\mathbf{u}(t)+\boldsymbol{\sigma}(t)\mathbf{x}) \geqslant \liminf_{t\to\infty} t\overline{F}(\mathbf{u}(t)+\boldsymbol{\sigma}(t)\mathbf{x}^\infty) \geqslant \Delta\overline{H}(\mathbf{x}_0^\infty),$$

for any $\Delta > 1$. Hence $t\overline{F}(\mathbf{u}(t)+\boldsymbol{\sigma}(t)\mathbf{x}) \to \infty$ for such $\mathbf{x}$, and thus if we define $-\log G(\mathbf{x}) = \infty$ for $\mathbf{x}$ which are not greater than $\inf_{\Delta\geqslant 1}\mathbf{x}_\Delta/\mathbf{c}_\Delta$ then (19) holds for all $\mathbf{x}$. Thus,

$$F(\mathbf{u}(t)+\boldsymbol{\sigma}(t)\mathbf{x})^t \to G(\mathbf{x}), \qquad \text{for all } \mathbf{x},$$

and hence $G(\mathbf{x})$ is a multivariate EV distribution, and it follows from the first part of the theorem that (5) holds, since $G(\mathbf{0}) = \mathrm{e}^{-\overline{H}(\mathbf{0})} = \mathrm{e}^{-1}$. □

Proof of Theorem 2.2. (i) Let $\mathbf{X}$ have distribution H. By definition, H is of the form (2) for some EV distribution G, so that

$$\begin{aligned} P(\mathbf{X} \nleq \mathbf{x}) &= 1 - \frac{1}{-\log G(\mathbf{0})} \log \frac{G(\mathbf{x})}{G(\mathbf{x}\wedge\mathbf{0})} \qquad (20)\\ &= \frac{1}{\log G(\mathbf{0})} \log \frac{G(\mathbf{x})G(\mathbf{0})}{G(\mathbf{x}\wedge\mathbf{0})}. \end{aligned}$$

Since G is max-stable, there exist continuous curves $\boldsymbol{\sigma}(t) > \mathbf{0}$, $\mathbf{u}(t)$ with $\boldsymbol{\sigma}(1) = \mathbf{1}$, $\mathbf{u}(1) = \mathbf{0}$ and $\mathbf{u}(t)$ strictly increasing, such that $G(\mathbf{u}(t)+\boldsymbol{\sigma}(t)\mathbf{x})^t = G(\mathbf{x})$. In particular, $G(\mathbf{u}(t)) = G(\mathbf{0})^{1/t}$. Further, by (17) and (20),

$$P(\mathbf{X_u} \leqslant \mathbf{x} | \mathbf{X_u} \nleq \mathbf{0}) = \frac{1}{-\log G(\mathbf{u}(t))} \log \frac{G(\mathbf{u}(t)+\boldsymbol{\sigma}(t)\mathbf{x})G(\mathbf{u}(t)+\boldsymbol{\sigma}(t)(\mathbf{x}\wedge\mathbf{0})\wedge\mathbf{0})}{G(\mathbf{u}(t)+\boldsymbol{\sigma}(t)\mathbf{x}\wedge\mathbf{0})G((\mathbf{u}(t)+\boldsymbol{\sigma}(t)\mathbf{x}\wedge\mathbf{0})}.$$

Since $(\mathbf{x}\wedge\mathbf{0})\wedge\mathbf{0} = \mathbf{x}\wedge\mathbf{0}$, it follows that

$$\begin{aligned} P(\mathbf{X_u} \leqslant \mathbf{x} | \mathbf{X_u} \nleq \mathbf{0}) &= \frac{t}{-\log G(\mathbf{0})} \log \frac{G(\mathbf{u}(t)+\boldsymbol{\sigma}(t)\mathbf{x})}{G(\mathbf{u}(t)+\boldsymbol{\sigma}(t)\mathbf{x}\wedge\mathbf{0})}\\ &= \frac{1}{-\log G(\mathbf{0})} \log \frac{G(\mathbf{u}(t)+\boldsymbol{\sigma}(t)\mathbf{x})^t}{G(\mathbf{u}(t)+\boldsymbol{\sigma}(t)\mathbf{x}\wedge\mathbf{0})^t}\\ &= \frac{1}{-\log G(\mathbf{0})} \log \frac{G(\mathbf{x})}{G(\mathbf{x}\wedge\mathbf{0})} = P(\mathbf{X} \leqslant \mathbf{x}). \end{aligned}$$

(ii) This is an easy consequence of Theorem 2.1 (ii). □

Proof of Theorem 2.3. (i) By Theorem 2.1 we have that $F \in D(G)$, and hence there are constants $\mathbf{u}_n$, $\boldsymbol{\sigma}_n$ such that $nP((\mathbf{X}_i - \mathbf{u}_n)/\boldsymbol{\sigma}_n \nleq \mathbf{x}) \to -\log G(\mathbf{x})$ on S. The conclusion of (i) then follows from Theorem 3.21 of Resnick (1987).

(ii) Again by Theorem 3.21 of Resnick (1987), it follows that $nP((\mathbf{X}_i - \mathbf{u}_n)/\boldsymbol{\sigma}_n \nleq \mathbf{x}) \to \mu((-\infty, \mathbf{x}]^c)$ for $\mathbf{x} > \mathbf{0}$. However, this is just a different way of writing (4) of Theorem 2.1, and hence condition (ii) of the theorem is satisfied, and the result follows. □

Acknowledgement

The authors thank the referee for very constructive comments.

References

Arnold, B.C. (1983) *Pareto Distributions*. Fairland, MD: International Co-operative Publishing House.

Balkema, A.A. and de Haan, L. (1974) Residual life time at high age. *Ann. Probab.*, **2**, 792–804.

Barnett, V. (1976) The ordering of multivariate data (with discussion). *J. Roy. Statist. Soc. Ser A*, **139**, 318–354.

Beirlant, J., Goetgebeur, Y., Segers, J. and Teugels, J. (2004) *Statistics of Extremes: Theory and Applications*. Chichester: Wiley.

Coles, S.G. (2001) *An Introduction to Statistical Modeling of Extreme Values*. London: Springer-Verlag.

Coles, S.G. and Tawn, J.A. (1991) Modelling multivariate extreme events. *J. Roy. Statist. Soc. Ser. B,* **53**, 377–392.

Davison, A.C. and Smith, R.L. (1990) Models for exceedances over high thresholds. *J. Roy. Statist. Soc. Ser. B,* **52**, 393–442.

Embrechts, P., Klüppelberg C. and Mikosch, T. (1997) *Modelling Extremal Events for Finance and Insurance*. Berlin: Springer-Verlag.

Falk, M. and Reiss, R.D. (2001) Estimation of canonical dependence parameters in a class of bivariate peaks-over-threshold models. *Statist. Probab. Lett.*, **52**, 233–242.

Falk, M. and Reiss, R.D. (2002) A characterization of the rate of convergence in bivariate extreme value models. *Statist. Probab. Lett.*, **59**, 341–351.

Falk, M. and Reiss, R.D. (2003a) Efficient estimation of the canonical dependence function. *Extremes*, **6**, 61–82.

Falk, M. and Reiss, R.D. (2003b) Efficient estimators and LAN in canonical bivariate POT models. *J. Multivariate Anal.*, **84**, 190–207.

Falk, M. and Reiss, R.D. (2005) On the distribution of Pickands coordinates in bivariate EV and GP models. *J. Multivariate Anal.*, **93**, 267–295.

Fougères, A.-L. (2004) Multivariate extremes. In B. Finkenstädt and H. Rootzén (eds), *Extreme Values in Finance, Telecommunications, and the Environment*. Boca Raton, FL: Chapman & Hall/CRC.

Joe, H., Smith, R.L. and Weissman, I. (1992) Bivariate threshold methods for extremes. *J. R. Statist. Soc. Ser. B*, **54**, 171–183.

Kotz, S. and Nadarajah, S. (2000) *Extreme Value Distributions: Theory and Applications*. London: Imperial College Press.

Kowaka, M. (1994). *An Introduction to Life Prediction of Plant Materials. Application of Extreme Value Statistical Methods for Corrosion Analysis*. New York: Allerton Press.

Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes*. Berlin: Springer-Verlag.

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

Pickands III, J. (1975) Statistical inference using extreme order statistics. *Ann. Statist.*, **3**, 119–131.

Pickands III, J. (1981) Multivariate extreme value distributions. In *Bulletin of the International Statistical Institute: Proceedings of the 43rd Session (Buenos Aires)*, Book 2, pp. 857—878. Voorburg, Netherlands: ISI.

Reiss, R. and Thomas, M. (2005) *Statistical Analysis of Extreme Values (for Insurance, Finance, Hydrology and Other Fields).* 3rd rev. edn. Basel: Birkhäuser.

Resnick, S.I. (1987) *Extreme Values, Regular Variation and Point Processes.* Berlin: Springer-Verlag.

Smith, R.L. (1985) Statistics of extreme values. In *Bulletin of the International Statistical Institute: Proceedings of the 45th Session (Amsterdam)*, Book 4, Paper 26.1, pp. 1—17. Voorburg,

Netherlands: ISI.
Smith, R.L. (1987) Estimating tails of probability distributions. *Ann. Statist.*, **15**, 1174–1207.
Smith, R.L., Tawn, J.A. and Yuen, H.K. (1990) Statistics of multivariate extremes. *Int. Statist. Inst. Rev.*, **58**, 47–58.
Smith, R.L., Tawn, J.A. and Coles, S.G. (1997) Markov chain models for threshold exceedances. *Biometrika*, **84**, 249–268.
Tajvidi, N. (1996) Multivariate generalized Pareto distributions. In *Characterisation and Some Statistical Aspects of Univariate and Multivariate Generalized Pareto Distributions*, PhD thesis, Department of Mathematics, Chalmers, Göteborg.
Tawn, J.A. (1988) Bivariate extreme value theory: Models and estimation. *Biometrika*, **75**, 397–415.
Tawn, J.A. (1990) Modelling multivariate extreme value distributions. *Biometrika*, **77**, 245–253.

Received October 2005 and revised January 2006