Multivariate generalized Pareto distributions

Holger Rootzén [*] and Nader Tajvidi [†]

Abstract

Statistical inference for extremes has been a subject of intensive research during the past couple of decades. One approach is based on modeling exceedances of a random variable over a high threshold with the Generalized Pareto (GP) distribution. This has shown to be an important way to apply extreme value theory in practice and is widely used. In this paper we introduce a multivariate analogue of the GP distribution and show that it is characterized by each of following two properties: (i) exceedances asymptotically have a multivariate GP distribution if and only if maxima asymptotically are Extreme Value (EV) distributed, and (ii) the multivariate GP distribution is the only one which is preserved under change of exceedance levels. We also give a number of examples and discuss lowerdimensional marginal distributions

Keywords: Generalized Pareto distribution, Multivariate extreme value theory, Multivariate Pareto distribution, Non-homogeneous Poisson process, Peaks over threshold method.

AMS 1991 subject classification: 62E20, 60F17, 62E10, 60G70, 62H12.

1 Introduction

Statistical modeling of extreme values has developed extensively during the last decades. This is witnessed by several recent books ((Coles (2001), Embrechts *et al.* (1998), Kotz and Nadarajah (2000), Kowaka (1994), Beirlant *et al.* (2005), Reiss and Thomas (2005)) and a large journal literature. The latter includes both theoretical papers and many articles which apply the methods to a wide range of important problems such as extreme windspeeds, waveheights, floods, insurance claims, price fluctuations, ... For references to some of this literature, see e.g. (Kotz and Nadarajah (2000), Beirlant *et al.* (2005)).

The main emphasis has been on univariate extremes, and so far the univariate results are the most complete and directly usable ones. Two main sets of methods, the Block Maxima

[*]Chalmers, SE-41296, Göteborg, Sweden, rootzen@math.chalmers.se

[†]Lund Institute of Technology, Box 118 SE-22100, Lund, Sweden, tajvidi@maths.lth.se

method and the Peaks over Thresholds method have been developed (Coles (2001)). We here 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 modeled by an Extreme Value (EV) distribution with distribution function (d.f.) $G(x) = \exp\{-(1+\gamma\frac{x-\mu}{\sigma})_+^{-1/\gamma}\}$. This choice of distribution is motivated by the two facts that (i) the EV distributions are the only ones which can appear as the limit of linearly normalized maxima, and that (ii) they are the only ones which are "max-stable", i.e. such that a change of block size only leads to a change of location and scale parameter in the distribution.

In the *Peaks over Thresholds* method one is instead supposed to have observed all values which are larger than some suitable threshold, e.g. all windspeeds higher than 20 m/s. These values are then assumed to follow the Generalized Pareto (GP) distribution with d.f. $H(x) = 1-(1+\gamma\frac{x}{\sigma})_+^{-1/\gamma}$. This choice of distributions is motivated by characterizations due to Balkema and de Haan (1974) and Pickands (1975). One characterization is (i) that the distribution of a scale normalized exceedance over a threshold asymptotically (as the threshold tends to the righthand 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 (ii) that the GP distributions are the only "stable" ones, i.e. 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 PoT 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 PoT method uses more of the data it can sometimes result in better estimation precision than the block maxima method. As a parenthesis, there are several other variants of the (one-dimensional) Pareto distribution, see e.g. Arnold (1983).

Multivariate extreme value distributions arise in connection with extremes of a random sample from a multivariate distribution. They are extensively discussed in the books by Resnick (1987), Kotz and Nadarajah (2000) and Beirlant *et al.* (2005) and in the review by Fougères in Finkenstadt and Rootzén (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 e.g. 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 $\{\mathbf{X}_i,\, i \geq 1\} = \{(X_i^{(1)}, \ldots, X_i^{(d)}),\, i \geq 1\}$, the maximum, $\mathbf{M}_n$, is defined by $\mathbf{M}_n = (M_n^{(1)}, \ldots, M_n^{(d)}) = (\bigvee_{i=1}^n X_i^{(1)}, \ldots, \bigvee_{i=1}^n X_i^{(d)})$. Here and hereafter $\bigvee$ denotes maximum. Under rather general conditions, the distribution of the linearly normalized $\mathbf{M_n}$ converges to a multivariate extreme value distribution. In applications $\mathbf{M_n}$ is often the vector of annual maxima, and block maxima methods can be applied similarly as when the observations are onedimensional. 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 don't show if the component maxima have occurred simultaneously or not.

The aim of this paper is to define the multivariate Generalized Pareto distributions and to prove that this definition indeed is the right one. The multivariate Generalized Pareto distribution should (a) be the natural distribution for exceedances of high thresholds by multivariate random vectors, and (b) it should 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, Section 2, 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, the 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. Lowerdimensional "marginals" of multivariate GP distributions may be thought of in different ways. This is discussed in Sections 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, 2003 and 2005) which introduced a class of distributions also which also were named Bivariate Generalized Pareto 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 disttributions has been discussed by the authors.

2 Multivariate Generalized Pareto Distributions

In this section we give the formal definition of the multivariate generalized Pareto distribution and reformulate the motivating characterizations into mathematical terms. Proofs of the characterizations given in Section 5. However, first some preliminaries.

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

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

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

$$\begin{aligned} &P[(M_n^{(i)} - u_n^{(i)})/\sigma_n^{(i)} \leq x^{(i)},\, 1 \leq i \leq d] = \\ &= F^n(\sigma_n^{(1)} x^{(1)} + u_n^{(1)}, \ldots, \sigma_n^{(d)} x^{(d)} + u_n^{(d)}) \longrightarrow G(\mathbf{x}) \end{aligned} \tag{1}$$

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-s except $x^{(i)}$ to $+\infty$ it is seen that each marginal G_i of G must be an EV d.f., so that $G_i(x) = \exp(-(1 + \gamma_i \frac{x-\mu_i}{\sigma_i})_+^{-1/\gamma_i})$. 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 $\exp(-\exp(-\frac{x-\mu_i}{\sigma_i}))$

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 an scalar, we have

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

and

$$\mathbf{xy} = (x_1y_1, x_2y_2) \quad ; \quad \mathbf{x}+\mathbf{y} = (x_1+y_1, x_2+y_2).$$

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

$$P((\mathbf{M}_n - \mathbf{u}_n)/\sigma_n \leq \mathbf{x}) \to G(\mathbf{x}).$$

For the definition we also use the convention that $0/0 = 1$. The definition has also independently been noticed by Beirlant *et al.* (2005, Ch 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 nondegenerate margins and with $0 < G(\mathbf{0}) < 1$. In particular, $H(\mathbf{x}) = 0$ for $\mathbf{x} < \mathbf{0}$ and $H(\mathbf{x}) = 1 - \frac{\log G(\mathbf{x})}{\log G(\mathbf{0})}$ 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 $G(\mathbf{0}) = e^{-1}$. 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 $\log \frac{G(\mathbf{x})^t}{G(\mathbf{x}\wedge\mathbf{0})^t}$, and by max-stability G^t again is an EV distribution, with $G(\mathbf{0})^t = \exp(-\frac{\log G(\mathbf{0})}{\log G(\mathbf{0})}) = e^{-1}$. However, for statistical application 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 $\mathbf{x} \not< \mathbf{0}$ whereas in the point process approach only values $\mathbf{x} > \mathbf{0}$ are modeled parametrically. This might prove to be an improvement for statistical analysis of extremes since then also negative x_is 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 has 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 distribution function F and write $\bar{F} = 1 - F$ for the tail function of a distribution F. Further, let $\{\mathbf{u}(t) \,|\, t \in [1, \infty)\}$ be a d-dimensional curve starting at $\mathbf{u}(1) = \mathbf{0}$, let $\sigma(\mathbf{u}) = \sigma(\mathbf{u}(t)) > \mathbf{0}$ be a function with values in $\mathbb{R}^d$, and let

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

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 2.2. *(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*

$$P(\mathbf{X_u} \leq \mathbf{x} | \mathbf{X_u} \not\leq \mathbf{0}) \longrightarrow \frac{1}{-\log G(\mathbf{0})} \log \frac{G(\mathbf{x})}{G(\mathbf{x} \wedge \mathbf{0})} \tag{3}$$

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

$$P(\mathbf{X}_u \leq \mathbf{x} | \mathbf{X}_u \not\leq \mathbf{0}) \quad \longrightarrow \quad H(\mathbf{x}), \tag{4}$$

for some function H, as $t \to \infty$, for $\mathbf{x} > \mathbf{0}$, where the marginals of H on $\mathbb{R}_+$ are nondegenerate. Then the lefthand 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 $G(\mathbf{0}) = e^{-1}$ such that

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

This G satisfies $G(\mathbf{x}) = e^{-\bar{H}(\mathbf{X})}$ 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 2.3. *(i) Suppose $\mathbf{X}$ has a multivariate generalized Pareto distribution. Then there exists an increasing continuous curve $\mathbf{u}$ with $P(\mathbf{X} \leq \mathbf{u}(t)) \to 1$ as $t \to \infty$, and a function $\sigma(\mathbf{u}) > \mathbf{0}$ such that*

$$P(\mathbf{X}_u \leq \mathbf{x} | \mathbf{X}_u \not\leq \mathbf{0}) \quad = \quad P(\mathbf{X} \leq \mathbf{x}), \tag{6}$$

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

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

An useful tool in extreme value theory is convergence of the point process of large values, see e.g. 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 i.i.d. with distribution function F.

Theorem 2.4. *(i) Suppose one of the conditions of Theorem 2.2 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*

$$\sum_{i=1}^{n} \varepsilon_{\left(\frac{1}{n}, \frac{\mathbf{X}_i - \mathbf{u}_n}{\sigma_n}\right)} \Longrightarrow \mathrm{PRM}(dt \times d\mu) \ \ \textit{on } S. \tag{7}$$

(ii) Suppose (7) holds on $\mathbf{x} \geq \mathbf{0}$ 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 2.2 hold, and hence also (i) above is satisfied.

3 Examples

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

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

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

with

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

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

This leads to the descriptions of the multivariate extreme value distribution G with arbitrary marginals as all distributions of the form

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

Here μ, σ and γ are d-dimensional vectors with potentially different entries. By 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 $\mathbf{x} \not\leq \mathbf{0}$*

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

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

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

The parameters μ, γ, and $\sigma > 0$ 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 $0 < G(\mathbf{0}) < 1$).

Now, the example:

Example 1 The *symmetric logistic model* has exponent measure

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

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

By transforming marginals to an arbitrary extreme value distribution we obtain the following bivariate GP distribution

$$H(x, y) =$$

$$\frac{\left(\left(1 + \frac{\gamma_x (x \wedge 0 - \mu_x)}{\sigma_x}\right)_+^{-r/\gamma_x} + \left(1 + \frac{\gamma_y (y \wedge 0 - \mu_y)}{\sigma_y}\right)_+^{-r/\gamma_y} \right)^{1/r} - \left(\left(1 + \frac{\gamma_x (x - \mu_x)}{\sigma_x}\right)_+^{-r/\gamma_x} + \left(1 + \frac{\gamma_y (y - \mu_y)}{\sigma_y}\right)_+^{-r/\gamma_y} \right)^{1/r}}{\left(\left(1 - \frac{\gamma_x \mu_x}{\sigma_x}\right)^{-r/\gamma_x} + \left(1 - \frac{\gamma_y \mu_y}{\sigma_y}\right)^{-r/\gamma_y} \right)^{1/r}}$$

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

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

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

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

$$H(x,y) =$$

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

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

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

Now, still considering the independent case, let X_∞ and Y_∞ be independent random variables where X_∞ has the GP d.f.

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

and Y_∞ is defined similarly. Further, let

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

Then H is the distribution function of a bivariate random variable which equals $(X_\infty, -\infty)$ with probability p and $(-\infty, Y_\infty)$ with probability $q = 1 - p$. Thus, in either case one of the components are degenerate at the lower bound of the distribution while the other one is a GP random variable.

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 Lowerdimensional marginal distributions

Interpreted in the usual way, lowerdimensional marginal distributions of multivariate GP distributions are not GP distributions. E.g., if $H(x, y)$ is the bivariate GP distribution from Example 1 above and $H_1(x)$ is the marginal distribution of the first component, then

$$H_1(x) = H(x,\infty) =$$

$$\frac{\left(\left(1+\frac{\gamma_x\,(x\wedge 0-\mu_x)}{\sigma_x}\right)_+^{-r/\gamma_x} + \left(1-\frac{\gamma_y\,\mu_y}{\sigma_y}\right)^{-r/\gamma_y}\right)^{1/r} - \left(\frac{1+\gamma_x\,(x-\mu_x)}{\sigma_x}\right)_+^{-1/\gamma_x}}{\left(\left(1-\frac{\gamma_x\mu_x}{\sigma_x}\right)^{-r/\gamma_x} + \left(1-\frac{\gamma_y\mu_y}{\sigma_y}\right)^{-r/\gamma_y}\right)^{1/r}}.$$

This is not a onedimensional generalized Pareto distribution. However, if X_1 has distribution H_1 then the conditional distribution of $X_1|X_1 > 0$ is generalized Pareto. 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 onedimensional GP distribution is the asymptotic conditional distribution of a random variable, given that it is large. In general (standard) lowerdimensional marginal distributions 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 lowerdimensional 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 $\mathbf{x} = (x_1, \dots x_{d-1})$ a $(d-1)$-dimensional vector let $G_{(i)}(\mathbf{x}) = G((x_1, \dots x_{i-1}, \infty, x_i, \dots x_{d-1}))$ be the $(d-1)$-dimensional marginal distribution, with the i-th component removed, of G. The $(d-1)$-dimensional *Generalized Pareto marginal distribution* $H_{(i)}^{GP}$ of H is defined to be

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

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

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

The 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 write $\mathbf{x}_{(i)}$ for the d-dimensional vector which is obtained from

$\mathbf{x}$ by inserting an ∞ at position i, i.e. $\mathbf{x}_{(i)} = (x_1, \ldots x_{i-1}, \infty, x_i, \ldots x_{d-1})$. Then,

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

Above 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 k times, and it is obvious that the resulting $(d-k)$-dimensional distributions do not depend on which order one steps down in dimension.

Acknowledgement: The authors thank the referee for very constructive comments.

5 Proofs

Proof of Theorem 2.2. (i) If $F \in D(G)$ then there exists $\sigma_n > 0, \mathbf{u}_n$ in $\mathbb{R}^d$ such that

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

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

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

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

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

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

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 $\sigma(\mathbf{u}(t)) = \sigma_n$ for $n \le t < n+1$. It then follows from (14) and (15) that

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

By straightforward argument

$$P(\mathbf{X_u} \le \mathbf{x} | \mathbf{X_u} \not\le \mathbf{0}) \quad = \quad \frac{P(\mathbf{X_u} \not\le \mathbf{x} \wedge 0) - P(\mathbf{X_u} \not\le \mathbf{x})}{P(\mathbf{X_u} \not\le \mathbf{0})}, \tag{17}$$

for $\mathbf{x} \not\leq \mathbf{0}$. Since $P(\mathbf{X_u} \not\leq \mathbf{x}) = \bar{F}(\sigma_\mathbf{u}\mathbf{x} + \mathbf{u})$, (3) now follows from (16).

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

$$\limsup_{t\to\infty} \frac{\bar{F}(\mathbf{u}(t-)}{\bar{F}(\mathbf{u}(t)} \leq \limsup_{t\to\infty} \frac{\bar{F}(\mathbf{u}(t))}{\bar{F}(\mathbf{u}(t) + \sigma(\mathbf{u}(t))\epsilon)} = \frac{1}{\bar{H}(\epsilon)}.$$

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

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

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

$$\begin{aligned} \lim_{t\to\infty} t\bar{F}(\mathbf{u}(t) + \sigma(\mathbf{u}(t))\mathbf{x}) &= \lim_{t\to\infty} \frac{\bar{F}(\mathbf{u}(t) + \sigma(\mathbf{u}(t))\mathbf{x})}{\bar{F}(\mathbf{u}(t)} \\ &= \lim_{t\to\infty} \frac{P(\mathbf{X_u} \not\leq \mathbf{x})}{P(\mathbf{X_u} \not\leq \mathbf{0})} = \bar{H}(\mathbf{x}). \end{aligned} \tag{18}$$

We next show that (18) holds also when $\mathbf{x}$ isn't positive. From (18) follows that $F(\mathbf{u}(t) + \sigma(t)\mathbf{x})^t \to e^{-\bar{H}(\mathbf{X})}$ for $\mathbf{x} > \mathbf{0}$. Further, $t\bar{F}(\mathbf{u}(\Delta t) + \sigma(\mathbf{u}(\Delta t))\mathbf{x}) \to \bar{H}(\mathbf{x})/\Delta$ for $\Delta > 0$ and hence

$$F(\mathbf{u}(\Delta t) + \sigma(\Delta t))^t \to e^{-\bar{H}(\mathbf{X})/\Delta}.$$

By the extremal types theorem (Leadbetter *et al.* (1983, p.7) there are $c_\Delta > 0$, $\mathbf{x}_\Delta$ with

$$\frac{\sigma(\Delta t)}{\sigma(t)} \to c_\Delta, \quad \frac{\mathbf{u}(\Delta t) - \mathbf{u}(t)}{\sigma(t)} \to x_\Delta.$$

Thus, for any $\mathbf{x} \geq -\mathbf{x}_\Delta/c_\Delta$,

$$\begin{aligned} \Delta t\bar{F}(\mathbf{u}(\Delta t) + \sigma(\Delta t)\mathbf{x}) &= \Delta t\bar{F}(\mathbf{u}(t) + \sigma(t)(\frac{\sigma(\Delta t)}{\sigma(t)}\mathbf{x} + \frac{\mathbf{u}(\Delta t) - \mathbf{u}(t)}{\sigma(t)}) \\ &\to \Delta\bar{H}(c_\Delta\mathbf{x} + \mathbf{x}_\Delta). \end{aligned}$$

This may be rephrased as

$$t\bar{F}(\mathbf{u}(t) + \sigma(t)\mathbf{x}) \to \Delta\bar{H}(c_\Delta\mathbf{x} + \mathbf{x}_\Delta).$$

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

$$t\bar{F}(\mathbf{u}(t) + \sigma(t)\mathbf{x}) \to -\log G(\mathbf{x}). \tag{19}$$

Suppose that one coordinate in $\mathbf{x}$ is less than the corresponding coordinate of $\inf_{\Delta\geq 1}\mathbf{x}_\Delta/c_\Delta$. 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

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

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

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

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

Proof of Theorem 2.3. (i) Let $\mathbf{X}$ have distribution H. By definition, H is of the form (2) for some extreme value distribution G, so that

$$\begin{aligned} P(\mathbf{X} \not\leq \mathbf{x}) &= 1 - \frac{1}{-\log G(\mathbf{0})}\log\frac{G(\mathbf{x})}{G(\mathbf{x}\wedge\mathbf{0})} \\ &= \frac{1}{\log G(\mathbf{0})}\log\frac{G(\mathbf{x})G(\mathbf{0})}{G(\mathbf{x}\wedge\mathbf{0})}. \end{aligned} \tag{20}$$

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

$$P(\mathbf{X_u} \leq \mathbf{x}|\mathbf{X_u} \not\leq \mathbf{0}) = \frac{1}{-\log G(\mathbf{u}(t))}\log\frac{G(\mathbf{u}(t)+\sigma(t)\mathbf{x})G((\mathbf{u}(t)+\sigma(t)(\mathbf{x}\wedge\mathbf{0}))\wedge\mathbf{0})}{G(\mathbf{u}(t)+\sigma(t)\mathbf{x}\wedge\mathbf{0})G((\mathbf{u}(t)+\sigma(t)\mathbf{x})\wedge\mathbf{0})}.$$

It can be seen that $(\mathbf{u}(t)+\sigma(t)(\mathbf{x}\wedge\mathbf{0}))\wedge\mathbf{0} = (\mathbf{u}(t)+\sigma(t)\mathbf{x})\wedge\mathbf{0}$, and hence

$$\begin{aligned} P(\mathbf{X_u} \leq \mathbf{x}|\mathbf{X_u} \not\leq \mathbf{0}) &= \frac{t}{-\log G(\mathbf{0})}\log\frac{G(\mathbf{u}(t)+\sigma(t)\mathbf{x})}{G(\mathbf{u}(t)+\sigma(t)\mathbf{x}\wedge\mathbf{0})} \\ &= \frac{1}{-\log G(\mathbf{0})}\log\frac{G(\mathbf{u}(t)+\sigma(t)\mathbf{x})^t}{G(\mathbf{u}(t)+\sigma(t)\mathbf{x}\wedge\mathbf{0})^t} \\ &= \frac{1}{-\log G(\mathbf{0})}\log\frac{G(\mathbf{x})}{G(\mathbf{x}\wedge\mathbf{0})} = P(\mathbf{X}\leq\mathbf{x}). \end{aligned}$$

(ii) This is an easy consequence of Theorem 2.2 (ii). □

Proof of Theorem 2.4. (i) By Theorem 2.2 we have that $F \in D(G)$, and hence there are constants $\mathbf{u}_n, \sigma_n$ such that $nP(\frac{\mathbf{X}_i-\mathbf{u}_n}{\sigma_n} \not\leq \mathbf{x}) \to -\log G(\mathbf{x})$ 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 $nP(\frac{\mathbf{X}_i-\mathbf{u}_n}{\sigma_n} \not\leq \mathbf{x}) \to \mu((-\infty,\mathbf{x}]^c)$ for $\mathbf{x} > \mathbf{0}$. However, this is just a different way of writing (4) of Theorem 2.2, and hence condition (ii) of the theorem is satisfied, and the result follows. □

References

[1] Arnold, B.C. (1983) *Pareto Distributions.* International Co-operative Publishing House.

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

[3] Barnett, V. (1976) The ordering of multivariate data (with discussion). *J. R. Statist. Soc. A* **139**, 318-354.

[4] Beirlant, J., Segers, J., and Teugels, J. (2005) *Statistics of extremes, theory and applications.* Wiley, Chichester.

[5] Coles, S. G. (2001) *An Introduction to Statistical Modeling of Extreme Values.* Springer, London.

[6] Coles, S. G. and Tawn, J.A. (1991) Modelling multivariate extreme events. *J. R. Statist. Soc. B* **53**, 377-392.

[7] Davison, A.C. and Smith, R.L. (1990) Models for exceedances over high thresholds. *J. R. Statist. Soc. B* **52**, 393-442.

[8] Embrechts, P., Klüppelberg C., and Mikosch, T. (1997) *Modelling extremal events.* Springer, New York.

[9] 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.

[10] 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.

[11] Falk, M. and Reiss, R. D. (2003) Efficient estimation of the canonical dependence function. *Extremes* **6**, 61-82.

[12] Falk, M. and Reiss, R. D. (2003) Efficient estimators and LAN in canonical bivariate POT models. *Journal of Multivariate Analysis* **84**, 190-207.

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

[14] Finkenstädt, B. and Rootzén, H., editors. (2004) *Extreme values in finance, telecommunications and the environment.* Chapman & Hall/CRC, Boca Raton.

[15] Joe, H., Smith, R.L. and Weissman, I.(1992) *Bivariate Threshold Methods for Extremes. J. R. Statist. Soc. B* **54**, 171-183.

[16] Kotz, S. and Nadarajah, S. (2000) *Extreme value distributions : theory and applications.* Imperial College Press, London.

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

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

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

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

[21] Pickands, J. (1981) Multivariate extreme value distributions. *Proc. 43rd Session I.S.I.*, 859-878.

[22] Reiss, R. and Thomas, M. (2005) *Statistical Analysis of Extreme Values (for Insurance, Finance, Hydrology and Other Fields).* 3rd revised edition, Birkhuser, Basel.

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

[24] Smith, R.L. (1985) Statistics of extreme values. *Proc. 45th Session I.S.I., Paper 26.1*, Amsterdam.

[25] Smith, R.L., (1987) Estimating tails of probability distributions. *Ann. Statist.* **15**, 1174-1207.

[26] Smith, R.L., Tawn, J.A. and Yuen, H.K. (1990) Statistics of multivariate extremes. *Int. Statist. Inst. Rev.* **58**, 47-58.

[27] Smith, R.L., Tawn, J.A. and Coles, S.G. (1997) Markov chain models for threshold exceedances. *Biometrika* **84**, 249-268.

[28] Tajvidi, N. (1996) Multivariate generalized Pareto distributions. Article in PhD thesis, Department of Mathematics, Chalmers, Göteborg.

[29] Tawn, J. A.(1988) Bivariate extreme value theory: Models and estimation. *Biometrika* **75**, 397-415.

[30] Tawn, J.A. (1990) Modelling multivariate extreme value distributions. *Biometrika* **77**, 245-253.