PREPRINT 2005:38

Multivariate Generalized Pareto Distributions

HOLGER ROOTZÉN
NADER TAJVIDI

Department of Mathematical Sciences
Division of Mathematical Statistics
CHALMERS UNIVERSITY OF TECHNOLOGY
GÖTEBORG UNIVERSITY
Göteborg Sweden 2005
Multivariate Generalized Pareto Distributions

Holger Rootzén and Nader Tajvidi

Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg, Sweden
Göteborg, November 2005
Multivariate generalized Pareto distributions

Holger Rootzén and Nader Tajvidi

Chalmers and Lund Institute of Technology
SE-41296 Göteborg, Sweden
Box 118, SE-22100 Lund, Sweden
rootzen@math.chalmers.se tajvidi@maths.lth.se

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 has a multivariate GP distribution if and only if maxima asymptotically are 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 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 during the last decades. This is witnessed by the appearance of 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 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.) \( G(x) = \exp\left\{ -\left(1 + \frac{\gamma x - \mu}{\sigma}\right)^{-1/\gamma} \right\} \). 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. \( H(x) = 1 - (1 + \gamma x \sigma)^{-1/\gamma} \). 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) and Davidson and Smith (1990). Since the PoT method includes more of the relevant 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). This reference also makes a comprehensive study of methods for statistical inference for various families of Pareto distributions.

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). Several recent papers (e.g. Joe *et al.* (1992), Coles and Tawn (1991), Tawn (1998, 1990), Smith *et al.* (1990)) have explored their statistical application. For another recent review, see the chapter by Fougères in Finkenstadt and Rootzén (2004).
There are several possibilities for ordering multivariate data, see e.g. the review by Barnett (1976). For multivariate extreme values the most widely used method is the marginal or M-ordering where the maximum of a set of vectors is defined by taking componentwise maxima. Then, for a series of independent and identically distributed vectors \( \{X_i, i \geq 1 \} = \{(X_i^{(1)}, \ldots, X_i^{(d)}), i \geq 1 \} \), the maximum, \( M_n \), is defined by \( M_n = (M_n^{(1)}, \ldots, M_n^{(d)}) = (\bigvee_{i=1}^n X_i^{(1)} , \ldots, \bigvee_{i=1}^n X_i^{(d)}) \). Here and hereafter \( \bigvee \) denotes maximum.

Under rather general conditions, the distribution of the suitably normalized \( M_n \) converges to a multivariate extreme value distribution. In applications \( M_n \) 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 available data and which can contribute to better estimation of parameters in the model. 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 has a multivariate GP distribution if and only if 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 which is 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). Section 2 also discusses this relation. In Section 3 we show how an ”explicit” formula for the multivariate EV distributions directly leads to corresponding explicit expression for the multivariate GP distributions. The section also illustrates the general results with a few concrete examples. Lowerdimensional “marginals” of multivariate GP distributions may be thought of in different ways. This is discussed in Sections 4. Proofs of the results in Section 2 are given in Section 5.

Finally, this paper is a further development of the results in Tajvidi (1996).
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 \( \{X_i, i \geq 1\} = \{(X_i^{(1)}, \ldots X_i^{(d)}), i \geq 1\} \) are independent, identically distributed \( d \)-dimensional random vectors with distribution function \( F \). As before, let \( M_n \) be the vector of componentwise maxima,

\[
M_n = (M_n^{(1)}, \ldots M_n^{(d)}) = (\bigvee_{i=1}^n X_i^{(1)}, \ldots \bigvee_{i=1}^n X_i^{(d)}).
\]

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 \( n \to \infty \)

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

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 \( G_i(x) = e^{-(1+\gamma_i^{-1} \frac{x-\mu_i}{\sigma_i})^{-1/\gamma_i}} \). Here \( \mu_i \) is a location parameter, \( \sigma_i > 0 \) is a scale parameter, \( \gamma_i \) 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

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

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 \( X \) and \( Y \) are bivariate vectors and \( \alpha \) is an scalar, we have

\[
\alpha X = (\alpha X_1, \alpha X_2) \quad ; \quad \alpha + X = (\alpha + X_1, \alpha + X_2)
\]
and
\[ \mathbf{X} \mathbf{Y} = (X_1Y_1, X_2Y_2) \quad ; \quad \mathbf{X} + \mathbf{Y} = (X_1 + Y_1, X_2 + Y_2). \]

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. As another example, with this notation and with \( \mathbf{u}_n = (u_n^{(1)}, \ldots u_n^{(d)}) \), \( \sigma_n = (\sigma_n^{(1)}, \ldots \sigma_n^{(d)}) \), the limit (1) can be written as
\[ P((\mathbf{M}_n - \mathbf{u}_n)/\sigma_n \leq \mathbf{x}) \rightarrow G(\mathbf{x}). \]

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
\[ H(\mathbf{x}) = \frac{1}{-\log G(0)} \log \frac{G(\mathbf{x})}{G(\mathbf{x} \land 0)} \tag{2} \]
for some extreme value distribution \( G \) with nondegenerate margins and with \( 0 < G(0) < 1 \). In particular, \( H(\mathbf{x}) = 0 \) for \( \mathbf{x} < \mathbf{0} \) and \( H(\mathbf{x}) = 1 - \frac{\log G(\mathbf{x})}{\log G(0)} \) for \( \mathbf{x} > \mathbf{0} \).

Perhaps somewhat more elegantly, the class of multivariate GP distributions could alternatively be taken to be all distributions of the form
\[ H(\mathbf{x}) = \log \frac{G(\mathbf{x})}{G(\mathbf{x} \land 0)}, \]
for \( G \) an EV distribution with \( G(0) = e^{-1} \). This is not less general than (2) since if we let \( t = 1/(\log G(0)) \) then the \( H \) in (2) is of the form \( \log G(t \mathbf{x})/G(t \mathbf{x} \land 0) \), and by max-stability \( G^t \) again is an EV distribution, with \( G(0)^t = \exp(-t \log G(0)) = e^{-1} \). However, for statistical applications one would typically want to parametrize \( G \) with (at least) a scale parameter, and then the form (2) is more convenient.

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 \( \{ \mathbf{u}(t) \mid t \in [1, \infty) \} \) 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
\[ \mathbf{X}_u = \frac{\mathbf{X} - \mathbf{u}}{\sigma(\mathbf{u})} \]

5
be the vector of normalized exceedances of the levels \( u \). 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. In the theorems the components of the curve \( \{u(t)\} \) 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 \( 0 < G(0) < 1 \). If \( F \in D(G) \) then there exists an increasing continuous curve \( u \) with \( F(u(t)) \to 1 \) as \( t \to \infty \), and a function \( \sigma(u) > 0 \) such that

\[
P(X_u \leq x | X_u \not\equiv 0) \rightarrow \frac{1}{-\log G(0)} \frac{G(x)}{G(x \wedge 0)}
\]

as \( t \to \infty \), for all \( x \).

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

\[
P(X_u \leq x | X_u \not\equiv 0) \rightarrow H(x),
\]

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

\[
H(x) = \log \frac{G(x)}{G(x \wedge 0)}.
\]

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

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

**Theorem 2.3.** (i) Suppose \( X \) has a multivariate generalized Pareto distribution. Then there exists an increasing continuous curve \( u \) with \( P(X \leq u(t)) \to 1 \) as \( t \to \infty \), and a function \( \sigma(u) > 0 \) such that

\[
P(X_u \leq x | X_u \not\equiv 0) = P(X \leq x),
\]

for \( t \in [1, \infty) \) and all \( x \).

(ii) If there exist an increasing continuous curve \( u \) with \( P(X \leq u(t)) \to 1 \) as \( t \to \infty \), and a function \( \sigma(u) > 0 \) such that (6) holds for \( x > 0 \), and \( X \) has nondegenerate margins, then \( X \) 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 \( X_1, X_2, \ldots \) 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 \( S = \{ x : G(x) \in (0, 1) \} \) and let \( \mu \) be the measure on \( S \) which is determined by \( \mu(-\infty, x]^c = -\log G(x) \). Then there exist \( d \)-dimensional vectors of constants \( u_n \) and \( \sigma_n > 0 \) such that

\[
\sum_{i=1}^{n} \epsilon_{\frac{X_i - u_n}{\sigma_n}} \Rightarrow \text{PRM}(dt \times d\mu) \tag{7}
\]

on \( S \).

(ii) Suppose (7) holds on \( x \geq 0 \) for some measure \( \mu \) with \( \mu(-\infty, 0]^c \in (0, \infty) \) and where the function \( \mu(-\infty, x]^c \) has non-degenerate marginals. Then the conditions of Theorem 2.2 hold, and hence also (i) above is satisfied.

In the next section we use a wellknown representation of multivariate EV distributions to give a general expression for the entire class of multivariate GP distributions and also give a number of examples of such distributions.

### 3 Examples

In this section we exhibit a general expression for the multivariate generalized Pareto distributions. We also discuss some specific families of distributions in the bivariate case.

Several authors have given equivalent characterizations of multivariate extreme value distributions assuming different marginal distribution, see e.g. Resnick (1987) and Pickands (1981). To obtain the general form of these distributions with arbitrary marginals one only needs simple marginal transformations.

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

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

with

\[
\mu_*[0, x]^c = \int_{S} \bigvee_{i=1}^{d} \left( \frac{a^{(i)}}{x^{(i)}} \right) S(da) \tag{9}
\]
Here $S$ is a finite measure defined on
\[ \mathcal{N} = \{ y \in E : \|y\| = 1 \}, \tag{10} \]
where $E = \mathbb{R}_+^d \setminus \{0\}$. Further $S$ must satisfy
\[ \int_{\mathbb{R}^d_o} a^{(i)} S(da) = 1, \quad 1 \leq i \leq d, \tag{11} \]
and $\| \cdot \|$ is an arbitrary norm in $\mathbb{R}^d$. In this characterization $\mu_*$ 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
\[ G(x) = G_\ast \left( 1 + \gamma \left( \frac{x - \mu}{\sigma} \right) \right)^{-\frac{1}{\gamma}}. \]

A corresponding expression for the multivariate GP distributions can be obtained by combining this result and Definition 2.1. This is summarized in the following proposition.

**Proposition 3.1.** $H(x)$ is a multivariate GP distribution if there exists a finite measure $S$ satisfying (11) such that for $x \not\leq 0$

\[
H(x) = \frac{\int_{\mathbb{R}^d_o} \bigg( a^{(i)} \left( 1 + \frac{\gamma^{(i)} (x^{(i)} - \mu^{(i)})}{\sigma^{(i)}} \right) \bigg) S(da) - \int_{\mathbb{R}^d_o} \bigg( a^{(i)} \left( 1 - \frac{\gamma^{(i)}}{\sigma^{(i)}} \right) \bigg) S(da)}{\int_{\mathbb{R}^d_o} \bigg( a^{(i)} \left( 1 - \frac{\gamma^{(i)}}{\sigma^{(i)}} \right) \bigg) S(da)}.
\]

The parameters $\mu$, $\gamma$, 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 $0 < G(0) < 1$).

We now consider the bivariate case and give a simple representation of bivariate generalized Pareto distributions. We also show some specific examples of bivariate GP distributions. When $d = 2$, $\mathcal{N}$ is one-dimensional and Relation (9) becomes

\[ \mu_*[0, (x, y)]^c = \int_0^1 \max\{w/x, (1 - w)/y\} S(dw) \]
\[ = \left( \frac{1}{x} + \frac{1}{y} \right) A \left( \frac{x}{x + y} \right) \tag{12} \]
where
\[ A(q) = \int_0^1 \max\{w(1-q), (1-w)q\} S(dw). \] (13)

The function \( A(q) \) is called the dependence function, see e.g. Tawn (1998). Here \( S \) is finite positive measure on interval \([0,1]\). In order that (11) is satisfied we need
\[ \int_0^1 wS(dw) = \int_0^1 (1-w)S(dw) = 1. \] (14)

For every \( w \in [0,1] \), \( \max\{w(1-q), (1-w)q\} \) is a convex function and hence \( A(q) \) is convex since convexity is preserved through maximization and summation. By (14) we have that \( A(0) = A(1) = 1 \). In addition from (13) we see that the function is bounded above by 1 and below by \( \max\{q, 1-q\} \). The first boundary corresponds to independent extremes. In this case the corresponding measure \( S \) puts mass one at each of the endpoints 0 and 1. The other boundary corresponds to completely dependent extremes, that is \( P(X = Y) = 1 \). The measure \( S \) then puts mass 2 at \( \frac{1}{2} \). Equation (14) gives that \( S(dw)/2 \) is a probability measure with mean \( \frac{1}{2} \). As we mentioned earlier there is no finite-dimensional parametric family for the dependence function.

We can now use dependence functions to give a representation of bivariate GP distributions which correspond to bivariate EV distributions with standard Gumbel marginals. Location and scale parameters can of course be added to each margin as in Proposition 3.1, and the case of general marginal distributions can be obtained by transformation.

**Proposition 3.2.** \( H(x, y) \) is a bivariate GP which corresponds to a bivariate EV distributions with marginal distributions \( \Lambda(x) = e^{-e^{-x}}, x \in R \) if there exists a convex function on \([0,1]\) satisfying \( A(0) = A(1) = 1 \) and \( \max\{w, 1-w\} \leq A(w) \leq 1 \) such that for \((x, y) \notin (0,0)\)
\[ H(x, y) = \frac{(e^{-(x\wedge 0)} + e^{-(y\wedge 0)})A\left(e^{x\wedge 0} + e^{y\wedge 0}\right) - (e^{-x} + e^{-y})A\left(e^{x} + e^{y}\right)}{2A(\frac{1}{2})}. \]

Next, a few examples:

**Example 1** The symmetric logistic model has exponent measure
\[ \mu_*([0, (x, y)]^c) = (x^{-r} + y^{-r})^{1/r}, \quad r \geq 1. \]
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

\[
H(x, y) = \left(1 + \frac{\gamma (x \wedge 0 - \mu)}{\sigma}\right)^{-r/\gamma} + \left(1 + \frac{\gamma (y \wedge 0 - \mu)}{\sigma}\right)^{-r/\gamma} - \left(1 + \frac{\gamma (x - \mu)}{\sigma}\right)^{-r/\gamma} + \left(1 + \frac{\gamma (y - \mu)}{\sigma}\right)^{-r/\gamma} \left(1 - \frac{2u}{\sigma}\right)^{-r/\gamma} \left(1 - \frac{2y}{\sigma}\right)^{-r/\gamma}^{1/r}.
\]

**Example 2** The asymmetric logistic model has dependence function

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

with exponent measure

\[
\mu_*([(0, 0), (x, y)]^c) = \frac{(1 - \phi)x + (1 - \theta)y + x (\phi^r x^r + \theta^r y^r)^{1/2} + y (\phi^r x^r + \theta^r y^r)^{1/2}}{xy}.
\]

For \(\theta = \phi = 1\) this model reduces to the corresponding symmetric logistic model which gives the diagonal case for \(r = +\infty\). Independence is obtained for \(\theta = 0\) and for \(\phi = 0\) or \(r = 1\).

The corresponding bivariate GP distribution is

\[
H(x, y) = \frac{\mu([(0, 0), ((x \wedge 0), (y \wedge 0))]^c) - \mu([(0, 0), (x, y)]^c)}{\mu([(0, 0), (0, 0)]^c)}.
\]

where

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

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

**Example 3** Suppose \(X\) and \(Y\) are independent and exponentially distributed and let \(\sigma_x, \sigma_y, p, q > 0\) with \(p + q = 1\). Taking \(\sigma = (\sigma_x^{-1}, \sigma_y^{-1})\) and \(u = (\log t/p, \log t/q)\) one can easily see that the normalized maxima converge,

\[
\frac{M_n - u_n}{\sigma} \to \exp (-pe^{-x/\sigma_x} - qe^{-y/\sigma_y})
\]
The corresponding bivariate GP distribution function $H(x, y)$ is shown in Figure 1. By differentiation one can see that this distribution has no density, and in fact it can be described as follows. Let $X_\infty$ and $Y_\infty$ be independent exponential variables with means $\sigma_x$ and $\sigma_y$. 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 $q$. This completely agrees with intuition: by independence, if $X > x/\sigma_x + \log t/p$ for a large $t$ (and $x > 0$) this doesn’t influence the value of $Y$ and hence $\sigma_y^{-1}(Y - \log t/q)$ is ”close” to $\sigma_y^{-1}\log t/q$, which tends to $-\infty$ as $t \to \infty$. The roles of $X$ and $Y$ can of course be interchanged in this argument.

Example 4 If we in the previous example instead assume complete dependence, i.e. $X = Y$, and take $\sigma = (\sigma_x^{-1}, \sigma_y^{-1})$ and $u = (\log tp, \log tq)$ we obtain

$$\frac{M_n - u_n}{\sigma} \to \exp(-\exp(\min(x/\sigma_x + \log p, y/\sigma_y + \log q)))$$

Figure 2 depicts the distribution function for the bivariate GP distribution in this case. This distribution of course is concentrated on the line $y = \sigma_y(x/\sigma_x + \log p/q)$.

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 3 in the previous section and $H_1(x)$ is the marginal distribution of the first component, then

$$H_1(x) = H(x, \infty) = 1 - pe^{-x\infty/\sigma}. \quad (16)$$

This is the distribution of a variable which is exponentially distributed with probability $p$ and which equals $-\infty$ with probability $q$. Hence it is not a onedimensional generalized Pareto distribution.

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

$$H(y) = \frac{1}{-\log G(0)} \log \frac{G(y)}{G(y \land 0)} \quad (17)$$

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

$$H_{GP}^{(i)}(x) = \frac{1}{-\log G(i)(0)} \log \frac{G(i)(x)}{G(i)(x \land 0)}. \quad (18)$$

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

The expression (18) is implicit since $i$ involves $G$. However the GP marginal distribution can also be expressed directly in terms of the parent GP distribution $H$. For $x$ a $(d - 1)$-dimensional vector write $x(i)$ for the $d$-dimensional vector which is obtained from $x$ by inserting an $\infty$ at position $i$, i.e. $x(i) = (x_1, \ldots, x_{i-1}, \infty, x_i, \ldots, x_{d-1})$. Then, as can be seen by straightforward computation,

$$H_{GP}^{(i)}(x) = \frac{1}{1 - \log H(0(i))} \log \frac{H(x(i))}{H((x \land 0)(i))}. \quad (19)$$
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.

### 5 Proofs

**Proof of Theorem 2.2.**

(i) By definition, if \(F \in D(G)\) then there exists \(\sigma_n > 0, u_n\) in \(\mathbb{R}^d\) such that

\[
F^n(\sigma_n x + u_n) \to G(x)
\]

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

By (20) also \(F^n(\sigma_{n+1} x + u_{n+1}) \to G(x)\) and by the convergence of types theorem (Leadbetter et al. (1983, p. 7)) applied to each marginal, it follows that

\[
\sigma_{n+1}/\sigma_n \to 1 \quad \text{and} \quad (u_{n+1} - u_n)/\sigma_n \to 0.
\]

Taking logarithms, it is seen that (20) is equivalent to

\[
n\bar{F}(\sigma_n x + u_n) \to -\log G(x).
\]

Now, define \(u\) by \(u(t) = u_n\) for \(t = n\) and by linear interpolation for \(n < t < n + 1\), and set \(\sigma(u(t)) = \sigma_n\) for \(n \leq t < n + 1\). It then follows from (21) and (22) that

\[
t\bar{F}(\sigma(u(t))x + u(t)) \to -\log G(x).
\]

By straightforward argument

\[
P(X_u \leq x | X_u \not\leq 0) = \frac{P(X_u \not\leq x \land 0) - P(X_u \not\leq x)}{P(X_u \not\leq 0)},
\]

for \(x \not\leq 0\). Since \(P(X_u \not\leq x) = \bar{F}(\sigma_n x + u)\), (3) now follows from (23).

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

\[
\limsup_{t \to \infty} \frac{\tilde{F}(u(t-))}{\tilde{F}(u(t))} \leq \limsup_{t \to \infty} \frac{\tilde{F}(u(t))}{\tilde{F}(u(t)) + \sigma(u(t))\epsilon} = \frac{1}{H(\epsilon)}.
\]

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

\[
t\tilde{F}(u(t)) \to 1.
\]

It follows from (4) and (24) that, for \( x > 0 \),

\[
\lim_{t \to \infty} \frac{\tilde{F}(u(t)) + \sigma(u(t))x}{\tilde{F}(u(t))} = \lim_{t \to \infty} \frac{P(X_u \leq x)}{P(X_u \leq 0)} = \tilde{H}(x).
\]

We next show that (25) holds also when \( x \) isn’t positive. From (25) follows that \( F(u(t) + \sigma(t)x)^t \to e^{-\tilde{H}(x)} \) for \( x > 0 \). Further, \( t\tilde{F}(u(\Delta t) + \sigma(u(\Delta t))x) \to \tilde{H}(x)/\Delta \) for \( \Delta > 0 \) and hence

\[
F(u(\Delta t) + \sigma(\Delta t))^{t} \to e^{-\tilde{H}(x)/\Delta}.
\]

It follows by the extremal types theorem (c Leadbetter et al. (1983, p.7) that there exist \( c_{\Delta} > 0 \), \( x_{\Delta} \) with

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

Thus, for any \( x \geq -x_{\Delta}/c_{\Delta} \),

\[
\Delta t\tilde{F}(u(\Delta t) + \sigma(\Delta t)x) = \Delta t\tilde{F}(u(t) + \sigma(t)(\frac{\sigma(\Delta t)}{\sigma(t)}x + \frac{u(\Delta t) - u(t)}{\sigma(t)}))
\]

\[
\to \Delta \tilde{H}(c_{\Delta}x + x_{\Delta}).
\]

This may be rephrased as

\[
t\tilde{F}(u(t) + \sigma(t)x) \to \Delta \tilde{H}(c_{\Delta}x + x_{\Delta}).
\]

Hence the limit does not depend on the choice of \( \Delta \), and we may uniquely define \( -\log G(x) \) as \( \Delta \tilde{H}(cx + x_{\Delta}) \) for any \( x \geq \inf_{\Delta \geq 1} x_{\Delta}/c_{\Delta} \) to obtain

\[
t\tilde{F}(u(t) + \sigma(t)x) \to -\log G(x),
\]

for such \( x \).
Further, suppose that one of the coordinates in a vector $x$ is less than the corresponding coordinate of $\inf_{\Delta \geq 1} x_{\Delta}/c_{\Delta}$. Then if we let $x^\infty$ be the vector which has all other components set to $\infty$, and $x_0^\infty$ the vector were all other components are set to $0$, we have that

$$\liminf_{t \to -\infty} t \bar{F}(u(t) + \sigma(t)x) \geq \liminf_{t \to -\infty} t \bar{F}(u(t) + \sigma(t)x^\infty) \geq \Delta \bar{H}(x_0^\infty),$$

for any $\Delta > 1$. It follows that $t \bar{F}(u(t) + \sigma(t)x) \to \infty$ for such $x$, and thus if we define $-\log G(x) = \infty$ for $x$ which are not greater than $\inf_{\Delta \geq 1} x_{\Delta}/c_{\Delta}$ then (26) holds for all $x$. Thus,

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

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

**Proof of Theorem 2.3.**

(i) Let $X$ have distribution $H$. By definition, $H$ is of the form (2) for some extreme value distribution $G$, so that

$$P(X \not\in x) = 1 - \frac{1}{-\log G(0)} \log \frac{G(x)}{G(x \land 0)} = \frac{1}{-\log G(0)} \log \frac{G(x)G(0)}{G(x \land 0)}. \quad (27)$$

Since $G$ is max-stable, there exist continuous curves $\sigma(t) > 0, u(t)$ with $\sigma(1) = 1, u(1) = 0$ and $u(t)$ strictly increasing, such that $G(u(t) + \sigma(t)x)^t = G(x)$. In particular, $G(u(t)) = G(0)^{1/t}$. Further, by (24) and (27),

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

Since $(x \land 0) \land 0 = x \land 0$ it follows that

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

(ii) This is an easy consequence of Theorem 2.2 (ii). \hfill \Box

**Proof of Theorem 2.4.**
(i) By Theorem 2.2 we have that $F \in D(G)$, and hence there are constants $u_n, \sigma_n$ such that $nP(\frac{X_i-u_n}{\sigma_n} \not\sim x) \to -\log G(x)$ 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 $nP(\frac{X_i-u_n}{\sigma_n} \not\sim x) \to \mu((-\infty, x]\cap c)$ for $x > 0$. 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


