Multivariate Extremes

For ease of writing we only consider the two-dimensional case. Consider i.i.d. random vectors $(X_1, Y_1),\ (X_2, Y_2), \ldots$ with distribution function F. Suppose that for suitabe sequences of constants $a_n, c_n > 0$ and $b_n, d_n \quad (n = 1, 2, \ldots)$ the sequence

$$\left(\frac{\max\limits_{i \le n} X_i - b_n}{a_n}, \ \frac{\max\limits_{i \le n} Y_i - d_n}{c_n} \right)$$

has a non-degenerate limit distribution, i.e.

$$\lim_{n \to \infty} F^n(a_n x + b_n,\ c_n y + d_n) =: G(x, y) \tag{1}$$

exists for all but countably many (x, y) and G the distribution function of a non-degenerate probability distribution.

1 Question: What limit distributions can occur?

1a. Via Fréchet distribution.

In order to answer that question, we first look at the marginal distributions of F. They should converge to a one–dimensional extreme–value distribution, i.e.

$$\begin{aligned} \lim_{n \to \infty} F^n(a_n x + b_n, \infty) &= \exp -(1 + \gamma_1 x)^{-1/\gamma_1} \\ \lim_{n \to \infty} F^n(\infty, c_n y + d_n) &= \exp -(1 + \gamma_2 x)^{-1/\gamma_2} \end{aligned} \tag{2}$$

for some $\gamma_1, \gamma_2 \in \mathbb{R}$. Moreover for $s > 0$

$$\begin{aligned} \frac{b_{[ns]} - b_n}{a_n} &\to \frac{s^{\gamma_1} - 1}{\gamma_1} \qquad \frac{a_{[ns]}}{a_n} \to s^{\gamma_1} \\ \frac{d_{[ns]} - d_n}{c_n} &\to \frac{s^{\gamma_2} - 1}{\gamma_2} \qquad \frac{c_{[ns]}}{c_n} \to s^{\gamma_2} \end{aligned} \tag{3}$$

$n \to \infty$ and all these relations hold locally uniformly on $(0, \infty)$.

Next we return to the two–dimensional distribution. For $s > 0$

$$G(x, y) \leftarrow F^{[ns]} \left(a_{[ns]} x + b_{[ns]}, c_{[ns]} y + d_{[ns]} \right) \sim F^{ns} \left(a_{[ns]} x + b_{[ns]}, c_{[ns]} y + d_{[ns]} \right)$$

$$= F^{ns} \left(a_n \left(\frac{a_{[ns]}}{a_n} x + \frac{b_{[ns]} - b_n}{a_n} \right) + b_n,\ c_n \left(\frac{c_{[ns]}}{c_n} y + \frac{d_{[ns]} - d_n}{c_n} \right) + d_n \right)$$

$$\longrightarrow G^s\left(s^{\gamma_1}x+\frac{s^{\gamma_1}-1}{\gamma_1},\ s^{\gamma_2}y+\frac{s^{\gamma_2}}{y}+\frac{s^{\gamma_2}-1}{\gamma_2}\right)$$

$n\to\infty$, by (1) and (3). Hence, replacing x by $\frac{u^{\gamma_1}-1}{\gamma_1}$ and y by $\frac{v^{\gamma_2}-1}{\gamma_1}$,, for $s,u,v>0$

$$G\left(\frac{u^{\gamma_1}-1}{\gamma_1},\ \frac{v^{\gamma_2}-1}{\gamma_2}\right)=G^s\left(\frac{(su)^{\gamma_1}-1}{\gamma_1},\ \frac{(sv)^{\gamma_2}-1}{\gamma_2}\right).$$

Hence it is useful to define

$$G_1(x,y):=G\left(\frac{x^{\gamma_1}-1}{\gamma_1},\ \frac{y^{\gamma_2}-1}{\gamma_2}\right). \tag{4}$$

Then for $a,x,y>0$

$$G_1^a(ax,ay)=G_1(x,y). \tag{5}$$

This is the basic homogeneity relation in multidimensional extreme–value theory.

In order to reveal the consequences of (5) we go back to (1): for (x,y) with $0<G(x,y)<1$

$$\begin{aligned}-\log G_1(x,y) &= \lim_{n\to\infty} -\log F^n(a_n\frac{x^{\gamma_1}-1}{\gamma_1}+b_n,\ c_n\frac{y^{\gamma_2}-1}{\gamma_2}+d_n)\\ &= \lim_{n\to\infty} -n\log F(a_n\frac{x^{\gamma_1}-1}{\gamma_1}+b_n,\ c_n\frac{y^{\gamma_2}-1}{\gamma_2}+d_n) \quad (6)\\ &= \lim_{n\to\infty} n\{1-F(a_n\frac{x^{\gamma_1}-1}{\gamma_1}+b_n,\ c_n\frac{y^{\gamma_2}-1}{\gamma_2}+d_n)\}.\end{aligned}$$

Define a measure ν_n (not a probability measure) by

$$\nu_n\{(s,t)\mid s>x \text{ or } t>y\}:=n\left\{1-F(a_n\frac{x^{\gamma_1}-1}{\gamma_1}+b_n,\ c_n\frac{y^{\gamma_2}-1}{\gamma_2}+d_n)\right\}$$

for $x,y>0$. Then (6) tells us that $\nu_n(I)$ converges for all sets of the form $I=\{(s,t)\mid s>x \text{ or } t>y\},\ x,y>0$. Hence the limit function must be a measure too, i.e. there exists a measure ν on $\mathbb{R}_+^2$ such that for $x,y>0$

$$\nu\{(s,t)\mid s>x \text{ or } t>y\}=-\log G_1(x,y).$$

From (5) it is clear that the measure ν is homogeneous: for any Borel set away from the origin and $a>0$

$$\nu(aB)=a^{-1}\nu(B) \tag{7}$$

where aB means the set of all elements aw with $w \in B$.

We use this relation in order to find the structure of the limit distribution function G_1. Take the Borel set

$$B_{\theta,r} := \left\{(u,v) \mid u^2 + u^2 > r^2 \ \& \ arctg, \ \frac{u}{v} \le \theta\right\}$$

for $r > 0$ and $0 \le \theta \le \pi/2$. Then $B_{\theta,r} = rB_{\theta,1}$, and hence

$$\nu(B_{\theta,r}) = \nu(r \cdot B_{\theta,1}) = r^{-1} \cdot \nu(B_{\theta,1}).$$

This product property determines G_1 : write

$$\Phi(\theta) := \nu(B_{\theta,1}). \tag{8}$$

Then Φ is the distribution function of a finite measure on $[0, \pi/2]$. In fact there is no other restriction on Φ: any finite measure gives rise to an extreme-value distribution as the following relation (de Haan and Resnick (1977)) shows. Any extreme value distribution (type) can be written as

$$G\left(\frac{x^{\gamma_1} - 1}{\gamma_1}, \ \frac{y^{\gamma_2} - 1}{\gamma_2}\right) = \exp - \int_0^{\pi/2} \left(\frac{\cos\theta}{x} \vee \frac{\sin\theta}{y}\right) \Phi(d\theta). \tag{9}$$

It is in fact easy to see that a distribution function of the type (8) satisfies (5) and hence has a non–empty domain of attraction.

So we see that the limit distribution G in (1) is determined by two real parameters (γ_1 and γ_2) and a finite measure on $[0, \pi/2]$ (i.e. Φ).
The measure associated with Φ is called the spectral measure (or one possible form of the spectral measure, cf. 1b).

Example–proof–motivation

Take i.i.d. random variables Y_1, and Y_2 with distribution function $\exp(-1/x)$ for $x > 0$ and compose for $\boldsymbol{r} = (r_1, r_2)$ and $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)$ with $0 \le \varphi_1, \varphi_2 \le \pi/2$

$$(X, Y) := (\max(r_1 Y_1 \cos\varphi_1, \ r_2 Y_2 \cos\varphi_2), \ \max(r_1 Y_1 \sin\varphi_1, \ r_2 Y_2 \sin\varphi_2)).$$

Then

$$P\{X \le x, \ Y \le y\} = P\{r_1 Y_1 \cos\varphi_1 \le x, \ r_2 Y_2 \cos\varphi_2 \le x, r_1 Y_1 \sin\varphi_1 \le y, r_2 Y_2 \sin\varphi_2 \le y\}$$

$$= P\left\{Y_1 \le \frac{x}{r_1 \cos\varphi_1} \wedge \frac{y}{r_1 \sin\varphi_1}, \ Y_2 \le \frac{x}{r_2 \cos\varphi_2} \wedge \frac{y}{r_2 \sin\varphi_2}\right\}$$

$$= \exp\left(-\frac{r_1\cos\varphi_1}{x} \vee \frac{r_1\sin\varphi_1}{y}\right)\exp\left(-\frac{r_2\cos\varphi_2}{x} \vee \frac{r_2\cos\varphi_2}{y}\right)$$

$$\exp - \sum_{i=1}^{2} r_i\left(\frac{\cos\varphi_i}{x} \vee \frac{\sin\varphi_i}{y}\right).$$

By starting from more and more variables Y_i, $i = 1, \cdots, k$, and taking the limit $k \to \infty$ we arrive at (8).

From this construction we can obtain easily the special cases:

a. If Φ is concentrated on the points $\{0\}$ and $\{\pi/2\}$, it means that $G(x, y) = G(x, \infty)G(\infty, y)$ i.e. G has independent components. It also means that $\nu\{(s, t) \mid s > x \text{ and } t > y\} = 0$ for $x, y > 0$, i.e. ν is concentrated on the coordinate axes.

b. Φ concentrates on the point $\theta_0 \in (0, \pi/2)$. Then there is full dependence i.e. one component of G determines the other.

1b. Via uniform distribution

In multidimensional extreme–value theory, unlike the theory of multidimensional stable laws, there is not one canonical set of conditions. There are many sets of conditions possible that only agree at a meta level. Let us consider one of the alternative ways. From (2) we know that we can choose b_n and d_n as in the one–dimensional case i.e.

$$\begin{aligned} b_n &= F_1^{\leftarrow}(1 - 1/n) =: U_1(n) \\ d_n &= F_2^{\leftarrow}(1 - 1/n) =: U_2(n) \end{aligned} \tag{10}$$

where F_1 and F_2 are the two marginal distribution functions of F. The arrow means: inverse function.

We have for $s, t > 0$ by (1) and (3)

$$\begin{aligned} F^n(b_{[ns]}, d_{[nt]}) &= F^n\left(b_n + a_n\frac{b_{[ns]} - b_n}{a_n}, d_n + c_n\frac{d_{[nt]} - d_n}{c_n}\right) \\ &\longrightarrow G\left(\frac{s^{\gamma_1} - 1}{\gamma_1}, \frac{t^{\gamma_2} - 1}{\gamma_2}\right). \end{aligned} \tag{11}$$

We now reason as before without indicating the details. By (10) and (11) for $s, t > 0$ (note (X, Y) is a random vector with distribution function F)

$$\begin{aligned} &nP\left\{n\left(1 - F_1(X)\right) < s \text{ or } n\left(1 - F_2(Y)\right) < t\right\} \\ =&nP\left\{X > U_1(n/s) \text{ or } Y > U_2(n/t)\right\} \\ &\longrightarrow -\log G\left(\frac{s^{-\gamma_1} - 1}{\gamma_1}, \frac{t^{-\gamma_2} - 1}{\gamma_2}\right) =: l(s, t) \quad (n \to \infty). \end{aligned} \tag{12}$$

More generally for a Borel set B away from (∞, ∞)

$$nP\left\{(1 - F_1(X),\ 1 - F_2(Y_2)) \in B\right\} \to \tau(B)$$

with τ a measure satisfying

$$\nu(a \cdot B) = a \cdot \nu(B) \tag{13}$$

for $a > 0$ and B a Borel set away from (∞, ∞), or, for the function l:

$$l(ax, ay) = al(x, y) \tag{14}$$

for $a, x, y > 0$, with $l(x, 0) = l(0, x) = x$. We have the approximate (for large x, y) equality (easy to remember)

$$1 - F(x, y) \approx l\left(1 - F_1(x),\ 1 - F_2(y)\right). \tag{15}$$

In a way analogous to what we did before we get

$$\tau\left\{(s, t) \mid s \wedge t < r,\ \frac{s}{t} \leq \tan\theta\right\} = r \cdot \Psi(\theta)$$

and Ψ is another form of the spectral measure.

1a. and **1b.** agree in that one first transforms to standard marginal distribution functions and then exploits the special structure of the joint distribution. They differ in two ways:

Firstly the standard distributions are different: Fréchet's distribution ($\gamma = 1$) in **1a.** and the uniform distribution ($\gamma = -1$) in **1b.**. Moreover the set on which the radial part of the measure (i.e. the spectral measure) is concentrated is different: for **1a.** this is the boundary of the unit circle (l_2 norm) and for **1b.** this is the boundary of the unit square (l_∞ norm). Another standard distribution that has been used is the exponential distribution ($\gamma = 0$). Another form of the spectral measure that has been used, is concentrated on the line $x + y = 1$ (l_1 norm).

2 Domains of attraction

The discussion in 1. leads to the following criteria for the domain of attraction.

Theorem 0.1. *Suppose G is a two–dimensional extreme value distribution characterized by the extreme value indices γ_1 and γ_2 and the spectral measure Φ. For a distribution function F with marginal distribution functions F_1 and F_2 the following statements are equivalent:*

1. *F is in the domain of attraction of G, i.e. (1) holds.*
2. *F_i is in the domain of attraction of an extreme–value distribution with index γ_i $(i = 1, 2)$ and*

$$\lim_{t\to\infty} t\left\{1 - F\left(U_1(tx),\ U_2(ty)\right)\right\} = -\log G\left(\frac{x^{\gamma_1}-1}{\gamma_1},\ \frac{y^{\gamma_2}-1}{\gamma_2}\right)$$

   for $x, y, > 0$. Here $U_i(z) := F_i^{\leftarrow}(1 - 1/z)$ for $i = 1, 2$.

3. *F_i is in the domain of attraction of an extreme–value distribution with index γ_i $(i = 1, 2)$ and for $r > 0,\ 0 \le \theta \le \pi/2$ (if (X, Y) is a random vector with distribution function F)*

$$\lim_{t\to\infty}\ tP\{Z_1^2 + Z_2^2 > t^2r^2 \ \textit{and arc tg}\frac{Z_2}{Z_1} \le \theta\} = r^{-1}\cdot\Phi(\theta)$$

   where $Z_1 := 1/(1 - F_1(X))$ and $Z_2 = 1/(1 - F_2(Y))$.

3 Statistics of multivariate extreme values

Suppose we have independent observations $(X_1, Y_1),\ (X_2, Y_2), \cdots, (X_n, Y_n)$ with common distribution function F in the domain of attraction of some extreme–value distribution G. Let F be an extreme failure set i.e. $X_i << x$ or $Y_i << y$ for any $(x, y) \in F$ and $i = 1, 2, \cdots, n$. Let P be the probability measure corresponding to F. We are interested in an estimator for $P(F)$.

We want to estimate $P(F)$ by using multidimensional extreme–value theory i.e. via an asymptotic method pretending that the number of observation increases without bound. Now a characteristic feature of the situation is that all the observations are quite far from F. Hence a first estimate is $P(F) < 1/n$. If we want to retain this characteristic feature as $n \to \infty$, we are forced to assume that F is not a fixed set put it depends on n in such a way that $P(F_n) \to o,\ n \to \infty$. In fact, in order to get a non–trivial result, we need to assume that F_n is of the form

$$F_n = \boldsymbol{a}(\frac{n}{k}) \frac{(c_n A)^{\boldsymbol{\gamma}} - 1}{\boldsymbol{\gamma}} + \boldsymbol{b}(\frac{n}{k}) \tag{16}$$

where $\{c_n\}$ is a sequence of positive numbers (typically $c_n \to \infty$), A a failure set in $(0,\infty)^2$, $\boldsymbol{a}(t) := (a(t), c(t))$, $\boldsymbol{b}(t) = (b(t), d(t))$ and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$. The sequence $k = k(n)$ is intermediate: $k(n) \to \infty,\ k(n)/n \to 0,\ n \to \infty$.

How do we estimate $p_n = P(F_n)$?

Let us go back to (6): if (X, Y) is a random vector with distribution function F,

$$\begin{aligned} &nP\left\{\left(1 + \gamma_1 \frac{X - b(n)}{a(n)}\right)^{1/\gamma_1} > x \text{ or } \left(1 + \gamma_2 \frac{Y - d(n)}{c(n)}\right)^{1/\gamma_2} > y\right\} \\ &\longrightarrow \nu\left\{(u, v) \mid u > x \text{ or } v > y\right\}. \end{aligned} \tag{17}$$

It is not difficult to see that this also holds with the running variable $n \to \infty$ replaced by a continuous variable $t \to \infty$. Hence also with n replaced by n/k.
Also (17) holds with the set

$$\{(u, v) \mid u > x \text{ or } v > y\}$$

replaced by an arbitrary Borel set in $(0,\infty)^2$, i.e. for $n \to \infty$

$$\begin{aligned} &\frac{n}{k} P\left\{\left(\left(1 + \gamma \cdot \frac{X - b(\frac{n}{k})}{a(\frac{n}{k})}\right)^{1/\gamma_1},\ \left(1 + \gamma_2 \frac{Y - d(\frac{n}{k})}{c(\frac{n}{k})}\right)^{1/\gamma_2}\right) \in B\right\} \\ &\longrightarrow \nu(B) \end{aligned} \tag{18}$$

Next observe that by (18) and (6)

$$\begin{aligned} p_n = P(F_n) &= P\left\{\left(\left(1 + \gamma_1 \frac{X - b(\frac{n}{k})}{a(\frac{n}{k})}\right)^{1/\gamma_1}, \left(1 + \gamma_2 \frac{Y - d(\frac{n}{k})}{c(\frac{n}{k})}\right)^{1/\gamma_2}\right) \in c_n A\right\} \\ &\approx \frac{k}{n} \nu(c_n A) \\ &= \frac{k}{n c_n} \nu(A). \end{aligned} \tag{19}$$

This leads us to propose the following estimator for p_n:

$$\hat{p}_n := \frac{k}{n\hat{c}_n}\hat{\nu}(\hat{A}).$$

This requires estimators for c_n, ν and A. They can be constructed (cf. de Haan and de Ronde (1998)) and under appropriate further conditions the asymptotic normality of

$$\frac{\sqrt{k}}{q_{\gamma_1 \wedge \gamma_2}(c_n)} \left(\frac{\hat{p}_n}{p_n} - 1\right)$$

can be established with $q_\gamma(x) := x^{-\gamma} \int_1^x y^{\gamma-1} \log y \; dy$. A visual explanation of the various steps in (19) is given in figures 1 and 2.

The first step is a standardization of the marginal distributions (transformation from F_n to $c_n A$).

The next step exploits the homogeneity property of the measure ν which enables us to shift a distant failure set with no observations in it ($c_n A$) to a set (A) which indeed contains some observations so that the measure of that set can be estimated by the empirical measure (counting the number of observations in the set).

This is the basis for estimating the probability of a remote failure set.