Topics on Multivariate

and

Infinite-Dimensional Extremes

Ana Ferreira

ISA, Universidade Técnica de Lisboa

http://home.isa.utl.pt/ anafh/anafh.html

Extremes Day

Faculty of Sciences, University of Lisbon, Portugal

February 22, 2006

A. Multivariate Extremes

MOTIVATION

How to determine the probability that an independent observation falls into the failure set C?

Suppose we have n i.i.d. observations of (X, Y), though well *far away* from the failure set C.

Failure set because its points (may) represent failure of the system.

Typical example: (X, Y) are sea level conditions (for instance wave heigh and surge) and $(x, y) \in C$ (may) leads to a flood.

So there are two (or more) random variables involved and we are concerned with joint extremes $\Rightarrow$ *Multivariate Extreme Value Theory*

de Haan and Resnick (1977); Resnick (1987); de Haan and Ferreira (2006)

THE LIMIT DISTRIBUTION

Let (X_i, Y_i) i.i.d. random vectors from

$$F(x,y) = P(X \leq x, Y \leq y).$$

If there exist $a_n, c_n > 0$, and b_n, d_n real (normalizing constants) and G is a d.f. with non-degenerate marginals such that, as $n \to \infty$,

$$\begin{aligned} &F^n(a_n x + b_n, c_n y + d_n) \\ &\qquad = P\left(\frac{\max(X_1, \ldots, X_n) - b_n}{a_n} \leq x, \right. \\ &\qquad\qquad \left. \frac{\max(Y_1, \ldots, Y_n) - d_n}{c_n} \leq y\right) \to^d G(x,y), \quad (1) \end{aligned}$$

then we have convergence (in distribution) of the joint (normalized) maxima and G is a *multivariate extreme value distribution*.

This implies convergence of the marginals F_i, $i = 1, 2$,

$$F_1^n(a_n x + b_n) \to^d \exp\left(-(1+\gamma_1 x)^{-1/\gamma_1}\right);$$

$$F_2^n(c_n x + d_n) \to^d \exp\left(-(1+\gamma_2 y)^{-1/\gamma_2}\right),$$

$1 + \gamma_1 x > 0$, $1 + \gamma_2 y > 0$, for some

$\gamma_i \in \mathbb{R}$ the (marginal) *extreme value indices.*

This is equivalent to, with $U_i := (1/(1-F_i))^{\leftarrow}$, $i = 1, 2$,

$$\frac{U_1(nx) - U_1(n)}{a_n} \to \frac{x^{\gamma_1} - 1}{\gamma_1} \qquad \frac{U_2(ny) - U_2(n)}{c_n} \to \frac{y^{\gamma_2} - 1}{\gamma_2}.$$

Let

$$x_n := \frac{U_1(nx) - U_1(n)}{a_n} \to \frac{x^{\gamma_1} - 1}{\gamma_1};$$

$$y_n := \frac{U_2(ny) - U_2(n)}{c_n} \to \frac{y^{\gamma_2} - 1}{\gamma_2}.$$

Convergence (1) implies (by uniform convergence type arguments and F_1, F_2 continuous), as $n \to \infty$,

$$\begin{aligned}
&F^n(a_n x_n + b_n, c_n y_n + d_n) \\
&\sim \quad F^n(U_1(nx), U_2(ny)) \\
&\sim \quad P\left(\max(X_1, \ldots, X_n) \le \left(\frac{1}{1-F_1}\right)^{\leftarrow}(nx),\right. \\
&\qquad\qquad \left.\max(Y_1, \ldots, Y_n) \le \left(\frac{1}{1-F_2}\right)^{\leftarrow}(ny)\right) \\
&\sim \quad P\left(\max\left(\frac{1}{1-F_1(X_1)}, \ldots, \frac{1}{1-F_1(X_n)}\right) \le nx,\right. \\
&\qquad\qquad \left.\max\left(\frac{1}{1-F_2(Y_1)}, \ldots, \frac{1}{1-F_2(Y_n)}\right) \le ny\right) \\
&\sim \quad G\left(\frac{x^{\gamma_1}-1}{\gamma_1}, \frac{y^{\gamma_2}-1}{\gamma_2}\right) =: G_0(x,y) \qquad (2)
\end{aligned}$$

Note the standardized Pareto $1/(1-F_i(x)) = 1 - 1/x$, $x \ge 1$, $i = 1, 2$, and the Fréchet $G_0(x, \infty) = G_0(\infty, x) = e^{-1/x}$, $x > 0$, marginals.

Focus now on the *dependence structure*.

Examples of (standardized) multivariate extreme value distributions:

$$G_0(x,y) = e^{-\sqrt{x^{-2}+y^{-2}}}, \quad x, y > 0 \qquad Geffroy(1958)$$

$$G_0(x,y) = e^{-\left(x^{-1}+y^{-1}-(x+y)^{-1}\right)}, \quad x, y > 0 \qquad Sibuya(1960)$$

EXPONENT MEASURE

From (2), for $x, y > 0$,

$$\lim_{n\to\infty} n\left\{1 - F(U_1(nx), U_2(ny))\right\} = -\log G_0(x,y).$$

Let

$$\begin{aligned}
&\lim_{n\to\infty} nP\left(\frac{1}{1-F_1(X_1)} > nx \text{ or } \frac{1}{1-F_2(Y)} > ny\right) \\
&= \lim_{n\to\infty} nP\left(\frac{1}{n(1-F_1(X_1))} > x \text{ or } \frac{1}{n(1-F_2(Y))} > y\right) \\
&= -\log G_0(x,y) =: \nu\{(s,t) : s > x \text{ or } t > y\}.
\end{aligned}$$

Note, e.g. when $\gamma_1 > 0$,

$$\lim_{x\to 0, y>0} G_0(x,y) = G\left(-\frac{1}{\gamma_1}, \frac{y^{\gamma_2}-1}{\gamma_2}\right) \le G\left(-\frac{1}{\gamma_1}, \infty\right) = 0.$$

More generally, for all Borel sets $A \subset [0,\infty)^2$ with $\inf_{x,y\in A} \max(x,y) > 0$ and $\nu(\partial A) = 0$

$$\nu(A) = \lim_{n\to\infty} nP\left\{\left(\frac{1}{n(1-F_1(X_1))}, \frac{1}{n(1-F_2(Y))}\right) \in A\right\}.$$

The exponent measure ν is finite on $[0,\infty)^2 \setminus [0,a]^2$, $\forall a > 0$.

The exponent measure may be a way to access to probabilities of extremes (in the limit).

Homogeneity property of the exponent measure:

$$\nu(cA) = c^{-1}\nu(A), \quad \forall c > 0.$$

From the exponent measure or related functions, and their properties, one finds ways to characterize multivariate extreme value distributions:

- the function L and the level sets Q_c (e.g. Huang (1992); de Haan and Ferreira (2006))

- the dependence functions from Sibuya (1960) and Pickands (1981).

Estimators of these can be constructed which lead to the estimation of G_0:

- $\widehat{L}$ and $\widehat{Q}_c$ (e.g. Huang (1992); de Haan and Ferreira (2006))

- estimation of Pickand's dependence function (e.g. Capéraà, Fougères and Genest (1997); Capéraà and Fougères (2000)).

The function L can be defined by

$$L(x,y) := -\log G_0\left(1/x, 1/y\right)$$

for $x, y > 0$. We can express the function L in terms of the exponent measure ν as

$$L(x,y) = \nu\left\{(s,t) \in [0,\infty)^2 \ : \ s > 1/x \text{ or } t > 1/y\right\} .$$

Proposition 1 (Properties of the function L). *1. Homogeneity of order 1: $L(ax, ay) = aL(x,y)$, for all $a, x, y > 0$.*

2. $L(x,0) = L(0,x) = x$, for all $x > 0$.

3. $x \vee y \leq L(x,y) \leq x + y$, for all $x, y > 0$.

4. Let (X,Y) be a random vector with distribution function G_0. If X and Y are independent, then $L(x,y) = x + y$, for $x, y > 0$. If on the other hand X and Y are completely positive dependent, i.e. $X = Y$ a.s., then $L(x,y) = \max(x,y)$ for $x, y > 0$.

5. L is continuous.

6. $L(x,y)$ is a convex function: $L(\lambda(x_1,y_1) + (1-\lambda)(x_2,y_2)) \leq \lambda L(x_1,y_1) + (1-\lambda)L(x_2,y_2)$, for all $x_1, y_1, x_2, y_2 > 0$ and $\lambda \in [0,1]$.

The function L leads to a characterization of the limit distribution in the following way. For $c > 0$ define the level sets Q_c by

$$Q_c := \{(x, y) \in [0, \infty)^2 \ : \ L(x, y) \leq c\} \ .$$

The sets Q_c have the following properties:

1. Q_c is a closed convex set.

2. The points $(0,0), (c,0)$ and $(0,c)$ are extreme points.

3. $Q_c = c\, Q_1$.

The convexity of the level set Q_1 is characteristic for a limit distribution as the following theorem shows.

Theorem 1. *For any simple max-stable distribution G_0 define the set Q_1 by*

$$Q_1 := \left\{(x, y) \in [0, \infty)^2 : -\log G_0\left(1/x, 1/y\right) \leq 1\right\} \ . \tag{3}$$

The set Q_1 is closed convex and the points $(0,0), (0,1)$ and $(1,0)$ are vertices. Conversely any closed convex set Q_1 with vertices $(0,0), (0,1)$ and $(1,0)$ gives rise to a limit distribution G_0 for which (3) holds. The mapping is one-to-one.

Sibuya (1960) (see also Geffroy (1958)) introduced for $t > 0$

$$\begin{aligned}\chi(t) &:= -\log G_0\,(1/t, 1) + \log G_0\,(1/t, \infty) + \log G_0\,(\infty, 1) \\ &= L(t, 1) - L(t, 0) - L(0, 1)\;.\end{aligned}$$

By the homogeneity of the function $-\log G_0$ the function χ determines the function G_0. The determining properties for the function χ are:

1. χ is convex,

2. $((-t) \vee (-1)) \leq \chi(t) \leq 0$ for $t > 0$.

Pickands (1981) introduced for $0 \le t \le 1$

$$A(t) := -\log G_0\left(\frac{1}{1-t}, \frac{1}{t}\right) = L(1-t, t)\ .$$

By the homogeneity of the function $-\log G_0$ the function A determines the function G_0. The determining properties of the function A are:

1. A is convex
2. $A(0) = A(1) = 1$
3. $((1-t) \vee t) \le A(t) \le 1$.

Any function A satisfying Properties (1)–(3) leads to a unique limit function G_0.

BACK TO THE APPLICATION

Estimate

$$
\begin{aligned}
p_n &= P((X,Y) \in C_n) \\
&= P\left(\left(\frac{1}{n(1-F_1(X_1))}, \frac{1}{n(1-F_2(Y))}\right) \in Q_n\right)
\end{aligned}
$$

with

$$
Q_n := \left\{\left(\frac{1}{n(1-F_1(x))}, \frac{1}{n(1-F_2(y))}\right) : (x,y) \in C_n\right\}.
$$

If $Q_n = c_n S$, $c_n \to \infty$, S fixed (open) Borel set in $[0,\infty)^2$ with $\inf_{x,y \in S} \max(x,y) > 0$ and $\nu(\partial S) = 0$, and

$$
\begin{aligned}
(n/k) P\Bigg(\Bigg(&\frac{1}{(n/k)(1-F_1(X_1))}, \\
&\frac{1}{(n/k)(1-F_2(Y))}\Bigg) \in c_n S\Bigg) \approx \nu(c_n S)
\end{aligned}
$$

$(n/k \to \infty)$ we have for the failure probability

$$
p_n \approx \frac{\nu(c_n S)}{n/k} = \frac{\nu(S)}{c_n n/k}.
$$

The proposed estimator is

$$
\begin{aligned}
&\hat{p}_n = \frac{k}{nc_n} \hat{\nu}(\hat{S}) \\
&= \frac{k}{nc_n} \frac{1}{k} \sum_{i=1}^{n} 1\left\{\left(\frac{k}{n\left(1-\hat{F}_1(X_i)\right)}, \frac{k}{n\left(1-\hat{F}_2(Y_i)\right)}\right) \in \frac{\hat{Q}_n}{c_n}\right\}
\end{aligned}
$$

where $\{(X_1, Y_1)\}_{i=1}^{n}$ is an i.i.d. sample of size n.

Consistency e.g. Ferreira, de Haan and Lin (2005).
Application e.g. Ferreira and de Haan (2005).

B. Extremes in $C[0,1]$

$C[0,1]$ - Space of continuous functions f on $[0,1]$ equipped with the supremum norm $\sup_{s\in[0,1]}|f(s)|$.

MOTIVATION

If f is a *deterministic function* on $[0,1]$ and X a *random function* on $[0,1]$ how can we estimate

$$p_n := P\{X(s) > f(s) \text{ for some } s \in [0,1]\}?$$

Suppose we have n i.i.d. observations of X, where none of it comes close to the boundary function f.

For instance f may represent the top of a dike along a coast and X the high tide water levels monitored along this coast.

Failure set because a breach in the dike at any place (may) leads to a flood.

Extremes in function space: we shall start with limit theory for *pointwise* maximum of i.i.d. random functions

$$\left\{\max_{1\le i\le n}\frac{X_i(s)-b_s(n)}{a_s(n)}\right\}_{s\in[0,1]}$$

where $a_s(n) > 0$ and $b_s(n)$ real are continuous functions in $[0,1]$.

Giné, Hahn and Vatan (1990); de Haan and Lin (2001)

THE LIMIT DISTRIBUTION

Suppose in $C[0,1]$,

$$\left\{ \max_{1\le i\le n} \frac{X_i(s) - b_s(n)}{a_s(n)} \right\}_{s\in[0,1]} \to^d \{Y(s)\}_{s\in[0,1]} . \qquad (4)$$

for the given normalizing functions. Let the marginal d.f.'s

$$F_s(x) := P(X(s) \le x),$$

be continuous for each $s \in [0,1]$. From (4),

$$\begin{aligned} & F_s^n \left(a_s(n)x + b_s(n)\right) \\ = \quad & P\left(\frac{\max(X_1(s),\dots,X_n(s)) - b_s(n)}{a_s(n)} \le x \right) \\ \to^d \quad & \exp\left(-\left(1+\gamma(s)x\right)^{-1/\gamma(s)} \right) \end{aligned}$$

with $\gamma(s) \in \mathbb{R}$ (in fact the *index function* is a continuous function in $s \in [0,1]$).

Moreover

$$\lim_{n\to\infty} n \left\{1 - F_s(a_s(n)x + b_s(n))\right\} = \left(1+\gamma(s)x\right)^{-1/\gamma(s)} \qquad (5)$$

for each $s \in [0,1]$, $1+\gamma(s)x > 0$.

Combining (4) and (5) (plus uniform convergence type arguments), in $C[0,1]$,

$$\left\{ \max_{1 \le i \le n} \frac{1}{n \{1 - F_s(X_i(s))\}} \right\}_{s \in [0,1]} \to^d \left\{ (1 + \gamma(s) Y(s))^{1/\gamma(s)} \right\}_{s \in [0,1]} .$$

Note, for $x > 0$ and each $s \in [0,1]$,

$$P\left((1 + \gamma(s) Y(s))^{1/\gamma(s)} \le x \right) = P\left(Y(s) \le \frac{x^{\gamma(s)} - 1}{\gamma(s)} \right)$$
$$= \exp\left(- \left(1 + \gamma(s) \frac{x^{\gamma(s)} - 1}{\gamma(s)} \right)^{-1/\gamma(s)} \right) = e^{-1/x}.$$

EXPONENT MEASURE

For every Borel set

$$A \subset \{ f \in C[0,1] : f \ge 0 \}$$

such that $\inf\{\sup_{s \in [0,1]} |f(s)| : f \in A\} > 0$ and $\nu(\partial A) = 0$,

$$\lim_{n \to \infty} nP\left(\left\{ \frac{1}{n \{1 - F_s(X_i(s))\}} \right\}_{s \in [0,1]} \in A \right) =: \nu(A)$$

Homogeneity property of the exponent measure:

$$\nu(cA) = c^{-1} \nu(A), \quad \forall c > 0.$$

To study the limit distribution of maxima in $C[0,1]$ is, basically, to study the limit relation in $C[0,1]$

$$\frac{1}{n} \max_{1 \le i \le n} \xi_i \to^d \eta \tag{6}$$

for i.i.d. stochastic processes $\{\xi_i(s)\}_{s \in [0,1]}$ in

$$C^+[0,1] := \{f \in C[0,1] : f > 0\} ,$$

and

$$P(\eta(s) \le 1) = e^{-1/x}, \quad x > 0,\ s \in [0,1].$$

The stochastic process $\{\eta(s)\}_{s \in [0,1]}$ is called *simple max-stable* and corresponds to the class of limit processes in (6).

EXAMPLE (Brown and Resnick, 1977)

Let $\{W_i\}_{i=1}^{\infty}$ be i.i.d. standard Brownian motions on $t \geq 0$,

$$Z_{n,i}(t) = \frac{W_i(1+t/b_n^2) - b_n\sqrt{1+t/b_n^2}}{\sqrt{1+t/b_n^2}/b_n}, \quad t \geq 0.$$

where the sequence b_n is choosen properly. Consider $C[0,\infty)$ the space of continuous functions with the topology of uniform convergence on bounded intervals. They showed that

$$\left\{\max_{1\leq i\leq n} Z_{n,i}(t)\right\}_{t\geq 0} \to^d \sup_{i\geq 1}\{T_i + B_i^*(t)\}_{t\geq 0}$$

where $\{T_i\}_{i=1}^{\infty}$ are points of a PPP on $\mathbb{R}$ with mean measure $e^{-x}dx$, $x \in \mathbb{R}$, and $B_i^*(t) = B_i(t) - t/2$, with B_i i.i.d. standard Brownian motions.

For the marginals note:

If $\{X_i\}_{i=1}^{\infty}$ are i.i.d. $N(0,\sigma^2)$, for e.g. with $b_n = \sqrt{2\log n - \log\log n - \log 4\pi}$

$$P\left(\frac{\max(X_1,\ldots,X_n) - b_n\sigma}{\sigma/b_n} \leq x\right) \to \exp\left(-e^{-x}\right), \quad x \in \mathbb{R}.$$

For fixed $t \geq 0$, $W_i(1+t/b_n^2) \cap N(0, 1+t/b_n^2)$ and indeed the marginal distributions of $\sup_{i\geq 1}\{T_i + B_i^*(t)\}_{t\geq 0}$ are $\exp\left(-e^{-x}\right)$, $x \in \mathbb{R}$.

BACK TO THE APPLICATION

Estimate

$$
\begin{aligned}
p_n &= P\left(X(s) > f_n(s) \text{ for some } s \in [0,1]\right) \\
&= P\Bigg(\frac{1}{(n/k)\{1 - F_s(X(s))\}} > \frac{1}{(n/k)\{1 - F_s(f_n(s))\}} \\
&\qquad\qquad \text{for some } s \in [0,1]\Bigg).
\end{aligned}
$$

If $(n/k)^{-1}\{1 - F_s(f_n(s))\}^{-1} = c_n h(s)$, $c_n \to \infty$, the set

$$\{g : g(s) > c_n h(s) \text{ for some } s \in [0,1]\}$$

satisfy the above conditions, $n/k \to \infty$ and

$$
\begin{aligned}
&(n/k)P\Bigg(\frac{1}{(n/k)\{1 - F_s(X(s))\}} > c_n h(s) \\
&\qquad\qquad \text{for some } s \in [0,1]\Bigg) \\
&\approx \nu\{g : g(s) > c_n h(s) \text{ for some } s \in [0,1]\}
\end{aligned}
$$

we have for the failure probability

$$
\begin{aligned}
p_n &\approx \frac{\nu\{g : g(s) > c_n h(s) \text{ for some } s \in [0,1]\}}{n/k} \\
&= \frac{k}{nc_n}\nu\{g : g(s) > h(s) \text{ for some } s \in [0,1]\}.
\end{aligned}
$$

The proposed estimator is

$$\widehat{p}_n = \frac{k}{nc_n}\widehat{\nu}\{g : g(s) > \widehat{h}(s) \text{ for some } s \in [0,1]\}$$

$$= \frac{k}{nc_n}\frac{1}{k}\sum_{i=1}^{n} 1\left\{\frac{k}{n\left(1-\widehat{F}_s(X_i(s))\right)} > \frac{k}{nc_n\left(1-\widehat{F}_s(f_n(s))\right)} \right.$$
$$\left. \text{for some } s \in [0,1]\right\}$$

where $\{X_i(s)\}_{s\in[0,1]}$, $i = 1,\ldots,n$, is an i.i.d. sample.

Consistency in Ferreira, de Haan and Lin (2005).

BIBLIOGRAPHY

Brown and Resnick: Extreme values of independent stochastic processes. J. Appl. Probab. **14**, 732–739 (1977)

Capéraà, Fougères and Genest: A nonparametric estimation procedure for bivariate extreme value copulas. Biometrika **84**, 567–577 (1997)

Capéraà and Fougères: Estimation of a bivariate extreme value distribution. Extremes **3**, 311–329 (2000)

Ferreira and de Haan: Estimação da probabilidade de ocorrência de um conjunto extremo. Actas do XII Congresso da Sociedade Portuguesa de Estatística. Edições SPE (2005)

Ferreira, de Haan and Lin: On the estimation of the probability of a failure set. Submetido (2005)

Geffroy: Contributions á la theorie des valeurs extrémes. Publ. Inst. Statist. Univ. Paris 7 **8**, 37–185 (1958)

Giné, Hahn and Vatan: Max-infinitely divisible and max-stable sample continuous processes. Probab. Th. Rel. Fields **73**, 139–165 (1990)

de Haan and Lin: On convergeance towards an extreme value distribution in C[0,1]. Ann. of Prob. **29**, 467–483 (2001)

de Haan and Resnick: Limit Theory for Multivariate Sample Extremes. Z. Wahrsch. verw. Gebiete. **40**, 317–337 (1977)

de Haan and Ferreira: Extreme Value Theory: An Introduction. Springer (2006)

Huang: Statistics of Bivariate Extreme Values. Thesis, Erasmus University Rotterdam (1992)

Pickands: Multivariate Extreme Value Distributions. Proceedings of 43rd Session of the International Statistical Institute. Book 2, Buenos Aires, Argentina, 859–878 (1981)

Resnick: Extreme Values, Regular Variation and Point Processes. Springer-Verlag, New York (1987)

Sibuya: Bivariate extreme statistics. Ann. Inst. Statist. Math. Tokyo, **11**, 195–210 (1960)