Topics on Multivariate and Infinite-Dimensional Extremes

Ana Ferreira

ISA, Department of Maths

Lisbon, Portugal

Abstract: We shall discuss a few main results and concepts in multivariate and infinite-dimensional extreme value theory, like limit distribution of normalized maxima and exponent measure. They will be introduced in the multivariate context, and hopefully the infinite-dimensional case will follow smoothly. The usefulness of the theory is exemplified with the problem of failure set probability estimation in both situations.

Key words and phrases: multivariate and infinite-dimensional extreme value theory, exponent measure, failure set estimation.

1 Introduction

One basic topic in multivariate extreme value theory is the limit distribution of normalized maxima. Let $(X, Y, \dots, Z)$ be a random vector with distribution function (d.f.) F. A common approach is to take componentwise maxima, and examine the convergence in distribution of the random vector,

$$\left(\frac{\max(X_1,\dots,X_n)-b_n}{a_n}, \frac{\max(Y_1,\dots,Y_n)-d_n}{c_n}, \cdots, \frac{\max(Z_1,\dots,Z_n)-f_n}{e_n}\right)$$

where $a_n, c_n, \dots, e_n > 0$ and $b_n, d_n, \dots, f_n$ real, are normalizing constants and $\{(X_i, Y_i, \dots, Z_i)\}_{i=1}^n$ is an independent and identically distributed (i.i.d.) sample from F (see e.g. de Haan and Resnick (1977) [4]; Resnick (1987) [9]; de Haan and Ferreira (2006) [5]).

Then the exponent measure can be obtained, a noteworthy feature in extreme value theory. For instance we shall see its usefulness in estimating the probability of failure set.

Similarly, extreme value theory in function space often starts with limit theory for normalized maximum of i.i.d. random functions. Let $C[0,1]$ be the space of continuous functions f on $[0,1]$ equipped with the supremum norm, $\sup_{s\in[0,1]} |f(s)|$, and X a random function with sample space in $C[0,1]$. Then we mean to examine the convergence in distribution in $C[0,1]$ of the random process

$$\left\{ \max_{1\le i\le n} \frac{X_i(s)-b_s(n)}{a_s(n)} \right\}_{s\in[0,1]} \tag{1.1}$$

where the maximum is taken pointwise for each $s \in [0,1]$ and, $a_s(n) > 0$ and $b_s(n)$ real, are continuous functions for $s \in [0,1]$ (see e.g. Giné, Hahn, and Vatan (1990) [3]; de Haan and Lin (2001) [6]; de Haan and Ferreira (2006) [5]). One can then obtain the exponent measure and we shall use it in the estimation of the probability of a failure set in function space.

2 Multivariate Extremes

In the multivariate setting we discuss the bivariate case. The extension to higher dimensions should de straightforward.

Let $\{(X_i, Y_i)\}_{i=1}^n$, be i.i.d. random vectors from $F(x,y) = P(X \le x, Y \le y)$. If there exist $a_n, c_n > 0$, and b_n, d_n real, and G is a d.f. with non-degenerate marginals such that,

$$\begin{aligned} &P\left(\frac{\max(X_1,\ldots,X_n)-b_n}{a_n} \le x, \frac{\max(Y_1,\ldots,Y_n)-d_n}{c_n} \le y\right) \\ &= F^n(a_n x + b_n, c_n y + d_n) \to^d G(x,y), \qquad n\to\infty, \end{aligned} \tag{2.1}$$

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

Consider the standardization to Pareto marginals, i.e. the random vector $(1-F_1(X))^{-1}, (1-F_2(Y))^{-1})$ where F_1, F_2 are the marginal d.f.'s of F supposed continuous. The exponent measure can be obtained as the result of the following limit, defined for all Borel sets $A \subset [0,\infty)^2$ with $\inf_{x,y\in A} \max(x,y) > 0$ and $\nu(\partial A) = 0$,

$$\nu(A) := \lim_{n\to\infty} nP\left\{ \left(\frac{1}{n(1-F_1(X_1))}, \frac{1}{n(1-F_2(Y))} \right) \in A \right\}. \tag{2.2}$$

It turns out that the exponent measure ν is finite on $[0,\infty)^2 \setminus [0,a]$, $\forall a > 0$ and, in particular, it possesses the following homogeneity property: for all sets A as above,

$$\nu(cA) = c^{-1}\nu(A), \quad \forall c > 0, \tag{2.3}$$

where the set cA is obtained by multiplying all elements of A by the constant c.

From the exponent measure or related functions, and their properties, one finds ways to characterize multivariate extreme value distributions. For instance via the function L and the level sets Q_c (Huang (1992) [7]; de Haan and Ferreira (2006) [5]) or the dependence functions from Sibuya (1960) [10] and Pickands (1981) [8].

Let us now see another application of the exponent measure. Suppose C_n is a given failure set, in particular containing large values of X or Y, and we want to estimate $p_n = P((X,Y) \in C_n)$ on the basis of an i.i.d. sample from F. Since we want to apply the previous asymptotic results, the failure set must depend on the sample size n in order to preserve the extreme nature of the problem. Then p_n can be rewritten as

$$p_n = P\left\{\left(\frac{1}{n(1-F_1(X_1))}, \frac{1}{n(1-F_2(Y))}\right) \in Q_n\right\}$$

where

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

and if $Q_n = c_n S$, for some $c_n \to \infty$ and S a fixed (open) Borel set in $[0,\infty)^2$ with $\inf_{x,y \in S} \max(x,y) > 0$ and $\nu(\partial S) = 0$, from (2.2) and (2.3),

$$p_n \approx \frac{\nu(c_n S)}{n} = \frac{\nu(S)}{c_n n}, \qquad \text{as } n \to \infty.$$

This motivates $\hat{p}_n = \hat{\nu}(\hat{S})/(nc_n)$. The proposal is then to use the empirical measure to estimate ν. While the set Q_n might not contain any observation, likewise the set C_n, the set S must contain enough observations which is possible via the sequence c_n. For the estimation of S one must deal with the estimation of the marginal d.f.'s.

The consistency of such an estimator, under appropriate additional conditions, was obtained in Ferreira, de Haan and Lin (2005) [2], and an application can be found in Ferreira and de Haan (2005) [1].

3 Extremes in $C[0,1]$

Let X be a random function in $C[0,1]$ with continuous marginal d.f.'s $F_s(x) := P\{X(s) \leq x\}$, for each $s \in [0,1]$. Suppose the process (1.1) converges in distribution to some limit process on $C[0,1]$, with non-degenerate marginals. Then the exponent measure can be obtained from the limit

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

for every Borel set $A \subset \{f \in C[0,1] : f \geq 0\}$ such that $\inf\{\sup_{s\in[0,1]} |f(s)| : f \in A\} > 0$ and $\nu(\partial A) = 0$. Note the similarity with the multivariate situation and, in particular, the same homogeneity property holds: $\nu(cA) = c^{-1}\nu(A)$, $\forall c > 0$, where the set cA is obtained by multiplying all elements of A by c.

Let us now apply the results to estimate $p_n := P\{X(s) > f_n(s) \text{ for some } s \in [0,1]\}$, on the basis of an i.i.d. sample and where f_n is a deterministic function on $C[0,1]$ possibly well far away from the observations. Similarly as before rewrite p_n as

$$P\left(\frac{1}{n\{1-F_s(X(s))\}} > \frac{1}{n\{1-F_s(f_n(s))\}} \text{ for some } s \in [0,1]\right)$$

and if $n^{-1}\{1-F_s(f_n(s))\}^{-1} = c_n h(s)$, for some $c_n \to \infty$,

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

Consequently the construction of the estimator is quite similar to the previous case.

The consistency of such an estimator, under appropriate additional conditions, was obtained in Ferreira, de Haan and Lin (2005) [2].

References

[1] Ferreira, A. and de Haan, L.(2005). Estimação da probabilidade de ocorrência de um conjunto extremo. Actas do XII Congresso da Sociedade Portuguesa de Estatıistica. Braumann, C.A., Infante, P., Oliveira, M.M., Alpzar-Jara, R. and Rosado, F., Statistical Portuguese Society.

[2] Ferreira, A., de Haan, L. and Lin, T. (2005). On the estimation of the probability of a failure set. Submited.

[3] Giné, E., Hahn, M. and Vatan, P. (1990). Max-infinitely divisible and max-stable sample continuous processes. *Probab. Th. Rel. Fields*, **73**, 139-165.

[4] de Haan, L. and Resnick, S. (1977). Limit Theory for Multivariate Sample Extremes. *Z. Wahrsch. verw. Gebiete*, **40**, 317-337.

[5] de Haan, L. and Ferreira, A. (2006). Extreme Value Theory: An Introduction. Springer.

[6] de Haan, L. and Lin, T. (2001). On convergence towards an extreme value distribution in C[0,1]. *Ann. Prob.*, **29**, 467-483.

[7] Huang, X. (1992). Statistics of Bivariate Extreme Values. Thesis, Erasmus University Rotterdam.

[8] Pickands, J. (1981). Multivariate Extreme Value Distributions. *Proceedings of 43rd Session of the International Statistical Institute.* Book 2, Buenos Aires, Argentina, 859-878.

[9] Resnick, S. (1987). Extreme Values, Regular Variation and Point Processes. Springer-Verlag, New York

[10] Sibuya, M. (1960). Bivariate extreme statistics. *Ann. Inst. Statist. Math. Tokyo*, **11**, 195-210.