Estimation of Pickands dependence function of bivariate extremes under mixing conditions

M. Boutahar, I. Kchaou and L. Reboul – Institut de Mathématiques de Marseille

Introduction

Modeling dependence structures of multivariate extremes is of great interest in many application fields . A well known way to model these structures is to use Pickands dependence function (Pickands, 1981). If $X = (X_1, X_2)$ is a bivariate vector of extremes with margins F_1 and F_2, Pickands dependence function A is defined via the extreme-value copula's type representation:

$$C(u,v) = \mathbb{P}(F_1(X_1) \le u, F_2(X_2) \le v) = \exp\left\{\log(uv) A\left(\frac{\log(u)}{\log(uv)}\right)\right\}, \ 0 \le u, v \le 1,$$

and totally characterizes the joint distribution $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$ of (X_1, X_2) of X knowing its marginal laws. It may be shown that $A : [0,1] \to [1/2, 1]$ is a convex function such that $A(0) = A(1) = 1$ and $\max(t, 1-t) \le A(t) \le 1$.
The problem of estimating Pickands dependence function by nonparametric methods has been extensively studied in the literature (see Zang et al., 2008). The underlying sequence of extremes observations is always assumed to be i.i.d., which excludes a possible serial correlation. This bias is to a certain extent supported by theoretical results on maxima of strictly stationary sequences (see e.g. Hsing, 1989). In practical situation however, the temporal independence of extremes is usually an unrealistic assumption. In the sequel, we propose to revisit the properties of the so called CFG estimator, a classical estimator of A (see Capéràa et al., 1997), when it is based on some kind of weakly dependent strictly stationary sequence of extremes, then to use these properties to build a test of independence of X's margins.
namely, we focus here on the class of absolutely regular sequences, which includes in particular the important class of linear processes and is defined as follows: Let $\mathcal{P}_0 = \sigma(X_t, t \le 0)$, $\mathcal{F}_m = \sigma(X_t, t \ge m)$ and define the decreasing sequence of coefficients of absolute regularity of X by

$$\beta(m) = \sup_{A_i \in \mathcal{P}_0, B_j \in \mathcal{F}_m} \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)|$$

where the supremum is taken over all pairs of partition $\{A_1, \dots, A_I\}$ and $\{B_1, \dots, A_J\}$ of a set Ω such that $A_i \in \mathcal{P}_0$ for each i and $B_j \in \mathcal{F}_m$ for each j.

We say that X is absolutely regular if i $\lim_{m \to +\infty} \beta(m) = 0$.

CFG estimator of absolutely regular sequences of bivariate extremes

Let $Z = (Z_1, Z_2)$ be defined from X by:

$$Z_1 = \frac{\log F_2(X_2)}{\log F_1(X_1) + \log F_2(X_2)}, \quad Z_2 = \frac{\log F_1(X_1)}{\log F_1(X_1) + \log F_2(X_2)} \qquad (1)$$

and denote by H_1 and H_2 its cdf. Then, it can be shown that (Zhang et al. 2008)

$$\log A(t) = \int_0^{1-t} \frac{H_1(z) - z}{z(1-z)} dz = \int_0^t \frac{H_2(z) - z}{z(1-z)}.$$

This naturally suggests the definition of the CFG estimator of A based on a sample $(X_i = (X_{i1}, X_{i2}))_{1 \le i \le n}$ of X:

- **Definition:** Let $(Z_i = (Z_{i1}, Z_{i2}))_{1 \le i \le n}$ by a sample of Z. The CFG estimator of A is defined by the weighted estimator

$$\hat{A}_n(s) = \left(\hat{A}_1(s)\right)^{\lambda(s)} \left(\hat{A}_2(s)\right)^{1-\lambda(s)}, \ s \in [0,1) \ \ \hat{A}_n(1) = 1, \qquad (2)$$

  where λ is an appropriately chosen nonnegative weight function in $[0,1]$ and $\hat{H}_1$ and $\hat{H}_2$ the empirical counterparts of H_1 et H_2.

  Empirical version of the CFG: The calculation of $\hat{A}$ relies on the knowledge of X margins through the definition of Z. However, margins are rarely known so that we can replace F_1 and F_2 by their empirical counterparts in (1). All the results below still hold for this version.

- **Asymptotic properties:** In the sequel we denote by $D[0,1]$ in the usual D space on $[0,1]$ with Skorokhod topology. Let B^* the bivariate centered Gaussian process with covariance function $\mathbb{E}(B^*(z)B^*(z')) = \sum_{i \in \mathbb{Z}} Cov(1_{Z_0 \le z}, 1_{Z_i \le z'})$ and define

$$\Gamma^*(s,t) = \sum_{i=1}^{2} \sum_{j=1}^{2} \lambda_i(s)\lambda_j(t) \int_0^{1-s_i} \int_0^{1-t_j} \frac{\mathbb{E}(B_i^*(z_1)B_j^*(z_2))}{z_1 z_2 (1-z_1)(1-z_2)} dz_1 dz_2, \ \ \Gamma^*(s) = \Gamma^*(s,s),$$

  with $s_1 = s$, $s_2 = 1-s$, $t_1 = t$, $t_2 = 1-t$, $\lambda_1 = 1 - \lambda_2$ and $B_1^*(z_j) = B^*(z_j, 1)$, $B_2^*(z_j) = B^*(1, z_j)$. Thus we have the following

  Theorem: Let $(X_i = (X_{i1}, X_{i2}))_{1 \le i \le n}$ be a strictly stationary absolutely regular bivariate sequence of extremes with sequence of coefficients of absolute regularity $(\beta(n))_{n > 0}$. Suppose that $A(s)$ has a bounded first derivative and that λ is a bounded function on $[0,1]$. Then

  1. If $\beta(n) = O(n^{-\theta})$ for some $\theta > 1 + 1/\sqrt{2}$ thus $\sup_{s \in [0,1]} |\hat{A}_n(s) - A(s)| \xrightarrow{P} 0$
  2. If $\beta(n) = O(n^{-\theta})$ for some $\theta \in (1, 2]$ then $\sqrt{n}(\log \hat{A}_n(s) - \log A(s)) \xrightarrow{D} \mathcal{N}(0, \Gamma^*(s))$, with $\Gamma^*(s) < \infty$

References

P. Capéraà, A.L. Fougères and C. Genest. *A nonparametric estimation procedure for bivariate extreme value copulas.* Biometrika, vol 84, n 3, p 567-577, (1997).
D. Zhang, M.T. Wells and L. Peng. *Nonparametric estimation of the dependence function for a multivariate extreme value distribution.* Journal of Multivariate Analysis, vol 99, n 4, p 577-588, (2008).

A test of independence for the margins of a bivariate vector of extreme

We wish to test H_0: X_1 and X_2 are independent against the alternative that they are not, based on a strictly stationary and absolutely regular sequence $(X_i = X_{i,1}, X_{i,2})_{1 \le i \le n}$ of X. Let A be the Pickands dependence function of X and $k = 2(1 - A(1/2))$.

- **Test:** It is shown in Segers (2007) that X_1 and X_2 are independent $\iff k = 0$. Thus, testing the independence of the margins of X amounts to test:

$$\begin{cases} H_0 : A(1/2) = 1 \\ H_1 : A(1/2) \ne 1 \end{cases}$$

- **Test statistic and Critical region:** Let $\widehat{\Gamma}_n^*$ be a suitable estimator of Γ^* and define the test statistic

$$U_n = \sqrt{\frac{n}{\widehat{\Gamma}_n^*\left(\frac{1}{2}\right)}} \log \hat{A}_n\left(\frac{1}{2}\right),$$

  Under the conditions of the Theorem (point 2.), $U_n \xrightarrow{D} N(0,1)$ under H_0 so that the test with critical region $\{U_n > q_{1-\alpha}\}$ as asymptotic level α.

Finite sample performances

We run a simulation study to investigate the finite sample performances of our test, based on bivariate logistic distributions. This model is known to be flexible enough to cover a wide range of dependence functions. The weight function used in (2) is set to $\lambda(s) = 1 - s$. The test is built from the empirical CFG estimator and an estimator of the limiting variance $\Gamma^*(1/2)$ based on the multivariate smoothed periodogram is computed. To see what is going on if we omit to take into account the within-dependence in the components of X, we compare our test with Capéràa et al.(1997) 's test of independence for i.i.d. sequences of extremes, based on the critical region $R_\alpha^* = \left\{(X_i)_{1 \le i \le n}, |T_n| > q_{1-\alpha/2}\right\}$, with $T_n = -(n/0.342)^{\frac{1}{2}} \log \hat{A}_n(1/2)$. Empirical levels are obtained as the percentage of rejection of the null over 1000 replications of the test statistic. We vary sample size $n \in \{25, 50, 100, 200, 300\}$ and nominal level $\alpha \in \{1\%, 5\%, 10\%\}$.

- **Models:** In order to generate an absolutely regular bivariate sequence of extremes, we first, build an i.i.d. bivariate sequence $(Y_1, \dots, Y_{n+k})$, $Y_i = (Y_{i,1}, Y_{i,2})$ for given $k \ge 1$ with margins G_1 and G_2 and symmetric logistic dependence function (Gumbel copula) :

$$A_Y(t) = (t^{\frac{1}{r}} + (1-t)^{\frac{1}{r}})^r, \ r \in (0,1],$$

  We then set

$$X_i = \begin{pmatrix} \max(Y_{i,1}, \dots, Y_{i-k,1}) \\ Y_{i,2} \end{pmatrix} \ 1 \le i \le n. \qquad (3)$$

  Proposition: $(X_i)_{i=1,\dots,n}$ given by (3) is a strictly stationary k-dependent bivariate sequence with marginal distributions $F_1 = G_1^{k+1}$, $F_2 = G_2$ and Pickands dependence function

$$A_X(t) = \left(1 - \frac{1}{k+1}\right)t + \left(\left(\frac{t}{k+1}\right)^{\frac{1}{r}} + (1-t)^{\frac{1}{r}}\right)^r, \ r \in (0,1], k \ge 0.$$

  Notice that a k-dependent is absolutely regular with mixing coefficient $\beta(m) = 0 \ \forall m \ge k+1$. The results below are obtained for standard Gumbel margins and the following parameters:

  1. Under H_0 (level): we use $k = 1$ and $r = 1$. Hence X is a 1-dependent process with independent components.
  2. Under H_1 (power) : we use $k = 1$ and $r = 0.5$. Hence X is a 1-dependent process with dependent components.

- **Empirical levels and powers:** Tables below highlight the good performances of our test. The empirical level tends to the nominal one with the sample size while the alternative test suffers from a large size distortion. The empirical powers tends to 100% and are large enough to think that the alternative is well detected even for moderate sample sizes.

$\alpha =$	$n = 25$	$n = 50$	$n = 100$	$n = 200$	$n = 300$
1%	4.6	3.2	2.1	1.4	1.0
5%	9.5	7.9	6.2	5.6	5.1
10%	16.7	13.3	11.9	11.1	10.2

Table 1: Empirical levels of the CFG test in %

$\alpha =$	$n = 25$	$n = 50$	$n = 100$	$n = 200$	$n = 300$
1%	7.2	6.2	10.4	16.9	27.7
5%	13	16	21.8	32.4	45
10%	18.2	23.8	30.7	42.3	53.4

Table 2: Empirical levels of the alternative test in %

$\alpha =$	$n = 25$	$n = 50$	$n = 100$	$n = 200$	$n = 300$
1%	64.9	93.1	99.7	100	100
5%	83.7	97.8	100	100	100
10%	90.1	99.5	100	100	100

Table 3: Empirical power of the CFG test in %

References

J. Pickands. *Multivariate extreme value distributions.*, p 859-878, (1981).
J. Segers. *Nonparametric inference for bivariate extreme-value copulas.* Topics in Extreme Values (M. Ahsanullah and S. N. U. A. Kirmani, eds.) Nova Science Publishers, New York, p 185-207, (2007).