Estimation de queues bivariées

Elena Di Bernardino, Véronique Maume-Deschamps, Clémentine Prieur

Estimating Bivariate Tails

Elena Di Bernardino

Université de Lyon, Université Lyon 1, ISFA, Laboratoire SAF.
E-mail : elena.di-bernardino@univ-lyon1.fr

Véronique Maume-Deschamps

Université de Lyon, Université Lyon 1, ISFA, Laboratoire SAF.
E-mail : veronique.maume@univ-lyon1.fr

Clémentine Prieur
Université Joseph Fourier, LJK/MOISE.
E-mail : clementine.prieur@imag.fr

Résumé

Ce papier traite de l'estimation de la queue d'une distribution bivariée. Nous développons une approche bidimensionnelle de la méthode de dépassement de seuil (Peaks Over Threshold) et une version bivariée du Théorème de Pickands-Balkema-de Hann. Nous démontrons des propriétés de convergence pour l'estimateur ainsi construit. La structure de dépendance entre les marges est modélisée par une copule. Enfin, nous proposons quelques simulations.

Abstract

In this paper we consider the general problem of estimating the tail of a bivariate distribution. An extension of the threshold method for extreme values is developed, using a two-dimensional version of the Pickands-Balkema-de Hann Theorem. We construct a two-dimensional tail estimator and we provide its asymptotic properties. The dependence structure between the marginals is described by a copula. Simulations are implemented.

Keywords: Extreme Value Theory, Bivariate Pickands-Balkema-de Hann Theorem, Peaks Over Threshold method, Copula, Tail Dependence.

We consider the general problem of estimating the tail of a distribution. We focus in this work on the study of bivariate distributions. We assume that the dependence between the marginals is described by a continuous symmetric copula C, which is assumed to be known or inferred from the data structure. Multivariate models with dependent marginals are of great importance in the applications. One main stake in the field of insurance for example is to build new risk indicators taking into account the dependence structure between lines of business. Our approach is based on the study of the excesses over a threshold. We obtain an estimator for the tail of bivariate distributions, for which we derive asymptotic results. The main ingredient is a generalization of the bivariate Pickands-Balkema-de Hann Theorem stated in [3].

To understand the heart of our method, we need first to recall more or less classical results for the estimation of the tail of univariate distributions. The univariate threshold method (or POT method, Peaks Over Threshold) is widely used for estimating extreme quantiles or tail distributions. This method is centered on the Generalized Pareto Distribution, as a model for distribution of excesses over a threshold, and it finds its justification in the classical Extreme Value Theory (EVT). An important result on which the method is based is the Pickands-Balkema-de Hann Theorem, stating that the generalized Pareto distribution (GPD), given by

$$V_{k,\sigma}(x) := \begin{cases} 1 - \left(1 - \frac{kx}{\sigma}\right)^{\frac{1}{k}}, & \text{if } k \neq 0,\ \sigma > 0 \\ 1 - \mathrm{e}^{\frac{-x}{\sigma}}, & \text{if } k = 0,\ \sigma > 0 \end{cases} \tag{1}$$

and $x \geq 0$ for $k \leq 0$ or $0 \leq x < \frac{\sigma}{k}$ for $k > 0$, appears as the limit distribution of scaled excesses over high thresholds u.

Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sequence with common distribution function F. From Fisher-Tippet Theorem [2, Theorem 3.2.3], we know that if there exists a non-degenerate distribution function $H_k(x)$ such that

$$\forall\, x \in \mathbb{R} \lim_{n\to\infty} \mathbb{P}\left[\frac{\max\{X_1, X_2, \ldots, X_n\} - b_n}{a_n} \leq x\right] = H_k(x), \tag{2}$$

with $(a_n)_{n>0} \in {\mathbb{R}_+^*}^{\mathbb{N}^*}$ and $(b_n)_{n>0} \in \mathbb{R}^{\mathbb{N}^*}$, then $H_k(x)$ is a member of the GEV (Generalized Extreme Value Distribution) family

$$H_k(x) = \begin{cases} \exp\left(-\left(1 - k\,x\right)^{\frac{1}{k}}\right), & \text{if } k \neq 0, \\ \exp\left(-\mathrm{e}^{-x}\right), & \text{if } k = 0, \end{cases} \tag{3}$$

where $1 - k\,x > 0$ and $k \in \mathbb{R}$ is a parameter determining the family to which the limit belongs to (Fréchet, Gumbel, Weibull). We then write $F \in MDA(H_k)$. This allows us to give the following precise formulation.

Pickands-Balkema-de Hann Theorem. [2, Theorem 3.4.13(b)]
Let $F_u(x) = \mathbb{P}[X - u \leq x \,|\, X > u]$ *and* $x_F := \sup\{x \in \mathbb{R} \,|\, F(x) < 1\}$, *(i.e.* x_F *is the right endpoint of* F*), then we have*

$$F \in MDA(H_k) \quad \Leftrightarrow \quad \lim_{u \to x_F} \sup_{0 \leq x \leq x_F - u} \mid F_u(x) - V_{k,\sigma(u)}(x) \mid = 0. \tag{4}$$

This result suggests that, for sufficiently high thresholds u, the distribution function of the excesses may be approximated by a GPD $V_{k,\sigma(u)}(x)$.

The use of the Pickands-Balkema-de Hahn Theorem to provide an estimator of the tail of univariate distributions is well-known, see for instance McNeil's results [4, 5] in the field of insurance. We generalize this approach to the bivariate frame.

If C denotes a symmetric continuous copula, we define for $(u_1, u_2) \in [0,1]^2$ the survival copula C^* by

$$C^*(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2). \tag{5}$$

We now need the following auxiliary result:

Upper Tail Dependence Copula Convergence Theorem. [3, Theorem 2.3]
Let C be a continuous symmetric copula. Assume that $C^*(1-u, 1-u) > 0$, for all $u > 0$, and that there exists an increasing continuous function $g : [0, \infty) \to [0, \infty)$ such that

$$\lim_{u \to 1} \frac{C^*(x(1-u), 1-u)}{C^*(1-u, 1-u)} = g(x), \quad x \in [0, \infty). \tag{6}$$

Then, there exists $\theta > 0$ such that $g(x) = x^\theta g\left(\frac{1}{x}\right)$ for all $x \in (0, \infty)$. Further, for all $(x, y) \in [0,1]^2$

$$\lim_{u \to 1} C_u^{up}(x, y) = x + y - 1 + G(g^{-1}(1-x), g^{-1}(1-y)) := C^{*\,G}(x, y), \tag{7}$$

where

$$G(x, y) := y^\theta g\left(\frac{x}{y}\right), \ \forall\,(x, y) \in (0,1]^2 \text{ and } G :\equiv 0 \text{ on } [0,1]^2 \setminus (0,1]^2. \tag{8}$$

We fix a threshold u and denote by N the binomial random variable of excesses (number of excesses above u). The distribution of N is the binomial with parameters n and $1 - F(u)$, $N \sim \mathrm{Bi}(n, 1 - F(u))$. The excesses are denoted by $Y_1, Y_2, \ldots, Y_N$ where $Y_j = X_i - u$ with i the index of the j^{th} exceedance.
Then, estimating σ and k by maximum likelihood estimators $\widehat{k}$ and $\widehat{\sigma}$, and following (4), Smith [7] proposed an estimator of the tail $F(y)$ in the one-dimensional case, of the form

$$1 - \widehat{F}^*(y) = \begin{cases} \frac{N}{n}\left(1 - \widehat{k}\,\frac{(y-u)}{\widehat{\sigma}}\right)^{\frac{1}{\widehat{k}}}, & \text{if } \widehat{k} \neq 0, \\ \frac{N}{n}\left(\mathrm{e}^{\frac{-(y-u)}{\widehat{\sigma}}}\right), & \text{if } \widehat{k} = 0, \end{cases} \tag{9}$$

for $u < y < \infty$ (if $\widehat{k} \leq 0$) or $u < y < \widehat{\sigma}/\widehat{k}$ (if $\widehat{k} > 0$).
Following this general approach, we propose a new tail estimator in the two-dimensional setting. Assuming that the dependence structure is described by some continuous symmetric copula C, and denoting by F_X and F_Y the distribution functions of the marginals (they are not assumed to be equal), we first introduce the following notations

$$\widehat{F}_1^*(u, y) = C\left(\widehat{F}_{n,X}(u), \widehat{F}_Y^*(y)\right) \tag{10}$$

$$\widehat{F}_2^*(x, u) = C\left(\widehat{F}_X^*(x), \widehat{F}_{n,Y}(u)\right), \tag{11}$$

where $\widehat{F}_{n,X}$ and $\widehat{F}_{n,Y}$ denote the empirical distribution functions of X and Y respectively. The estimates $\widehat{F}_X^*$ and $\widehat{F}_Y^*$ are defined by (9).

Define

$$\widehat{F}_{n,(X,Y)}(s,t) = (1/n)\sum_{i=1}^{n} 1_{\{X_i \le s, Y_i \le t\}},$$

$$\overline{\widehat{F}}(s,t) = (1/n)\sum_{i=1}^{n} 1_{\{X_i > s,\, Y_i > t\}}.$$

Denote $V_{\widehat{k}_X,\widehat{\sigma}_X}$ by V^X and $V_{\widehat{k}_Y,\widehat{\sigma}_Y}$ by V^Y. Denote for all $u \in \mathbb{R}$, $u_Y = F_Y^{-1}(F_X(u))$.

Denoting now by $\widehat{k}_X$, $\widehat{\sigma}_X$ (resp. $\widehat{k}_Y$, $\widehat{\sigma}_Y$) the maximum likelihood estimators obtained from the sample $\{X_i\}_{i=1,\dots,N_u}$ (resp. $\{Y_i\}_{i=1,\dots,N_{u_Y}}$), we define the estimator $\widehat{F}^*(x,y)$ for the bivariate tail by

$$\overline{\widehat{F}}(u,u_Y)\left(1 - g(1-V^X(x-u)) - g(1-V^Y(y-u_Y)) + G\left(1-V^X(x-u), 1-V^Y(y-u_Y)\right)\right)$$
$$+\widehat{F}_1^*(u,y) + \widehat{F}_2^*(x,u_Y) - \widehat{F}_{n,(X,Y)}(u,u_Y). \tag{12}$$

We obtain Theorem 1 below as an extension of [3, Theorem 4.1].

Theorem 1. *Let X, Y be continuous, univariate random variables, with different marginal distributions, respectively F_X, F_Y, and symmetric copula C. Suppose that $F_X \in MDA(H_{k_1})$, $F_Y \in MDA(H_{k_2})$. Assume that C satisfies the hypotheses of the* **Upper Tail Dependence Copula Convergence Theorem** *for some g. Then*

$$\sup_{\substack{0<x<x_{F_Y}-u,\\ 0<y<x_{F_Y}-F_Y^{-1}(F_X(u))}} \Bigg|\mathbb{P}\left[X-u \le x, Y-u_Y \le y \,\middle|\, X>u, Y>u_Y\right]$$
$$- C^{*G}\big(1-g(1-V_{k_1,a_1(u)}(x)), 1-g(1-V_{k_2,a_2(u_Y)}(y))\big)\Bigg| \xrightarrow{u\to x_{F_X}} 0, \tag{13}$$

where $V_{k_i,a_i(\cdot)}$ is the GPD with parameters $k_i, a_i(\cdot)$, defined in (1)*, and $x_{F_X} := \sup\{x \in \mathbb{R} \,|\, F_X(x) < 1\}$, $x_{F_Y} := \sup\{y \in \mathbb{R} \,|\, F_Y(y) < 1\}$.*

Our estimator (12) is only valid for $x > u$ and $y > u$, when u is *large enough.* We will see how to interpret the expression u *large enough* and to provide a careful selection of thresholds in order to have convergence results for (12).

In order to provide bivariate convergence results, we first prove a new one-dimensional convergence theorem. Let $(u_n)_{n\ge0}$ be a sequence of thresholds in $\mathbb{R}_+$ increasing to infinity, and $(z_n)_{n\ge0}$ an auxiliary sequence in $\mathbb{R}_+$ increasing to infinity. The role of $(z_n)_{n\ge0}$ is to control how far we can extrapolate into the tail. Under suitable regularity conditions on F and on the respective speed of convergence of $(u_n)_{n\ge0}$ and $(z_n)_{n\ge0}$, we prove

$$\left[F(u_n z_n) - \widehat{F}^*(u_n z_n)\right] \xrightarrow[n\to\infty]{\mathbb{P}} 0. \tag{14}$$

This result is one of the essential tools to demonstrate our convergence result for the bivariate tail estimator (12). It is a corollary of a new central limit theorem which allows us to build confidence intervals for univariate tails.

We consider a continuous symmetric copula C and marginal distributions F_X and F_Y. Let $N_X = N_{n,X}$ (resp. $N_Y = N_{n,Y}$) denote the random number of excesses above some threshold $u_1(n)$ (resp. $u_2(n)$)

in a total sample of size n. We are interested in providing a convergence theorem unconditionally on the number of excesses.

Under suitable relations on u_1 and u_2 and under regularity assumptions on F_X, F_Y and C, we obtain the following result:

Theorem 2. *If all the assumptions of Theorem 1 and of the* **Upper Tail Dependence Copula Convergence Theorem** *are verified, then*

$$\sup_{x>u_1(n),y>u_2(n)} \left[F(x,y) - \widehat{F}^*(x,y)\right] \xrightarrow[n\to\infty]{\mathbb{P}} 0. \tag{15}$$

We propose simulations to illustrate the general convergence result in the bi-dimensional case (Theorem 2 above). We choose the marginal distributions and symmetric copula as follows

$$C(u,v) = u + v - 1 + [(1-u)^{-1} + (1-v)^{-1} - 1]^{-1} \text{ (Survival Clayton copula)} \tag{16}$$

$$F_X(x) = 1 - (1+x)^{-1}, \;\; F_Y(y) = 1 - (1+y^2)^{-1} \;\; \text{(different Burr distributions)} \tag{17}$$

In the simulations, we use the general method proposed in [1], *conditional sampling*, to sample from a bivariate distribution.

This work has been partially supported by the french research national agency (ANR) under the reference ANR-08BLAN-0314-01.

References

[1] U. Cherubini, E. Luciano, W. Vecchiato, *Copula Methods in Finance*, Wiley, New York, (2004).

[2] P. Embrechts, C. Kluppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance*, Berlin, Heidelberg: Springer-Verlag, (1997).

[3] A. Juri, M.V. Wüthrich, *Tail Dependence From a Distributional Point of View*, Extremes 6, Vol. 6, no. 3, (2003), 213-246.

[4] A. J. McNeil, *Estimating the Tails of loss severity distributions using extreme value theory*, ASTIN Bulletin 27, (1997), 1117-1137.

[5] A. J. McNeil, *Extreme value theory for risk managers*, preprint, ETH, Zurich, (1999).

[6] A. J. McNeil, R. Frey, P. Embrechts, *Quantitative risk management. Concepts, techniques and tools*, Princeton Series in Finance, (2005)

[7] R. Smith, *Estimating tails of probability distributions*, Ann. Statist., no. 15, (1987), 1174-1207.

[8] M. V. Wüthrich, *Bivariate extension of the Pickands-Balkema-de Hann Theorem*, Ann. Inst. H. Poincare Probab. Statist. 40, no. 1, 33-41 (2004)