Estimating Bivariate Tail

Elena Di Bernardino, Véronique Maume-Deschamps, Clémentine Prieur

ESTIMATING BIVARIATE TAIL

Elena Di Bernardino[1] , Véronique Maume-Deschamps[1] and Clémentine Prieur[2]

Université Lyon 1[1], and Université Joseph Fourier[2]

Abstract: This paper deals with the problem of estimating the tail of a bivariate distribution function. To this end we develop a general extension of the POT (Peaks-Over-Threshold) method, mainly based on a two-dimensional version of the Pickands-Balkema-de Hann Theorem. We construct a two-dimensional tail estimator and study its asymptotic properties. We also present a simulation study which illustrates our theoretical results.

Key words and phrases: Extreme Value Theory, Peaks Over Threshold method, Pickands-Balkema-de Hann Theorem, tail dependence.

1 Introduction

The univariate POT (Peaks-Over-Threshold) method is common for estimating extreme quantiles or tail distributions. A key idea of this method is the use of the generalized Pareto distribution to approximate the distribution of excesses over thresholds. Its justification can be found in the classical Extreme Value Theory (EVT). The main point is given by the Pickands-Balkema-de Hann Theorem, stating that the generalized Pareto distribution (GPD), whose distribution function is given by

$$V_{k,\sigma}(x) := \begin{cases} 1-\left(1-\frac{kx}{\sigma}\right)^{\frac{1}{k}}, & \text{if } k\neq 0,\ \sigma>0, \\ 1-\mathrm{e}^{\frac{-x}{\sigma}}, & \text{if } k=0,\ \sigma>0, \end{cases} \tag{1.1}$$

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. A precise formulation of this result is given in Theorem B below. To be self contained, we first recall the statement of the Fisher-Tippet Theorem (Theorem A below).

Theorem A *(Embrechts, Kluppelberg and Mikosch (1997), Theorem 3.2.3) Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sequence with common distribution function F. If there exist a sequence of positive numbers $(a_n)_{n>0}$ and a sequence $(b_n)_{n>0}$ of real numbers such that*

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{\max\{X_1, X_2, \ldots, X_n\} - b_n}{a_n} \leq x\right] = H_k(x), \quad x \in \mathbb{R}, \tag{1.2}$$

for a non-degenerate distribution function $H_k(x)$, then $H_k(x)$ is a member of the GEV (Generalized Extreme Value Distribution) family

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

where $1 - k\,x > 0$ and $k \in \mathbb{R}$ is a parameter determining the family to which the limit belongs to (Fréchet for $k < 0$, Gumbel for $k = 0$ and Weibull for $k > 0$).

Definition 1 *[Maximum Domain of Attraction] We say that F belongs to the Maximum Domain of Attraction of a generalized extreme value distribution if there exist a sequence of positive numbers $(a_n)_{n>0}$ and a sequence $(b_n)_{n>0}$ of real numbers such that (1.2) holds. We write $F \in MDA(H_k)$.*

We can now state a precise formulation of the Pickands-Balkema-de Hann Theorem.

Theorem B *(Embrechts, Kluppelberg and Mikosch (1997), Theorem 3.4.13(b))*

$$F \in MDA(H_k) \quad \Leftrightarrow \quad \lim_{u\to x_F} \sup_{0\leq x<x_F-u} \left|F_u(x) - V_{k,\sigma(u)}(x)\right| = 0,$$

where $F_u(x) = \mathbb{P}[X - u \leq x \,|\, X > u]$ and $x_F := \sup\{x \in \mathbb{R} \,|\, F(x) < 1\}$, (i.e. x_F is the right endpoint of F).

This result suggests that, for sufficiently high thresholds u, the distribution function of the excesses may be approximated by a GPD $V_{k,\sigma(u)}(x)$. Its use for estimating the tail of a univariate distribution is now classical (see for example McNeil (1997), McNeil (1999) and references therein).

In this paper, we are interested in the problem of fitting the joint distribution of bivariate observations exceeding high thresholds. To this end, we develop a

two-dimensional extension of the POT method. This extension is general in the sense that it allows to consider a two-dimensional structure of dependence between both random components X and Y. The modeling of this dependence is done via a continuous symmetric copula C, which is supposed to be known or inferred from the data structure.

The paper is organized as follows. To be self-contained, we recall main existing results (Theorems A to D) leading to an extension of the Pickands-Balkema-de Hann Theorem in dimension 2 (Theorem E). We then give a generalization of this theorem in the case of bivariate distributions with non necessarily equal marginals (Theorem 1). To estimate the tail of two-dimensional distributions, we need first to introduce a one-dimensional estimator for the tail. It is the purpose of Section 3. In Section 4 we present the construction of a new bivariate tail estimator, both in the case with equal marginals ($F_X = F_Y$, see Section 4.1) and in a general case ($F_X \neq F_Y$, see Section 4.2). The study of the asymptotic properties of our estimator makes use of new convergence results in the univariate case (Theorem 6 and Corollary 7 of Section 5) dealing with asymptotic properties of the absolute error between the theoretical distribution function and its tail estimator. Our main results, concerned with the asymptotic properties of our two-dimensional tail estimator, are stated in Section 6 (Theorems 9 and 10). Examples with simulated data are presented in Section 7. Some auxiliary results and more technical proofs are postponed to the Appendix.

Remark 1

▪ Assume we observe $X_1, \ldots, X_n$ i.i.d. with common distribution function F. If we fix some high threshold u, let N denote the number of excesses above u. In the following, two approaches will be considered. In the first one, we work conditionally on N. If n is the sample size and u_n the associated threshold, the number of excesses is m_n, with $\lim_{n\to\infty} m_n = \infty$ and $\lim_{n\to\infty} \frac{m_n}{n} = 0$. The second approach considers the number of excesses N_n as a binomial random variable (which is the case in the simulations), $N_n \sim \text{Bi}\,(n, 1 - F(u_n))$ with $\lim_{n\to\infty} 1 - F(u_n) = 0$ and $\lim_{n\to\infty} n(1 - F(u_n)) = \infty$.

▪ Our bivariate results are proved for continuous random variables X and Y (in particular F_X and F_Y are assumed to be continuous). The structure of dependence between X and Y is described by a continuous and symmetric copula

C. F_X and F_Y are a priori unknown, with regularity properties specified in the statement of our theorems.

2 On the two-dimensional Pickands-Balkema-de Hann Theorem

A central result in tail estimation is the so-called Pickand-Balkema-de Hann Theorem. In the introduction we recalled its statement in dimension 1 (Theorem B). As our aim is the estimation of bivariate tails, we are interested in two-dimensional extensions of Theorem B. Such a generalization (Theorem E below) can be found in the literature (Juri and Wüthrich (2003) and Wüthrich (2004)) with the assumption $F_X = F_Y$. In the paper by Juri and Wüthrich (2003), the non-exchangeable case $F_X \neq F_Y$ is discussed in a remark. We provide here a precise formulation and proof of the Pickands-Balkema-de Hann Theorem when $F_X \neq F_Y$ (Theorem 1 below). We first introduce some notations and recall results from Juri and Wüthrich (2003), McNeil, Frey and Embrechts (2005) and Nelsen (1999), which we will need later.

Definition 2 *[Copula] A 2-dimensional copula is a function $C(u, v)$ which is the distribution function of a random vector (U, V) with uniform marginals on $[0, 1]$.*

Remark 2

▪ For a given copula C and given marginal distributions F_X and F_Y

$$F(x, y) = C(F_X(x), F_Y(y)),\ \forall\,(x, y) \in \mathbb{R}^2 \tag{2.1}$$

is a bivariate distribution function with marginals F_X and F_Y.

▪ In the following, for a distribution function F, we will denote by F^{-1} its generalized left continuous inverse defined by $F^{-1}(p) = \inf\{x \in \mathbb{R},\ F(x) \geq p\},\ p \in (0, 1)$. Then, for a joint distribution F with marginals F_X and F_Y, there is always a copula C satisfying (2.1). This copula is not necessarily unique, but it is if F_X and F_Y are continuous in which case (*Sklar's Theorem*, Nelsen (1999), Theorem 2.3.3)

$$\forall\,(u_1, u_2) \in [0, 1]^2, \quad C(u_1, u_2) = F(F_X^{-1}(u_1), F_Y^{-1}(u_2)). \tag{2.2}$$

Definition 3 *[Survival Copula] The Survival Copula C^* of a copula C is defined by:*

$$\forall\,(u_1, u_2) \in [0, 1]^2, \quad C^*(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2). \tag{2.3}$$

A Survival Copula (2.3) is a copula. Moreover, if (X, Y) is a random vector in $\mathbb{R}^2$ with joint distribution F, copula C and marginal distributions F_X and F_Y,

$$\text{we have:} \quad C^*(1-F_X(x), 1-F_Y(y)) = \mathbb{P}[X > x, Y > y], \quad \text{for } (x, y) \in \mathbb{R}^2. \tag{2.4}$$

In a first time, let us assume that X and Y are uniformly distributed on $[0, 1]$, and that the copula associated to (X, Y) is continuous and symmetric. Let us fix a threshold $u \in [0, 1)$ such that $\mathbb{P}[X > u, Y > u] > 0$, i.e. such that $C^*(1-u, 1-u) > 0$. We consider the distribution of X conditioned on $\{X > u, Y > u\}$:

$$\forall x \in [0, 1], \quad \overline{F_u}(x) := \mathbb{P}[X \leq x \,|\, X > u, Y > u] = 1 - \frac{C^*(1 - x \vee u, 1 - u)}{C^*(1-u, 1-u)}. \tag{2.5}$$

Note that the continuity of the copula C implies that $\overline{F_u}(\cdot)$ is also continuous. Its symmetry implies that Equation (2.5) also defines the distribution of Y conditioned on $\{X > u, Y > u\}$ (recall that X and Y are uniformly distributed).

Definition 4 *[Upper-tail dependence copula] Let X and Y be uniformly distributed on $[0, 1]$. Assume that the copula associated to (X, Y) is symmetric. For a threshold $u \in [0, 1)$ satisfying $C^*(1-u, 1-u) > 0$, we define the upper-tail dependence copula at level $u \in [0, 1)$ relative to the copula C by*

$$\forall\, (x, y) \in [0, 1]^2, \quad C_u^{up}(x, y) := \mathbb{P}[X \leq \overline{F_u}^{-1}(x), Y \leq \overline{F_u}^{-1}(y) \,|\, X > u, Y > u],$$

where $\overline{F_u}$ is given from (2.5).

Note that $\mathbb{P}[X \leq x, Y \leq y \,|\, X > u, Y > u]$ obviously defines a two dimensional distribution function whose marginals are both given by $\overline{F_u}$. Hence we deduce from (2.2) that $C_u^{up}(x, y)$ is a copula. Moreover, the asymptotic behavior of C_u^{up} for u around 1 describes the dependence structure of X, Y in their upper tails. More precisely, Juri and Wüthrich (2003) demonstrate the following result:

Theorem C *(Juri and Wüthrich (2003), Theorem 2.3) Let C be a symmetric Copula such that $C^*(1-u, 1-u) > 0$, for all $u > 0$. Furthermore, assume that there is a strictly increasing continuous function $g : [0, \infty) \to [0, \infty)$ such that*

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

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

$$\lim_{u\to 1} C_u^{up}(x,y) = x + y - 1 + G(g^{-1}(1-x), g^{-1}(1-y)) := C^{*\,G}(x,y), \qquad (2.6)$$

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

Remark 3

▪ From (2.3), we note that $C^{*\,G}(x,y)$ defined in (2.6) is the Survival Copula of the Copula $C^G(x,y) := G(g^{-1}(x), g^{-1}(y))$ (for more details see Juri and Wüthrich (2003), Section 2) and thus, in particular, is a Copula.

▪ G is continuous, symmetric with strictly increasing marginals (for more details about properties of G see Juri and Wüthrich (2003), Section 2.3).

Proposition D below is the last step before the statement of the central results of this section.

Proposition D *(Embrechts, Kluppelberg and Mikosch (1997), Theorem 3.4.5)* $F \in MDA(H_k)$ *is equivalent to the existence of a positive measurable function* $a(\cdot)$ *such that, for* $1 - k\,x > 0$ *and* $k \in \mathbb{R}$,

$$\lim_{u\to x_F} \frac{1 - F(u + x\,a(u))}{1 - F(u)} = \begin{cases} (1 - k\,x)^{\frac{1}{k}}, & \textit{if } k \neq 0, \\ \mathrm{e}^{-x}, & \textit{if } k = 0. \end{cases} \qquad (2.7)$$

Combining Proposition D with Theorem C, Juri and Wüthrich (2003) demonstrate the following Theorem, which can be interpreted as a two-dimensional extension of the Pickands-Balkema-de Hann Theorem in the case $F_X = F_Y$.

Theorem E *(Juri and Wüthrich (2003), Theorem 4.1) Let* X *and* Y *be two continuous real valued random variables, with identical distribution function* F *and with symmetric copula* C. *Suppose that* $F \in MDA(H_k)$ *and that* C *satisfies the hypotheses of Theorem C for some function* g. *Then*

$$\lim_{u\to x_F} \sup_{0 < x,\,y \le x_F - u} \Big| \mathbb{P}\big[X - u \le x,\, Y - u \le y \,\big|\, X > u, Y > u\big] - C^{*\,G}\big(1 - g(1 - V_{k,a(u)}(x)), 1 - g(1 - V_{k,a(u)}(y))\big)\Big| = 0,$$

where $V_{k,a(u)}$ is the GPD with parameters k, $a(u)$ defined in (1.1)*, $a(\cdot)$ is as in* (2.7)*; and $x_F := \sup\{x \in \mathbb{R} \,|\, F(x) < 1\}$.*

From (2.6), $C^{*G}\big(1 - g(1 - V_{k,a(u)}(x)),\, 1 - g(1 - V_{k,a(u)}(y))\big)$ is equal to

$$1 - g(1 - V_{k,a(u)}(x)) - g(1 - V_{k,a(u)}(y)) + G\big(1 - V_{k,a(u)}(x), 1 - V_{k,a(u)}(y)\big), \quad (2.8)$$

where $a(u) = \sigma(u)$, with $\sigma(u)$ as in Theorem B.
Theorem E above is concerned with exchangeable random variables. Indeed it is proved in Nelsen (1999), Theorem 2.7.4, that two continuous random variables identically distributed and with symmetric copula are exchangeable. The case of non-exchangeable random variables is mentioned in the remark following Theorem 4.1 of Juri and Wüthrich (2003). We give below a rigorous result for this more general case.

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

$$\sup_{\mathcal{A}} \Big| \mathbb{P}\big[X - u \le x, Y - F_Y^{-1}(F_X(u)) \le y \big| X > u, Y > F_Y^{-1}(F_X(u))\big] - C^{*G}\big(1 - g(1 - V_{k_1,a_1(u)}(x)), 1 - g(1 - V_{k_2,a_2(F_Y^{-1}(F_X(u)))}(y))\big) \Big| \xrightarrow[u \to x_{F_X}]{} 0, \quad (2.9)$$

where $V_{k_i,a_i(\cdot)}$ is the GPD with parameters $k_i, a_i(\cdot)$ defined in (1.1)*, $a_i(\cdot)$ is as in* (2.7)*, for $i = 1, 2$, $x_{F_X} := \sup\{x \in \mathbb{R} \,|\, F_X(x) < 1\}$, $x_{F_Y} := \sup\{y \in \mathbb{R} \,|\, F_Y(y) < 1\}$ and $\mathcal{A} := \{(x, y) : 0 < x \le x_{F_X} - u, 0 < y \le x_{F_Y} - F_Y^{-1}(F_X(u))\}$.*

The proof of Theorem 1 is postponed to the Appendix.

3 Estimating the tail of univariate distributions

The estimation of bivariate tail distributions requires first the estimation of one-dimensional tail. Let us describe a way to fit one-dimensional tail, starting from the Pickands-Balkema-de Hann Theorem (see for example McNeil (1997), McNeil (1999)). Fix a threshold u. For $x > u$, decompose F as

$$F(x) = \mathbb{P}[X \le x] = (1 - \mathbb{P}[X \le u])\, F_u(x - u) + \mathbb{P}[X \le u], \quad (3.1)$$

where $F_u(x) = \mathbb{P}[X \leq x \,|\, X > u]$. Let $X_1, X_2, \ldots$ be a sequence of i.i.d random variables with unknown distribution function F. Using Theorem B, and estimating $F(u) = \mathbb{P}[X \leq u]$ by the empirical distribution function $\widehat{F}_X(u) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq u\}}$ we obtain the univariate tail estimate

$$F^*(x) = (1 - \widehat{F}_X(u))\, V_{k,\sigma}(x-u) + \widehat{F}_X(u). \tag{3.2}$$

Usually the parameters k and σ of the GPD are unknown and must be estimated from the data. In this paper, we estimate these parameters by the maximum likelihood estimators $\hat{k}$, $\hat{\sigma}$ (MLE) based on the excesses above u for which there exist interesting asymptotic convergence results (see Section 5). We then write (3.2) as

$$\widehat{F}^*(x) = (1 - \widehat{F}_X(u)) V_{\widehat{k},\widehat{\sigma}}(x-u) + \widehat{F}_X(u), \quad \text{for } x > u. \tag{3.3}$$

Using (3.3) we get the estimator, proposed by Smith (1987)

$$1 - \widehat{F}^*(y) = \begin{cases} \frac{N}{n}\left(1 - \widehat{k}\,\frac{(y-u)}{\widehat{\sigma}}\right)^{\frac{1}{\widehat{k}}}, & \text{if } \widehat{k} \neq 0, \\ \frac{N}{n}\left(\mathrm{e}^{\frac{-(y-u)}{\widehat{\sigma}}}\right), & \text{if } \widehat{k} = 0, \end{cases} \tag{3.4}$$

with $u < y < \infty$ (if $\widehat{k} \leq 0$) or $u < y < \frac{\widehat{\sigma}}{\widehat{k}}$ (if $\widehat{k} > 0$) and N the random number of excesses.

4 Estimating the tail of bivariate distributions

In this section we present the main construction of this paper. We propose indeed a new tail estimator for the two-dimensional distribution function $F(x, y)$. Asymptotic properties for this estimator are stated and proved in Section 6. This construction generalizes the one-dimensional construction stated in the previous section. The two main steps are the following:

Step 1: We approximate $F(x, y)$ by $F^*(x, y)$, using one of both two-dimensional versions of the Pickands-Balkema-de Hann Theorem (Theorem E if $F_X = F_Y$, and Theorem 1 if $F_X \neq F_Y$). An important quantity to define $F^*(x, y)$ is the survival copula C^{*G}, introduced in Section 2.

Step 2: We define $\widehat{F}^*(x, y)$ by replacing the parameters k and σ of the GPD by the maximum likelihood estimators.

For sake of clarity, we first give the construction of our estimator when $F_X = F_Y$. We then give the generalization for $F_X \neq F_Y$.

4.1 **Case with same marginal distributions**

Let X and Y be two real valued random variables with same marginal distribution F. The structure of dependence between X and Y is represented by some continuous and symmetric copula C. Let u be a threshold such that $\overline{F}(u,u) := \mathbb{P}[X > u, Y > u] > 0$. The bivariate tail estimator that we propose, for $x > u,\ y > u$, is

$$\widehat{F}^*(x,y) = \left(\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i>u,\, Y_i>u\}}\right)\Big(1-g(1-V_{\widehat{k},\widehat{\sigma}}(x-u))-g(1-V_{\widehat{k},\widehat{\sigma}}(y-u)) + G\big(1-V_{\widehat{k},\widehat{\sigma}}(x-u), 1-V_{\widehat{k},\widehat{\sigma}}(y-u)\big)\Big) + \widehat{F}_1^*(u,y) + \widehat{F}_2^*(x,u) - \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i\le u,\, Y_i\le u\}}.$$

Description of the construction: Given a threshold u, we introduce the distribution of excesses above u: $F_u(x,y) := \mathbb{P}[\, X-u \le x, Y-u \le y \,|\, X > u, Y > u\,]$. We have $F_u(x,y) = [F(x+u,y+u) - F(u,y+u) - F(x+u,u) + F(u,u)]\,\overline{F}(u,u)^{-1}$. So we obtain, for $x > u,\ y > u$,

$$F(x,y) = (\overline{F}(u,u)) \cdot F_u(x-u,y-u) + F(u,y) + F(x,u) - F(u,u). \qquad (4.1)$$

From Theorem E we now know that for u around x_F (that is for high thresholds u) we can approximate $F_u(x-u,y-u)$ with $C^{*\,G}\big(1-g(1-V_{k,a(u)}(x-u)), 1-g(1-V_{k,a(u)}(y-u))\big)$. Then, we estimate $F(u,u)$ and $\overline{F}(u,u)$ in (4.1) from the data $\{X_i,Y_i\}_{i=1,\dots,n}$, using the empirical distribution estimates

$$\widehat{F}(u,u) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i\le u,\, Y_i\le u\}}, \quad \widehat{\overline{F}}(u,u) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i>u,\, Y_i>u\}}. \qquad (4.2)$$

We estimate $F(u,y)$ and $F(x,u)$ in (4.1) by

$$F_1^*(u,y) = C(\widehat{F}_X(u), F_Y^*(y)) \quad \text{and} \quad F_2^*(x,u) = C(F_X^*(x), \widehat{F}_Y(u)), \qquad (4.3)$$

where $\widehat{F}_X(u)$ and $\widehat{F}_Y(u)$ are the classical empirical estimators of F_X and F_Y, evaluated at u. $F_X^*(x)$ (resp. $F_Y^*(y)$) in (4.3) is the one-dimensional tail estimator (defined by (3.2)) of the distribution function $F_X(x)$ (resp. $F_Y(y)$) for $x > u$

(resp. $y > u$) and u large. In the case $k \neq 0$, recall that

$$\begin{aligned} F_X^*(x) &= (1-\widehat{F}_X(u))V_{k,\sigma}(x-u)+\widehat{F}_X(u) \\ &= \left(\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i>u\}}\right)\left(1-\left(1-\frac{k(x-u)}{\sigma}\right)^{\frac{1}{k}}\right)+\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i\leq u\}}, \quad (4.4) \end{aligned}$$

$$\begin{aligned} F_Y^*(y) &= (1-\widehat{F}_Y(u))V_{k,\sigma}(y-u)+\widehat{F}_Y(u) \\ &= \left(\frac{1}{n}\sum_{i=1}^{n} 1_{\{Y_i>u\}}\right)\left(1-\left(1-\frac{k(y-u)}{\sigma}\right)^{\frac{1}{k}}\right)+\frac{1}{n}\sum_{i=1}^{n} 1_{\{Y_i\leq u\}}. \quad (4.5) \end{aligned}$$

Since we assume $F_X = F_Y$, the parameters k and σ for the GPD_X and the GPD_Y are the same (see (4.4) and (4.5)). Now, using (4.2) to (4.5), we can approximate $F(x,y)$ for $x > u$, $y > u$ and u large by $F^*(x,y)$ defined by

$$\begin{aligned} \big(\widehat{\overline{F}}(u,u)\big)\, C^{*\,G}\big(1-g(1-V_{k,\sigma}(x-u)), 1-g(1-V_{k,\sigma}(y-u))\big) + F_1^*(u,y) \\ + F_2^*(x,u) - \widehat{F}(u,u). \end{aligned}$$

Using (2.8) we obtain

$$\begin{aligned} F^*(x,y) = \left(\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i>u,\, Y_i>u\}}\right)\Big(1-g(1-V_{k,\sigma}(x-u))-g(1-V_{k,\sigma}(y-u)) \\ +G\big(1-V_{k,\sigma}(x-u), 1-V_{k,\sigma}(y-u)\big)\Big)+F_1^*(u,y)+F_2^*(x,u)-\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i\leq u,\, Y_i\leq u\}}. \end{aligned}$$

In practice we estimate k and σ by MLE based on the excesses of X.

Remark 4 Note that $\widehat{F}^*(x,y)$, is only valid for $x > u$ and $y > u$, when u is *large enough*. The expression large enough is a fundamental problem of the POT method (Peak Over Threshold). The choice of the threshold u is indeed a compromise: u has to be large for the GPD approximation to be valid, but if it is too large, the estimation of the parameters k and σ will suffer from a lack of observations over the threshold. The compromise will be explained in Section 6.

4.2 Case with different marginal distributions

Using the same procedure as the one of Section 4.1, we propose a new bivariate tail estimator, in the case with continuous symmetric copula C and different

marginal distributions F_X and F_Y. For $x > u,\ y > F_Y^{-1}(F_X(u)) := u_Y$ we define

$$\widehat{F}^*(x,y) = \left(\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i>u,\, Y_i>u_Y\}}\right)\big(1 - g(1 - V_{\widehat{k}_X,\widehat{\sigma}_X}(x-u)) - g(1 - V_{\widehat{k}_Y,\widehat{\sigma}_Y}(y-u_Y)) + G\big(1 - V_{\widehat{k}_X,\widehat{\sigma}_X}(x-u), 1 - V_{\widehat{k}_Y,\widehat{\sigma}_Y}(y-u_Y)\big)\big) + \widehat{F}_1^*(u,y) + \widehat{F}_2^*(x,u_Y) - \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i\le u,\, Y_i\le u_Y\}}, \quad (4.6)$$

where $\widehat{k}_X$, $\widehat{\sigma}_X$ (resp. $\widehat{k}_Y$, $\widehat{\sigma}_Y$) are MLE based on the excesses of X (resp. Y). We refer to Remark 4 for the choice of u.

Description of the construction: The main difference, between the case with same marginals and the case with different marginals, is represented by the theorem of approximation of the distribution of excesses. The central idea, in this case, is to use the bi-dimensional extension of Pickands-Balkema-de Hann Theorem in case with continuous symmetric copula C and different marginal distributions, that we have rigorously proved in Theorem 1. From this approach, following a construction similar to Section 4.1, we obtain (4.6), with

$$\widehat{F}_1^*(u,y) = C(\widehat{F}_X(u), \widehat{F}_Y^*(y)),\ \ \widehat{F}_2^*(x,u_Y) = C(\widehat{F}_X^*(x), \widehat{F}_Y(u_Y)). \quad (4.7)$$

In (4.7), $\widehat{F}_X(u)$ and $\widehat{F}_Y(u_Y)$ are the empirical univariate estimators, evaluated at respective thresholds, and $\widehat{F}_X^*(x)$ and $\widehat{F}_Y^*(y)$ are one-dimensional tail estimators of the marginal distribution functions, as proposed in (4.4)-(4.5). Finally, in Section 7.3, we also remark that $F_Y^{-1}(F_X(u))$ may be estimated by its empirical version.

5 Convergence results in the univariate case

Our main contributions (Theorem 6, Corollary 7) in this section are new one dimensional convergence results, needed in Section 6 to derive asymptotic properties of the bivariate tail estimator (4.6). Incidentally we get asymptotic confidence intervals for the unknown theoretical function $F(x)$, using Theorem 4. Cox and Hinkley (1974) (Chapter 9) and Smith (1987) develop theorems of convergence for maximum likelihood estimators, $\widehat{k}$ and $\widehat{\sigma}$, and for the univariate tail estimator $\widehat{F}^*$ defined in (3.4). In particular they show that it is possible to obtain accurate results, with expressions of mean and variance assuming specific

domains of attraction for the theoretical distribution F. From now on we assume that the tail of F decays like a power function, i.e. is in the domain of attraction of Fréchet. This class of distributions is quite large and includes the Pareto, Burr, Cauchy and t-distributions and is called the class of *heavy tailed distributions*. Let us recall the following result:

Theorem F *(Embrechts, Kluppelberg and Mikosch (1997), Theorem 3.3.7) The distribution function F belongs to the maximum domain of attraction of Fréchet*

$$\Phi_\alpha(x) = \begin{cases} 0, & \text{if } x \leq 0, \\ \mathrm{e}^{-x^{-\alpha}}, & \text{if } x > 0, \end{cases}$$

for some $\alpha > 0$, ($F \in MDA(\Phi_\alpha)$), if and only if $\overline{F}(x) = x^{-\alpha}L(x)$ for some slowly varying function $L(x)$.

We remark that in this case the endpoint $x_F := \sup\{x \in \mathbb{R} \,|\, F(x) < 1\} = +\infty$. As in Smith (1987), Section 3, we shall assume that L satisfies one of the following conditions:

- SR1: $\frac{L(tx)}{L(x)} = 1 + O(\phi(x)),\ \forall\ t > 0, \quad \text{as} \quad x \to \infty,$
- SR2: $\frac{L(tx)}{L(x)} = 1 + k(t)\phi(x) + o(\phi(x)),\ \forall\ t > 0, \quad \text{as} \quad x \to \infty,$

where $\phi(x) > 0$ and $\phi(x) \to 0$ as $x \to \infty$. Let R_ρ be the set of $\rho-$regularly varying functions. In the case SR2, excluding trivial cases, $\phi \in R_\rho$, for some $\rho \leq 0$, and $k(t) = c\, h_\rho(t)$, with $h_\rho(t) = \int_1^t u^{\rho-1}\mathrm{d}u$; (for more details see Smith (1987), Section 3). In the following we recall (Theorem G, H) important results in Smith (1987). Those theorems are written conditionally on $N = m_n$ and N denoting the number of excesses above the threshold. In practice we work with some threshold u and N is considered as random. Therefore we give the analogues (Corollary 2, 3) of Theorem G and H working unconditionally on N. Let us now introduce the GDP density $g(y, \sigma, k) = \frac{\partial}{\partial y}V(y, \sigma, k)$ and the log likelihood function $L_{m_n}(\sigma, k) = \sum_{i=1}^{m_n} \log(g(Y_i, \sigma, k))$.
An important result for maximum likelihood estimators $\widehat{k}$ and $\widehat{\sigma}$ of the Generalized Pareto distribution is

Theorem G *(Smith (1987), Theorem 3.2) Suppose L satisfies SR2. Let $Y_1, \ldots,$ Y_{m_n} i.i.d from an unknown distribution function $F_{u_{m_n}}$ where $\lim_{n\to\infty} m_n = \infty$,*

$\lim_{n\to\infty} \frac{m_n}{n} = 0$. *For each* m_n *we define a threshold* $u_{m_n} := \overline{f}(m_n) \xrightarrow[n\to\infty]{} \infty$ *such that*

$$\frac{\sqrt{m_n}\, c\, \phi(\overline{f}(m_n))}{\alpha - \rho} \xrightarrow[n\to\infty]{} \mu \in (-\infty, \infty). \tag{5.1}$$

We define $k = -\alpha^{-1}$ *and* $\sigma_{m_n} = \overline{f}(m_n)\alpha^{-1}$. *Then there exists, with probability tending to 1, a local maximum* $(\widehat{\sigma}_{m_n}, \widehat{k}_{m_n})$ *of the GPD log likelihood function* L_{m_n}, *such that*

$$\sqrt{m_n} \begin{pmatrix} \frac{\widehat{\sigma}_{m_n}}{\sigma_{m_n}} - 1 \\ \widehat{k}_{m_n} - k \end{pmatrix} \xrightarrow[n\to\infty]{d} \mathcal{N}\left(\begin{pmatrix} \frac{\mu(1-k)(1+2k\rho)}{1-k+k\rho} \\ \frac{\mu(1-k)k(1+\rho)}{1-k+k\rho} \end{pmatrix}; \begin{pmatrix} 2(1-k) & (1-k) \\ (1-k) & (1-k)^2 \end{pmatrix} \right).$$

If L *satisfies only SR1, with* ϕ *non-increasing, and if* $\sqrt{m_n}\, \phi(\overline{f}(m_n)) \xrightarrow[n\to\infty]{} 0$, *then the same result holds with* $\mu = 0$.

Corollary 2 *Suppose* L *satisfies SR2. Let* n *be the sample size and* $u_n := \overline{f}(n)$ *the threshold, such that* $\overline{f}(n) \xrightarrow[n\to\infty]{} \infty$. *Let* $N = N_n$ *denote the random number of excesses of* u_n. *If*

$$n(1 - F(u_n)) \xrightarrow[n\to\infty]{} \infty, \tag{5.2}$$

$$\sqrt{n(1 - F(u_n))}\, c\, \phi(u_n) \xrightarrow[n\to\infty]{} \mu(\alpha - \rho), \tag{5.3}$$

then Theorem G holds also unconditionally on N.

Proof: If (5.2) and (5.3) hold then $N(n\,(1 - F(u_n)))^{-1} \xrightarrow[n\to\infty]{\mathbb{P}} 1$, and (5.1) holds in probability, i.e.

$$\frac{\sqrt{N}\, c\, \phi(u_n)}{\alpha - \rho} = \frac{\sqrt{N}\, c\, \phi(\overline{f}(n))}{\alpha - \rho} \xrightarrow[n\to\infty]{\mathbb{P}} \mu \in (-\infty, \infty).$$

Hence we conclude with a Skorohod-type construction of probability spaces on which (5.1) holds almost surely. □

We now recall an important result of convergence of the univariate tail estimator.

Theorem H *(Smith (1987), Theorem 8.1) Suppose* L *satisfies SR2 and that the other conditions of Theorem G hold. Let* z_{m_n} *be a sequence such that*

$$z_{m_n} := f(m_n) \xrightarrow[n\to\infty]{} \infty, \quad \frac{\log(z_{m_n})}{\sqrt{m_n}} \xrightarrow[n\to\infty]{} 0, \tag{5.4}$$

$$\frac{\sqrt{m_n}}{\log(z_{m_n})} \left[\frac{m_n}{n\,(1 - F(u_{m_n}))} - 1 \right] \xrightarrow[n\to\infty]{} 0, \tag{5.5}$$

$$(z_{m_n})^{-s\,\rho}\frac{\phi(u_{m_n}(z_{m_n})^s)}{\phi(u_{m_n})} = f(m_n)^{-s\,\rho}\frac{\phi(\overline{f}(m_n)f(m_n)^s)}{\phi(\overline{f}(m_n))} \xrightarrow[n\to\infty]{} 1, \quad \textit{for } s\in[0,1].$$

Then

$$\frac{\sqrt{m_n}}{\log(f(m_n))}\left[\frac{1-\widehat{F}^*(\overline{f}(m_n)\,f(m_n))}{1-F(\overline{f}(m_n)\,f(m_n))}-1\right] \xrightarrow[n\to\infty]{d} \mathcal{N}(\nu,\tau^2), \tag{5.6}$$

where $\nu=0$ *if* $\rho=0$, $\nu=\frac{\mu\alpha(\alpha+1)(1+\rho)}{1+\alpha-\rho}$ *for* $\rho<0$ *and* $\tau^2=\alpha^2(1+\alpha)^2$.

Corollary 3 *Suppose L satisfies SR2. Let n be the sample size,* $u_n:=\overline{f}(n)\xrightarrow[n\to\infty]{}\infty$ *and* $z_n:=f(n)\xrightarrow[n\to\infty]{}\infty$ *such that* $(z_n)^{-s\,\rho}\frac{\phi(u_n(z_n)^s)}{\phi(u_n)}\xrightarrow[n\to\infty]{}1$, *for* $s\in[0,1]$. *Let* $N=N_n$ *denote the random number of excesses of* u_n. *If conditions* (5.2) *and* (5.3) *hold and moreover*

$$\frac{\log(z_n)}{\sqrt{n(1-F(u_n))}}\xrightarrow[n\to\infty]{}0, \tag{5.7}$$

then Theorem H holds also unconditionally on N.

Proof: If (5.2), (5.3) and (5.7) hold, then (5.4) and (5.5) hold in probability, i.e

$$\frac{\log(z_n)}{\sqrt{N}}\xrightarrow[n\to\infty]{\mathbb{P}}0, \qquad \frac{\sqrt{N}}{\log(z_n)}\left[\frac{N}{n\,(1-F(u_n))}-1\right]\xrightarrow[n\to\infty]{\mathbb{P}}0.$$

We conclude as for Corollary 2. □

In order to obtain a confidence interval from Theorem H, we have to estimate the unknown quantity $1-F(\overline{f}(m_n)\,f(m_n))$. Define

$$\widetilde{r}_{m_n}:=\frac{\sqrt{m_n}}{\log(f(m_n))}\left(\frac{1}{1-\frac{1}{n}\sum_{i=1}^n 1_{(X_i\le\overline{f}(m_n)f(m_n))}}\right).$$

We then get the following result:

Theorem 4 *Let F in the domain of attraction of* Φ_α *for some* $\alpha>0$, *and suppose L satisfies SR2. If*

$$\frac{(z_{m_n})^\alpha}{\sqrt{m_n}}\xrightarrow[n\to\infty]{}0, \tag{5.8}$$

and the other conditions of Theorems G and H hold, then

$$\widetilde{r}_{m_n}\left[F(\overline{f}(m_n)\,f(m_n))-\widehat{F}^*(\overline{f}(m_n)\,f(m_n)\right]\xrightarrow[n\to\infty]{d}\mathcal{N}(\nu,\tau^2), \tag{5.9}$$

where $\nu=0$ *if* $\rho=0$ *and* $\nu=\frac{\mu\alpha(\alpha+1)(1+\rho)}{1+\alpha-\rho}$ *for* $\rho<0$ *and* $\tau^2=\alpha^2(1+\alpha)^2$.

The proof of Theorem 4 is postponed to the Appendix. A straightforward consequence of Theorem 4 is

Corollary 5 *If all conditions of Theorem 4 hold then we have the following convergence in probability*

$$\Big[F(\overline{f}(m_n)\,f(m_n)) - \widehat{F}^*(\overline{f}(m_n)\,f(m_n)\Big] \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Let us give the unconditionally version of Theorem 4 and Corollary 5.

Theorem 6 *Suppose L satisfies SR2. Let n be the sample size, $u_n := \overline{f}(n) \xrightarrow[n\to\infty]{} \infty$ and $z_n := f(n) \xrightarrow[n\to\infty]{} \infty$ such that $(z_n)^{-s\,\rho}\frac{\phi(u_n(z_n)^s)}{\phi(u_n)} \xrightarrow[n\to\infty]{} 1$, for $s \in [0,1]$. Let $N = N_n$ denote the random number of excesses of u_n. If conditions (5.2), (5.3) and (5.7) hold and if $(z_n)^{\alpha}(n(1-F(u_n)))^{-1/2} \xrightarrow[n\to\infty]{} 0$, then*

$$\frac{\sqrt{N}}{\log(f(n))\,\widehat{\overline{F}}_n(\overline{f}(n)\,f(n))}\Big[F(\overline{f}(n)\,f(n)) - \widehat{F}^*(\overline{f}(n)\,f(n)\Big] \xrightarrow[n\to\infty]{d} \mathcal{N}(\nu,\tau^2). \quad (5.10)$$

Remark 5 It is important to notice that (5.10) yields asymptotic confidence intervals for $F(\overline{f}(n)\,f(n))$.

Corollary 7 *Under assumptions of Theorem 6 we have*

$$\Big[F(\overline{f}(n)\,f(n)) - \widehat{F}^*(\overline{f}(n)\,f(n))\Big] \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Theorem 6 and Corollary 7 are proved from Theorem 4 and Corollary 5. The proofs will be omitted here.

We can even obtain a more general result for the absolute error.

Theorem 8 *Suppose F belongs to the maximum domain of attraction of Fréchet and L satisfies SR2. Assume that the threshold $u_n := \overline{f}(n) \xrightarrow[n\to\infty]{} \infty$, then if (5.2) and (5.3) hold we get*

$$\sup_{x>\overline{f}(n)}\Big|F(x) - \widehat{F}^*(x)\Big| \xrightarrow[n\to\infty]{\mathbb{P}} 0. \quad (5.11)$$

Proof: From (3.1) and (3.3) we can rewrite (5.11) as

$$\sup_{x>\overline{f}(n)}\left|\overline{F}(\overline{f}(n))F_u(x-\overline{f}(n))+F(\overline{f}(n))-\left[\widehat{\overline{F}}(\overline{f}(n))V_{\widehat{k},\widehat{\sigma}}(x-\overline{f}(n))+\widehat{F}(\overline{f}(n))\right]\right|$$

$$\leq \sup_{x>\overline{f}(n)}\left|\overline{F}(\overline{f}(n))\,F_u(x-\overline{f}(n))-\widehat{\overline{F}}(\overline{f}(n))V_{\widehat{k},\widehat{\sigma}}(x-\overline{f}(n))\right|+\left|F(\overline{f}(n))-\widehat{F}(\overline{f}(n))\right|.$$

To conclude we use Pickands-Balkema-de Hann Theorem B, univariate Glivenko-Cantelli Theorem (8.9) and an auxiliary result (Proposition 12), postponed to the end of the paper. □

6 Convergence results in the bivariate case

In this section we extend the univariate convergence results of Section 5 to the two-dimensional tail estimator (4.6). We work here in the general setting with non necessarily equal marginal distributions F_X and F_Y. The copula C is still assumed to be continuous and symmetric. We work unconditionally on N, which is more interesting for applications.

Remark 6 Let n be the sample size. We choose, from Theorem 1,

$$u_{1\,n}:=\overline{f}_1(n)\xrightarrow[n\to\infty]{}\infty \quad \text{threshold for } X,$$

$$u_{2\,n}=\overline{f}_2(n)=F_Y^{-1}(F_X(\overline{f}_1(n)))\xrightarrow[n\to\infty]{}\infty \quad \text{threshold for } Y.$$

We now introduce our main result of convergence in probability for the tail estimator (4.6). To state the following result we need a two-dimensional version of the Glivenko-Cantelli Theorem (Theorem J in Appendix).

Theorem 9 *Suppose F_X and F_Y belong to the maximum domain of attraction of Fréchet and L_X, L_Y satisfy the condition SR2. Assume that the copula C is continuous and symmetric. Then under assumptions of Theorems C and 1, if sequences $\overline{f}_1(n)$, $\overline{f}_2(n)$, defined by Remark 6, satisfy the conditions of Theorem 8 and if the conditions of Theorem J hold then*

$$\sup_{x>\overline{f}_1(n),\,y>\overline{f}_2(n)}\left|F(x,y)-\widehat{F}^*(x,y)\right|\xrightarrow[n\to\infty]{\mathbb{P}}0. \tag{6.1}$$

The proof of Theorem 9 is postponed to the Appendix.

In Section 7.3 we will propose an estimator for $\overline{f}_2(n)$ denote by $\widehat{\overline{f}}_2(n)$. This

approach is useful in applications and Theorem 9 remains valid when replacing $\overline{f_2}(n)$ by $\widehat{\overline{f}}_2(n)$ (see discussions in Section 7.3).

In Theorem 9 we are assuming the following structure of dependence between X and Y:

- X, Y are continuous random variables, univariate, with distribution functions F_X, F_Y, and with continuous symmetric copula C, which represents the structure of dependence between X and Y;
- $F_X \in \mathrm{MDA}(\Phi_{\alpha_1})$ and $F_Y \in \mathrm{MDA}(\Phi_{\alpha_2})$, with $\alpha_1, \alpha_2 > 0$ (domain of attraction of Fréchet);
- We consider $\{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n), \ldots\}$, i.i.d. with distribution function $F_{X,Y}$.

For clarity, we recall the expressions of the two components of (6.1). From (4.1), we have, for $x > \overline{f}_1(n)$ and $y > \overline{f}_2(n)$,

$$F(x,y) = \overline{F}\left(\overline{f}_1(n), \overline{f}_2(n)\right) \cdot F_u(x - \overline{f}_1(n), y - \overline{f}_2(n)) + F(\overline{f}_1(n), y) + F(x, \overline{f}_2(n)) - F(\overline{f}_1(n), \overline{f}_2(n)), \quad (6.2)$$

and from (4.6) we obtain

$$\widehat{F}^*(x,y) = \left(\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i > \overline{f}_1(n),\, Y_i > \overline{f}_2(n)\}}\right) C^{*G}\big(1 - g(1 - V_{\widehat{k}_X, \widehat{\sigma}_X}(x - \overline{f}_1(n))),$$
$$1 - g(1 - V_{\widehat{k}_Y, \widehat{\sigma}_Y}(y - \overline{f}_2(n)))\big) + \widehat{F}_1^*(\overline{f}_1(n), y) + \widehat{F}_2^*(x, \overline{f}_2(n)) - \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \le \overline{f}_1(n)),\, Y_i \le \overline{f}_2(n)\}}, \quad (6.3)$$

where $\widehat{F}_1^*(\overline{f}_1(n), y) = C\left(\widehat{F}_X(\overline{f}_1(n)),\, \widehat{F}_Y^*(y)\right)$ is equal to

$$C\left(\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \le \overline{f}_1(n)\}}, \left(\frac{1}{n}\sum_{i=1}^{n} 1_{\{Y_i > \overline{f}_2(n)\}}\right) \cdot V_{\widehat{k}_Y, \widehat{\sigma}_Y}\left(y - \overline{f}_2(n)\right) + \frac{1}{n}\sum_{i=1}^{n} 1_{\{Y_i \le \overline{f}_2(n)\}}\right),$$

and $\widehat{F}_2^*(x, \overline{f}_2(n)) = C\left(\widehat{F}_X^*(x),\, \widehat{F}_Y(\overline{f}_2(n))\right)$ is equal to

$$C\left(\left(\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i > \overline{f}_1(n)\}}\right) \cdot V_{\widehat{k}_X, \widehat{\sigma}_X}\left(x - \overline{f}_1(n)\right) + \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \le \overline{f}_1(n)\}}, \frac{1}{n}\sum_{i=1}^{n} 1_{\{Y_i \le \overline{f}_2(n)\}}\right).$$

Remark 7 Since the sequence $u_{2\,n}$ is obtained from $u_{1\,n}$ (see Remark 6), the asymptotic properties of u_2 can be deduced from those of u_1, with additional regularity assumptions. Assume that $u_{1\,n}$ satisfies the conditions of Theorem 8. So if $\lim_{n\to\infty} u_{1\,n} = \infty$, then also $\lim_{n\to\infty} u_{2\,n} = \infty$, and if $u_{1\,n}$ satisfies (5.2) then also $\lim_{n\to\infty} n(1 - F_Y(u_{2\,n})) = \infty$. If $u_{1\,n}$ satisfies (5.3) and if $\lim_{n\to\infty} \frac{\phi_2(u_{2\,n})}{\phi_1(u_{1\,n})} = L \in (-\infty, \infty)$, then u_2 satisfies

$$\lim_{n\to\infty} \sqrt{n(1 - F_Y(u_{2\,n}))}\, c_2\, \phi_2(u_{2\,n})(\alpha_2 - \rho_2)^{-1} = \mu_2 \in (-\infty, \infty).$$

Remark 8 Assume that the marginal distributions are the same. We choose

$$u_{1\,n} = u_{2\,n} := \overline{f}(n) \xrightarrow[n\to\infty]{} \infty \quad \text{threshold for } X, Y.$$

Theorem 9 can be reformulated as follows:

Theorem 10 *Suppose F belongs to the maximum domain of attraction of Fréchet and L satisfies the condition SR2. Assume that the copula C is continuous and symmetric. Then under assumptions of Theorems C and E, if sequences $\overline{f}(n)$ satisfies the conditions of Theorem 8 and if the conditions of Theorem J hold then*

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

7 Simulation Study

We include three examples for the estimation of the bivariate tail (Theorem 9 and 10). Note that one of the characteristics of our estimator is that it is easy to implement.

7.1 Case with identical marginal distributions

Model: We choose the marginal distributions F_X, F_Y and the continuous symmetric copula C as follows:

$$C(u,v) = u+v-1+[(1-u)^{-1}+(1-v)^{-1}-1]^{-1} \text{ (Survival Clayton copula)}, \quad (7.1)$$

$$F_X(x) = 1-(1+x)^{-1}, \quad F_Y(y) = 1-(1+y)^{-1} \text{ (same Burr distributions)}. \quad (7.2)$$

To draw our sample $(X_i, Y_i)_{i=1,\dots,n}$ we use the conditional sampling approach well described in Cherubini, Luciano and Vecchiato (2004). We estimate σ and k by MLE based on the excesses of X.

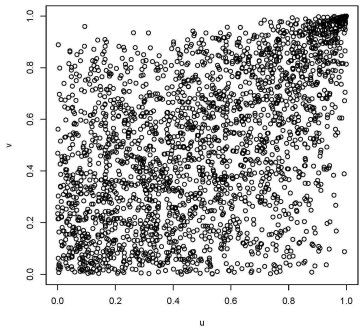

Figure 7.1: Copula Survival Clayton (7.1).

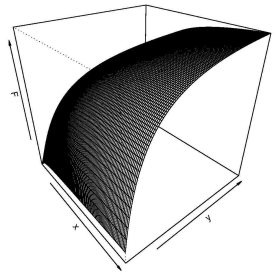

Figure 7.2: Bivariate distribution function $F_{X,Y}(x,y)$, with $F_X = F_Y$, for $x > 0, y > 0$, as in (7.2).

Threshold: We choose $\overline{f}(n) = \frac{n^{\frac{1}{3}}}{3} \xrightarrow[n\to\infty]{} \infty$. Then the assumptions of Theorem 10 are satisfied hence

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

We call t the number of simulations of $n-$samples. We define for each $x > \overline{f}(n)$, $y > \overline{f}(n)$

$$ERR_{i,\,abs} = \left| \widehat{F}^*(x,y) - F(x,y) \right|, \text{ for } i = 1, \ldots, t \text{ and } \overline{ERR_{abs}} = \frac{1}{t} \sum_{i=1}^{t} ERR_{i,\,abs}, \tag{7.3}$$

$$ERR_{i,\,rel} = \left| \frac{\widehat{F}^*(x,y) - F(x,y)}{F(x,y)} \right|, \text{ for } i = 1, \ldots, t \text{ and } \overline{ERR_{rel}} = \frac{1}{t} \sum_{i=1}^{t} ERR_{i,\,rel}. \tag{7.4}$$

Let us fix $t = 100$, $(x_0,\, y_0) = (10, 10)$, so that both x_0 and y_0 are above the threshold $\overline{f}(n)$ for $n = 1000$, 2000, 5000 or 10000. We then compute for each value of n the mean absolute and relative errors, $\overline{ERR_{abs}}$ and $\overline{ERR_{rel}}$, as far as the empirical variance of $\widehat{F}^*(x_0, y_0)$ (see Table 7.1 below). It gives us an idea of the ponctual convergence of our estimator as $n \to \infty$ for $(x_0, y_0) = (10, 10)$.

The last column of Table 7.1 gives the mean number of excesses above the threshold (computed on the $t = 100$ samples). Recall that the choice of the

threshold satisfies a compromise: the GPD approximation is valid only for high thresholds but the number of excesses decreases rather fastly with the threshold.

n	$\overline{ERR_{abs}}$	$\mathrm{var}(\widehat{F}^*(x_0,y_0))_t$	$\overline{ERR_{rel}}$	$\overline{f}(n)$	$\mathrm{mean(Excesses)}_t$
1000	0.0112	$1.95e^{-04}$	0.0129	3.333	233
2000	0.0086	$8.87e^{-05}$	0.0099	4.199	384
5000	0.0051	$4.14e^{-05}$	0.0059	5.699	745
10000	0.0034	$2.16e^{-05}$	0.0039	7.181	1223

Table 7.1: Errors (7.3)-(7.4) and empirical variance for $(x_0,\, y_0) = (10, 10)$, $t = 100$, case with same marginals.

Now, for $n = 1000$ and $t = 10$, we discretize the set $[\overline{f}(n)+1, 250]^2 = [4.333, 250]^2$ with a grid of 62500 points. On this grid we get $\max\left(\overline{ERR_{abs}}\right) = 0.0171$, $\max\left(\overline{ERR_{rel}}\right) = 0.0195$ and $\max\left(\mathrm{var}(\widehat{F}^*(x,y))\right) = 3.4e^{-04}$. For $n = 2000$ and $t = 10$, if we discretize the set $[\overline{f}(n) + 1, 250]^2 = [5.199, 250]^2$, we obtain $\max\left(\overline{ERR_{abs}}\right) = 0.0104$, $\max\left(\overline{ERR_{rel}}\right) = 0.0121$ and $\max\left(\mathrm{var}(\widehat{F}^*(x,y))\right) = 1.9e^{-04}$. So the quality of our estimate is good for any point above the threshold.

7.2 **Case with different marginal distributions**

Model: We choose the marginal distributions F_X, F_Y and the continuous symmetric copula C as follows:

$$C(u,v) = u+v-1+[(1-u)^{-1}+(1-v)^{-1}-1]^{-1} \text{ (Survival Clayton copula)}, \quad (7.5)$$

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

Still using conditional sampling we draw a sample $(X_i, Y_i)_{i=1,\dots,n}$. We estimate the parameters by $\widehat{\sigma}_X$, $\widehat{k}_X$ and $\widehat{\sigma}_Y$, $\widehat{k}_Y$, as described in Section 4.2.

Thresholds: We choose $\overline{f}_1(n) = \frac{n^{\frac{1}{3}}}{3}$ and $\overline{f}_2(n) = F_Y^{-1}(F_X(\overline{f}_1(n))) = \sqrt{\frac{n^{\frac{1}{3}}}{3}}$. In our particular case the marginals are continuous and invertible functions. From Theorem 9 we have

$$\sup_{x>\overline{f}_1(n),\, y>\overline{f}_2(n)} \left|F(x,y) - \widehat{F}^*(x,y)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

We call t the number of simulations of $n-$samples. We choose the point $(x_0, y_0) = (10, 10)$, with $x_0 > \overline{f}_1(n)$, $y_0 > \overline{f}_2(n)$, we fix $t = 100$ and we calculate the errors in (7.3)-(7.4) and the empirical variance of $\widehat{F}^*(x_0, y_0)$ (see Table 7.2 below).

n	$\overline{ERR_{abs}}$	$\text{var}(\widehat{F}^*(x_0,y_0))_t$	$\overline{ERR_{rel}}$	$\overline{f}_1(n)$, $\overline{f}_2(n)$	mean(Excesses)_t
1000	0.0086	$1.2e^{-04}$	0.0094	3.333, 1.825	231
2000	0.0061	$4.6e^{-05}$	0.0067	4.199, 2.049	385
5000	0.0040	$2.1e^{-05}$	0.0044	5.699, 2.387	744
10000	0.0027	$1.1e^{-05}$	0.003	7.181, 2.679	1221

Table 7.2: Errors (7.3)-(7.4) and empirical variance for $(x_0,\, y_0) = (10, 10)$, $t = 100$, case with different marginals.

If we calculate $\overline{ERR_{abs}}$ and $\overline{ERR_{rel}}$ for $n = 1000$ and $t = 10$, in a grid of 62500 points in the region $[\overline{f}_1(n)+1, 250]\times[\overline{f}_2(n)+1, 250] = [4.333, 250]\times[2.825, 250]$ we get $\max\left(\overline{ERR_{abs}}\right) = 0.0123$, $\max\left(\overline{ERR_{rel}}\right) = 0.0143$ and $\max\left(\text{var}(\widehat{F}^*(x,y))\right) = 2.5e^{-04}$. For $n = 2000$ and $t = 10$, if we discretize the set $[\overline{f}_1(n) + 1, 250] \times [\overline{f}_2(n) + 1, 250] = [5.199, 250] \times [3.049, 250]$ with a grid of 62500 points we get $\max\left(\overline{ERR_{abs}}\right) = 0.0093$, $\max\left(\overline{ERR_{rel}}\right) = 0.0111$ and $\max\left(\text{var}(\widehat{F}^*(x,y))\right) = 1.1e^{-04}$.

7.3 **Estimation of** $\overline{f}_2(n)$

In the previous section the threshold $\overline{f}_2(n)$ for the second marginal distribution was obtained as $\overline{f}_2(n) = F_Y^{-1}(F_X(\overline{f}_1(n)))$. In practice F_X and F_Y are unknown so $\overline{f}_2(n)$ has to be estimated. Let us be more precise. Let us draw a sample $(X_i, Y_i)_{i=1,\dots,n}$ from the bivariate distribution $F_{X,Y}$. We estimate $\overline{f}_2(n)$ by $\widehat{\overline{f}}_2(n) = \widehat{F}_Y^{-1}(\widehat{F}_X(\overline{f}_1(n)))$, with $\widehat{F}_X(\overline{f}_1(n)) = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i\le\overline{f}_1(n)\}}$ and $\widehat{F}_Y^{-1}$ the empirical quantile function.

Let us justify it. We denote $U(t) = F_Y^{-1}\left(1-\frac{1}{t}\right)$ so that the quantile y_p is defined by $y_p = U\left(\frac{1}{1-p}\right)$. Denoting $U_n(t) = \widehat{F}_Y^{-1}\left(1-\frac{1}{t}\right)$,

$$U_n\left(\frac{n}{k-1}\right) = \widehat{F}_Y^{-1}\left(1-\frac{k-1}{n}\right) = Y_{k,n}, \quad \text{for} \;\; k = 1,\dots,n.$$

Hence $Y_{k,n}$ appears as a natural estimator of the $\left(1-\frac{k-1}{n}\right)$-quantile. The range $[Y_{n,n},\; Y_{1,n}]$ of the data allows one to make a within-sample estimation up to the $(1-\frac{1}{n})-$quantile. From a mathematical point of view the difference between high quantiles within and outside the sample can for instance be described as follows. For $n\to\infty$

- high quantiles within the sample: $p = p_n \nearrow 1$ and $n(1-p_n)\to c\in(1,\infty]$,
- high quantiles without the sample: $p = p_n \nearrow 1$ and $n(1-p_n)\to c\in[0,1)$.

In the first case with $c = \infty$ there are classical results that basically tell us that we can just use the empirical quantile function for estimating the quantile y_p as follows:

Theorem 11 *(Dekkers, de Haan (1989), Theorem 3.1) Suppose $Y_1, Y_2, \ldots, Y_n$ is an i.i.d. sample from $F_Y \in MDA(\Phi_\alpha)$, $\alpha > 0$, and F_Y has a positive density $f_Y \in R_{-1-\alpha}$. Write $p = p_n$ and $k = k(n) = [n(1-p_n)]$, where $[x]$ denotes the integer part of x. If the conditions $p_n \xrightarrow[n\to\infty]{} 1$ and $n(1-p_n) \xrightarrow[n\to\infty]{} \infty$ hold then*

$$\sqrt{2k}\,\frac{Y_{k,n} - y_p}{Y_{k,n} - Y_{2k,n}} \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, \frac{2^{\frac{2}{\alpha}+1}\frac{1}{\alpha^2}}{(2^{\frac{1}{\alpha}}-1)^2}\right). \tag{7.6}$$

From (7.6) we have $Y_{k,n} - y_p \xrightarrow{\mathbb{P}} 0$, for $n \to \infty$.

Model: For the simulations we keep the model of Section 7.2.
Thresholds: We choose $\overline{f}_1(n) = \frac{n^{\frac{1}{3}}}{3}$ and $\widehat{\overline{f}}_2(n) = \widehat{F}_Y^{-1}(\widehat{F}_X(\overline{f}_1(n)))$. In our case $p = p_n = \widehat{F}_X(\overline{f}_1(n)))$. From condition (5.2) and univariate Glivenko-Cantelli Theorem we have $\lim_{n\to\infty} n(1 - \widehat{F}_X(\overline{f}_1(n))) = \infty$. In that case assumptions of Theorem 11 are satisfied hence we have

$$\left[\widehat{\overline{f}}_2(n) - \overline{f}_2(n)\right] \xrightarrow[n\to\infty]{\mathbb{P}} 0. \tag{7.7}$$

Using (7.7) and the fact that the derivative of the Generalized Pareto distribution, in the Fréchet case, is bounded it can be proved that Theorem 9 is true when replacing $\overline{f}_2(n)$ by $\widehat{\overline{f}}_2(n)$.
We call t the number of simulated $n-$samples. We choose $(x_0, y_0) = (10, 10)$ with $x_0 > \overline{f}_1(n)$, $y_0 > \widehat{\overline{f}}_2(n)$. For $t = 100$ we calculate the errors in (7.3)-(7.4) and the empirical variance of $\widehat{F}^*(x_0, y_0)$ (see Table 7.3 below).

Now, for $n = 1000$ and $t = 10$, we discretize the set $[\overline{f}_1(n) + 1, 250] \times [\text{mean}(\widehat{\overline{f}}_2(n))_t + 1, 250] = [4.333, 250] \times [2.814, 250]$ with a grid of 62500 points. On this grid we get $\max\left(\overline{ERR_{abs}}\right) = 0.0164$, $\max\left(\overline{ERR_{rel}}\right) = 0.0195$ and $\max\left(\text{var}(\widehat{F}^*(x,y))\right) = 3.9e^{-04}$. If we calculate $\overline{ERR_{abs}}$ and $\overline{ERR_{rel}}$ for $n = 2000$ and $t = 10$, in a grid of 62500 points in the region $[\overline{f}_1(n) + 1, 250] \times [\text{mean}(\widehat{\overline{f}}_2(n))_t + 1, 250] = [5.199, 250] \times [3.054, 250]$ we obtain $\max\left(\overline{ERR_{abs}}\right) = 0.0094$, $\max\left(\overline{ERR_{rel}}\right) = 0.011$ and $\max\left(\text{var}(\widehat{F}^*(x,y))\right) = 1.3e^{-04}$.

n	$\overline{ERR_{abs}}$	$\mathrm{var}(\widehat{F}^*(x_0,y_0))_t$	$\overline{ERR_{rel}}$	$\overline{f}_1(n)$, $\mathrm{mean}(\widehat{\overline{f}}_2(n))_t$	$\mathrm{mean(Excesses)}_t$
1000	0.0086	$1.1e^{-04}$	0.0095	3.333, 1.814	229
2000	0.0068	$5.3e^{-05}$	0.0075	4.199, 2.054	386
5000	0.0039	$2.1e^{-05}$	0.0043	5.699, 2.389	747
10000	0.0031	$1.4e^{-05}$	0.0034	7.181, 2.676	1225

Table 7.3: Errors (7.3)-(7.4) and the empirical variance calculate in $(x_0,\, y_0) = (10, 10)$, in the model with different marginal distributions, estimated threshold $\overline{f}_2(n)$ and $t = 100$.

8 Appendix: proofs and auxiliary results

Proof [**Theorem 1**] : From (2.7) we obtain the existence of $a_1(\cdot)$ and $a_2(\cdot)$ such that, for $p := u + xa_1(u)$ and $q := u_Y + ya_2(u_Y)$

$$V_{k_1,1}(x) = \lim_{u\to x_{F_X}} 1 - \frac{1-F_X(p)}{1-F_X(u)} = \lim_{u\to x_{F_X}} \mathbb{P}[X \le p | X > u], \tag{8.1}$$

$$V_{k_2,1}(y) = \lim_{u_Y\to x_{F_Y}} 1 - \frac{1-F_Y(q)}{1-F_Y(u_Y)} = \lim_{u_Y\to x_{F_Y}} \mathbb{P}[Y \le q | Y > u_Y]. \tag{8.2}$$

From $X \overset{d}{=} F_X^{-1}(F_Y(Y))$ or symmetrically $Y \overset{d}{=} F_Y^{-1}(F_X(X))$, we take $u_Y = F_Y^{-1}(F_X(u))$ and from (8.1)-(8.2) we have, as $u \to x_{F_X}$

$$1 - (1 - V_{k_1,1}(x))(1 - F_X(u)) \sim F_X(u + x\, a_1(u)),$$

$$1 - (1 - V_{k_2,1}(y))(1 - F_Y(F_Y^{-1}(F_X(u)))) \sim F_Y(F_Y^{-1}(F_X(u)) + y\, a_2(F_Y^{-1}(F_X(u)))).$$

It follows from (2.4) that

$$\begin{aligned}
&\lim_{u\to x_{F_X}} \mathbb{P}\left[\frac{X-u}{a_1(u)} > x, \frac{Y - F_Y^{-1}(F_X(u))}{a_2(F_Y^{-1}(F_X(u)))} > y \,\middle|\, X > u, Y > F_Y^{-1}(F_X(u))\right] \\
&= \lim_{u\to x_{F_X}} \frac{C^*\big(1 - F_X(u + x\, a_1(u)), 1 - F_Y(F_Y^{-1}(F_X(u)) + y\, a_2(F_Y^{-1}(F_X(u))))\big)}{C^*\big(1 - F_X(u), 1 - F_Y(F_Y^{-1}(F_X(u)))\big)} \\
&= \lim_{u\to x_{F_X}} \frac{C^*\big((1 - V_{k_1,1}(x))(1 - F_X(u)), (1 - V_{k_2,1}(y))(1 - F_Y(F_Y^{-1}(F_X(u))))\big)}{C^*\big(1 - F_X(u), 1 - F_Y(F_Y^{-1}(F_X(u)))\big)} \\
&= \lim_{\nu\to 1} \frac{C^*\big((1 - V_{k_1,1}(x))(1-\nu), (1 - V_{k_2,1}(y))(1-\nu)\big)}{C^*\big(1-\nu, 1-\nu\big)}.
\end{aligned} \tag{8.3}$$

We can rewrite $1 - V_{k_1,1}(x) = V_{1,1}(h)$, with $h := (1 - k_1\, x)^{\frac{1}{k_1}}$, if $k_1 \neq 0$, or $h := \mathrm{e}^{-x}$, if $k_1 = 0$, and $1 - V_{k_2,1}(y) = V_{1,1}(w)$ with $w := (1 - k_2\, y)^{\frac{1}{k_2}}$, if $k_2 \neq 0$,

or $w := \mathrm{e}^{-y}$, if $k_2 = 0$. So (8.3) becomes

$$\lim_{\nu \to 1} \frac{C^*\big(V_{1,1}(h)(1-\nu), V_{1,1}(w)(1-\nu)\big)}{C^*\big(1-\nu, 1-\nu\big)}.$$

Because C satisfies the hypotheses of Theorem C we have that the above limit equals

$$G(V_{1,1}(h), V_{1,1}(w)) = G(1 - V_{k_1,1}(x), 1 - V_{k_2,1}(y)).$$

From (2.4) we have finally that

$$\begin{aligned}\lim_{u \to x_{F_X}} \mathbb{P}\left[\frac{X-u}{a_1(u)} \le x,\ \frac{Y - F_Y^{-1}(F_X(u))}{a_2(F_Y^{-1}(F_X(u)))} \le y \,\middle|\, X > u, Y > F_Y^{-1}(F_X(u))\right] \\ = 1 - g(1 - V_{k_1,1}(x)) - g(1 - V_{k_2,1}(y)) + G(1 - V_{k_1,1}(x), 1 - V_{k_2,1}(y)) \\ = C^{*\,G}\big(1 - g(1 - V_{k_1,1}(x)), 1 - g(1 - V_{k_2,1}(y))\big).\end{aligned}$$

Since the limit is a continuous distribution function (because $C^{*\,G}$, g and the GPD are) it follows that the convergence can the strengthened to uniform convergence (see for example Embrechts, Kluppelberg and Mikosch (1997), pag. 552). So

$$\begin{aligned}\sup_{\mathcal{A}} \Bigg| \mathbb{P}\left[\frac{X-u}{a_1(u)} \le x,\ \frac{Y - F_Y^{-1}(F_X(u))}{a_2(F_Y^{-1}(F_X(u)))} \le y \,\middle|\, X > u, Y > F_Y^{-1}(F_X(u))\right] \\ - C^{*\,G}\big(1 - g(1 - V_{k_1,1}(x)), 1 - g(1 - V_{k_2,1}(y))\big) \Bigg| \xrightarrow[u \to x_{F_X}]{} 0, \quad (8.4)\end{aligned}$$

where $\mathcal{A} := \{(x,y) : 0 < x \le x_{F_X} - u, 0 < y \le x_{F_Y} - F_Y^{-1}(F_X(u))\}$. Then (2.9) is a straightforward consequence of (8.4). □

Proof [Theorem 4]:
To prove Theorem 4, we need to prove the following convergence in probability

$$\frac{\widetilde{r}_{m_n}}{r_{m_n}} = \frac{\overline{F}(\overline{f}(m_n)\, f(m_n))}{\widehat{\overline{F}}(\overline{f}(m_n)\, f(m_n))} \xrightarrow[n \to \infty]{\mathbb{P}} 1. \quad (8.5)$$

If we are able to prove (8.5), using (5.6) and by direct application of Slutsky's Theorem, we obtain convergence (5.9). To prove (8.5) we use the following result.

Proposition I *(Einmhal (1990), Corollary 1) Let a sequence of i.i.d random variables $X_1, X_2, \ldots$ from a distribution function F. We denote with $\{k_n\}_{n=1}^{\infty}$*

an arbitrary sequence of positive numbers, such that $k_n \le n$ *and* $k_n \to \infty$, $\lim_{n\to\infty} \frac{k_n}{n} = 0$. *Let* $\{\gamma_n\}_{n=1}^{\infty}$ *be a sequence of positive numbers, such that* $\lim_{n\to\infty} \frac{\gamma_n}{\sqrt{k_n}} = \infty$, *then*

$$\sup_{t \ge F^{-1}(1-\frac{k_n}{n})} \left(\frac{n}{\gamma_n}\right) \left|\widehat{\overline{F}}(t) - \overline{F}(t)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Then we will take, following Proposition I, an arbitrary sequence $\{k_n\}_{n=1}^{\infty}$ as

$$\{k_n\}_{n=1}^{\infty} := \{m_n\}_{n=1}^{\infty} \text{ (number of excesses on a sample of size } n), \text{ such that}$$

$$m_n \le n, \quad \lim_{n\to\infty} m_n = \infty \quad \text{and} \quad \lim_{n\to\infty} \frac{m_n}{n} = 0.$$

The sequence of positive numbers $\{\gamma_n\}_{n=1}^{\infty}$ of Proposition I is taken as

$$\{\gamma_n\}_{n=1}^{\infty} := \{\sqrt{m_n}\,\alpha_n\}_{n=1}^{\infty},$$

where α_n is an arbitrary sequence of positive numbers such that $\lim_{n\to\infty} \alpha_n = \infty$. Then, using Proposition I we have, for $\overline{f}(m_n)\, f(m_n) \ge F^{-1}(1 - \frac{m_n}{n})$,

$$\left(\frac{n}{\sqrt{m_n}\,\alpha_n}\, \overline{F}(\overline{f}(m_n)\, f(m_n))\right) \left|\frac{\widehat{\overline{F}}(\overline{f}(m_n)\, f(m_n)) - \overline{F}(\overline{f}(m_n)\, f(m_n))}{\overline{F}(\overline{f}(m_n)\, f(m_n))}\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0. \tag{8.6}$$

To prove (8.5), starting from (8.6), it is sufficient to choice α_n such that for n large enough, there exists $c > 0$ such that

$$0 < \frac{\sqrt{m_n}\,\alpha_n}{n\,\overline{F}(\overline{f}(m_n)\, f(m_n))} \le c. \tag{8.7}$$

Recalling that, in the Fréchet case $L(x) = x^{\alpha}\overline{F}(x)$, for $\alpha > 0$, and $\frac{L(tx)}{L(x)} = 1 + k(t)\phi(x) + o(\phi(x))$, $\forall\, t > 0$, as $x \to \infty$. Using the assumptions of Theorem H we obtain

$$\begin{aligned}\frac{\overline{F}(\overline{f}(m_n)\, f(m_n))}{\overline{F}(\overline{f}(m_n))} &= \frac{L(\overline{f}(m_n)\, f(m_n))\, f(m_n)^{-\alpha}}{L(\overline{f}(m_n))} \\ &= f(m_n)^{-\alpha}\left[1 + k(f(m_n))\phi(\overline{f}(m_n)) + o(\phi(\overline{f}(m_n)))\right].\end{aligned}$$

So we have

$$\begin{aligned}&\frac{n\,\overline{F}(\overline{f}(m_n)\, f(m_n))}{\sqrt{m_n}} = \frac{n}{\sqrt{m_n}}\, \overline{F}\left(\overline{f}(m_n)\right)\big[f(m_n)^{-\alpha}\big(1 + k(f(m_n))\phi(\overline{f}(m_n)) \\ &+ o(\phi(\overline{f}(m_n)))\big)\big] \approx \sqrt{m_n} f(m_n)^{-\alpha}\left(1 + k(f(m_n))\phi(\overline{f}(m_n)) + o(\phi(\overline{f}(m_n)))\right).\end{aligned} \tag{8.8}$$

Then, recalling (5.8) and the properties of k and ϕ we prove that the right hand side of (8.8) increases to infinity hence one can choose α_n satisfying (8.7). It concludes the proof. □

Proof [Theorem 9]:
To prove (6.1) we divide the problem in the study of four different uniform convergences in probability. So, using (6.2)-(6.3), we want to prove

1. $\sup_{x>\overline{f}_1(n),\, y>\overline{f}_2(n)} \Big| \overline{F}(\overline{f}_1(n),\overline{f}_2(n)) \cdot F_u(x-\overline{f}_1(n), y-\overline{f}_2(n)) - \widehat{\overline{F}}(\overline{f}_1(n),\overline{f}_2(n)) \cdot$
$C^{*\,G}\Big(1-g(1-V_{\widehat{k}_X,\widehat{\sigma}_X}(x-\overline{f}_1(n))), 1-g(1-V_{\widehat{k}_Y,\widehat{\sigma}_Y}(y-\overline{f}_2(n)))\Big)\Big| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$
2. $\sup_{y>\overline{f}_2(n)} \Big| F(\overline{f}_1(n), y) - \widehat{F}_1^*(\overline{f}_1(n), y)\Big| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$
3. $\sup_{x>\overline{f}_1(n)} \Big| F(x, \overline{f}_2(n)) - \widehat{F}_2^*(x, \overline{f}_2(n))\Big| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$
4. $\Big| \frac{1}{n}\sum_{i=1}^n 1_{\{X_i\le\overline{f}_1(n),\, Y_i\le\overline{f}_2(n)\}} - F(\overline{f}_1(n),\overline{f}_2(n))\Big| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$

Convergence 1: We remark that, in our case (the Fréchet case), the endpoint $x_F = +\infty$. To prove convergence 1 we will use Proposition 12 and the uniform convergence (almost surely or in probability) for the bivariate empirical estimator $\widehat{F}(x,y)$ introduced in Theorem J (see below).
If we call, for $x > \overline{f}_1(n)$ and $y > \overline{f}_2(n)$

$A := \frac{1}{n}\sum_{i=1}^n 1_{\{X_i>\overline{f}_1(n),\, Y_i>\overline{f}_2(n)\}}$,

$B := C^{*\,G}\big(1-g(1-V_{\widehat{k}_X,\widehat{\sigma}_X}(x-\overline{f}_1(n))), 1-g(1-V_{\widehat{k}_Y,\widehat{\sigma}_Y}(y-\overline{f}_2(n)))\big)$;

$a := \overline{F}(\overline{f}_1(n),\overline{f}_2(n))$ and $b := F_u(x-\overline{f}_1(n), y-\overline{f}_2(n))$.

So by uniform convergences of Theorem 1, Proposition 12 and Theorem J we have that uniformly

$$|A-a| \xrightarrow[n\to\infty]{\mathbb{P}} 0;\quad |B-b| \xrightarrow[n\to\infty]{\mathbb{P}} 0;\quad |AB-ab| \le |a(B-b)| + |B(A-a)| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

So we have proved the first uniform convergence.
Convergence 2: By definition of copula we know that $F(x,y) := C(F_X(x), F_Y(y))$

so $F(\overline{f}_1(n), y) := C(F_X(\overline{f}_1(n)), F_Y(y))$, where C is a continuous function.
If we replace $C(F_X(\overline{f}_1(n)), F_Y(y))$ with $C\big(\widehat{F}_X(\overline{f}_1(n)), \widehat{F}^*_Y(y)\big)$, then we can use to prove convergence 2 the uniform convergence of $\widehat{F}_X(x)$ to $F_X(x)$ (univariate Glivenko-Cantelli Theorem)

$$\sup_x \left|\widehat{F}_X(x) - F_X(x)\right| = \left|\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \le x\}} - F_X(x)\right| \xrightarrow[n\to\infty]{(a.s.)} 0, \tag{8.9}$$

and the uniform convergence of Theorem 8

$$\sup_{y > \overline{f}_2(n)} \left|F_Y(y) - \widehat{F}^*_Y(y)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Finally we know that convergence in probability is preserved under continuous transformations and the copula C is continuous, by assumption, on a compact set, then C is uniformly continuous. We have proved the convergence 2.
Convergence 3: The structure of this proof is equal to convergence 2.
Convergence 4: To prove the convergence 4 for the empirical distribution function $\widehat{F}(x, y)$ to $F(x, y)$ we use Theorem J below. This theorem shows that, under suitable assumptions, $\widehat{F}(x, y)$ converges almost surely to $F(x, y)$. So we can obtain the desired convergence in probability:

$$\sup_{(x,\,y)} \left|\widehat{F}(x, y) - F(x, y)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

With these four uniform convergences we have proved (6.1). □

Auxiliary results

Proposition 12 *Let $V_{k,\sigma}(x)$ the Generalized Pareto Distribution (GPD), defined by (1.1), and $\widehat{k}_N$, $\widehat{\sigma}_N$, the maximum likelihood estimators of the parameters $k = -\alpha^{-1} < 0$ and $\sigma = u_n\alpha^{-1}$, in the case unconditionally on N. If all the conditions of Corollary 2 hold then*

$$\sup_{x\,\in[0,+\infty)} \left|V_{\widehat{k}_N,\widehat{\sigma}_N}(x) - V_{k,\sigma}(x)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Proof: From Corollary 2, we obtain

$$(\widehat{\sigma}_N - \sigma) \xrightarrow[n\to\infty]{\mathbb{P}} 0, \quad \left(\widehat{k}_N - k\right) \xrightarrow[n\to\infty]{\mathbb{P}} 0. \tag{8.10}$$

So, from (8.10) and with $g(k,\sigma,x) = V_{k,\sigma}(x)$ we have the following pointwise convergence in probability for each point $x \in [0,+\infty)$

$$\left[V_{\widehat{k}_N,\widehat{\sigma}_N}(x) - V_{k,\sigma}(x)\right] = \left[\left(1-\frac{k\,x}{\sigma}\right)^{\frac{1}{k}} - \left(1-\frac{\widehat{k}_N\,x}{\widehat{\sigma}_N}\right)^{\frac{1}{\widehat{k}_N}}\right] \xrightarrow[n\to\infty]{\mathbb{P}} 0. \qquad (8.11)$$

We remark that the random variable $T := \frac{\sigma(U^{-k}-1)}{k} \sim V_{k,\sigma}(x)$, with $U \sim$ Uniform$[0,1]$. Finally, we apply the Polya's Theorem (Rao (1965), Theorem (vi), pag. 120) and the general ideas of Horowitz (2001) and we obtain that pointwise convergence (almost surely or in probability) of a sequence of distribution functions to a continuous function implies that the sequence converges (almost surely or in probability) to this function uniformly. Since the limit $V_{k,\sigma}(x)$ is a continuous distribution function it follows that the convergence in (8.11) can be strengthened to uniform convergence (about this, see for example Embrechts, Kluppelberg and Mikosch (1997), pag. 552). □

Extension of Glivenko-Cantelli Theorem

To prove Theorem 9, we need a preliminary result of uniform convergence (almost surely or in probability) for the bivariate empirical estimator $\widehat{F}(x,y)$, which represents an extension of the Glivenko-Cantelli Theorem in higher dimensions.

Theorem J *(Blum (1955), Theorem 1) If there exists*

1. *an absolutely continuous probability measure μ on the Borel sets of two dimensional Euclidean space $\mathbb{R}^2$ and*
2. *a sequence of two-dimensional independent random vectors $\{\overline{X}_n\} := \{(X_1,Y_1), (X_2,Y_2), \ldots, (X_n,Y_n), \ldots\}$, distributed according of μ*

$$\textit{then} \qquad \mathbb{P}\left[\lim_{n\to\infty} \sup_{(x,y)} \left|\widehat{F}(x,y) - F(x,y)\right| = 0\right] = 1. \qquad (8.12)$$

DeHardt (1970) constructs a counterexample to show that, without the assumption of absolute continuity of the measure μ, the convergence (8.12) may not hold.

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

References

Blum, J.R. (1955). On the convergence of empiric distribution functions. *Ann. Math. Statist.* **26**, 527-529.

Cox, D.R. and Hinkley, D.V. (1974). Theoretical Statistics. Chapman and Hall, London.

Cherubini, U. and Luciano, E. and Vecchiato, W. (2004). Copula Methods in Finance. Wiley, New York.

DeHardt, J. (1970). A necessary condition for Glivenko-Cantelli convergence in E_n, *The Annals of Mathematical Statistics 41*, **6**, 2177-2178.

Dekkers, A.L.M. and de Haan, L. (1989). On the Estimation of the Extreme-Value Index and Large Quantile Estimation, *Ann. Statist.* **4**, 1795-1832.

Einmhal, J.H.J. (1990). The empirical distribution function as a tail estimator, *Statistica Neerlandica 44*, **2**, 79-82.

Embrechts, P. and Kluppelberg, C. and Mikosch, T. (1997). Modelling Extremal Events for Insurance and Finance. Berlin, Springer, Heidelberg.

Horowitz, J. L. (2001). The Bootstrap. In *Handbook of Econometrics*, (Ed. by J.J. Heckman and E.E. Leamer), 3159-3228, Elsevier Science B.V.

Juri, A. and Wüthrich, M.V. (2003). Tail Dependence From a Distributional Point of View, *Extremes 6*, **3**, 213-246.

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

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

McNeil, A. J. and Frey, R. and Embrechts, P. (2005). Quantitative risk management. Princeton Series in Finance.

Nelsen, R.B. (1999). An Introduction to Copulas. Springer, New York.

Rao, C.R. (1965). Linear Statistical Inference and Its Applications. Wiley, New York.

Smith, R. (1987). Estimating tails of probability distributions, *Ann. Statist.* **15**, 1174-1207.

Wüthrich, M. V. (2004). Bivariate extension of the Pickands-Balkema-de Hann Theorem, *Ann. Inst. H. Poincare Probab. Statist. 40*, **1**, 33-41.

E-mail: elena.di-bernardino@univ-lyon1.fr
Université de Lyon, Université Lyon 1, ISFA, Laboratoire SAF
50 avenue Tony Garnier, 69366 Lyon, France
Tel.: +33 4 37 28 74 46, Fax: +33 4 37 28 76 32

E-mail: veronique.maume@univ-lyon1.fr
Université de Lyon, Université Lyon 1, ISFA, Laboratoire SAF
50 avenue Tony Garnier, 69366 Lyon, France
Tel.: +33 4 37 28 74 28, Fax: +33 4 37 28 76 32

E-mail: clementine.prieur@imag.fr
Université Joseph Fourier,
Tour IRMA, MOISE-LJK B.P. 53 38041 Grenoble, France
Tel.: +33 4 76 63 59 63, Fax : +33 4 76 63 12 63