



HAL
open science

Estimating Bivariate Tail

Elena Di Bernardino, Véronique Maume-Deschamps, Clémentine Prieur

► **To cite this version:**

Elena Di Bernardino, Véronique Maume-Deschamps, Clémentine Prieur. Estimating Bivariate Tail. 2010. hal-00475386v2

HAL Id: hal-00475386

<https://hal.science/hal-00475386v2>

Preprint submitted on 11 Oct 2010 (v2), last revised 25 May 2011 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimating Bivariate Tail

Elena Di Bernardino · Véronique
Maume-Deschamps · Clémentine Prieur

the date of receipt and acceptance should be inserted later

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 Haan Theorem. We introduce a new parameter that describes the nature of the tail dependence and we provide a way to estimate it. We construct a two-dimensional tail estimator and study its asymptotic properties. We also present real data examples which illustrate our theoretical results.

Keywords Extreme Value Theory · Peaks Over Threshold method · Pickands-Balkema-de Haan Theorem · Tail dependence.

Mathematics Subject Classification (2000) 62H12 · 62H05 · 60G70.

1 Introduction

The univariate POT (Peaks-Over-Threshold) method is common for estimating extreme quantiles or tail distributions (see e.g. McNeil 1997, 1999 and references therein). A key idea of this method is that a distribution is in the

Elena Di Bernardino

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: elena.di-bernardino@univ-lyon1.fr

Véronique Maume-Deschamps

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: veronique.maume@univ-lyon1.fr

Clémentine Prieur

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
E-mail: clementine.prieur@imag.fr

domain of attraction of an extreme value distribution if and only if the distribution of excesses over high thresholds is asymptotically generalized Pareto (GPD) (e.g. Balkema and de Haan, 1974, Pickands, 1975):

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

and $x \geq 0$ for $\xi \leq 0$ or $0 \leq x < \frac{\sigma}{\xi}$ for $\xi > 0$. This univariate modelling is well understood, and has been discussed by Davison (1984), Davison and Smith (1990) and other papers of these authors.

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, mainly based on a version of the Pickands-Balkema-de Haan Theorem in dimension 2 (Theorem 1). This extension is general in the sense that it allows to consider a two-dimensional structure of dependence between both continuous random components X and Y . The modelling of this dependence is done via a continuous symmetric copula C , which is supposed to be unknown, with regularity properties specified in the statement of our theorems. An overview of multivariate-threshold methods based on multivariate Generalized Pareto distribution can be found e.g. in Tajvidi (1996).

We recall here some classical bivariate threshold models, based on a characterization of the joint tail by Resnick (1987). Let F denote the joint distribution of (Y_1, Y_2) with marginals F_j , $j = 1, 2$. Define $Z_j = -1/\log(F_j(Y_j))$, $j = 1, 2$, i.e. each Y_j is transformed to a unit Fréchet variable and $\mathbb{P}(Z_j \leq z) = \exp^{-1/z}$, for $0 < z < \infty$. Let F_* denote the joint distribution of (Z_1, Z_2) , we have $F(y_1, y_2) = F_*(z_1, z_2)$. The assumption F is in the maximum domain of attraction (MDA) of a bivariate extreme value distribution G is equivalent to F_* being in the domain of attraction of a bivariate extreme value distribution G_* , where the marginals of G_* are unit Fréchet. The characterization of Resnick (1987) can be written as

$$\lim_{t \rightarrow \infty} \frac{\log(F_*(tz_1, tz_2))}{\log(F_*(t, t))} = \lim_{t \rightarrow \infty} \frac{1 - F_*(tz_1, tz_2)}{1 - F_*(t, t)} = \frac{\log(G_*(z_1, z_2))}{\log(G_*(1, 1))}. \quad (2)$$

Equating the left and the right-hand terms for large t leads to the following model for the joint tail of F (see Ledford and Tawn, 1996):

$$\mathcal{F}_1(y_1, y_2) = \exp\{-l(-\ln(F_{Y_1}(y_1)), -\ln(F_{Y_2}(y_2)))\}, \quad (3)$$

for $y_j > u_j$, where u_j are high thresholds for the marginal distributions and with l the stable tail dependence function of the limiting extreme value distribution G_* . Furthermore the first-order expansion of (3) gives

$$\mathcal{F}_2(y_1, y_2) = 1 - l(1 - F_{Y_1}(y_1), 1 - F_{Y_2}(y_2)), \quad (4)$$

for $1 - F_{Y_j}(y_j)$ sufficiently small. However (3) is preferable over (4) because $1 - l(\cdot, \cdot)$ undervalues the probability of joint exceedances (for details see e.g. Beirlant *et al.*, 2004, Section 8).

Then approximations (3)-(4) can be estimated by

$$\widehat{\mathcal{F}}_1^*(y_1, y_2) = \exp\{-\widehat{l}(-\ln(\widehat{F}_{Y_1}^*(y_1)), -\ln(\widehat{F}_{Y_2}^*(y_2)))\}, \quad (5)$$

$$\widehat{\mathcal{F}}_2^*(y_1, y_2) := 1 - \widehat{l}(1 - \widehat{F}_{Y_1}^*(y_1), 1 - \widehat{F}_{Y_2}^*(y_2)), \quad (6)$$

for high values of y_1 and y_2 , where $\widehat{F}_{Y_1}^*(y_1)$ (resp. $\widehat{F}_{Y_2}^*(y_2)$) is an estimator for the marginal tail of Y_1 (resp. Y_2). For instance $\widehat{F}_{Y_1}^*(y_1)$ (resp. $\widehat{F}_{Y_2}^*(y_2)$) comes from the univariate POT method described in Section 3. In (5)-(6) \widehat{l} is an estimator of the stable tail dependence function (see Drees and Huang, 1998, Draisma *et al.*, 2004, Einmahl *et al.*, 2008). For another approach, based on the estimation of the so-called univariate dependence function of Pickands (Pickands, 1981) see for instance Capéraà and Fougères (2000). Problems arise with both these bivariate techniques when (Y_1, Y_2) are asymptotically independent i.e.

$$\lambda := \lim_{t \rightarrow 0} \mathbb{P}[F_{Y_1}^{-1}(Y_1) > 1 - t \mid F_{Y_2}^{-1}(Y_2) > 1 - t] = 0. \quad (7)$$

When the data exhibit positive or negative association that only gradually disappears at more and more extreme levels these methods produce a significant bias. Basically this happens because a bivariate extreme value type dependence structure is assumed to hold in the joint tail above the marginal thresholds.

In order to overcome this problem, Ledford and Tawn (1996, 1997, 1998) introduced a model in which the tail dependence is characterized by a coefficient $\eta \in (0, 1]$. In these works the joint survival distribution function of a bivariate random vector (Z_1, Z_2) with unit Fréchet marginals is assumed to satisfy $\mathbb{P}[Z_1 > z, Z_2 > z] \sim L(z)\mathbb{P}[Z_1 > z]^{\frac{1}{\eta}}$, where L is a slowly varying function at infinity. Several methods to estimate this coefficient η are proposed in Peng (1999), Draisma *et al.* (2004), Beirlant *et al.* (2010). For some counter examples of the Ledford and Tawn's model see Schlather (2001).

Contrary to this approach, we will propose a model based on regularity conditions of the copula and on the explicit description of the dependence structure in the joint tail (see condition (11) in Theorem A of Juri and Wüthrich, 2003). In our model the structure of the tail dependence will be described by a parameter θ coming from Theorem A. The general idea is to decompose the estimation of $\mathbb{P}(X \leq x, Y \leq y)$, for x, y above some marginal thresholds u_X, u_Y , in the estimation of different bivariate regions. For the joint upper tail in $[u_X, x] \times [u_Y, y]$ we will use the non parametric estimators coming from Theorem 1 (see Section 4). For the *lateral regions* $[-\infty, x] \times [u_Y, y]$ and $[-\infty, u_X] \times [-\infty, y]$ we will approximate the distribution function F using (3).

The stability of our estimation compared to the one of $\widehat{\mathcal{F}}_1^*$ is analyzed in some real cases (Section 8) which have been studied in some classical works (e.g. Coles and Tawn, 1994, Ledford and Tawn 1997, Ramos and Ledford, 2009, Lescourret and Robert, 2006). Therefore our estimator, in a different way from the Ledford and Tawn's method, covers situations less restrictive than dependence or perfect independence above the thresholds.

Note also that our method is free from the pre-treatment of data because we can work directly with the original general samples without the transformation in Fréchet marginal distributions.

The paper is organized as follows. We introduce an extension of the Pickands-Balkema-de Haan Theorem in dimension 2 in the case of bivariate distributions with different marginals (Theorem 1). In Section 4 we provide a new non parametric estimator for the dependence structure of a bivariate random sample in the upper tail. In Section 5 we use Theorem 1 in order to build a new estimator for the tail of the bivariate distribution. The study of the asymptotic properties of our estimator makes use of a new convergence result in the univariate case (Theorem 5 of Section 6) dealing with asymptotic behavior of the absolute error between the theoretical distribution function and its tail estimator. Our main result is the gaussian approximation of this error with its speed of convergence (Theorem 6). Examples with real data are presented in Section 8. Some auxiliary results and more technical proofs are postponed to the Appendix.

Remark 1 Assume we observe X_1, \dots, 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 $\lim_{n \rightarrow \infty} m_n = \infty$ and $\lim_{n \rightarrow \infty} \frac{m_n}{n} = 0$. 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 $\lim_{n \rightarrow \infty} 1 - F(u_n) = 0$ and $\lim_{n \rightarrow \infty} n(1 - F(u_n)) = \infty$.

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

A central one dimensional result in tail estimation is the so-called Pickand-Balkema-de Haan Theorem. As our aim is the estimation of bivariate tails, we are interested in two-dimensional extensions of this theorem. Such a two dimensional generalization of this result can be found in the literature (Juri and Wüthrich 2003 and Wüthrich 2004) with the assumption $F_X = F_Y$. We provide here a precise formulation and proof of the Pickands-Balkema-de Haan Theorem when $F_X \neq F_Y$ (Theorem 1 below). We first introduce some notations and recall results from Juri and Wüthrich (2003), McNeil *et al.* (2005) and Nelsen (1999), which we will need later.

Definition 1 [*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 $[0, 1]$.

Definition 2 [*Survival Copula*] The Survival copula C^* of a copula C is defined

$$\text{by: } \forall (u_1, u_2) \in [0, 1]^2, \quad C^*(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2). \quad (8)$$

A Survival copula (8) 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 ,

$$\text{we have: } C^*(1 - F_X(x), 1 - F_Y(y)) = \mathbb{P}[X > x, Y > y], \quad \text{for } (x, y) \in \mathbb{R}^2. \quad (9)$$

Let us assume the copula associated with (X, Y) is continuous and symmetric and, in a first time, that X and Y are uniformly distributed on $[0, 1]$. 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 $\{X > u, Y > u\}$:

$$\forall x \in [0, 1], \quad \overline{F}_u(x) := \mathbb{P}[X \leq x | X > u, Y > u] = 1 - \frac{C^*(1 - x \vee u, 1 - u)}{C^*(1 - u, 1 - u)}. \quad (10)$$

Note that the continuity of the copula C implies that $\overline{F}_u(\cdot)$ is also continuous. Its symmetry implies that (10) also defines the distribution of Y conditioned on $\{X > u, Y > u\}$ (recall that X and Y are uniformly distributed).

Definition 3 [*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 (10).

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 . We remark 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 A (*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 $g : [0, \infty) \rightarrow [0, \infty)$ such that

$$\lim_{u \rightarrow 1} \frac{C^*(x(1 - u), 1 - u)}{C^*(1 - u, 1 - u)} = g(x), \quad x \in [0, \infty). \quad (11)$$

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

$$\lim_{u \rightarrow 1} C_u^{up}(x, y) = x + y - 1 + G(g^{-1}(1 - x), g^{-1}(1 - y)) := C^{*G}(x, y), \quad (12)$$

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

Remark 2

- From (8), we note that $C^{*G}(x, y)$ defined in (12) 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).
- In this work, the copula is assumed to be symmetric in the sense that $C(u, v) = C(v, u)$ (see e.g. Klement and Mesiar, 2006). This assumption is needed for the proof of Theorem A by Juri and Wüthrich (2003). Hence a preliminary task in the application is to test for the validity of this assumption. In Section 8 the symmetry of the copula seems reasonable. If the present results could be extended to the non symmetric setting is an interesting point but outside the scope of the present paper.

de Haan (1970) proves that $F \in MDA(H_\xi)$ is equivalent to the existence of a positive measurable function $a(\cdot)$ such that, for $1 - \xi x > 0$ and $\xi \in \mathbb{R}$,

$$\lim_{u \rightarrow x_F} \frac{1 - F(u + x a(u))}{1 - F(u)} = \begin{cases} (1 - \xi x)^{\frac{1}{\xi}}, & \text{if } \xi \neq 0, \\ e^{-x}, & \text{if } \xi = 0. \end{cases} \quad (13)$$

This result is useful to give below a rigorous formulation of the two-dimensional Pickands-Balkema-de Haan Theorem in a 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 $F_X \in MDA(H_{\xi_1})$, $F_Y \in MDA(H_{\xi_2})$ and that C satisfies the hypotheses of Theorem A for some g . Then*

$$\sup_{\mathcal{A}} \left| \mathbb{P}[X - u \leq x, Y - F_Y^{-1}(F_X(u)) \leq y | X > u, Y > F_Y^{-1}(F_X(u))] - C^{*G}(1 - g(1 - V_{\xi_1, a_1(u)}(x)), 1 - g(1 - V_{\xi_2, a_2(F_Y^{-1}(F_X(u)))}(y))) \right| \xrightarrow{u \rightarrow x_{F_X}} 0, \quad (14)$$

where $V_{\xi_i, a_i(\cdot)}$ is the GPD with parameters $\xi_i, a_i(\cdot)$ defined in (1), $a_i(\cdot)$ is as in (13), 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 $\mathcal{A} := \{(x, y) : 0 < x \leq x_{F_X} - u, 0 < y \leq x_{F_Y} - F_Y^{-1}(F_X(u))\}$.

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 (McNeil 1997, 1999). Fix a threshold u and define $F_u(x) = \mathbb{P}[X \leq x | X > u]$. Let X_1, X_2, \dots be a sequence of i.i.d random variables with unknown distribution function F and $\hat{F}_X(u)$ the empirical distribution

function evaluated at the threshold u . Recall that the univariate tail may be estimated by

$$\widehat{F}^*(x) = (1 - \widehat{F}_X(u))V_{\widehat{\xi}, \widehat{\sigma}}(x - u) + \widehat{F}_X(u), \quad \text{for } x > u, \quad (15)$$

where $\widehat{\xi}, \widehat{\sigma}$ are the maximum likelihood estimators (MLE) based on the excesses above u . Using (15) we get the estimator, proposed by Smith (1987)

$$1 - \widehat{F}^*(y) = \begin{cases} \frac{N}{n} \left(1 - \widehat{\xi} \frac{(y-u)}{\widehat{\sigma}}\right)^{\frac{1}{\widehat{\xi}}}, & \text{if } \widehat{\xi} \neq 0, \\ \frac{N}{n} \left(e^{-\frac{(y-u)}{\widehat{\sigma}}}\right), & \text{if } \widehat{\xi} = 0, \end{cases} \quad (16)$$

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

4 Tail dependence in the bivariate framework

As well known a bivariate distribution function F with continuous marginal distribution functions F_X, F_Y is said to have a *stable tail dependence function* l if for $x \geq 0$ and $y \geq 0$ the following limit exists:

$$\lim_{t \rightarrow 0} \frac{1}{t} \mathbb{P}[1 - F_X(X) \leq tx \text{ or } 1 - F_Y(Y) \leq ty] := l(x, y) \quad (17)$$

or similarly

$$\lim_{t \rightarrow 0} \frac{1}{t} \mathbb{P}[1 - F_X(X) \leq tx, 1 - F_Y(Y) \leq ty] := R(x, y) = x + y - l(x, y),$$

see Huang (1992). If F_X, F_Y are in the maximum domain of attraction of two extreme value distributions H_X, H_Y and if (17) holds then F is in the domain of attraction of an extreme value distribution H with marginals H_X, H_Y and with copula C determined by l because (17) is equivalent to

$$\lim_{t \rightarrow 0} \frac{1}{t} (1 - C(1 - tx, 1 - ty)) = l(x, y), \quad \text{for } x \geq 0, y \geq 0. \quad (18)$$

Note that the upper tail dependence coefficient defined in (7) is such that $\lambda = R(1, 1)$.

4.1 Asymptotic dependent case

If X and Y are asymptotically dependent ($\lambda > 0$) we can introduce an estimator for $g(x)$ which will be used later to estimate the tail of the bivariate distribution function. Using (17)-(18), we write (11) and (12) as

$$g(x) = \frac{x + 1 - l(x, 1)}{2 - l(1, 1)} = \frac{R(x, 1)}{R(1, 1)}, \quad G(x, y) = \frac{x + y - l(x, y)}{2 - l(1, 1)} = \frac{R(x, y)}{R(1, 1)},$$

and in this case $\theta = 1$. As $\eta \in (0, 1]$ in the Ledford and Tawn's model (see Ledford and Tawn 1996, 1997, 1998), θ describes the nature of the tail dependence, it does not depend on the marginal distribution functions.

In order to estimate g and G , we can use the non parametric estimator of l (see Einmahl *et al.*, 2008):

$$\widehat{l}(x, y) = \frac{1}{k_n} \sum_{i=1}^n 1_{\{R(X_i) > n - k_n x + 1 \text{ or } R(Y_i) > n - k_n y + 1\}}, \quad (19)$$

where $k_n \rightarrow \infty$, $\frac{k_n}{n} \rightarrow 0$ and $R(X_i)$ is the rank of X_i among (X_1, \dots, X_n) , $R(Y_i)$ is the rank of Y_i among (Y_1, \dots, Y_n) , $i = 1, \dots, n$.

We obtain

$$\widehat{g}(x) = \frac{\sum_{i=1}^n \frac{1}{k_n} 1_{\{R(X_i) > n - k_n x + 1; R(Y_i) > n - k_n + 1\}}}{\sum_{i=1}^n \frac{1}{k_n} 1_{\{R(X_i) > n - k_n + 1; R(Y_i) > n - k_n + 1\}}}, \quad (20)$$

$$\widehat{G}(x, y) = \frac{\sum_{i=1}^n \frac{1}{k_n} 1_{\{R(X_i) > n - k_n x + 1; R(Y_i) > n - k_n y + 1\}}}{\sum_{i=1}^n \frac{1}{k_n} 1_{\{R(X_i) > n - k_n + 1; R(Y_i) > n - k_n + 1\}}}. \quad (21)$$

Using (20)-(21) we get for any $x > 0$, $x \neq 1$ a non parametric estimator of θ :

$$\widehat{\theta}_x = \frac{\log(\widehat{g}(x)) - \log(\widehat{g}(\frac{1}{x}))}{\log(x)}. \quad (22)$$

We observe the sensitivity of $\widehat{\theta}_x$ to the sequence k_n for different copulas: Survival Clayton and Logistic copulas (Figure 1). We draw the mean curve on 100 samples of size $n = 1000$ (full line) and the empirical standard deviation (dashed lines). On simulations it seemed to us that for each value of x one could exhibit a range of values of k_n under which our estimate well behaved. In the following we fixed x for each simulation and made vary k_n . The choice of k_n does not appear to be crucial for $\widehat{\theta}_x$. In Figure 2 the mean squared error for $\widehat{\theta}_x$ is calculated on 100 samples of size $n = 1000$.

4.2 Asymptotic independent case

We say that X and Y are asymptotically independent if $\lambda = R(1, 1) = 0$. In term of copula this means that $C(u, u) = 1 - 2(1 - u) + o(1 - u)$, for $u \rightarrow 1$. In that case we deduce Proposition 2 below.

Proposition 2 *If $\lambda = 0$ and C is a twice continuously differentiable symmetric copula with the determinant of the Hessian matrix of C at $(1, 1)$ different to zero, then*

$$\lim_{u \rightarrow 1} \frac{C^*(x(1-u), 1-u)}{C^*(1-u, 1-u)} = \frac{b/2(x^2 + 1) + cx}{b + c} \quad \forall x \in [0, \infty),$$

and $\theta = 2$, with $b = \frac{\partial^2 C}{\partial u^2}(1, 1)$, $c = \frac{\partial C}{\partial u \partial v}(1, 1)$.

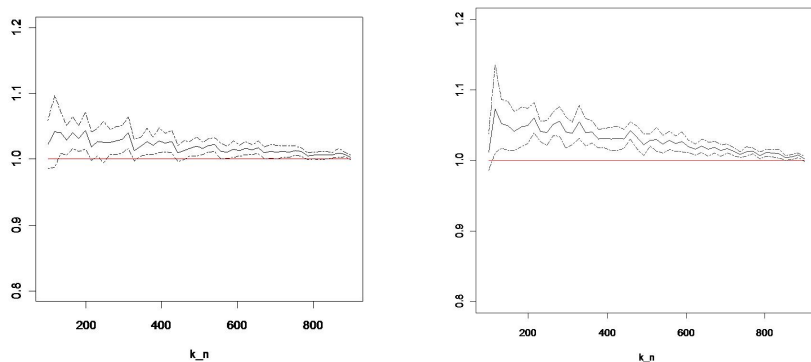


Figure 1 Estimators for $\theta, (k, \hat{\theta}_x)$ (left) $x = 0.07$, Survival Clayton copula with parameter 1 (right) $x = 5$, Logistic copula with parameter 0.5

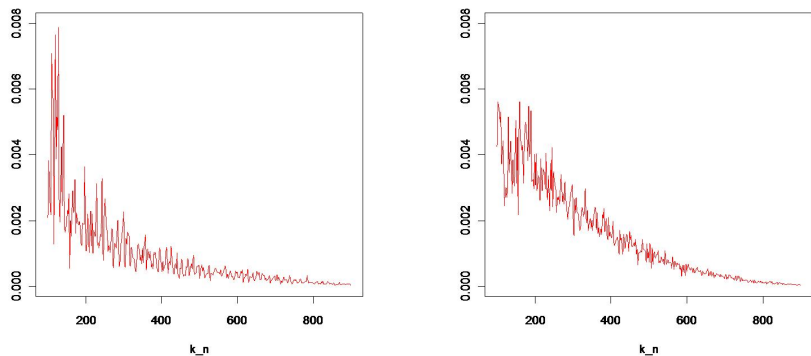


Figure 2 Mean squared error for $\hat{\theta}_x$ (left) $x = 0.07$, Survival Clayton copula with parameter 1 (right) $x = 5$, Logistic copula with parameter 0.5

Details of the proof will be omitted here. The main ingredient is the second-order development of copula C .

The assumptions of Proposition 2 are satisfied, for example, in the case of Ali Mikhail-Haq, Frank, Clayton with $a \geq 0$, Independent and Fairlie-Gumbel-Morgenstern copulas. Figure 3 show the mean curve on 100 samples of size $n = 1000$ (full line) and the empirical standard deviation (dashed lines) in the case of Independent and Clayton copulas. In Figure 4 the mean squared error for $\hat{\theta}$ are calculated in 100 samples.

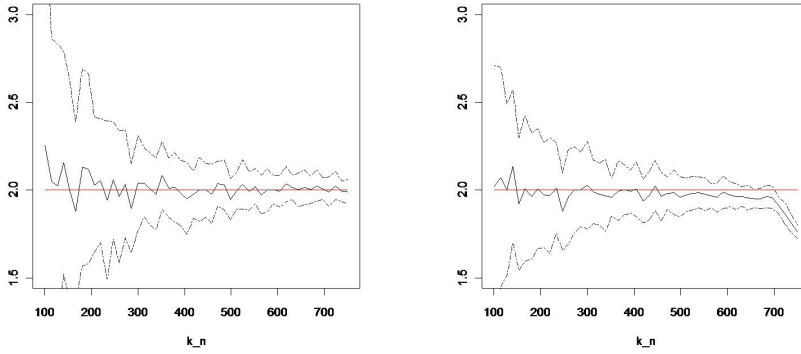


Figure 3 Estimators for θ , $(k, \hat{\theta}_x)$ (left) $x = 0.8$, Independent copula (right) $x = 0.7$, Clayton copula with parameter 0.05.

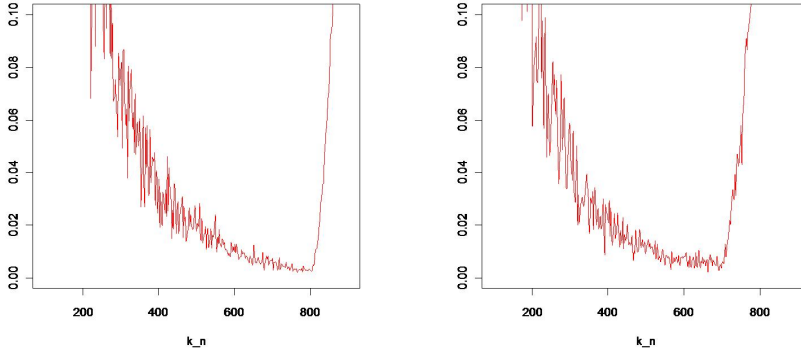


Figure 4 Mean squared error for $\hat{\theta}_x$ (left) $x = 0.8$, Independent copula (right) $x = 0.7$, Clayton copula with parameter 0.05.

5 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 7.

This construction generalizes the one-dimensional construction stated in Section 3. Let X and Y be two real valued random variables with different continuous marginal distributions F_X and F_Y . The structure of dependence between X and Y is represented by some continuous and symmetric copula C . For $x > u$, $y > F_Y^{-1}(F_X(u)) := u_Y$ we define

$$\begin{aligned} \widehat{F}^*(x, y) &= \left(\frac{1}{n} \sum_{i=1}^n 1_{\{X_i > u, Y_i > u_Y\}} \right) (1 - \widehat{g}(1 - V_{\widehat{\xi}_X, \widehat{\sigma}_X}(x - u)) \\ &- \widehat{g}(1 - V_{\widehat{\xi}_Y, \widehat{\sigma}_Y}(y - u_Y)) + \widehat{G}(1 - V_{\widehat{\xi}_X, \widehat{\sigma}_X}(x - u), 1 - V_{\widehat{\xi}_Y, \widehat{\sigma}_Y}(y - u_Y))) \\ &+ \widehat{F}_1^*(u, y) + \widehat{F}_2^*(x, u_Y) - \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq u, Y_i \leq u_Y\}}, \quad (23) \end{aligned}$$

where $\widehat{\xi}_X, \widehat{\sigma}_X$ (resp. $\widehat{\xi}_Y, \widehat{\sigma}_Y$) are MLE based on the excesses of X (resp. Y).

Description of the construction: Given a high threshold u and $u_Y := F_Y^{-1}(F_X(u))$, we introduce the distribution of excesses: $F_u(x, y) := \mathbb{P}[X - u \leq x, Y - u_Y \leq y | X > u, Y > u_Y]$. Using (3) for large value of u and $x > u, y > u_Y$, we can approximate $F(u, y)$ and $F(x, u_Y)$ as

$$F_1^*(u, y) = e^{\{-l(-\lg(F(u)), -\lg(F(y)))\}}, \quad F_2^*(x, u_Y) = e^{\{-l(-\lg(F(x)), -\lg(F(u_Y)))\}}, \quad (24)$$

where l is the stable tail dependence function defined by (17). We recall that behind approximation (24) we suppose that the data structure is characterized by dependence (or perfect independence) in the lateral regions $[-\infty, x] \times [-\infty, u_Y]$ and $[-\infty, u_X] \times [-\infty, y]$.

From Theorem 1 we now know that, for u around x_F , we can approximate the distribution of excesses with C^*G . So we obtain, for $x > u, y > u_Y$,

$$\begin{aligned} F^*(x, y) &= (\overline{F}(u, u_Y)) \cdot C^*G(1 - g(1 - V_{\xi_X, \sigma_X}(x - u)), 1 - g(1 - V_{\xi_Y, \sigma_Y}(y - u_Y))) \\ &+ F_1^*(u, y) + F_2^*(x, u_Y) - F(u, u_Y). \quad (25) \end{aligned}$$

Then, we estimate $F(u, u_Y)$ and $\overline{F}(u, u_Y)$ in (25) from the data $\{X_i, Y_i\}_{i=1, \dots, n}$, using the empirical distribution estimates

$$\widehat{F}(u, u_Y) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq u, Y_i \leq u_Y\}}, \quad \widehat{\overline{F}}(u, u_Y) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i > u, Y_i > u_Y\}}. \quad (26)$$

From (24) and using the non parametric estimator (19) we obtain

$$\widehat{F}_1^*(u, y) = \exp\{-\widehat{l}(-\lg(\widehat{F}_X(u)), -\lg(\widehat{F}_Y^*(y)))\}, \quad (27)$$

$$\widehat{F}_2^*(x, u_Y) = \exp\{-\widehat{l}(-\lg(\widehat{F}_X^*(x)), -\lg(\widehat{F}_Y(u_Y)))\}, \quad (28)$$

where $\widehat{F}_X(u)$ 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, defined by (15). Now, using (26)-(28), we can approximate $F^*(x, y)$ for $x > u, y > u_Y$ and u large by (23). Our tail estimator, in the case with same marginal distributions, is a particular case of (23) with same threshold u for X and Y .

Remark 3 Note that $\widehat{F}^*(x, y)$, is only valid for $x > u$ and $y > u_Y = F_Y^{-1}(F_X(u))$, when u is *large enough*. The expression large enough is a fundamental problem of the POT method. 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 ξ_X, ξ_Y and σ_X, σ_Y will suffer from a lack of observations over the thresholds u, u_Y . The compromise will be explained in Section 6 and 7.

6 Convergence results in the univariate case

In order to study asymptotic properties of our bivariate tail estimator $\widehat{F}^*(x, y)$ we present in this section some extensions of one dimensional convergence results in Smith (1987; Theorems 3.2. and 8.1.). Incidentally we get asymptotic confidence intervals for the unknown theoretical function $F(x)$, using Theorem 5. 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 i.e. $\overline{F}(x) = x^{-\alpha}L(x)$ for some slowly varying function $L(x)$, $\alpha > 0$. As in Smith (1987), Section 3, we shall assume that L satisfies the following condition

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

where $\phi(x) > 0$ and $\phi(x) \rightarrow 0$ as $x \rightarrow \infty$. Let R_ρ be the set of ρ -regularly varying functions. Condition SR2 implies, excluding trivial cases, $\phi \in R_\rho$, for some $\rho \leq 0$, and $k(t) = ch_\rho(t)$, with $h_\rho(t) = \int_1^t u^{\rho-1}du$; (for more details see Smith (1987), Section 3). Theorems 3.2. and 8.1. in Smith (1987) are written conditionally on $N = m_n$, 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 analogue of Theorems 3.2. (*resp.* Theorem 8.1.), Corollary 3 (*resp.* Corollary 4).

Corollary 3 *Suppose L satisfies SR2. Let n be the sample size and $u_n := \overline{f}(n)$ the threshold, such that $\overline{f}(n) \xrightarrow[n \rightarrow \infty]{} \infty$. Let $N = N_n$ denote the random number of excesses of u_n . We define $\xi = -\alpha^{-1}$ and $\sigma_n = \overline{f}(n)\alpha^{-1}$. If*

$$n(1 - F(u_n)) \xrightarrow[n \rightarrow \infty]{} \infty, \quad (29)$$

$$\sqrt{n(1 - F(u_n))}c\phi(u_n) \xrightarrow[n \rightarrow \infty]{} \mu(\alpha - \rho), \quad (30)$$

then there exists, with probability 1, a local maximum $(\widehat{\sigma}_n, \widehat{\xi}_n)$ of the GPD log likelihood function, such that

$$\sqrt{N} \begin{pmatrix} \frac{\widehat{\sigma}_n}{\sigma_n} - 1 \\ \widehat{\xi}_n - \xi \end{pmatrix} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N} \left(\begin{pmatrix} \frac{\mu(1-\xi)(1+2\xi\rho)}{1-\xi+\xi\rho} \\ \frac{\mu(1-\xi)\xi(1+\rho)}{1-\xi+\xi\rho} \end{pmatrix}; \begin{pmatrix} 2(1-\xi) & (1-\xi) \\ (1-\xi) & (1-\xi)^2 \end{pmatrix} \right).$$

Proof: If (29) and (30) hold then $N(n(1 - F(u_n)))^{-1} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 1$, and (3.2) in Smith (1987) holds in probability, i.e.

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

Hence we conclude with a Skorohod-type construction of probability spaces on which (3.2) in Smith (1987) holds almost surely. \square

Corollary 4 *Suppose all the assumptions of Corollary 3 are satisfied. Let n be the sample size, $u_n := \bar{f}(n) \xrightarrow[n \rightarrow \infty]{} \infty$ and $z_n := f(n) \xrightarrow[n \rightarrow \infty]{} \infty$ such that $(z_n)^{-s\rho} \frac{\phi(u_n(z_n)^s)}{\phi(u_n)} \xrightarrow[n \rightarrow \infty]{} 1$, for $s \in [0, 1]$. Let $N = N_n$ denote the random number of excesses above u_n .*

$$\text{If } \frac{\lg(z_n)}{\sqrt{n(1 - F(u_n))}} \xrightarrow[n \rightarrow \infty]{} 0, \quad (31)$$

then

$$\frac{\sqrt{N}}{\lg(f(n))} \left[\frac{1 - \widehat{F}^*(\bar{f}(n) f(n))}{1 - F(\bar{f}(n) f(n))} - 1 \right] \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(\nu, \tau^2), \quad (32)$$

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

Proof: If (29), (30) and (31) hold, then (8.7), (8.8) and (8.11) in Smith (1987) hold in probability, i.e

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

We conclude as for Corollary 3. \square

Let us now give the central result of this section.

Theorem 5 *Assume that all the assumptions of Corollary 4 are satisfied. We use the same notation.*

$$\text{If } (z_n)^\alpha (n(1 - F(u_n)))^{-1/2} \xrightarrow[n \rightarrow \infty]{} 0, \text{ then} \quad (33)$$

$$\frac{\sqrt{N}}{\lg(f(n)) \widehat{F}_n(\bar{f}(n) f(n))} \left[F(\bar{f}(n) f(n)) - \widehat{F}^*(\bar{f}(n) f(n)) \right] \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(\nu, \tau^2), \quad (34)$$

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

The proof of Theorem 5 is postponed to the Appendix.

As a byproduct, it is possible to construct from (34) asymptotic confidence intervals for $F(\bar{f}(n) f(n))$.

7 Convergence results in the bivariate case

In this section we provide a convergence result for our two-dimensional tail estimator (23), using Theorem 5 above. Copula C is still assumed to be continuous and symmetric. We work unconditionally on N .

Remark 4 Let n be the sample size. We choose, from Theorem 1,

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

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

7.1 Asymptotic dependent case

Our main result is a gaussian approximation of the absolute error of our bivariate tail estimator (23). Its proof makes use of a result by Einmahl *et al.* (2006) (see Theorem C in the Appendix) which specifies the asymptotic behaviour of $\hat{l}(x, y)$ uniformly in $0 \leq x, y \leq 1$.

Theorem 6 *Suppose F_X and F_Y belong to the maximum domain of attraction of Fréchet, L_X, L_Y satisfy the condition SR2 and that exists stable tail dependence function l defined by (17). Assume that copula C is continuous, symmetric, $\lambda > 0$ and that the assumptions of Theorem C are satisfied. Then under assumptions of Theorem 1 and if sequences $f_1(n), f_2(n), u := \bar{f}_1(n), u_Y := \bar{f}_2(n)$, defined by Remark 4, satisfy the conditions of Theorem 5 then*

$$|\sqrt{k_n}(F^*(x_n, y_n) - \widehat{F}^*(x_n, y_n)) + B(x_n^*, y_n^*)| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0, \quad (35)$$

with $x_n = \bar{f}_1(n)f_1(n)$, $y_n = \bar{f}_2(n)f_2(n)$, $k_n \rightarrow \infty$, $\frac{k_n}{n} \rightarrow 0$, $k_n/N_X \xrightarrow{\mathbb{P}} 0$, $k_n/N_Y \xrightarrow{\mathbb{P}} 0$, $k_n = o(n^{\frac{2\alpha}{1+2\alpha}})$, $\alpha > 0$, $B(x_n^*, y_n^*) = B(-\lg(\widehat{F}_X(\bar{f}_1(n))), -\lg(\widehat{F}_Y^*(y_n))) + B(-\lg(\widehat{F}_X^*(x_n)), -\lg(\widehat{F}_Y(\bar{f}_2(n))))$, where $B(\cdot, \cdot)$ is the gaussian process defined in Theorem C.

The proof of Theorem 6 is postponed to the Appendix.

7.2 Asymptotic independent case

In the asymptotic independent case, the problem is that Equality $g(x) = \frac{R(x,1)}{R(1,1)}$ has no sense as $R(1,1) = \lambda = 0$. In order to overcome this difficulty, we need to introduce second order assumptions on copula C . More precisely we refine condition (11). Let us assume

$$\lim_{t \rightarrow 0} \frac{\frac{C^*(tx, ty)}{C^*(t, t)} - G(x, y)}{q_1(t)} := Q(x, y) \quad (36)$$

with $x, y \geq 0$, $x + y > 0$, where q_1 is some positive function and Q is neither a constant nor a multiple of G . Under (36) and using Lemma 6.1 in Draisma *et al.* (2004) we can prove a result of Gaussian approximation.

8 Real data

In this section we present three real cases (see Figure 5-8) for which we estimate bivariate tail probabilities to illustrate the finite sample properties of our estimator. The stability with respect to k_n of our estimation is compared to the one of \mathcal{F}_1^* (defined by (5)). In order to study these data we will follow some classical models which make use of a symmetric copula structure.

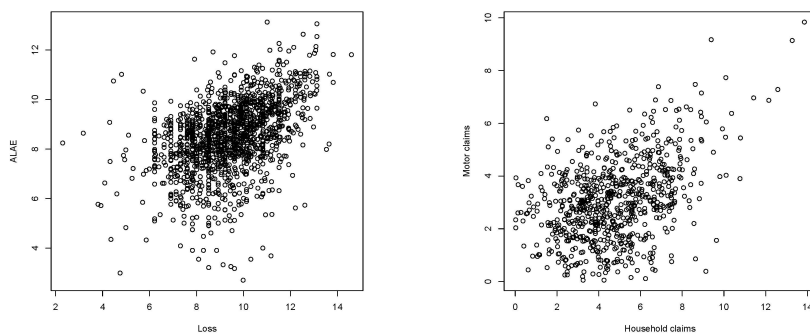


Figure 5 Logarithmic scale (left) ALAE versus Loss (right) Storm damages.

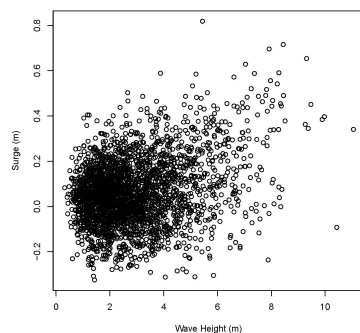


Figure 6 Wave Height (m) versus Surge (m).

We consider the **Loss-ALAE data** following the approach in Frees and Valdez (1998). Each claim consists of an indemnity payment (the loss, X) and

an allocated loss adjustment expense (ALAE, Y). We estimate $F(2.10^5, 10^5)$. The empirical probability (defined by (26)) is 0.9506667 and the survival empirical probability is 0.006 (see Beirlant *et al.*, 2010 for a comparison). Figure 7 shows the sensitivity of $\hat{\theta}$, \hat{F}^* with k_n .

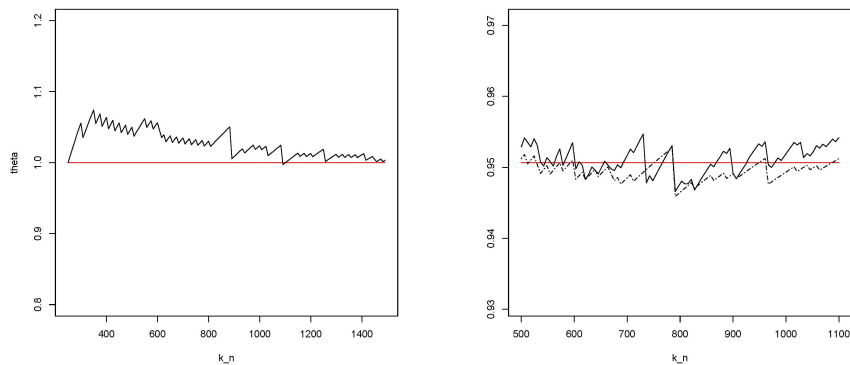


Figure 7 (left) $\hat{\theta}_{0.04}$ (right) $\hat{F}^*(2.10^5, 10^5)$ (full line), $\hat{\mathcal{F}}_1^*(2.10^5, 10^5)$ (dashed line), with the empirical probability indicated with a horizontal line.

Now we consider an **example from storm insurance**: aggregate claims of motor policies (Y) and aggregate claims of household policies (X) from a French insurance portfolio for 736 storm events (see Lescouret and Robert, 2006 for a detailed description). We estimate $F(8000, 950)$. The empirical probability is 0.96875 and the survival empirical probability is 0.014. The stability of our estimation compared to the one of $\hat{\mathcal{F}}_1^*$ is presented in Figure 8.

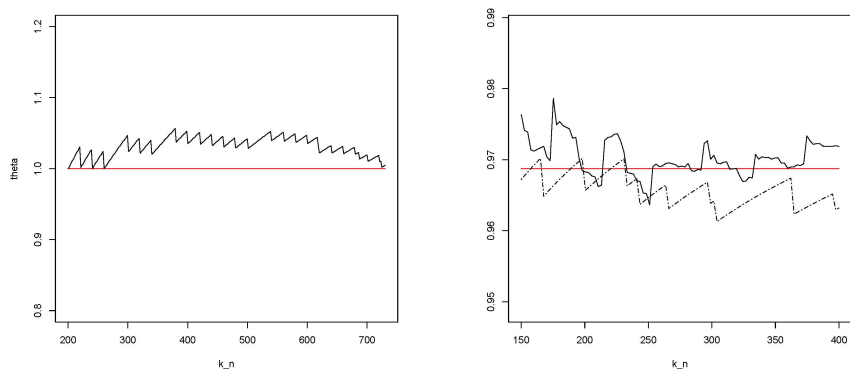


Figure 8 (left) $\hat{\theta}_{0.05}$ (right) $\hat{F}^*(8.10^3, 950)$ (full line), $\hat{\mathcal{F}}_1^*(8.10^3, 950)$ (dashed line), with the empirical probability indicated with a horizontal line.

We analyze the **wave surge data** comprising 2894 bivariate events that occurred during 1971 – 1977 in Cornwall (England) (for details see e.g. Coles and Tawn, 1994 or Ramos and Ledford, 2009). We estimate $F(8.32, 0.51)$. The empirical probability is 0.9903 and the survival empirical probability is 0.00069. The sensitivity of $\hat{\theta}$, \hat{F}^* with k_n is presented in Figure 9.

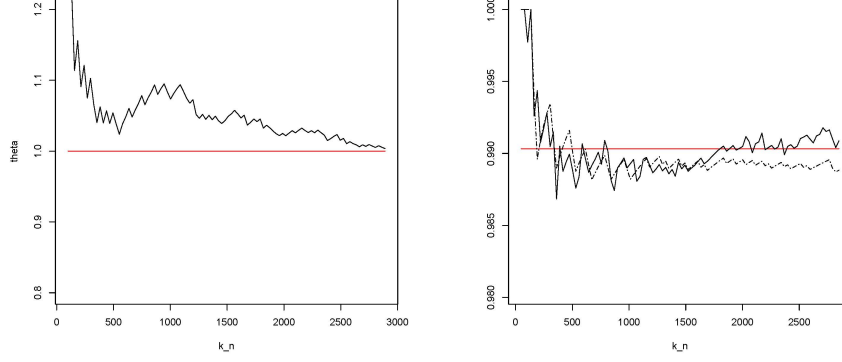


Figure 9 (left) $\hat{\theta}_{0,02}$ (right) $\hat{F}^*(8.32, 0.51)$ (full line), $\mathcal{F}_1^*(8.32, 0.51)$ (dashed line), with the empirical probability indicated with a horizontal line.

9 Appendix: proofs and auxiliary results

Proof [Theorem 1] : From (13) we obtain the existence of $a_1(\cdot)$ and $a_2(\cdot)$ such that, for $p := u + xa_1(u)$ and $q := u_Y + ya_2(u_Y)$

$$V_{\xi_{1,1}}(x) = \lim_{u \rightarrow x_{F_X}} 1 - \frac{1 - F_X(p)}{1 - F_X(u)} = \lim_{u \rightarrow x_{F_X}} \mathbb{P}[X \leq p | X > u], \quad (37)$$

$$V_{\xi_{2,1}}(y) = \lim_{u_Y \rightarrow x_{F_Y}} 1 - \frac{1 - F_Y(q)}{1 - F_Y(u_Y)} = \lim_{u_Y \rightarrow x_{F_Y}} \mathbb{P}[Y \leq q | Y > u_Y]. \quad (38)$$

From $Y \stackrel{d}{=} F_Y^{-1}(F_X(X))$, we take $u_Y = F_Y^{-1}(F_X(u))$ and from (37)-(38) we have

$$1 - (1 - V_{\xi_{1,1}}(x))(1 - F_X(u)) \sim_{u \rightarrow x_{F_X}} F_X(u + xa_1(u)),$$

$$1 - (1 - V_{\xi_{2,1}}(y))(1 - F_Y(F_Y^{-1}(F_X(u)))) \sim F_Y(F_Y^{-1}(F_X(u)) + ya_2(F_Y^{-1}(F_X(u)))).$$

It follows from (9) that

$$\begin{aligned} & \lim_{u \rightarrow x_{F_X}} \mathbb{P} \left[\frac{X - u}{a_1(u)} > x, \frac{Y - F_Y^{-1}(F_X(u))}{a_2(F_Y^{-1}(F_X(u)))} > y \mid X > u, Y > F_Y^{-1}(F_X(u)) \right] \\ &= \lim_{u \rightarrow x_{F_X}} \frac{C^*(1 - F_X(u + xa_1(u)), 1 - F_Y(F_Y^{-1}(F_X(u)) + ya_2(F_Y^{-1}(F_X(u)))))}{C^*(1 - F_X(u), 1 - F_Y(F_Y^{-1}(F_X(u))))} \end{aligned}$$

$$\begin{aligned}
&= \lim_{u \rightarrow x_{F_X}} \frac{C^*((1 - V_{\xi_1,1}(x))(1 - F_X(u)), (1 - V_{\xi_2,1}(y))(1 - F_Y(F_Y^{-1}(F_X(u))))))}{C^*(1 - F_X(u), 1 - F_Y(F_Y^{-1}(F_X(u))))} \\
&= \lim_{\nu \rightarrow 1} \frac{C^*((1 - V_{\xi_1,1}(x))(1 - \nu), (1 - V_{\xi_2,1}(y))(1 - \nu))}{C^*(1 - \nu, 1 - \nu)}. \quad (39)
\end{aligned}$$

Now, if $h := (1 - \xi_1 x)^{\frac{1}{\xi_1}}$, $\xi_1 \neq 0$ or if $h := e^{-x}$, $\xi_1 = 0$ then $1 - V_{\xi_1,1}(x) = V_{1,1}(h)$. So (39) becomes $\lim_{\nu \rightarrow 1} C^*(V_{1,1}(h)(1 - \nu), V_{1,1}(w)(1 - \nu))/C^*(1 - \nu, 1 - \nu)$.

As C satisfies hypotheses of Theorem A, the above limit is equal to

$$G(V_{1,1}(h), V_{1,1}(w)) = G(1 - V_{\xi_1,1}(x), 1 - V_{\xi_2,1}(y)). \text{ From (9) we get}$$

$$\begin{aligned}
\lim_{u \rightarrow x_{F_X}} \mathbb{P} \left[\frac{X - u}{a_1(u)} \leq x, \frac{Y - F_Y^{-1}(F_X(u))}{a_2(F_Y^{-1}(F_X(u)))} \leq y \mid X > u, Y > F_Y^{-1}(F_X(u)) \right] \\
= C^{*G}(1 - g(1 - V_{\xi_1,1}(x)), 1 - g(1 - V_{\xi_2,1}(y))). \quad (40)
\end{aligned}$$

Since the limit is a continuous distribution function (as C^{*G} , g and the GPD are), (40) can be strengthened to uniform convergence (see e.g. Embrechts *et al.* 1997, p. 552). Then (14) follows. \square

Proof [Theorem 5]:

To begin with, we work conditionally on $N_n = m_n$. We need first to prove the following convergence in probability

$$\frac{\overline{F}(\overline{f}(m_n) f(m_n))}{\widehat{\overline{F}}(\overline{f}(m_n) f(m_n))} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 1. \quad (41)$$

Then, using Theorem 8.1. in Smith (1987), we obtain

$$\tilde{r}_{m_n} \left[F(\overline{f}(m_n) f(m_n)) - \widehat{F}^*(\overline{f}(m_n) f(m_n)) \right] \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(\nu, \tau^2), \quad (42)$$

$$\text{with } \tilde{r}_{m_n} = \frac{\sqrt{m_n}}{\lg(\overline{f}(m_n))} \left(\frac{1}{1 - \frac{1}{n} \sum_{i=1}^n 1_{(X_i \leq \overline{f}(m_n) f(m_n))}} \right).$$

Now, to prove (41) we use the following result:

Proposition B (*Einmahl, 1990; Corollary 1*) *Let a sequence of i.i.d random variables X_1, X_2, \dots from a distribution function F . We denote with $\{k_n\}_{n=1}^{\infty}$ an arbitrary sequence of positive numbers, such that $k_n \leq n$ and $k_n \rightarrow \infty$, $\lim_{n \rightarrow \infty} \frac{k_n}{n} = 0$. Let $\{\gamma_n\}_{n=1}^{\infty}$ be a sequence of positive numbers, such that $\lim_{n \rightarrow \infty} \frac{\gamma_n}{\sqrt{k_n}} = \infty$, then $\sup_{t \geq F^{-1}(1 - \frac{k_n}{n})} \left(\frac{n}{\gamma_n} \right) \left| \widehat{F}(t) - \overline{F}(t) \right| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$.*

We choose an arbitrary sequence $\{k_n\}_{n=1}^{\infty} := \{m_n\}_{n=1}^{\infty}$ (number of excesses on a sample of size n) such that $m_n \leq n$, $\lim_{n \rightarrow \infty} m_n = \infty$ and $\lim_{n \rightarrow \infty} \frac{m_n}{n} = 0$. We take $\{\gamma_n\}_{n=1}^{\infty} := \{\sqrt{m_n} \alpha_n\}_{n=1}^{\infty}$, where α_n is an arbitrary sequence of

positive numbers such that $\lim_{n \rightarrow \infty} \alpha_n = \infty$.

Then, using Proposition B we have, for $\bar{f}(m_n) f(m_n) \geq F^{-1}(1 - \frac{m_n}{n})$,

$$\left(\frac{n}{\sqrt{m_n} \alpha_n} \bar{F}(\bar{f}(m_n) f(m_n)) \right) \left| \frac{\widehat{\bar{F}}(\bar{f}(m_n) f(m_n)) - \bar{F}(\bar{f}(m_n) f(m_n))}{\bar{F}(\bar{f}(m_n) f(m_n))} \right| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0. \quad (43)$$

We choose α_n such that for large n

$$\exists c > 0 : 0 < \frac{\sqrt{m_n} \alpha_n}{n \bar{F}(\bar{f}(m_n) f(m_n))} \leq c. \quad (44)$$

In the Fréchet case we have $L(x) = x^\alpha \bar{F}(x)$, for $\alpha > 0$ and $\forall t > 0$, $\frac{L(tx)}{L(x)} = 1 + k(t)\phi(x) + o(\phi(x))$ for $x \rightarrow \infty$. Using then Assumptions (8.7) and (8.8) of Theorem 8.1. in Smith (1987) we obtain

$$\frac{\bar{F}(\bar{f}(m_n) f(m_n))}{\bar{F}(f(m_n))} = f(m_n)^{-\alpha} [1 + k(f(m_n))\phi(\bar{f}(m_n)) + o(\phi(\bar{f}(m_n)))] .$$

Hence $\frac{n \bar{F}(\bar{f}(m_n) f(m_n))}{\sqrt{m_n}}$ is equal to

$\frac{n}{\sqrt{m_n}} \bar{F}(\bar{f}(m_n)) [f(m_n)^{-\alpha} (1 + k(f(m_n))\phi(\bar{f}(m_n)) + o(\phi(\bar{f}(m_n))))]$ which can be approximated by

$$\sqrt{m_n} f(m_n)^{-\alpha} (1 + k(f(m_n))\phi(\bar{f}(m_n)) + o(\phi(\bar{f}(m_n)))) . \quad (45)$$

Assume now $\frac{(z_{m_n})^\alpha}{\sqrt{m_n}} \xrightarrow[n \rightarrow \infty]{} 0$ (that is the analogue of (33) conditionally on $N_n = m_n$). Then the properties of k and ϕ insure that the right hand side of (45) increases to infinity hence one can choose α_n satisfying (44). To conclude the proof, we use assumption (33) and a Skorohod type argument. \square

Proof [Theorem 6]:

To prove (35) we first observe, using Corollary 7, Proposition 8 and the analogue of Kolmogorov-Smirnov Theorem in dimension 2 (see e.g. Donker, 1952 or Dudley, 1966), that

$$\sqrt{k_n} \left| C^{*G} \left(1 - g(1 - V_{\xi_X, \sigma_X}(\bar{f}_1(n) f_1(n) - \bar{f}_1(n))), 1 - g(1 - V_{\xi_Y, \sigma_Y}(\bar{f}_2(n) f_2(n) - \bar{f}_2(n))) \right) \cdot \bar{F}(\bar{f}_1(n), \bar{f}_2(n)) - \widehat{\bar{F}}(\bar{f}_1(n), \bar{f}_2(n)) \cdot C^{*G} \left(1 - \widehat{g}(1 - V_{\widehat{\xi}_X, \widehat{\sigma}_X}(\bar{f}_1(n) f_1(n) - \bar{f}_1(n))), 1 - \widehat{g}(1 - V_{\widehat{\xi}_Y, \widehat{\sigma}_Y}(\bar{f}_2(n) f_2(n) - \bar{f}_2(n))) \right) \right| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

Furthermore $r_n \left| \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq \bar{f}_1(n), Y_i \leq \bar{f}_2(n)\}} - F(\bar{f}_1(n), \bar{f}_2(n)) \right| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$,

with $r_n \ll \sqrt{n}$. At last using Corollary 7, Theorem 5, we obtain the convergence (35) with

$$B(x_n^*, y_n^*) = B(-\lg(\widehat{F}_X(\bar{f}_1(n))), -\lg(\widehat{F}_Y^*(y_n))) + B(-\lg(\widehat{F}_X^*(x_n)), -\lg(\widehat{F}_Y(\bar{f}_2(n)))),$$

where $B(\cdot, \cdot)$ is the gaussian process defined in Theorem C.

Auxiliary results

Theorem C (*Einmahl et al. 2006; Theorem 2.2*) Assume that exists the limit $R(x, y)$ such that, for some $\alpha > 0$

$$\frac{1}{t} \mathbb{P}(1 - F_X(X) \leq tx, 1 - F_Y(Y) \leq ty) - R(x, y) = O(t^\alpha), \text{ as } t \rightarrow 0,$$

uniformly for $\max(x, y) \leq 1, x, y \geq 0$. Let $k = o(n^{\frac{2\alpha}{1+2\alpha}})$. If $R_1(x, y) := \frac{\partial R(x, y)}{\partial x}$ and $R_2(x, y) := \frac{\partial R(x, y)}{\partial y}$ are continuous then we have

$$\sup_{0 < x, y \leq 1} |\sqrt{k}(\widehat{l}(x, y) - l(x, y)) + B(x, y)| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0,$$

where $B(x, y) := W(x, y) - R_1(x, y)W_1(x) - R_2(x, y)W_2(y)$, with W a Gaussian process on $[0, x] \times [0, y]$, $W_1(x) = W([0, x] \times [0, \infty])$, $W_2(y) = W([0, \infty] \times [0, y])$.

In the case where C is twice differentiable continuous copula, Theorem C holds for any $\alpha \geq 1$.

Corollary 7 Under conditions of Theorem C if we have v_n such that $\frac{v_n}{\sqrt{k_n}} \rightarrow 0$, for $n \rightarrow \infty$, and $\lambda > 0$ we obtain

$$v_n \sup_{0 < x, y \leq 1} |\widehat{l}(x, y) - l(x, y)| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0,$$

$$v_n \sup_{0 < x, y \leq 1} |\widehat{G}(x, y) - G(x, y)| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0, \quad v_n \sup_{0 < x \leq 1} |\widehat{g}(x) - g(x)| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0,$$

with $\widehat{g}(x)$ as in (20), $\widehat{G}(x, y)$ as in (21) and $k_n \rightarrow \infty$, $\frac{k_n}{n} \rightarrow 0$ and $k_n = o(n^{\frac{2\alpha}{1+2\alpha}})$.

Proposition 8 Let $V_{\xi, \sigma}(x)$ the Generalized Pareto Distribution (GPD) and $\widehat{\xi}_N, \widehat{\sigma}_N$, the maximum likelihood estimators of the parameters $\xi = -\alpha^{-1} < 0$ and $\sigma = u_n \alpha^{-1}$, in the case unconditionally on N . If all the conditions of Corollary 3 hold then

$$p_{nx} \sup_{x \in [0, +\infty)} |V_{\widehat{\xi}_N, \widehat{\sigma}_N}(x) - V_{\xi, \sigma}(x)| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0, \quad \text{where } \frac{p_{nx}}{\sqrt{N_x}} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

Proof: Using Corollary 3 we obtain for each point $x \in [0, +\infty)$,

$$p_{nx} \left[V_{\widehat{\xi}_X, \widehat{\sigma}_X}(x) - V_{\xi, \sigma}(x) \right] = p_{nx} \left[\left(1 - \frac{\xi x}{\sigma} \right)^{\frac{1}{\xi}} - \left(1 - \frac{\widehat{\xi}_X x}{\widehat{\sigma}_X} \right)^{\frac{1}{\widehat{\xi}_X}} \right] \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0, \quad (46)$$

where $\frac{p_{nx}}{\sqrt{N_x}} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$. Finally, applying a stochastic version of Polya's Theorem (see Horowitz, 2001), as $V_{\xi, \sigma}(x)$ is a continuous distribution function, the convergence in (46) holds uniformly on $[0, +\infty)$. \square

Acknowledgements This work has been partially supported by the french research national agency (ANR) under the reference ANR-08BLAN-0314-01. The authors thank an anonymous referee for pointing a gap in a previous version. They also thank Anne-Laure Fougères and José R. León for fruitful discussions.

References

1. Balkema A. A. and de Haan L., Residual life time at great age. *Annals of Probability* **2**, 792-804, (1974).
2. Beirlant J., Dierckx G. and Guillou A., Reduced bias estimators in joint tail modelling, *submitted*, (2010).
3. Beirlant J., Goegebeur Y., Segers J. and J. Teugels. *Statistics of Extremes: Theory and Applications*. Chichester: Wiley, (2004).
4. Capéreaux P. and Fougères A.-L., Estimation of a bivariate extreme value distribution. *Extremes* **3**, 311-329, (2000).
5. Coles, S. G. and Tawn, J. A., Statistical methods for multivariate extremes: an application to structural design (with discussion). *Appl. Statist.*, **43**, 1-48, (1994).
6. Davison A. C., A statistical model for contamination due to long-range atmospheric transport of radionuclides. PhD Thesis. Department of Mathematics, Imperial College of Science and Technology, London, (1984).
7. Davison A. C. and Smith R. L., Models for exceedances over high thresholds. *Journal of the Royal Statistical Society, Series B*, **52**, 393-442, (1990).
8. Dekkers A.L.M. and de Haan L., On the estimation of the extreme-value index and large quantile estimation, *Annals of Statistics*. **17**, 1795-1832, (1989).
9. Drees H. and Huang X., Best attainable rates of convergence for estimators of the stable tail dependence function. *J. Multivariate Anal.* **64**, 25-47, (1998).
10. Draisma G., Drees H., Ferreira A. and de Haan L., Bivariate tail estimation: dependence in asymptotic independence, *Bernoulli*, **10**, 251-280, (2004).
11. Einmahl J.H.J., The empirical distribution function as a tail estimator, *Statistica Neerlandica*, **44**, 79-82, (1990).
12. Einmahl J.H.J., Krajina A. and Segers J., A Method of Moments Estimator of Tail Dependence, *Bernoulli*, **14**, 1003-1026, (2008).
13. Embrechts P., Kluppelberg C. and Mikosch T., *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin Heidelberg, (1997).
14. Frees E.W. and Valdez E.A. Understanding relationships using copulas. *N. Am. Actuar. J.* **2**, 1-15, (1998).
15. de Haan L., On regular variation and its application to the weak convergence of sample extremes, *Mathematical Centre Tract*, **32**, Amsterdam, (1970).
16. Horowitz J. L., The Bootstrap. In *Handbook of Econometrics*, (Ed. by J.J. Heckman and E.E. Leamer), Elsevier Science B.V, 3159-3228, (2001).
17. Huang X., *Statistics of bivariate extreme values*, Thesis, Erasmus University Rotterdam, Tinbergen Institute. Research Series 22, (1992).
18. Juri A. and Wüthrich M.V., Tail Dependence From a Distributional Point of View, *Extremes*, **3**, 213-246, (2003).
19. Klement E.P. and Mesiar R., How non-symmetric can a copula be?, *Commentationes Mathematicae Universitatis Carolinae*, **1**, 141-148, (2006).
20. Ledford A.W. and Tawn J.A., Statistics for near independence in multivariate extreme values, *Biometrika*, **83**, 169-187, (1996).
21. Ledford A.W. and Tawn J.A., Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society, Series B*, **59**, 475-499, (1997).
22. Ledford A.W. and Tawn J.A., Concomitant Tail Behaviour for Extremes. *Advances in Applied Probability* **30**, 197-215, (1998).
23. Lescourret L. and Robert C.Y., Extreme dependence of multivariate catastrophic losses, *Scandinavian Actuarial Journal*, **4**, 203-225, (2006).
24. McNeil A. J., Estimating the Tails of loss severity distributions using extreme value theory, *ASTIN Bulletin*, **27**, 1117-1137, (1997).

25. McNeil A. J., Extreme value theory for risk managers, preprint, ETH, Zurich, (1999).
26. McNeil A. J., Frey R. and Embrechts P., Quantitative risk management: Concepts, Techniques and Tools. Princeton Series in Finance, (2005).
27. Nelsen R.B., An Introduction to Copulas. Springer-Verlag, New York, (1998).
28. Peng L., Estimation of the coefficient of tail dependence in bivariate extremes. *Statistics and Probability Letters*, **43**, 399-409, (1999).
29. Pickands J., Multivariate extreme value distributions. *Bulletin of International Statistical Institute*, Buenos Aires, 859-878, (1981).
30. Pickands J., Statistical inference using extreme order statistics. *Annals of Statistics* **3**, 119-131, (1975).
31. Ramos A. and Ledford A., A new class of models for bivariate joint tails. *Journal Of The Royal Statistical Society Series B*, **1**, 219-241, (2009).
32. Rao C.R., Linear Statistical Inference and Its Applications. Wiley, New York, (1965).
33. Resnick S., Extreme Values, Regular Variation and Point Processes. Springer-Verlag, (1987).
34. Schlather M., Examples for the coefficient of tail dependence and the domain of attraction of a bivariate extreme value distribution. *Statistics and Probability Letters* **53**, 325-329, (2001).
35. Smith R., Estimating tails of probability distributions, *Annals of Statistics* **15**, 1174-1207, (1987).
36. Tajvidi N., Characterisation and Some Statistical Aspects of Univariate and Multivariate Generalized Pareto Distributions, PhD thesis, Department of Mathematics, Chalmers, Göteborg, (1996).
37. Wüthrich M. V., Bivariate extension of the Pickands-Balkema-de Haan Theorem, *Ann. Inst. H. Poincaré Probability and Statistics* **40**, **1**, 33-41, (2004).