Estimating a bivariate tail: A copula based approach

Elena Di Bernardino\textsuperscript{a}, Véronique Maume-Deschamps\textsuperscript{b,}\textsuperscript{*}, Clémentine Prieur\textsuperscript{c}

\textsuperscript{a} Conservatoire National des Arts et Métiers, IMATH, 292 Rue Saint Martin, EA4629 Paris, France
\textsuperscript{b} Université de Lyon, Université Lyon 1, ISFA, Laboratoire SAF, 50 avenue Tony Garnier, 69366 Lyon, France
\textsuperscript{c} Université Joseph Fourier, Tour IRMA, MOISE-LJK B.P. 53 38041 Grenoble, France

\textbf{A B S T R A C T}

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.

© 2013 Elsevier Inc. All rights reserved.

\textbf{1. Introduction}

The univariate POT (peaks-over-threshold) method is commonly adopted for estimating extreme quantiles or tail distributions (see e.g. [34,35] and references therein). A key idea of this method is that a distribution is in the domain of attraction of an extreme value distribution if and only if the distribution of excesses over high thresholds is asymptotically generalized Pareto (GPD; see e.g. [1,39]):

\begin{equation}
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}
\end{equation}

and \( x \geq 0 \) for \( \xi \leq 0 \) or \( 0 \leq x < \frac{\sigma}{\xi} \) for \( \xi > 0 \). This univariate modeling is well understood, and has been discussed in Davison [9], Davison and Smith [10] and other papers by 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 bivariate estimation procedure, mainly based on a version of the Pickands–Balkema–de Haan Theorem in dimension 2 (Theorem 2.1). This extension allows us to consider a two-dimensional structure of
dependence between the two continuous random components $X$ and $Y$. This dependence is modeled via a copula $C$, which is supposed to be unknown.

We recall here some classical bivariate threshold models, based on a characterization of the joint tail by Resnick [43]. Let $F$ denote the joint distribution of $(Y_1, Y_2)$ with marginals $F_j, j = 1, 2$. Define $Z_j = -1 / \log(F(Y_j)), j = 1, 2$; i.e. each $Y_j$ is transformed to a unit Fréchet variable and $\mathbb{P}(Z_j \leq z) = e^{-1/z},$ for $0 < z < \infty$. Let $F_\eta$ denote the joint distribution of $(Z_1, Z_2)$; we have $F(y_1, y_2) = F_\eta(z_1, z_2)$. The assumption that $F$ is in the maximum domain of attraction (MDA) of a bivariate extreme value distribution $G$ is equivalent to assuming $F_\eta$ to be in the domain of attraction of a bivariate extreme value distribution $G_\eta$, where the marginals of $G_\eta$ are unit Fréchet. The characterization of Resnick [43] can be written as

$$
\lim_{t \to \infty} \frac{\log(F_\eta(t, z_2))}{\log(F_\eta(t, t))} = \lim_{t \to \infty} \frac{1 - F_\eta(t, z_2)}{1 - F_\eta(t, t)} = \frac{\log(G_\eta(z_1, z_2))}{\log(G_\eta(1, 1))}.
$$

Equating the left-hand and the right-hand terms for large $t$ leads to the following model for the joint tail of $F$ (see [30]):

$$
\mathcal{F}_1(y_1, y_2) = \exp\{-l (- \log(F(y_1)), - \log(F(y_2)))\},
$$

for $y_j > u_j$, where the $u_j$ are high thresholds for the marginal distributions and $l$ is the stable tail dependence function of the limiting extreme value distribution $G_\eta$. Then approximation (3) can be estimated by

$$
\tilde{\mathcal{F}}_1^*(y_1, y_2) = \exp\{-\tilde{l} (- \log(\tilde{F}(y_1)), - \log(\tilde{F}(y_2)))\},
$$

for high values of $y_1$ and $y_2$, where $\tilde{F}(y_1)$ (resp. $\tilde{F}(y_2)$) is an estimator for the marginal tail of $Y_1$ (resp. $Y_2$). For instance $\tilde{F}(y_1)$ comes from the univariate POT method described in Section 4.1. In (4), $\tilde{l}$ is an estimator of the stable tail dependence function (see [15, 13, 19]). For another approach, based on the estimation of the so-called univariate dependence function of Pickands [40], see for instance Capéraà and Fougères [5]. Problems arise with both of these bivariate techniques when $(Y_1, Y_2)$ are asymptotically independent, i.e.,

$$
\lambda := \lim_{t \to 0} \mathbb{P}[F_Y^{-1}(Y_1) > 1 - t | F_Y^{-1}(Y_2) > 1 - t] = 0.
$$

When the data exhibit positive or negative association that only gradually disappears at more and more extreme levels, these methods produce a significant bias. In order to overcome this problem, Ledford and Tawn [30–32] introduced a model in which the tail dependence is characterized by a coefficient $\eta \in (0, 1]$. In Ledford and Tawn [32] the joint survival distribution function of a bivariate random vector $(Z_1, Z_2)$ with unit Fréchet marginals is assumed to satisfy

$$
\tilde{F}(z_1, z_2) = L(z_1, z_2) z_1^{-c_1} z_2^{-c_2}, \quad \text{with } z_1 \text{ and } z_2 \text{ large enough},
$$

with $c_1, c_2 > 0, c_1 + c_2 = 1/\eta$, and $L$ a quasi-symmetric bivariate slowly varying function, i.e.,

$$
\lim_{t \to 0} \frac{L(tx, ty)}{L(t, t)} = \tilde{g}(x, y),
$$

where $\tilde{g}$ is a function such that $\tilde{g}(cx, cy) = g(x, y)$, for each $c > 0$. Various methods for estimating this coefficient $\eta$ are proposed in Peng [38], Draiisma et al. [13], and Beirlant et al. [2]. For some counter-examples for the Ledford and Tawn model, see Schlather [45].

In the present paper, we propose a new alternative estimation procedure. The model that we propose for the tail, in contrast to the one in Ledford and Tawn [32] (see Eqs. (6)–(7)), does not assume Fréchet marginal distributions. Thus it is free from pre-treatment of the data, and we can indeed work directly with the original samples. Moreover, we will prove later that the Ledford and Tawn model [32] is a particular case of ours (see Eq. (10)).

The general idea of our procedure is to decompose the estimation of $\mathbb{P}(X < x, Y < y)$, for $x, y$ above some marginal thresholds $u_x, u_y$, in the estimation of three different regions. For the joint upper tail region $[u_x, x] \times [u_y, y]$ we use a new non-parametric estimator coming from Theorem 2.1 (see Section 2). This new estimator adopts a distributional point of view for the tail, following Juri and Wüthrich [29] (see also Charpentier and Juri [6]). It is important to notice that the study of the tail dependence from a distributional point of view by means of appropriate copulas has received attention in the past decade. The interested reader is referred to Juri and Wüthrich [28, 29], Wüthrich [47], Charpentier and Juri [6], Charpentier and Segers [7], and Javid [27].

For the remaining two lateral regions $[\infty, x] \times [\infty, u_y]$ and $[\infty, u_x] \times [\infty, y]$, we approximate the distribution function $F$ using the estimator in (3), well known in the extreme value literature.

Our new tail estimator covers situations less restrictive than dependence or perfect independence above thresholds. The consistency of the proposed estimator is proved both in the asymptotically dependent case (Theorem 6.1) and in the asymptotically independent one (Theorem 6.2).

The stability of our estimation compared to that of $\tilde{F}_1$ is analyzed for some real cases (Section 7) which have been studied in other papers (e.g. Beirlant et al. [2], Frees and Valdez [23], and Lescourret and Robert [33]).
Finally, we recall that, in the past decade, bivariate extensions of the POT method via the generalized Pareto distribution have been developed in a series of papers by Falk and Reiss [22] and references therein and in [42, Chapter 13]. Recently a multivariate generalization was treated in [3,44,36]. The role of multivariate generalized Pareto distributions in the framework of extreme value theory is still under scrutiny. In contrast to the situation for the univariate case, it is not intuitively clear how exceedances over high thresholds are to be defined. Our paper makes a contribution to this part of the recent literature. To the best of our knowledge the POT procedure that we propose in this paper cannot be directly deduced from the POT methods proposed in works cited above. Moreover we provide an estimation of bivariate tails such that this type of estimation is not obtained in the papers cited above. However, some ingredients for a comparison are investigated in Theorem 4.2 in Juri and Wüthrich [29].

The paper is organized as follows. In Section 2 we state an extension of the Pickands–Balkema–de Haan Theorem for the case of bivariate distributions with different marginals (Theorem 2.1). In Section 3 we provide a new non-parametric estimator for the dependence structure of a bivariate random sample in the upper tail. In Section 4 we recall the POT procedure for univariate distributions and we use Theorem 2.1 in order to construct a new estimator for the tail of the bivariate distribution. The study of the asymptotic properties of our estimator makes use of a convergence result in the univariate case (Theorem 5.1) dealing with asymptotic behavior of the absolute error between the theoretical distribution and its estimator. In Section 6 we present the consistency result for our estimator with its convergence rate both in the asymptotically dependent case (Theorem 6.1) and in the asymptotically independent one (Theorem 6.2).

Examples with real data are presented in Section 7. Some auxiliary results and more technical proofs are postponed to the Appendix.

**Remark 1.** Assume that 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 exceedances 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 exceedances is $m_n$, with $\lim_{n \to \infty} m_n = \infty$ and $\lim_{n \to \infty} m_n/n = 0$. The second approach considers the number of exceedances $N_n$ as a binomial random variable (which is the case in the simulations): $N_n \sim Bi(n, 1 - F(u_n))$ with $\lim_{n \to \infty} 1 - F(u_n) = 0$ and $\lim_{n \to \infty} n(1 - F(u_n)) = \infty$. Keeping these considerations in mind will be useful in the following (in particular in Section 5).

**2. On the two-dimensional Pickands–Balkema–de Haan Theorem**

A central one-dimensional result in univariate tail estimation is the so-called Pickands–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 can be found in the literature (e.g. see Juri and Wüthrich, [29]; Wüthrich, [47]) with the assumption $F_X = F_Y$. Starting from Theorem 4.1 in Juri and Wüthrich [29] and Theorem 3.1 in [6], we provide here a precise formulation and proof of a general bivariate Pickands–Balkema–de Haan Theorem (Theorem 2.1 below). We first introduce some notation and recall results from Juri and Wüthrich [29] and Nelsen [37], which we will need later.

We consider a two-dimensional copula $C(u, v)$ and the associated survival copula $C^*(u, v)$. At first, we assume 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 $X$ and $Y$ conditioned on $\{X > u, Y > u\}$:

$$\forall x \in [0, 1], \quad F_{X, u}(x) := \mathbb{P}[X \leq x | X > u, Y > u] = 1 - \frac{C^*(1 - x \vee u, 1 - u)}{C^*(1 - u, 1 - u)}, \quad (8)$$

$$\forall y \in [0, 1], \quad F_{Y, u}(y) := \mathbb{P}[Y \leq y | X > u, Y > u] = 1 - \frac{C^*(1 - u, 1 - y \vee u)}{C^*(1 - u, 1 - u)}. \quad (9)$$

Note that the continuity of the copula $C$ implies that $F_{X, u}$ and $F_{Y, u}$ are also continuous.

**Definition 2.1.** Let $X$ and $Y$ be uniformly distributed on $[0, 1]$. Assume that for a threshold $u \in [0, 1)$, $C^*(1 - u, 1 - u) > 0$. We define the upper tail dependence copula at level $u \in [0, 1)$ relative to the copula $C$ by

$$C^{up}(x, y) := \mathbb{P}[X \leq F_{X, u}^{-1}(x), Y \leq F_{Y, u}^{-1}(y) | X > u, Y > u],$$

$\forall (x, y) \in [0, 1]^2$, where $F_{X, u}, F_{Y, u}$ are given by (8)-(9).

Note that $\mathbb{P}[X \leq x, Y \leq y | X > u, Y > u]$ obviously defines a two-dimensional distribution function whose marginals are given by $F_{X, u}$ and $F_{Y, u}$. We remark that $C^{up}(x, y)$ is a copula and from the continuity of $F_{X, u}$ and $F_{Y, u}$ we obtain the uniqueness of $C^{up}$. Moreover, the asymptotic behavior of $C^{up}$ for $u$ around 1 describes the dependence structure of $X$, $Y$ in their upper tails.

In order to provide an explicit form for $\lim_{u \to 1} C^{up}(x, y)$, we state Proposition 2.1 below, which is a modification of Theorem 3.1 in [6]. More precisely we adapt Theorem 3.1 in [6] in the case of the upper tail dependence copula, assuming that $C$ satisfies a suitable regularity condition under the direction $(1 - u, 1 - u)$ (see the limit in (10)). For comparisons, we refer the reader to Section 3 in [6].
Proposition 2.1. Assume that $\partial C^*(1-u, 1-v)/\partial u < 0$ and $\partial C^*(1-u, 1-v)/\partial v < 0$ for all $u, v \in [0, 1)$. Furthermore, assume that there is a positive function $G$ such that
\[ \lim_{u \to 1} \frac{C^*(x(1-u), y(1-u))}{C^*(1-u, 1-u)} = G(x, y), \quad \text{for all} \ x, y > 0. \] (10)

Then for all $(x, y) \in [0, 1]^2$,
\[ \lim_{u \to 1} C_{\text{app}}^*(x, y) = x + y - 1 + G(g_1^{-1}(1-x), g_2^{-1}(1-y)) := C^\ast G(x, y), \] (11)
where $g_1(x) := G(x, 1)$, $g_2(y) := G(1, y)$. Moreover there is a constant $\theta > 0$ such that, for $x > 0$,
\[ G(x, y) = \begin{cases} x^\theta g_1 \left( \frac{y}{x} \right) & \text{for} \ y \in [0, 1], \\ y^\theta g_2 \left( \frac{x}{y} \right) & \text{for} \ y \in (1, \infty). \end{cases} \] (12)

The proof of Proposition 2.1 is postponed to the Appendix. We adapt to our setting the proof of Theorem 3.1 by Charpentier and Juri [6]. Since $\partial C^*(1-u, 1-v)/\partial u < 0$ and $\partial C^*(1-u, 1-v)/\partial v < 0$ for all $u, v \in [0, 1)$, we have $C^*(1-u, 1-u) > 0$, for all $u > 0$, i.e. $C_{\text{app}}^*$ is well defined for all $u > 0$. Then we ask that the joint survival distribution function of $X$ and $Y$, uniformly distributed on $[0, 1]$, be strictly decreasing in each coordinate. As in Remark 3.2 in [6], one can prove that the convergence in (11) is uniform in $[0, 1]^2$. From Proposition 2.1, functions $G$, $g_1$, and $g_2$ characterize the asymptotic behavior of the dependence structure for extremal events.

Remark 2. Under the approximation of the Ledford and Tawn model [32] (see Eqs. (6)–(7)) we obtain (with arguments similar to those in Section 4.2 in Juri and Wüthrich [29])
\[ \lim_{u \to 1} \frac{C^*(x(1-u), y(1-u))}{C^*(1-u, 1-u)} = \lim_{u \to 1} \frac{L \left( \frac{1}{x}, \frac{1}{y} \right) \left( \frac{1}{x} \right)^{\frac{1}{n}} \left( \frac{1}{y} \right)^{\frac{1}{n}}}{L \left( \frac{1}{1-u} \right)^{\frac{1}{n}} \left( \frac{1}{1-u} \right)^{\frac{1}{n}}} = \tilde{G} \left( \frac{1}{x}, \frac{1}{y} \right) x^{\frac{1}{n}} y^{\frac{1}{n}}. \]
This means that models of type (6)–(7) satisfy the assumption of Proposition 2.1 with $G(x, y) = \tilde{G} \left( \frac{1}{x}, \frac{1}{y} \right) x^{\frac{1}{n}} y^{\frac{1}{n}}$. Using regularity properties of $\tilde{G}$ and $G$, we can deduce $\theta = 1/\eta$. Then the extremal dependence parameter $\theta$ inherits the properties of $\eta$ in the Ledford and Tawn model. Thus it describes not only the type but also the intensity of the asymptotic dependence. For explicit examples, the interested reader is referred to Section B.1 in the Ph.D. Thesis of Di Bernardino [12] (available online). In Section 3 we propose an estimate of the parameter $\theta$.

Remark 3. We note that $C^*(x, y)$ defined in (11) is the survival copula of the copula $C^G(x, y) := G(g_1^{-1}(x), g_2^{-1}(y))$ and thus, in particular, is a copula (for more details see Section 3 in [6]).

- In the case of a symmetric copula, i.e. $C(u, v) = C(v, u)$ for all $u$ and $v$, the limit in (10) is continuous and symmetric, with marginals $G(x, 1) = G(1, x) = g(x)$, where $g : [0, \infty) \to [0, \infty)$ is a strictly increasing function and $g(x) = x^\theta g(1/x)$ for all $x \in (0, \infty)$ (for more details about properties of $G$ in the symmetric case, see Section 2 in [29]).

For in the univariate setting, de Haan [11] proves that $F \in \text{MDA}(H_\xi)$ is equivalent to the existence of a positive measurable function $a(\cdot)$ such that, for $1 - \xi x > 0$ and $\xi \in \mathbb{R}$,
\[ \lim_{u \to x} \frac{1 - F(u + x a(u))}{1 - F(u)} = \begin{cases} (1 - \xi x)^2, & \text{if} \ \xi \neq 0, \\ e^{-x}, & \text{if} \ \xi = 0, \end{cases} \] (13)
where $x_F := \sup \{x \in \mathbb{R} | F(x) < 1 \}$. This allows us to state below a rigorous formulation of the two-dimensional Pickands–Balkema–de Haan Theorem for the general case.

Theorem 2.1. Let $X$ and $Y$ be two continuous real valued random variables, with different marginal distributions, respectively $F_X$, $F_Y$, and copula $C$. Suppose that $F_X \in \text{MDA}(H_{\xi_1})$, $F_Y \in \text{MDA}(H_{\xi_2})$ and that $C$ satisfies the assumptions of Proposition 2.1. Then
\[ \sup_{a'} \left\{ P \left[ X - u \leq x, Y - F_X^{-1}(F_X(u)) \leq y | X > u, Y > F_Y^{-1}(F_Y(u)) \right] \right\} \rightarrow 0, \] (14)
where $V_{\xi_i,a_i}(\cdot)$ is the GPD with parameters $\xi_i$, $a_i(\cdot)$ defined in (1), $a_i(\cdot)$ as in (13), for $i = 1, 2$, and $a' := \{(x, y) : 0 < x \leq x_{F_X} - u, 0 < y \leq x_{F_Y} - F_Y^{-1}(F_Y(u)) \}$, with $x_{F_X} := \sup \{x \in \mathbb{R} | F_X(x) < 1 \}$, $x_{F_Y} := \sup \{y \in \mathbb{R} | F_Y(y) < 1 \}$.

The proof of Theorem 2.1 is postponed to the Appendix.
3. Estimating dependence structure in the bivariate framework

It is well known that 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:

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

(15)

or similarly

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

(16)

see e.g. [26]. If $F_X, F_Y$ are in the maximum domain of attraction of two extreme value distributions $H_X, H_Y$ and if (15) holds then $F$ is in the domain of attraction of an extreme value distribution $H$ with marginals $H_X, H_Y$ and with copula determined by $l$. Furthermore (15) is equivalent to

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

(17)

Note that the upper tail dependence coefficient defined in (5) is such that $\lambda = R(1, 1)$. We introduce the non-parametric estimators for $l$ and $R$ (see [18]):

$$
\hat{l}(x, y) = \frac{1}{kn} \sum_{i=1}^{n} I_{\left[\left(\frac{X_i}{Y_i} > n - knx + 1 \text{ or } \frac{Y_i}{X_i} > n - kny + 1\right]\right]},
$$

(18)

$$
\hat{R}(x, y) = \frac{1}{kn} \sum_{i=1}^{n} I_{\left[\left(\frac{X_i}{Y_i} > n - knx + 1, \frac{Y_i}{X_i} > n - kny + 1\right]\right]},
$$

(19)

where $k_n \to \infty, k_n/n \to 0$ and $R(X_i)$ is the rank of $X_i$ among $(X_1, \ldots, X_n)$, $R(Y_i)$ is the rank of $Y_i$ among $(Y_1, \ldots, Y_n)$, for $i = 1, \ldots, n$.

3.1. The asymptotically dependent case

If $X$ and $Y$ are asymptotically dependent ($\lambda > 0$) we introduce an estimator for $G, g_X$ and $g_Y$ which will be used later to estimate the tail of the bivariate distribution function. Using (15)-(17), we write

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

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

Using (12), as $R$ is homogeneous of order 1 then $\theta = 1$. As $\eta \in (0, 1]$ in the Ledford and Tawn model (see [30–32]), $\theta$ describes the nature of the tail dependence; it does not depend on the marginal distribution functions.

In order to estimate $g_X, g_Y$ and $G$, we use the non-parametric estimator for $R$ in (19) and we obtain

$$
\hat{g}_X(x) = \frac{\hat{R}(x, 1)}{\hat{R}(1, 1)}, \quad \hat{g}_Y(x) = \frac{\hat{R}(1, y)}{\hat{R}(1, 1)}, \quad \text{and} \quad \hat{G}(x, y) = \frac{\hat{R}(x, y)}{\hat{R}(1, 1)}.
$$

(20)

Using (20) we get a non-parametric estimator for $\theta$; for $x > 0$,

$$
\hat{\theta}_x = \begin{cases} 
\frac{\log(\hat{G}(x, y)) - \log(\hat{g}_Y(x))}{\log(x)} & \text{if } \frac{y}{x} \in [0, 1], \\
\frac{\log(\hat{G}(x, y)) - \log(\hat{g}_X(x))}{\log(y)} & \text{if } \frac{y}{x} \in (1, \infty).
\end{cases}
$$

(21)

Following Remark 3, in the case of a symmetric copula, using $g_X(x) = g_Y(x) = g(x) = x^\theta g(1/x)$ for $x > 0$, we get

$$
\hat{\theta}_x = \frac{\log(\hat{g}(x)) - \log(\hat{g}(1/x))}{\log(x)}.
$$

(22)
Fig. 1. Estimator for \( \theta, (k, \hat{\theta}) \) (left), \( x = 0.07 \), survival Clayton copula with parameter 1 (right), \( x = 5 \), and logistic copula with parameter 0.5.

Fig. 2. Mean squared error for \( \hat{\theta} \) (left), \( x = 0.07 \), survival Clayton copula with parameter 1 (right), \( x = 5 \), and logistic copula with parameter 0.5.

Using a slightly adapted version of Theorem 2.2 in [18] (see Theorem A.1 in the Appendix), we state the following consistency result for \( \hat{G}, \hat{g}_X \) and \( \hat{g}_Y \):

**Corollary 3.1.** Under the assumptions of Theorem A.1, if we have \( v_n \) such that \( v_n / \sqrt{k_n} \to 0 \), for \( n \to \infty \), and \( \lambda > 0 \), we obtain

\[
\begin{align*}
\sup_{0 < x, y \leq 1} | \hat{G}(x, y) - G(x, y) | & \xrightarrow{p} 0, \\
v_n \sup_{0 < x \leq 1} | \hat{g}_X(x) - g_X(x) | & \xrightarrow{p} 0, \\
v_n \sup_{0 < y \leq 1} | \hat{g}_Y(y) - g_Y(y) | & \xrightarrow{p} 0
\end{align*}
\]

with \( \hat{g}_X(x), \hat{g}_Y(y) \) and \( \hat{G}(x, y) \) as in (20), \( k_n \to \infty \), \( k_n/n \to 0 \) and \( k_n = o(n^{\frac{1}{1+2\alpha}}) \).

We now provide an illustration for two different copulae: survival Clayton and logistic copulae. We remark that they are two symmetric copulae with \( \lambda > 0 \). In particular we observe the sensitivity of \( \hat{\theta}_X \) in (22) to the sequence \( k_n \) (Fig. 1). We draw the mean curve on 100 samples of size \( n = 1000 \) (full line) and the empirical standard deviation (dashed lines).

In 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 was well behaved. In the following we fix \( x \) for each simulation and may vary \( k_n \). The choice of \( k_n \) does not appear to be crucial for \( \hat{\theta}_X \). In Fig. 2 the mean squared error for \( \hat{\theta}_X \) is calculated on 100 samples of size \( n = 1000 \).
3.2. The asymptotically independent case

We say that $X$ and $Y$ are asymptotically independent if $\lambda = R(1, 1) = 0$. In terms of a copula this means that $C(u, u) = 1 - 2(1 - u) + o(1 - u)$, for $u \to 1$. The problem, with respect to Section 3.1, is that $g_x(x) = \frac{R(x, 1)}{R(1, 1)}$ and $g_y(y) = \frac{R(y, 1)}{R(1, 1)}$ makes no sense, as $\lambda = 0$ and $R(x, y) = x + y - I(x, y) = 0$, $\forall x, y$.

We thus need to introduce a second-order refinement of the condition of (10). More precisely, as in [13], we shall assume that

$$\lim_{t \to 0} \frac{C^*(u,v)}{C^*(t,t)} = G(x, y)$$

for all $x, y \geq 0$, $x + y > 0$, where $q_1$ is some positive function and $Q$ is neither a constant nor a multiple of $G$. Moreover we assume that convergence in (23) is uniform on $[x^2 + y^2 = 1]$. Let $q(t) := \mathbb{P}[1 - F_X(x) < t, 1 - F_Y(y) < t]$ and $q^{-1}$ be its inverse function. Then, using (23), we obtain the following consistency result for $G$, $g_x$, and $g_y$:

**Proposition 3.1.** Suppose (10) and (23) hold. We assume that $\lim_{t \to 0} q(t)/t = \lambda = 0$. Then, for a sequence $k_n$ such that $a_n := n q(k_n/n) \to \infty$ (this implies that $k_n \to \infty$), $k_n/n \to 0$, $\sqrt{n}q_1(q^-(a_n/n)) \to 0$, it holds that

$$\psi_n \sup_{0 \leq x, y \leq 1} |\hat{G}(x, y) - G(x, y)| \to 0,$$

$$\psi_n \sup_{0 \leq x, y \leq 1} |\hat{g}_x(x) - g_x(x)| \to 0,$$

$$\psi_n \sup_{0 \leq y \leq 1} |\hat{g}_y(y) - g_y(y)| \to 0,$$

with $\psi_n \ll \sqrt{n} \hat{g}_x(x), \hat{g}_y(y)$ and $\hat{G}(x, y)$ as in (20).

Details of the proof are postponed to the Appendix. It is mainly based on Lemma 6.1 in [13]. Let us note that in this case (asymptotic independence) the function $q_1$, regularly varying at zero, appears in the convergence rate.

In Proposition 3.2 below, by assuming some regularity properties on $C$, we deduce specific forms for $G$, $g_x$, $g_y$ and $\theta$.

**Proposition 3.2.** If $\lambda = 0$ and $C$ is a copula that is twice continuously differentiable with the determinant of the Hessian matrix of $C$ at $(1, 1)$ different from zero, then

$$\lim_{u \to 1} \frac{C^*(x(1-u), y(1-u))}{C^*(1-u, 1-u)} = xy, \quad \forall x, y > 0,$$

$$g_x(x) = g_y(x) = x \quad \text{and} \quad \theta = 2.$$

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

The assumptions of Proposition 3.2 are satisfied for a large class of asymptotically independent copulae: Ali Mikhail–Haq, Frank, Clayton with $a \geq 0$, independent and Fairlie–Gumbel–Morgenstern copulae. An example of a non-symmetric copula that satisfies the assumptions of Proposition 3.2 is $C(x, y) = xy + \frac{1}{2}(1 - |2x - 1|)(1 - 2y - 1^2)$. This type of asymmetric copula is proposed in [4] to model the evolution of price spread between electricity and gas prices.

We introduce some examples of asymptotically independent copulae that do not satisfy the assumptions of Proposition 3.2.

We consider the Ledford and Tawn model (e.g. see [30]): $2u - 1 + C(1-u, 1-u) = (1-u)^2 L(1-u)$, with $L$ a slowly varying function at zero and $\eta \in (0, 1)$. Then, for $\eta > 1/2$, $\lim_{u \to 1} (C(1, 1) - C(1-u, 1) - C(1, 1-u) - C(1-u, 1-u))/(1-u)^2 = \infty$. Thus $C_{\min}^{(u)}$ does not exist at the point $(1, 1)$. In particular this is the case for the Gaussian copula with correlation parameter $\rho > 0$. However, from Theorem 5.3 in [29], for a Gaussian copula with $|\rho| < 1$ it holds that $\lim_{u \to 1} C_{\min}^{(u)}(x, y) = xy$, for $(x, y) \in [0, 1]^2$.

Let $C(x, y) = xy - \frac{1}{2}(1 - |2x - 1|)(1 - 2y - 1^2)$ (for further details see [4]). In this case, $\frac{\partial^2 C}{\partial u \partial v}(1, 1) = 0$. However, we can calculate the limit in (10), and using (12) we obtain

$$G(x, y) = xy^2, \quad g_x(x) = x, \quad g_y(y) = y^2, \quad \theta = 3.$$

We now provide an illustration for a Clayton copula. In particular we observe the sensitivity of $\hat{\theta}_n$ in (22) to the sequence $k_n$ (Fig. 3). We draw the mean curve on 100 samples of size $n = 1000$ (full line) and the empirical standard deviation (dashed lines). Furthermore the mean squared error for $\hat{\theta}_n$ is calculated on 100 samples of size $n = 1000$. 
4. Estimating tail distributions

4.1. Estimating the tails of univariate distributions

The estimation of the tail of a bivariate distribution requires first the estimation of a one-dimensional tail [35,20]. Fix a threshold \( u \) and define \( F_U(x) = \mathbb{P}[X \leq x \mid X > u] \). Let \( X_1, X_2, \ldots \) be a sequence of i.i.d. random variables with unknown distribution function \( F \) and \( \hat{F}_U(u) \) the empirical distribution function evaluated at the threshold \( u \). Recall that the univariate tail may be estimated by

\[
1 - \hat{F}^*_1(u, y) = \frac{1}{n} \left( 1 - \frac{\hat{\xi}}{\hat{\sigma}} \right)^{\frac{1}{\hat{\sigma}}}, \quad \text{if} \; \hat{\xi} \neq 0,
\]

\[
1 - \hat{F}^*_2(x, u_Y) = \frac{1}{n} \left( \hat{\xi} \right), \quad \text{if} \; \hat{\xi} = 0,
\]

where \( \hat{\xi}, \hat{\sigma} \) are the maximum likelihood estimators (MLE) based on the excesses above \( u \). Using (24) we get the estimator, proposed by Smith [46],

\[
1 - \hat{F}^*(y) = \begin{cases}
\frac{N}{n} \left( 1 - \frac{\hat{\xi}}{\hat{\sigma}} \right)^{\frac{1}{\hat{\sigma}}}, & \text{if} \; \hat{\xi} \neq 0, \\
\frac{N}{n} \left( \hat{\xi} \right), & \text{if} \; \hat{\xi} = 0,
\end{cases}
\]

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

4.2. Estimating the tails of bivariate distributions

In this section we present the main construction of this paper. We do indeed propose a POT procedure in order to estimate 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 POT construction stated in Section 4.1. Let \( X \) and \( Y \) be two real valued random variables with different continuous marginal distributions \( F_X \) and \( F_Y \). The structure of the dependence between \( X \) and \( Y \) is represented by copula \( C \).

Construction of the tail estimator:

Given a high threshold \( u \) and \( u_Y := F_Y^{-1}(F_X(u)) \), we introduce the distribution of excesses: \( F_0(x, y) := \mathbb{P}[X - u \leq x, Y - u_Y \leq y \mid X > u, Y > u_Y] \). Using (3) for a large value of \( u \) and \( x > u, y > u_Y \), we can approximate \( F(u, y) \) and \( F(x, u_Y) \) as

\[
F_1^*(u, y) = e^{-\left(-l(-\log(F_X(u)), -\log(F_Y(y)))\right)},
\]

\[
F_2^*(x, u_Y) = e^{-\left(-l(-\log(F_X(x)), -\log(F_Y(u_Y)))\right)},
\]

where \( l \) is the stable tail dependence function defined by (15). We recall that behind approximations (26)-(27), in order to avoid significant bias, 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 2.1 we now know that, for \( u \) around \( x_F \), we can approximate the distribution of excesses with \( C^* G \). So we obtain, for \( x > u, y > u_Y \),

\[
F^*(x, y) = \hat{F}(u, u_Y) \cdot C^* G \left( 1 - g_{\hat{X}, \hat{\sigma}_X} (x - u), 1 - g_{\hat{Y}, \hat{\sigma}_Y} (y - u_Y) \right) \\
+ \hat{F}_1^*(u, y) + \hat{F}_2^*(x, u_Y) - \hat{F}(u, u_Y). 
\]

(28)

Then, we estimate \( F(u, u_Y) \) and \( \hat{F}(u, u_Y) \) in (28) from the data \( \{X_i, Y_i\}_{i=1,...,n} \), using the empirical distribution estimates

\[
\hat{F}(u, u_Y) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq u, Y_i \leq u_Y\}}, \quad \hat{F}(u, u_Y) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i > u, Y_i > u_Y\}}. 
\]

(29)

From (26)–(27) and using the non-parametric estimator (18) we obtain

\[
\hat{F}_1^*(u, y) = \exp\left[-\hat{\ell}(\log(\hat{F}_X(u)), -\log(\hat{F}_Y(y)))\right], \\
\hat{F}_2^*(x, u_Y) = \exp\left[-\hat{\ell}(\log(\hat{F}_X(x)), -\log(\hat{F}_Y(u_Y)))\right],
\]

where \( \hat{F}_X(u) \) and \( \hat{F}_Y(u_Y) \) are the empirical univariate estimators evaluated at respective thresholds, and \( \hat{F}_X^*(x) \) and \( \hat{F}_Y^*(y) \) are one-dimensional POT tail estimators of the marginal distribution functions, defined by (24). Now, using (29)–(31), we can approximate \( F^*(x, y) \), for \( x > u, y > u_Y \), by

\[
\hat{F}^*(x, y) = \left( \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i > u, Y_i > u_Y\}} \right) \left( 1 - \hat{g}_{\hat{X}} (1 - V_{\hat{X}, \hat{\sigma}_X} (x - u)) \right) \\
- \hat{g}_{\hat{Y}} (1 - V_{\hat{X}, \hat{\sigma}_X} (y - u_Y)) + \hat{G} (1 - V_{\hat{X}, \hat{\sigma}_X} (x - u), 1 - V_{\hat{Y}, \hat{\sigma}_Y} (y - u_Y)) \\
+ \hat{F}_1^*(u, y) + \hat{F}_2^*(x, u_Y) - \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq u, Y_i \leq u_Y\}}.
\]

(32)

where \( \hat{g}_{\hat{X}} , \hat{\sigma}_X \) (resp. \( \hat{g}_{\hat{Y}} , \hat{\sigma}_Y \)) are MLE based on the excesses of \( X \) (resp. \( Y \)). Finally we remark that the second threshold in (32) depends on the unknown marginal distribution functions \( F_X \) and \( F_Y \). Then, in order to compute in practice \( \hat{F}^*(x, y) \), we propose to estimate \( u_Y \) by \( \hat{u}_Y = \hat{F}_Y^{-1}(\hat{F}_Y(u)) \), with \( \hat{F}_Y(u) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq u\}} \) and \( \hat{F}_Y^{-1} \) the empirical quantile function of \( Y \).

So we obtain, from (32), the tail estimator for the two-dimensional distribution function for \( x > u \) and \( y > \hat{u}_Y \):

\[
\hat{F}^*(x, y) = \left( \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i > u, Y_i > \hat{u}_Y\}} \right) \left( 1 - \hat{g}_{\hat{X}} (1 - V_{\hat{X}, \hat{\sigma}_X} (x - u)) \right) \\
- \hat{g}_{\hat{Y}} (1 - V_{\hat{X}, \hat{\sigma}_X} (y - \hat{u}_Y)) + \hat{G} (1 - V_{\hat{X}, \hat{\sigma}_X} (x - u), 1 - V_{\hat{Y}, \hat{\sigma}_Y} (y - \hat{u}_Y)) \\
+ \hat{F}_1^*(u, y) + \hat{F}_2^*(x, \hat{u}_Y) - \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq u, Y_i \leq \hat{u}_Y\}}.
\]

(33)

In the case with the same marginal distributions we have a particular case of (32), with the same threshold \( u \) for \( X \) and \( Y \), and we do not need to estimate the second threshold.

**Remark 5.** Note that \( \hat{F}^*(x, y) \) in (33), is only valid for \( x > u \) and \( y > \hat{u}_Y \), when \( u \) is large enough. Having 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 \( \hat{g}_{\hat{X}}, \hat{g}_{\hat{Y}}, \hat{\sigma}_X, \hat{\sigma}_Y \) will suffer from a lack of observations over the thresholds. The compromise will be studied in Sections 5 and 6. More precisely, in Theorem 5.1 we state some sufficient conditions on thresholds \( u_n, \hat{u}_Y \), in terms of the sample size \( n \), in order to obtain consistency properties for our bivariate tail estimator \( \hat{F}^*(x, y) \) (i.e. Theorems 6.1 and 6.2).

### 5. Convergence results in the univariate case

In order to study asymptotic properties of our bivariate tail estimator we present in this section some slight modifications of one-dimensional convergence results from [46, Theorems 3.2 and 8.1]. Incidentally we get asymptotic confidence intervals for the unknown theoretical univariate function \( F(x) \), using Theorem 5.1. 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. \( F(x) = x^{-\alpha}L(x) \), for some slowly varying function \( L(x) \), with \( \alpha > 0 \).

As in [46, Section 3], we shall assume that \( L \) satisfies the following condition:

- **SR2**: \( \frac{L(tx)}{L(t)} = 1 + k(t)\phi(x) + o(\phi(x)) \), as \( x \to \infty \), \( \forall t > 0 \),
where \( \phi(x) > 0 \) and \( \phi(x) \to 0 \) as \( x \to \infty \). Let \( R_\rho \) be the set of \( \rho \)-regularly varying functions. Condition SR2 implies, excluding trivial cases, \( \phi \in R_\rho \), for some \( \rho \leq 0 \), and \( k(t) = c h_\rho(t) \), with \( h_\rho(t) = \int_1^t u^{\rho-1} du \) (for more details see Section 3 in [46] or [24]).

The study of the asymptotic properties of the maximum likelihood estimators of the scale and shape parameters of the generalized Pareto distribution in the POT method has received attention in the literature. For instance, the asymptotic normality of \( \hat{\xi} \) and \( \hat{\sigma} \) in the case of a random threshold in the POT procedure is studied in depth in [14]. Smith [46] examines a slightly different version of the MLE that is based on the excesses over a deterministic threshold and on the second-order condition SR2. For details about the difference between these two approaches see, for instance, Remark 2.3 in [14]. In this paper we follow the approach proposed in [46]. In particular, Theorems 3.2 and 8.1 in [46] are written conditionally on \( N = m_n \), with \( N \) denoting the number of excesses above the threshold. In practice we work with some deterministic threshold \( u \), and \( N \) is considered as random (see Remark 1 in Section 1). Therefore we give the version of Theorem 3.2 in [46]) (resp. Theorem 8.1), Corollary 5.1 (resp. Corollary 5.2), unconditionally on \( N \).

**Corollary 5.1.** Suppose \( L \) satisfies SR2. Let \( n \) be the sample size and \( u_n := \bar{f}(n) \) the threshold, such that \( \bar{f}(n) \to \infty \), for \( n \to \infty \). Let \( N = N_n \) denote the random number of excesses of \( u_n \). We define \( \xi = -\alpha^{-1} \) and \( \sigma_n = \bar{f}(n) \alpha^{-1} \). If

\[
\begin{align*}
 n(1 - F(u_n)) & \xrightarrow{n \to \infty} \infty, \\
 \sqrt{n(1 - F(u_n))c \phi(u_n)} & \xrightarrow{n \to \infty} \mu(\alpha - \rho).
\end{align*}
\]

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

\[
\sqrt{n} \left( \frac{\hat{\sigma}_n}{\sigma_n} - 1 \right) \xrightarrow{n \to \infty} N \left( \frac{\mu(1 - \xi)(1 + 2\xi\rho)}{\mu(1 - \xi)(1 + \rho)}, \frac{2(1 - \xi)}{(1 - \xi)^2} \right).
\]

**Proof.** If (34) and (35) hold then \( N(n(1 - F(u_n))^{-1} \xrightarrow{n \to \infty} 1 \), and (3.2) in [46] holds in probability, i.e.,

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

Hence we conclude with a Skorohod-type construction of probability spaces on which (3.2) in [46] holds almost surely. \( \square \)

**Corollary 5.2.** Suppose all the assumptions of Corollary 5.1 are satisfied. Let \( n \) be the sample size, \( u_n := \bar{f}(n) \to \infty \) and \( z_n := f(n) \to \infty \), for \( n \to \infty \), such that \( (z_n)^{-1} \rho \frac{\phi(\bar{f}(z_n))}{\phi(u_n)} \to 1 \), for \( n \to \infty \) and \( s \in [0, 1] \). Let \( N = N_n \) denote the random number of excesses above \( u_n \). If

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

then

\[
\frac{\sqrt{N}}{\log(f(n))} \left[ 1 - \frac{\hat{F}^+(\bar{f}(f(n)) f(n))}{1 - F(\bar{f}(f(n)) f(n))} \right] \xrightarrow{n \to \infty} N(\nu, \tau^2),
\]

where \( \hat{F}^+ \) as in (25), \( \nu = 0 \) 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 (34)–(36) hold, then (8.7), (8.8) and (8.11) in [46] hold in probability, i.e.,

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

We conclude as for Corollary 5.1. \( \square \)

Note that, in simple cases, we often have \( \phi(x) = x^\rho \); in that case, \( (z_n)^{-1} \rho \frac{\phi(\bar{f}(z_n))}{\phi(u_n)} \to 1 \), for \( n \to \infty \), is automatic satisfied. From Corollary 5.2 the following result can be obtained.

**Theorem 5.1.** Assume that all the assumptions of Corollary 5.2 are satisfied. We use the same notation. If

\[
(z_n)^\rho (n(1 - F(u_n)))^{-1/2} \xrightarrow{n \to \infty} 0,
\]

(37)
then
\[
\frac{\sqrt{N}}{\log(f(n))} \left[ F(\tilde{f}(n)f(n)) - \tilde{F}^*(\tilde{f}(n)f(n)) \right] \xrightarrow{d} N(\nu, \tau^2),
\]
(38)
where \( \tilde{F} \) is the univariate empirical survival function, \( \tilde{F}^* \) is as in (25), \( \nu = 0 \) if \( \rho = 0 \), \( \nu = \frac{\alpha(\alpha+1)(1+\rho)}{1+\alpha-\rho} \) for \( \rho < 0 \) and \( \tau^2 = \alpha^2(1+\alpha)^2 \).

The proof of Theorem 5.1 is postponed to the Appendix. As a byproduct, from (38) it is possible to construct in practice asymptotic confidence intervals for \( F(\tilde{f}(n)f(n)) \).

6. Convergence results in the bivariate case

In this section we provide our main result: the consistency property of our bivariate tail estimator (33) with the convergence rate. We consider:

Remark 6. Let \( n \) be the sample size. We choose, from Theorem 2.1,
\[
u_1(n) := \tilde{F}_1(n) \xrightarrow{n \to \infty} \infty \quad \text{threshold for } X,
\]
\[
u_2(n) := \tilde{F}_2(n) = F^{-1}_Y(\tilde{F}(\tilde{f}(n))) \xrightarrow{n \to \infty} \infty \quad \text{threshold for } Y.
\]

Remark 7. As in Section 4.2, in the following we propose to estimate the second threshold \( \nu_2(n) \) by \( \tilde{F}_2(n) := \tilde{F}_Y^{-1}(\tilde{F}(\tilde{f}(n))) \), with \( \tilde{F}_Y(\tilde{f}(n)) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq \tilde{f}(n)\}} \) and \( \tilde{F}_Y^{-1} \) the empirical quantile function of \( Y \).

In the following we state and prove our consistency result separately in the asymptotically dependent case (Theorem 6.1) and in the asymptotically independent one (Theorem 6.2).

6.1. The asymptotically dependent case

The proof of Theorem 6.1, below, makes use of a result of Einmahl et al. [18] which specifies the asymptotic behavior of \( \hat{L}(x, y) \) uniformly in \( 0 \leq x, y \leq 1 \) and provides a convergence rate (see Theorem A.1 in the Appendix). More precisely, in the asymptotically dependent case, using (20) and applying Corollary 3.1, we obtain the following main result:

Theorem 6.1. Suppose \( F_X \) and \( F_Y \) belong to the maximum domain of attraction of Fréchet, and \( L_X, L_Y \) satisfy Condition SR2. Assume that \( \lambda > 0 \) and that the assumptions of Theorem 2.1 and Corollary 3.1 are satisfied. If sequences \( f_1(n), f_2(n), \tilde{F}_1(n), \tilde{F}_2(n), \) defined by Remark 6, satisfy the conditions of Theorem 5.1 then
\[
\left| \sqrt{n} \left( F^*(x_0, y_0) - F^*(x_n, y_n) \right) \right| \xrightarrow{P} 0,
\]
(39)
with \( x_n = \tilde{f}_1(n), y_n = \tilde{f}_2(n) \). Moreover, if \( \tilde{F}_2(n) \) satisfies the conditions of Theorem 5.1 in probability then
\[
\left| \sqrt{n} \left( F^*(x_0, y_n) - F^*(x_n, y_0) \right) \right| \xrightarrow{P} 0,
\]
(40)
with \( y_n = \tilde{f}_2(n) \). In (39)-(40) we have \( k_n \to \infty, k_n/n \to 0, k_n/N_X \xrightarrow{P} 0, k_n/N_Y \xrightarrow{P} 0, k_n = o(n^{\frac{4(\alpha+1)}{4\alpha}}), \alpha > 0 \).

The proof of Theorem 6.1 is postponed to the Appendix.

Remark 8. Let us study, for a class of examples, the assumption of Theorem 6.1. First if we suppose that the function \( \phi(x) \) in Condition SR2 (Section 5) has the general form \( \phi(x) = x^\phi \), with \( \rho \leq 0 \), then
\[
(z_n)^{-\rho} \frac{\phi(\tilde{F}_2(n)z_n^\phi)}{\phi(\tilde{f}_2(n))} = 1, \quad \forall s \in [0, 1].
\]

For instance this is the case for Burr or Fréchet univariate distributions. Furthermore if we assume that \( F_Y \) belongs to the maximum domain of attraction of Fréchet \((i.e. F_Y(y) = y^{-\alpha}L(y))\), \( F_Y \) has positive density \( f_Y \in R_{1-\alpha} \) and \( \tilde{F}_2(n) \) satisfies the conditions in (34)-(37) then also the estimated second threshold \( \tilde{F}_2(n) \) satisfies, in probability, the conditions in (34)-(37).

We remark that \( \tilde{F}_X(\tilde{f}_1(n)) \) is indeed a high quantile within the sample (see [21]), i.e. \( \tilde{F}_X(\tilde{f}_1(n)) \xrightarrow{P} 1 \) and \( n(1 - \tilde{F}_X(\tilde{f}_1(n))) \xrightarrow{P} \infty \). Then, using Theorem 6.4.14 in [21] and a Skorohod-type construction of probability spaces, we obtain \( \tilde{F}_2(n) \xrightarrow{P} 1 \).
Furthermore, using Condition SR2,
\[
\frac{F_y(f_2(n))}{F_Y(f_2(n))} = \frac{f_2(n)^{-\alpha} L(f_2(n))}{f_2(n)^{-\alpha} L(f_2(n))} = 1 + k \left( \frac{\hat{f}_2(n)}{f_2(n)} \right) \phi(f_2(n)) + o(\phi(f_2(n)))
\]

Using properties of \( k \) and \( \phi \) (see Section 5), we obtain
\[
F_y(f_2(n)) \xrightarrow{\mathbb{P}} 1.
\]

Then \( \hat{f}_2(n) \) satisfies, in probability, the condition of (34):
\[
n(1 - F_y(\hat{f}_2(n))) = F_y(\hat{f}_2(n)) - F_Y(\hat{f}_2(n)) \xrightarrow{\mathbb{P}} 0.
\]

The proof for the conditions in (35)–(37) is completely analogous to that of the condition in (34).

### 6.2. The asymptotically independent case

As noticed in Section 3.2 in the asymptotically independent case we need to introduce a second-order refinement of condition in (10). Under the condition in (23) we obtain the following main result:

**Theorem 6.2.** Suppose \( F_X \) and \( F_Y \) belong to the maximum domain of attraction of Fréchet, and \( L_X, L_Y \) satisfy Condition SR2. Assume that the assumptions of Theorem 2.1, Proposition 3.1 and Corollary A.1 are satisfied. If the sequences \( f_1(n), f_2(n), \hat{f}_1(n), \hat{f}_2(n) \), defined by Remark 6, satisfy the conditions of Theorem 5.1 then
\[
\left| \sqrt{a_n} (F^*(x_n, y_n) - \tilde{F}^*(x_n, y_n)) \right| \xrightarrow{\mathbb{P}} 0,
\]
where \( x_n = f_1(n)f_1(n), y_n = f_2(n)f_2(n) \). Moreover, if \( \hat{f}_2(n) \) satisfies the conditions of Theorem 5.1 in probability then
\[
\left| \sqrt{a_n} (F^*(x_n, y_n) - \tilde{F}^*(x_n, y_n)) \right| \xrightarrow{\mathbb{P}} 0,
\]
with \( \tilde{y}_n = \hat{f}_2(n)f_2(n) \). In (41)–(42) we have \( a_n = n q(k_n/n) \to \infty \) (this implies that \( k_n \to \infty \), \( k_n/n \to 0 \), \( \sqrt{a_n} q_1(q^{-1}(a_n/n)) \to 0, k_n/N_X \to 0, k_n/N_Y \to 0, \) and \( k_n = o(n^{2\alpha/3}) \), for some \( \alpha > 0 \).

The proof of Theorem 6.2 is postponed to the Appendix.

### 7. Illustrations with real data

In this section we present four real cases (see Figs. 4–9) for which we estimate bivariate tail probabilities to illustrate the finite sample properties of our estimator. We analyze the stability of our estimation compared to that of \( \tilde{F}_1^* \), as well as the stability of the estimation of parameter \( \theta \).
7.1. Asymptotically dependent real data

The data sets presented in this section are considered in the recent literature as asymptotically dependent data. Firstly, we consider the loss–ALAE data (for details see Frees and Valdez [23]). Each claim consists of an indemnity payment (the loss, $X$) and an allocated loss adjustment expense (ALAE, $Y$) (see Fig. 4, left). The second data set is 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 (for a detailed description see Lescourret and Robert [33]; see Fig. 4, right).

For the loss–ALAE data we estimate $F(2.10^5, 10^5)$. The empirical probability, defined by Eq. (29), is 0.9506667 and the survival empirical probability is 0.006 (for a comparison using the Ledford and Tawn model, see Beirlant et al. [2]). Fig. 5 shows the sensitivity of $\hat{\theta}$ and $\hat{F}^*$ to the sequence $k_n$ and provides a comparison with the estimator $\hat{F}_1^*$.

For the example from the storm insurance 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 that of $\hat{F}_1^*$, and the estimation of parameter $\theta$ are presented in Fig. 6.

For these two examples our new estimate behaves well, although it seems to have a greater variance than the classical estimator ($\hat{F}_1^*$).
7.2. Asymptotically independent real data

The data set presented in this section is considered in the recent literature as asymptotically independent data. We analyze the wave height versus water level data, recorded during 828 storm events spread over 13 years in front of the Dutch coast near the town of Patten (for details, see Draisma et al. [13]; see Fig. 7).

For the wave height versus water level data we estimate \( F(5.93, 1.87) \). The empirical probability is 0.97584 and the survival empirical probability is 0.00604. The sensitivity of \( \hat{\theta} \) and \( \hat{F}^* \) to the sequence \( k_n \) and the estimation of \( \theta \) are presented in Fig. 8. In view of the results obtained by Draisma et al. [13], it seems plausible to assume asymptotic independence between the wave heights and the water level. In particular, it seems that the coefficient \( \eta \) of the Ledford and Tawn model for these data is smaller than 1. Then, from Remark 2, we expect \( \theta \geq \frac{1}{\eta} > 1 \). In Fig. 8 (left) we do indeed notice that \( \hat{\theta} > 1 \).

In this asymptotically independent case study (Fig. 7), the bias of our new estimate is significantly reduced with respect to that based on the stable tail dependence function (\( \hat{F}_I^* \)).

7.3. A more complex case

We now consider the wave–surge data comprising 2894 bivariate events that occurred during 1971–1977 in Cornwall (UK) (for details see Coles and Tawn [8] or Ramos and Ledford [41]; see Fig. 9).
Fig. 9. Wave–surge data.

Fig. 10. Left: \( \hat{\theta}_{0.02} \); right: \( \hat{F}^*(8.32, 0.51) \) (full line), \( \hat{F}^*_1(8.32, 0.51) \) (dashed line), with the empirical probability indicated with a horizontal line; wave–surge data.

The asymptotic (in)dependence of this data set has been discussed in a great number of papers (for details see Coles and Tawn [8] or Ramos and Ledford [41], Reiss and Thomas [42]). In particular, in the book by Reiss and Thomas [42] the extremal dependence of the two components is tested. Table 13.1 in Reiss and Thomas [42] presents the p-values with respect to the threshold \( c \). Having in mind a significance level of 5%, the null hypothesis (i.e., extremal tail dependence) is always accepted. The p-value increases as \( c \) gets close to zero (that is, equivalent to our \( u_X, u_Y \) becoming larger). However, as \( u_X, u_Y \) tend to infinity we may nevertheless decide to accept the null hypothesis, entailing the accepting of an extremal tail dependence.

For this data set we estimate \( F(8.32, 0.51) \). The empirical probability is 0.9903 and the survival empirical probability is 0.00069. The sensitivity of \( \hat{\theta} \) and \( \hat{F}^* \) to the sequence \( k_n \) and the estimation of \( \theta \) are presented in Fig. 10.

In this case we remark that the estimation of \( \theta \) is not stable with respect to \( k_n \) (see Fig. 10, left). However, the estimation of the bivariate tail seems to have a smaller bias using our procedure (see Fig. 10, right).

8. Conclusion

The paper deals with the inference of the tails of bivariate distribution functions, both in the asymptotically dependent case and in the asymptotically independent one. In contrast to many other papers in the literature, ours does not assume that the margins have the same distribution. Consistency results for our estimator are provided in Theorems 6.1 and 6.2. We also present the results that we obtained for real data examples.
To estimate the tails of bivariate distributions we decomposed the problem into the estimations of the marginal distributions and the dependence structure (i.e. the copula function). The dependence structure is estimated using the stable tail dependence function $t$, well known in the extreme value literature. Our estimation procedure depends on the estimate $\bar{t}$ of the stable tail dependence function. This estimate is known to be biased in the asymptotically independent case. The impact of this bias on our estimation procedure is an interesting open problem which is beyond the scope of this paper.

Acknowledgments

This work was partially supported by the French research national agency (ANR) under the reference ANR-08BLAN-0314-01. The authors thank two anonymous referees for relevant remarks and constructive comments on a previous version of the paper. They also thank Anne-Laure Fougères, Ragnar Norberg and José R. León for fruitful discussions.

Appendix. Proofs and auxiliary results

Proof of Proposition 2.1. We know that

$$C_{u}^{\text{up}}(x, y) = 1 - \frac{C^{\ast}(1 - F_{X, u}^{-1}(x), 1 - u)}{C^{\ast}(1 - u, 1 - u)} - \frac{C^{\ast}(1 - u, 1 - F_{Y, u}^{-1}(y))}{C^{\ast}(1 - u, 1 - u)} + \frac{C^{\ast}(1 - F_{X, u}^{-1}(x), 1 - F_{Y, u}^{-1}(y))}{C^{\ast}(1 - u, 1 - u)}.$$  

Then

$$\lim_{u \to 1} C_{u}^{\text{up}}(x, y) = \lim_{u \to 1} \left[ x + y - 1 + \frac{C^{\ast}(1 - F_{X, u}^{-1}(x), 1 - F_{Y, u}^{-1}(y))}{C^{\ast}(1 - u, 1 - u)} \right].$$ (43)

We introduce the following lemma.

Lemma A [Charpentier and Juri [6, Lemma 6.1)]. Suppose that the random vectors $(X_n, Y_n)$ have continuous, strictly increasing marginals and are such that $\lim_{n \to \infty} (X_n, Y_n) = (X, Y)$ in distribution for some $(X, Y)$. Then

$$\lim_{n \to \infty} \|C_n - C\|_{\infty} = 0,$$

where $C_n$ and $C$ denote the copulas of $(X_n, Y_n)$ and $(X, Y)$, respectively.

Let $(X, Y)$ have distribution function $C$. Note that

$$\mathbb{P}[X > x(1 - u)|X > u, Y > u] = \frac{C^{\ast}(1 - x(1 - u), 1 - u)}{C^{\ast}(1 - u, 1 - u)},$$ (44)

$$\mathbb{P}[Y > y(1 - u)|X > u, Y > u] = \frac{C^{\ast}(1 - u, 1 - y(1 - u))}{C^{\ast}(1 - u, 1 - u)},$$ (45)

$$\mathbb{P}[X > x(1 - u), Y > y(1 - u)|X > u, Y > u] = \frac{C^{\ast}(1 - x(1 - u), 1 - y(1 - u))}{C^{\ast}(1 - u, 1 - u)}.$$ (46)

The distributions in (44)-(46) are respectively the survival conditional distributions of $(\frac{X}{1-u}, \frac{Y}{1-u})$ and $(\frac{X}{1-u}, \frac{Y}{1-u})$, given that $X > u$ and $Y > u$. Since $\partial C^{\ast}(1 - u, 1 - v)/\partial u < 0$ and $\partial C^{\ast}(1 - u, 1 - v)/\partial v < 0$, for all $u, v \in [0, 1)$, it follows that the distributions in (44)-(45) are continuous and strictly increasing.

By hypothesis, we have

$$\lim_{u \to 1} \frac{C^{\ast}(x(1 - u), y(1 - u))}{C^{\ast}(1 - u, 1 - u)} = G(x, y), \quad \text{for all } x, y > 0.$$ (47) 

implying that the expressions in (44)-(45) respectively converge to $g_X(1 - x)$ and $g_Y(1 - y)$ as $u \to 1$, with $g_X(x) := G(x, 1)$, $g_Y(y) := G(1, y)$.

Since copulas are invariant under strictly increasing transformations of the underlying variables, it follows that we can use the conditional distributions in (44)-(45), instead of $F_{X, u}$ and $F_{Y, u}$, to construct $C_{u}^{\text{up}}(x, y)$. Then, from (43) and using Lemma A, Lemma 6.1 in Charpentier and Juri [6], we have

$$\lim_{u \to 1} C_{u}^{\text{up}}(x, y) = \lim_{u \to 1} \left[ x + y - 1 + \frac{C^{\ast}(g_{X}^{-1}(1 - x)(1 - u), g_{Y}^{-1}(1 - y)(1 - u))}{C^{\ast}(1 - u, 1 - u)} \right]$$

$$= x + y - 1 + G(g_{X}^{-1}(1 - x), g_{Y}^{-1}(1 - y)).$$

As in the proof of Theorem 3.1 in [6], the limit in (47) implies that there is a $\theta > 0$ such that $G$ is homogeneous of order $\theta$, i.e. for all $t, x, y > 0$.

$$G(t x, t y) = t^{\theta} G(x, y).$$ (48)
By a discussion of the general solution of functional (48) we obtain the explicit form of $G$:

$$G(x, y) = \begin{cases} x \, g_y \left( \frac{y}{x} \right) & \text{for } x, y \in [0, 1], \\ y \, g_x \left( \frac{x}{y} \right) & \text{for } x, y \in (1, \infty). \end{cases}$$

For this part of the proof we refer the interested reader to Theorem 3.1 in [6].

**Proof of Theorem 2.1.** From (13) we obtain the existence of $a_1(\cdot)$ and $a_2(\cdot)$ such that, for $p := u + x a_1(u)$ and $q := u + y a_2(u)$,

$$V_{\xi_1}(x) = \lim_{u \to x_{\xi_1}} \frac{1 - \frac{F_X(p)}{1 - F_X(u)}}{1 - \frac{F_X(u)}{1 - F_X(p)}} = \lim_{u \to x_{\xi_1}} \mathbb{P}[X \leq p | X > u],$$

$$V_{\xi_2}(y) = \lim_{u \to y_{\xi_2}} \frac{1 - \frac{F_X(q)}{1 - F_X(u)}}{1 - \frac{F_X(u)}{1 - F_X(q)}} = \lim_{u \to y_{\xi_2}} \mathbb{P}[Y \leq q | Y > u].$$ (49) (50)

From $Y \overset{d}{=} F_Y^{-1}(F_X(X))$, we take $u_Y = F_Y^{-1}(F_X(u))$, and from (49)–(50), as $u \to x_{\xi_1}$, we have

$$1 - (1 - V_{\xi_1}(x))(1 - F_X(u)) \sim F_X(u + x a_1(u)),$n

$$1 - (1 - V_{\xi_2}(y))(1 - F_Y(F_Y^{-1}(F_X(u)))) \sim F_Y(F_X^{-1}(F_X(u)) + y a_2(F_Y^{-1}(F_X(u)))).$$

Then

$$\lim_{u \to x_{\xi_1}} \mathbb{P} \left[ \frac{X - u}{a_1(u)} > x, \frac{Y - F_Y^{-1}(F_X(u))}{a_2(F_Y^{-1}(F_X(u)))} > y \right| X > u, Y > F_Y^{-1}(F_X(u)) \right] = \lim_{u \to x_{\xi_1}} \frac{C^*(1 - F_X(u, u + x a_1(u), 1 - F_Y(F_Y^{-1}(F_X(u))) + y a_2(F_Y^{-1}(F_X(u))))}{C^*(1 - F_X(u), 1 - F_Y(F_Y^{-1}(F_X(u))))}$$

$$= \lim_{u \to x_{\xi_1}} \frac{C^*(1 - V_{\xi_1}(x))(1 - F_X(u), (1 - V_{\xi_2}(y))(1 - F_Y(F_Y^{-1}(F_X(u))))}{C^*(1 - F_X(u), 1 - F_Y(F_Y^{-1}(F_X(u))))}$$

$$= \lim_{v \to 1} \frac{C^*(1 - V_{\xi_1}(x)(1 - v), (1 - V_{\xi_2}(y))(1 - v))}{C^*(1 - v, 1 - v)}. (51)$$

Now, if $h := (1 - \xi_1) \xi_1^{-1}$, $\xi_1 \neq 0$ or if $h := e^{-x}$, $\xi_1 = 0$ then $1 - V_{\xi_1}(x) = V_{1,1}(h)$. So (51) becomes $\lim_{v \to 1} C^*(V_{1,1}(h)(1 - v), V_{1,1}(1)(1 - v))/C^*(1 - v, 1 - v)$.

As $C$ satisfies the hypotheses of Proposition 2.1, the above limit is equal to $G(V_{1,1}(h), V_{1,1}(1)) = G(1 - V_{\xi_1}(x), 1 - V_{\xi_2}(y))$. Then

$$\lim_{u \to x_{\xi_1}} \mathbb{P} \left[ \frac{X - u}{a_1(u)} \leq x, \frac{Y - F_Y^{-1}(F_X(u))}{a_2(F_Y^{-1}(F_X(u)))} \leq y \right| X > u, Y > F_Y^{-1}(F_X(u)) \right] = C^*(1 - g_X - V_{\xi_1}(x), 1 - g_Y(1 - V_{\xi_2}(y))). (52)$$

Since the limit is a continuous distribution function (as $C^*$, g and the GPD are), (52) can be strengthened to uniform convergence (see e.g. [21, p. 552]). Then (14) follows. \hfill \Box

**Proof of Theorem 5.1.** To begin with, we work conditionally on $N_n = m_n$. First we have to prove that

$$\overline{t}_{m_n} \left[ F(u_{m_n} z_{m_n}) - F^*(u_{m_n} z_{m_n}) \right] \overset{a}{\longrightarrow} \mathcal{N}(\nu, \tau^2),$$

(53)

with $\overline{t}_{m_n} = \sqrt{m_n} \left( \frac{1}{1 - \frac{1}{n} \sum_{i=1}^n 1_{X_i \in z_{m_n}}} \right) = \frac{\sqrt{m_n}}{\log(z_{m_n}) F(u_{m_n} z_{m_n})}$. To this end we need to prove that

$$\frac{F(u_{m_n} z_{m_n})}{F(u_{m_n} z_{m_n})} \overset{p}{\longrightarrow} 1,$$ (54)

and then, using Theorem 8.1 in Smith [46] and the Slutsky theorem, we obtain Eq. (53). To prove (54) we use Corollary 1 in Einmahl [17]. To apply this result we have to choose two sequences of positive numbers.
In particular, we choose \( \{m_n\}_{n=1}^{\infty} \) (number of excesses on a sample of size \( n \)) such that \( m_n \leq n, \lim_{n \to \infty} m_n = \infty, \lim_{n \to \infty} \frac{m_n}{n} = 0 \) (see Remark 1 in Section 1) and \( \{\sqrt{m_n} \alpha_n\}_{n=1}^{\infty} \), where \( \alpha_n \) is an arbitrary sequence of positive numbers such that \( \lim_{n \to \infty} \alpha_n = \infty \).

Then, using Corollary 1 in Einmahl [17], we have, for \( u_m \), \( z_{mn} \geq F^{-1}(1 - \frac{m_n}{n}) \),

\[
\left( \frac{n}{\sqrt{m_n} \alpha_n} \right) \frac{f(u_m z_{mn})}{F(u_m z_{mn})} \rightarrow 0.
\]

We choose \( \alpha_n \) such that for large \( n \),

\[
\exists \ c > 0 \ : \ 0 < \frac{\sqrt{m_n} \alpha_n}{n F(u_m z_{mn})} \leq c.
\]

In the Fréchet case we have \( L(x) = x^\alpha F(x), \) for \( \alpha > 0 \) and \( \forall t > 0, \frac{L(t)}{L(x)} = 1 + k(t) \phi(x) + o(\phi(x)) \) for \( x \to \infty \). Then, using assumptions (8.7) and (8.8) of Theorem 8.1 in [46], we obtain

\[
\frac{F(u_m z_{mn})}{F(u_m)} = z_m^{-\alpha} [1 + k(z_m) \phi(u_m) + o(\phi(u_m))].
\]

Hence \( \frac{n F(u_m z_{mn})}{\sqrt{m_n}} \) is equal to \( \frac{n F(u_m)}{\sqrt{m_n}} [z_m^{-\alpha} (1 + k(z_m) \phi(u_m) + o(\phi(u_m))) \right] \) which, for \( n \) large, can be approximated by

\[
\sqrt{m_n} z_m^{-\alpha} (1 + k(z_m) \phi(u_m) + o(\phi(u_m))).
\]

Assume that now \( \frac{z_m}{\sqrt{m_n}} \nrightarrow 0 \), that is the analogue of the condition of (37) conditionally on \( N_n = m_n \). Then the properties of \( k \) and \( \phi \) insure that the right-hand side of (56) increases to infinity; hence one can choose \( \alpha_n \) satisfying (55). To conclude the proof, we use assumption (37) and a Skorohod-type argument. \( \square \)

**Proof of Theorem 6.1.** To prove (39) we first observe, using Corollary 3.1, Proposition A.1 and the analogue of the Kolmogorov–Smirnov Theorem in dimension 2 (e.g. see [16]), that, for \( n \to \infty \),

\[
\sqrt{k_n} \left| C^x G^x \left( 1 - g_x(1 - V_{g_x, \alpha_x} J_1(n) f_1(n) - J_1(n)) \right) - 1 - g_y(1 - V_{g_y, \alpha_y} J_2(n) f_2(n) - J_2(n)) \right| \rightarrow 0.
\]

Furthermore \( r_n \left| \sum_{i=1}^{n} \left( X_i J_1(n), Y_i J_2(n) \right) - F(J_1(n), J_2(n)) \right| \rightarrow 0 \), with \( r_n \ll \sqrt{n} \). Finally, using Corollary 3.1, Theorem 5.1, we obtain convergence (39). If \( J_2(n) \) satisfies conditions of Theorem 5.1 in probability then with the same proof structure we obtain (40). \( \square \)

**Proof of Proposition 3.1.** Under assumptions of Proposition 3.1, as in the proof of Lemma 6.1 in [13] we obtain

\[
\sup_{0 < x, y \leq 1} \left| \frac{1}{\alpha_n} \left( \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(R_i > -k_0 x + 1: R_i > -k_0 y + 1) - G(x, y) \right) - W(x, y) \right| \rightarrow_{\text{a.s.}} 0,
\]

where \( \alpha_n = n q(k_0/n) \) and \( W(x, y) \) is a zero-mean Gaussian process with \( \mathbb{E}(W(x_1, y_1) W(x_2, y_2)) = G(x_1 \wedge x_2, y_1 \wedge y_2) \). Then, in particular,

\[
\psi_n \sup_{0 < x, y \leq 1} \left| \frac{1}{\alpha_n} \left( \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(R_i > -k_0 x + 1: R_i > -k_0 y + 1) - G(x, y) \right) \right| \rightarrow_{\text{p}} 0,
\]

with \( \psi_n \ll \sqrt{\alpha_n} = \sqrt{n} q(k_0/n) \) and \( \hat{G} \) as in (20). Finally for the marginals \( g_x \) and \( g_y \) we have

\[
\psi_n \sup_{0 < x \leq 1} \left| \hat{g}_x(x) - g_x(x) \right| \rightarrow_{\text{p}} 0, \quad \psi_n \sup_{0 < y \leq 1} \left| \hat{g}_y(y) - g_y(y) \right| \rightarrow_{\text{p}} 0,
\]

with \( \hat{g}_x \) and \( \hat{g}_y \) as in (20). \( \square \)
Proof of Theorem 6.2. Under assumptions of Theorem 6.2 and Proposition 3.1 we obtain asymptotic convergence results for \( \widehat{G}(x, y) \), \( \widehat{G}_x(x) \) and \( \widehat{G}_y(y) \), with convergence rate \( \psi_n \ll \sqrt{n q(k_0/n)} \) and \( \widehat{G}_x, \widehat{G}_y, \widehat{G} \) as in (20).

With the same proof structure as Theorem 6.1, using Corollary A.1 and Proposition A.1 we obtain convergence (41). Moreover, if \( \widehat{f}_2(n) \) satisfies the conditions of Theorem 5.1 in probability then we obtain (42). \( \square \)

Auxiliary results

For the sake of clarity, in the following we state a slightly adapted version of Theorem 2.2 of Einmahl et al. [18]. This result plays a central role in the statement of our consistency properties.

Theorem A.1 (Adapted Version of Theorem 2.2 in Einmahl et al. [18]). Assume that there exists a limit \( R(x, y) \) in (16) such that, for some \( \alpha > 0 \),

\[
\frac{1}{t} P\left(1 - F_X(X) \leq tx, 1 - F_Y(Y) \leq ty \right) - R(x, y) = O(t^\alpha), \quad \text{as } t \to 0, \tag{57}
\]

uniformly for \( \max(x, y) \leq 1, x, y \geq 0 \). Let \( k_n \to \infty, k_n/n \to 0 \) and \( k_n = o(n^{2\alpha}) \). 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

\[
\sup_{0 < x, y \leq 1} \left| \sqrt{k_n} \hat{l}(x, y) - l(x, y) \right| + B(x, y) \xrightarrow{P} 0, \quad n \to \infty,
\]

where \( B(x, y) := W(x, y) - R_1(x, y)W_1(x) - R_2(x, y)W_2(y) \), with \( W \) a continuous mean-zero Gaussian process on \([0, x] \times [0, y]\) with covariance structure \( \mathbb{E}(W(x_1, y_1)W(x_2, y_2)) = R(x_1 \wedge x_2, y_1 \wedge y_2) \) and with marginal processes defined by \( W_1(x) = W([0, x] \times [0, \infty)), W_2(y) = W([0, \infty] \times [0, y]) \).

Note that (57) is a second-order condition quantifying the speed of convergence in (16), and condition \( k_n = o(n^{2\alpha}) \) gives an upper bound on the speed with which \( k_n \) can grow to infinity. This upper bound is related to the speed of convergence in (57) by \( \alpha \). If \( C \) is a copula that is twice continuously differentiable on \([0, 1]^2\) then (57) holds for any \( \alpha \geq 1 \). Furthermore, it is easily seen that \( \hat{l}(x, y) + \hat{R}(x, y) = \frac{[k_n x]^{1/2} [k_n y]^{1/2}}{[k_n x]^{1/2} [k_n y]^{1/2}} \) almost surely, for each \( 0 < x, y \leq 1 \), where \( [z] \) is the smallest integer \( \geq z \). Then under the assumption of Theorem A.1, we can easily obtain a Gaussian approximation for \( R(x, y) \approx \hat{R}(x, y) \).

Note that the asymptotic variance of \( \sqrt{k_n} (\hat{l}(x, y) - l(x, y)) \), in Theorem A.1, vanishes in the asymptotically independent case. Then, Theorem A.1 in the case \( \lambda = 0 \), can be rewritten as:

Corollary A.1. Assume that, for some \( \alpha > 0 \),

\[
\frac{1}{t} P\left(1 - F_X(X) \leq tx, 1 - F_Y(Y) \leq ty \right) - O(t^\alpha), \quad \text{as } t \to 0, \tag{57}
\]

uniformly for \( \max(x, y) \leq 1, x, y \geq 0 \). Let \( k_n \to \infty, k_n/n \to 0 \) and \( k_n = o(n^{2\alpha}) \). Then it holds that

\[
\sup_{0 < x, y \leq 1} \left| \sqrt{k_n} (\hat{l}(x, y) - l(x, y)) \right| \xrightarrow{P} 0, \quad n \to \infty.
\]

The following result is used in the proof of Theorems 6.1 and 6.2. For the sake of clarity, we have postponed to here the statement and the demonstration of this result.

Proposition A.1. Let \( V_{\xi, \sigma}(x) \) the generalized Pareto distribution (GPD) and \( \hat{\xi}_n, \hat{\sigma}_n \) the maximum likelihood estimators of the parameters \( \xi = -\alpha^{-1} < 0 \) and \( \sigma = u_0 \alpha^{-1} \), in the case that is not conditional on \( N \). If all the conditions of Corollary 5.1 hold then

\[
P_{\max} \sup_{x \in [0, +\infty)} \left| V_{\hat{\xi}_n, \hat{\sigma}_n}(x) - V_{\xi, \sigma}(x) \right| \xrightarrow{P} 0, \quad \text{where } P_{\max} \xrightarrow{P} 0.
\]

Proof. Using Corollary 5.1 we obtain, for each point \( x \in [0, +\infty) \),

\[
P_{\max} \left[ V_{\hat{\xi}_n, \hat{\sigma}_n}(x) - V_{\xi, \sigma}(x) \right] = P_{\max} \left[ \left( 1 - \frac{\xi x}{\sigma} \right)^{\frac{1}{\xi}} \left( 1 - \frac{\xi x}{\hat{\sigma}_n} \right)^{\frac{1}{\xi}} \right] \xrightarrow{P} 0, \tag{58}
\]

where \( P_{\max} \xrightarrow{P} 0 \). Finally, applying a stochastic version of Polya’s Theorem (see [25]), as \( V_{\xi, \sigma}(x) \) is a continuous distribution function, the convergence in (58) holds uniformly on \([0, +\infty)\). \( \square \)