ESAIM: Probability and Statistics Will be set by the publisher
URL: http://www.emath.fr/ps/

PLUG-IN ESTIMATION OF LEVEL SETS IN A NON-COMPACT SETTING WITH APPLICATIONS IN MULTIVARIATE RISK THEORY

Elena Di Bernardino[1], Thomas Laloë[2], Véronique Maume-Deschamps[3] and Clémentine Prieur[4]

Abstract. This paper deals with the problem of estimating the level sets $L(c) = \{F(x) \geq c\}$, with $c \in (0,1)$, of an unknown distribution function F on $\mathbb{R}^2_+$. A plug-in approach is followed. That is, given a consistent estimator F_n of F, we estimate $L(c)$ by $L_n(c) = \{F_n(x) \geq c\}$. In our setting no compactness property is a priori required for the level sets to estimate. We state consistency results with respect to the Hausdorff distance and the volume of the symmetric difference. Our results are motivated by applications in multivariate risk theory. In this sense we also present simulated and real examples which illustrate our theoretical results.

1991 Mathematics Subject Classification. 62G05,62G20,60E05,91B30.

The dates will be set by the publisher.

Introduction

The problem: background and motivations

The problem of estimating level sets of an unknown function (for instance of a density function and more recently a regression function) has received many attention in the last decade. In particular the estimation of density level sets has been studied in Polonik [26], Tsybakov [34], Baíllo *et al.* [3], Baíllo [2], Cadre [8], Rigollet and Vert [28]. The estimation of regression level sets in a compact setting has been analyzed in Cavalier [10], Biau *et al.* [6], Laloë [22].
An alternative approach, based on the geometric properties of the compact support sets, has been presented by Hartigan [20], Cuevas and Fraiman [12], Cuevas and Rodríguez-Casal [14]. The problem of estimating general level sets under compactness assumptions has been discussed by Cuevas *et al.* [13]. The asymptotic behaviour of minimum volume sets and of a generalized quantile process is analyzed by Polonik [27].

The motivation of this research field lies in a large number of possible applications: mode estimation (Müller and Sawitzki [24]; Polonik [26]), clustering (Biau *et al.* [6]), abnormal behavior in a system (Baíllo *et al.* [4]; Baíllo *et*

Keywords and phrases: Level sets, distribution function, plug-in estimation, Hausdorff distance, Conditional Tail Expectation.

[1] Université de Lyon, Université Lyon 1, ISFA, Laboratoire SAF, 50 Avenue Tony Garnier, 69366 Lyon, France. e-mail: elena.di-bernardino@univ-lyon1.fr

[2] Université de Nice Sophia-Antipolis, Laboratoire J-A Dieudonné, Parc Valrose, 06108 Nice Cedex 02. e-mail: laloe@unice.fr

[3] Université de Lyon, Université Lyon 1, ISFA, Laboratoire SAF, 50 avenue Tony Garnier, 69366 Lyon, France. e-mail: veronique.maume@univ-lyon1.fr

[4] Université Joseph Fourier, Tour IRMA, MOISE-LJK B.P. 53 38041 Grenoble, France. e-mail: clementine.prieur@imag.fr

al. [3]; Baíllo [2]), study of spherical data (Cuevas *et al.* [13]), set estimation (Cuevas and Rodríguez-Casal [14]).

Non compact case: our proposal

In this paper we consider the problem of estimating the level sets of a bivariate distribution function F. As in the density or regression case, this research field is motivated by practical applications. In particular the purpose of this paper is to use the estimation of level sets of F in order to estimate some new bivariate risk measures.

First, we provide a consistent estimator of the level set

$$L(c) = \{F(x) \geq c\}, \quad \text{for } c \in (0, 1).$$

To this end we consider a *plug-in* approach (e.g. see Baíllo *et al.* [3]; Rigollet and Vert [28]; Cuevas *et al.* [13]), that is, $L(c)$ is estimated by

$$L_n(c) = \{F_n(x) \geq c\}, \quad \text{for } c \in (0, 1),$$

where F_n is a consistent estimator of F.
The regularity properties of F and F_n as well as the consistency properties of F_n will be specified in the statements of our theorems. A special issue in our work is that we do not assume any compactness property for the level sets we estimate. This requires special attention in the statement of the problem.
In order to provide consistency results we define two proximity criteria between sets. A standard choice is the volume of the symmetric difference. Another natural criterion is given from the Hausdorff distance that corresponds to an intuitive notion of "physical proximity" between sets. Our consistency properties are stated with respect to these two criteria under reasonable assumptions on F and F_n (Theorems 2.1 - 3.1). These results are based on a slight modification of Proposition 3.1 in the PhD Thesis of Rodríguez-Casal [29] (Proposition 1.1).

Concerning our application in bivariate risk theory we recall first the definition of the bivariate Value-at-Risk (for details see Embrechts and Puccetti [17]). Then we introduce a new definition for the bivariate Conditional Tail Expectation. We propose an estimator for this new risk measure using plug-in approach for level sets and provide consistency result (Theorem 4.1).

The paper is organized as follows. We introduce some notation, tools and technical assumptions in Section 1. Consistency and asymptotic properties of our estimator of $L(c)$ are given in Sections 2 and 3. In Section 4 we present some applications in the field of multivariate risk theory. Illustrations with simulated and real data are presented in Section 5. Section 5.2.2 summarizes and briefly mentions directions for future research. Finally, some auxiliary results and more technical proofs are postponed to Section 6.

1. Notation and preliminaries

In this section we introduce some notation and tools which will be useful later.

Let $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$, ${\mathbb{R}^2_+}^* = \mathbb{R}^2_+ \setminus (0, 0)$, $\mathcal{F}$ the set of continuous distribution functions $\mathbb{R}^2_+ \to [0, 1]$ and $F \in \mathcal{F}$. Given an *i.i.d* sample $\{X_i\}_{i=1}^n$ in $\mathbb{R}^2_+$ with distribution function F, we denote by $F_n(\cdot) = F_n(X_1, X_2, \dots, X_n, \cdot)$ an estimator of F based on the finite sample $(X_1, X_2, \dots, X_n)$. We restrict ourselves to $\mathbb{R}^2_+$ for convenience but the following results are completely adaptable in $\mathbb{R}^2$. This choice is motivated essentially by our application in risk theory proposed in Section 4, where random variables will be losses then defined in a positive support.

Define, for $c \in (0, 1)$, the *upper c-level set* of $F \in \mathcal{F}$ and its plug-in estimator

$$L(c) = \{x \in \mathbb{R}^2_+ : F(x) \geq c\}, \quad L_n(c) = \{x \in \mathbb{R}^2_+ : F_n(x) \geq c\},$$

and

$$\{F = c\} = \{x \in \mathbb{R}_+^2 : F(x) = c\}.$$

In addition, given $T > 0$, we set

$$L(c)^T = \{x \in [0,T]^2 : F(x) \geq c\}, \;\; L_n(c)^T = \{x \in [0,T]^2 : F_n(x) \geq c\},$$

$$\{F = c\}^T = \{x \in [0,T]^2 : F(x) = c\}.$$

Furthermore, for any $A \subset \mathbb{R}_+^2$ we denote by ∂A its boundary.

Note that $\{F = c\}$ can be a portion of quadrant $\mathbb{R}_+^2$ instead of a set of Lebesgue measure null in $\mathbb{R}_+^2$ (that is the presence of plateau at level c). In the following we will introduce suitable conditions in order to avoid the presence of plateau in the graph of F.

In the metric space $(\mathbb{R}_+^2, d)$, where d stands for the Euclidean distance, we denote by $B(x, \rho)$ the closed ball centered on x and with positive radius ρ. Let $B(S, \rho) = \bigcup_{x \in S} B(x, \rho)$, with S a closed set of $\mathbb{R}_+^2$. For $r > 0$ and $\zeta > 0$, define

$$E = B(\{x \in \mathbb{R}_+^2 : \mid F(x) - c \mid \leq r\}, \zeta),$$

and, for a twice differentiable function F,

$$m^{\triangledown} = \inf_{x \in E} \|(\nabla F)_x\|, \qquad M_H = \sup_{x \in E} \|(HF)_x\|,$$

where $(\nabla F)_x$ is the gradient vector of F evaluated at x and $\|(\nabla F)_x\|$ its Euclidean norm, $(HF)_x$ the Hessian matrix evaluated in x and $\|(HF)_x\|$ its matrix norm induced by the Euclidean norm.

For sake of completeness, we recall that if A_1 and A_2 are compact sets in $(\mathbb{R}_+^2, d)$, the Hausdorff distance between A_1 and A_2 is defined by

$$d_H(A_1, A_2) = \inf\{\rho > 0 : A_1 \subset B(A_2, \rho), A_2 \subset B(A_1, \rho)\},$$

or equivalently by

$$d_H(A_1, A_2) = \max\left\{\sup_{x \in A_1} d(x, A_2), \sup_{x \in A_2} d(x, A_1)\right\},$$

where $d(x, A_2) = \inf_{y \in A_2} \parallel x - y \parallel$.

Finally, we introduce the following assumption (e.g. see Tsybakov [34]; Cuevas *et al.* [13]).

H: There exist $\gamma > 0$ and $A > 0$ such that, if $\mid t - c \mid \leq \gamma$ then $\forall \;\; T > 0$ such that $\{F = c\}^T \neq \emptyset$ and $\{F = t\}^T \neq \emptyset$,

$$d_H(\{F = c\}^T, \{F = t\}^T) \leq A \mid t - c \mid.$$

Assumption **H** is satisfied under mild conditions. Proposition 1.1 below is a slight modification of Proposition 3.1 in the PhD Thesis of Rodríguez-Casal [29] in order to deal with non-compact sets.

Proposition 1.1. *Let $c \in (0,1)$. Let $F \in \mathcal{F}$ be twice differentiable on $\mathbb{R}_+^{2*}$. Assume there exist $r > 0$, $\zeta > 0$ such that $m^{\triangledown} > 0$ and $M_H < \infty$. Then F satisfies Assumption* **H**, *with $A = \frac{2}{m^{\triangledown}}$.*

The proof is postponed to Section 6.

Remark 1. Under assumptions of Proposition 1.1, F is continuous and $m^{\triangledown} > 0$, there is no plateau in the graph of F for each level t such that $\mid t - c \mid \leq r$. Furthermore from Theorem 1 in Rossi [31] we know that each half-line in $\mathbb{R}_+^2$, parallel to one of the axis, meets $\{F = t\}$ in a connected set of points. As a consequence we obtain that $\{F = t\}$ is a curve in the quadrant $\mathbb{R}_+^2$. In particular, from $m^{\triangledown} > 0$, for a fixed x we have to consider all corresponding values y in a specific interval (we refer the interested reader to Remark 2.1 in Rossi [31] or Remark 3.1 in Rossi [30]). In this case the plane curve $\{F = t\}$ is not the graph of a function but it has the following monotonic property. We consider (x, y), $(x', y') \in \{F = t\}$, if $x < x'$ then $y \geq y'$, if $y < y'$ then $x \geq x'$. In particular if we suppose that each component of $(\nabla F)_x$ is greater than zero in E then $\{F = t\}$ is a monotone decreasing curve in $\mathbb{R}^2_+$. Finally under assumptions of Proposition 1.1 we obtain $\partial L(c)^T = \{F = c\}^T = \{F = c\} \cap [0, T]^2$ (we refer for details to Theorem 3.2 in Rodríguez-Casal [29]).

2. Consistency in terms of the Hausdorff distance

In this section we study the consistency properties of $L_n(c)^T$ with respect to the Hausdorff distance between $\partial L_n(c)^T$ and $\partial L(c)^T$. The metric d_H is not always completely successful in capturing the shape properties: two sets can be very close in d_H and still have quite different shapes. Following Cuevas and Rodríguez-Casal [14] and Cuevas *et al.* [13], a way to reinforce the notion of visual proximity between two sets provided by d_H is to impose the proximity of the respective boundaries.

From now on we note, for $n \in \mathbb{N}^*$,

$$\|F - F_n\|_\infty = \sup_{x \,\in\, \mathbb{R}^2_+} \mid F(x) - F_n(x) \mid,$$

and for $T > 0$

$$\|F - F_n\|^T_\infty = \sup_{x \,\in\, [0,T]^2} \mid F(x) - F_n(x) \mid .$$

The following result can be considered an adapted version of Theorem 2 in Cuevas *et al.* [13].

Theorem 2.1. *Let $c \in (0, 1)$. Let $F \in \mathcal{F}$ be twice differentiable on $\mathbb{R}^{2*}_+$. Assume that there exist $r > 0$, $\zeta > 0$ such that $m^\nabla > 0$ and $M_H < \infty$. Let $T_1 > 0$ such that for all $t : \mid t - c \mid \leq r$, $\partial L(t)^{T_1} \neq \emptyset$. Let $(T_n)_{n \in \mathbb{N}^*}$ be an increasing sequence of positive values. Assume that, for each n, F_n is continuous with probability one and that*

$$\|F - F_n\|_\infty \to 0, \quad a.s.$$

Then

$$d_H(\partial L(c)^{T_n}, \partial L_n(c)^{T_n}) = O(\|F - F_n\|_\infty), \qquad a.s.$$

Theorem 2.1 states that $d_H(\partial L(c)^{T_n}, \partial L_n(c)^{T_n})$ converges to zero at least at the same rate as $\|F - F_n\|_\infty$.

Remark 2. Theorem 2.1 provides an asymptotic result for a fixed level c. In particular following the proof of Theorem 2.1 (postponed to Section 6) we remark that, for n large enough,

$$d_H(\partial L(c)^{T_n}, \partial L_n(c)^{T_n}) \leq \; 6\, A \, \|F - F_n\|^{T_n}_\infty, \quad a.s.,$$

where $A = \frac{2}{m^\nabla}$. Note that in the case c is close to one the constant A could be large. In this case, we will need a large number of data to get a reasonable estimation.

Note that the empirical distribution estimator does not satisfy continuity assumption imposed in Theorem 2.1. However, in order to overcome this problem it can be considered a smooth version of this estimator (e.g. see Chaubey and Sen [11]).

3. L_1 consistency

The previous section was devoted to the consistency of L_n in terms of the Hausdorff distance. We consider now another consistency criterion: the consistency of the volume (in the Lebesgue measure sense) of the symmetric difference between $L(c)^{T_n}$ and $L_n(c)^{T_n}$. This means that we define the distance between two subsets A_1 and A_2 of $\mathbb{R}_2^+$ by

$$d_\lambda(A_1, A_2) = \lambda(A_1 \triangle A_2),$$

where λ stands for the Lebesgue measure on $\mathbb{R}^2$ and $\triangle$ for the symmetric difference.

Let us introduce the following assumption:

A1 There exists a positive increasing sequence v_n such that $v_n \underset{n\to\infty}{\to} \infty$ and

$$v_n \, \|F - F_n\|_\infty \underset{n\to\infty}{\to} 0, \quad a.s.$$

We now establish our consistency result in terms of the volume of the symmetric difference.

Theorem 3.1. *Let $c \in (0,1)$. Let $F \in \mathcal{F}$ be twice differentiable on $\mathbb{R}_+^{2*}$. Assume that there exist $r > 0$, $\zeta > 0$ such that $m^\nabla > 0$ and $M_H < \infty$. Assume that for each n, with probability one, F_n is measurable and that Assumption* **A1** *is satisfied. Let p_n be an increasing positive sequence such that $p_n = o(v_n)$. Then for any increasing positive sequence $(T_n)_{n\in\mathbb{N}^*}$ such that for all $t : |\, t - c \,| \leq r$, $\partial L(t)^{T_1} \neq \emptyset$ and $T_n = o(v_n/p_n)$, it holds that*

$$p_n \, d_\lambda(L(c)^{T_n}, L_n(c)^{T_n}) \underset{n\to\infty}{\to} 0, \quad a.s.$$

Theorem 3.1 provides a convergence rate, which is closely related to the choice of the sequence T_n. A convergence rate p_n close to (but slower than) v_n implies choosing a sequence T_n whose divergence rate is small. Remark that Theorem 3.1 does not require continuity assumption on F_n.

Remark 3. Assumptions of Theorems 2.1 and 3.1 are satisfied for a quite large class of classical bivariate distribution functions; for instance independent copula and exponential marginals, Farlie-Gumbel-Morgenstern (FGM), Clayton or Survival Clayton copulas and Burr marginals.

4. Application in bivariate risk theory

From the usual definition in the univariate setting we know that the quantile function Q_X provides a point which accumulates a probability α to the left tail and $1-\alpha$ to the right tail. More precisely, given an univariate continuous and strictly monotonic loss distribution function F_X,

$$Q_X(\alpha) = F_X^{-1}(\alpha), \quad \forall\, \alpha \in (0,1). \tag{1}$$

The notion of univariate quantile function Q_X is used in risk theory to define an univariate measure of risk: the Value-at-Risk (VaR). This measure is defined as

$$\mathrm{VaR}_\alpha(X) = Q_X(\alpha), \quad \forall\, \alpha \in (0,1). \tag{2}$$

In the last decade several attempts have been made to provide a multidimensional generalization of univariate quantile. We refer to Serfling [32] for a complete review on the topic. For example, using (1), Massé and Theodorescu [23] defined multivariate quantiles as halfplanes and Koltchinskii [21] provided a general treatment of multivariate quantiles as inversions of mappings. Tibiletti [33], Fernández-Ponce and Suárez-Lloréns [18] and Belzunce *et al.* [5] defined multivariate quantiles as level curves.

Following the general ideas of Embrechts and Puccetti [17] and Nappo and Spizzichino [25] an intuitive generalization of the VaR measure in the case of a bidimensional loss distribution function F is given by its α-quantile curves. More precisely:

Definition 4.1. *For $\alpha \in (0,1)$ and $F \in \mathcal{F}$, the bidimensional Value-at-Risk at probability level α is the boundary of its α-level set, i.e. $VaR_\alpha(F) = \partial L(\alpha)$.*

As well described in Tibiletti [33], imposing $\alpha = \frac{1}{2}$, we get a natural extension of bidimensional median. For details about a parametric formulation of the quantile curve $\partial L(\alpha)$ see Belzunce *et al.* [5]. For details about its properties see for instance Fernández-Ponce and Suárez-Lloréns [18] (and references therein) and Nappo and Spizzichino [25].
Then, using a bivariate estimator F_n as in Sections 2 - 3, we can define our estimator of the bivariate Value-at-Risk by

$$\mathrm{VaR}_\alpha(F_n) = \partial L_n(\alpha).$$

Moreover, under assumptions of Theorem 2.1, we obtain a consistency result, with respect to the Hausdorff distance, for the $\mathrm{VaR}_\alpha(F_n)$ on the quadrant $\mathbb{R}^2_+$ i.e.

$$d_H(\mathrm{VaR}_\alpha(F)^{T_n}, \mathrm{VaR}_\alpha(F_n)^{T_n}) = O(\|F - F_n\|_\infty), \quad a.s.$$

As in the univariate case, the bidimensional VaR at a predetermined level α does not give any information about thickness of the upper tail of the distribution function. This is a considerable shortcoming of VaR measure because in practice we are not only concerned with frequency of the default but also with the severity of loss in case of default. In order to overcome this problem, another risk measure has recently received growing attention in insurance and finance literature: Conditional Tail Expectation (CTE). Following Artzner *et al.* [1] and Dedu and Ciumara [15], for a continuous loss distribution function F_X the CTE at level α is defined by

$$\mathrm{CTE}_\alpha(X) = \mathbb{E}[\, X \,|\, X \geq \mathrm{VaR}_\alpha(X)\,],$$

where $\mathrm{VaR}_\alpha(X)$ is the univariate VaR in (2). For a comprehensive treatment and for references to the extensive literature on $\mathrm{VaR}_\alpha(X)$ and $\mathrm{CTE}_\alpha(X)$ one may refer to Denuit *et al.* [16].

Several bivariate generalizations of the classical univariate CTE have been proposed; mainly using as conditioning events the total risk or some univariate extreme risk in the portfolio. We recall for instance:

$$\mathbb{E}[\,(X,Y)\,|\, X+Y > Q_{X+Y}(\alpha)\,], \quad \mathbb{E}[\,(X,Y)\,|\, \min\{X,Y\} > Q_{\min\{X,Y\}}(\alpha)\,]$$
$$\text{and} \quad \mathbb{E}[\,(X,Y)\,|\, \max\{X,Y\} > Q_{\max\{X,Y\}}(\alpha)\,].$$

The interested reader is referred to Cai and Li [9] for further details. Starting from these general considerations we propose to study a new bivariate version of the Conditional Tail Expectation (Definition 4.2 below).

Let us first introduce the following assumption:

- **A2**: (X,Y) is a positive random vector with absolutely continuous distribution (with respect to the Lebesgue measure λ on $\mathbb{R}^2$) with density $f_{(X,Y)} \in L^{1+\epsilon}(\lambda)$, with $\epsilon > 0$ and $\mathbb{E}(X^2) < \infty$, $\mathbb{E}(Y^2) < \infty$.

Definition 4.2. *Consider a random vector (X,Y) satisfying Assumption* **A2***, with associate distribution function $F \in \mathcal{F}$. For $\alpha \in (0,1)$, we define*

(1) *the bivariate α-Conditional Tail Expectation*

$$\mathrm{CTE}_\alpha(\mathrm{X},\mathrm{Y}) = \mathbb{E}[(\mathrm{X},\mathrm{Y})|(\mathrm{X},\mathrm{Y}) \in \mathrm{L}(\alpha)] = \begin{pmatrix} \mathbb{E}[\,X \,|\, (X,Y) \in L(\alpha)\,] \\ \mathbb{E}[\,Y \,|\, (X,Y) \in L(\alpha)\,] \end{pmatrix}.$$

(2) *the estimated bivariate α-Conditional Tail Expectation*

$$\widehat{\mathrm{CTE}}_\alpha(X,Y) = \begin{pmatrix} \dfrac{\sum_{i=1}^n X_i 1_{\{(X_i,Y_i)\in L_n(\alpha)\}}}{\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)\}}} \\ \dfrac{\sum_{i=1}^n Y_i 1_{\{(X_i,Y_i)\in L_n(\alpha)\}}}{\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)\}}} \end{pmatrix}. \tag{3}$$

Note that this bivariate Conditional Tail Expectation is a natural extension of the univariate one. Moreover, if X and Y are identically distributed with a symmetric copula then $\mathbb{E}[\,X \mid (X,Y) \in L(\alpha)\,] = \mathbb{E}[\,Y \mid (X,Y) \in L(\alpha)\,]$.

Contrarily to the generalizations of the classical univariate CTE above, our $\mathrm{CTE}_\alpha(X,Y)$ does not use an aggregate variable in order to analyze the bivariate risk's issue. Conversely, with a geometric approach, $\mathrm{CTE}_\alpha(X,Y)$ deals with the simultaneous joint damages considering the bivariate dependence structure of data in a specific risk's area ($L(\alpha)$).

Let $\alpha \in (0,1)$ and $F \in \mathcal{F}$. We introduce truncated versions of the theoretical and estimated CTE_α:

$$\mathrm{CTE}_\alpha^T(X,Y) = \mathbb{E}[(X,Y)|(X,Y) \in L(\alpha)^T],$$

$$\widehat{\mathrm{CTE}}_\alpha^T(X,Y) = \begin{pmatrix} \dfrac{\sum_{i=1}^n X_i 1_{\{(X_i,Y_i)\in L_n(\alpha)^T\}}}{\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)^T\}}} \\ \dfrac{\sum_{i=1}^n Y_i 1_{\{(X_i,Y_i)\in L_n(\alpha)^T\}}}{\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)^T\}}} \end{pmatrix},$$

where $L(\alpha)^T$ and $L_n(\alpha)^T$ are the truncated versions of theoretical and estimated upper α-level set defined in Section 1.

Theorem 4.1. *Under Assumption* **A2***, Assumptions of Theorem 3.1 and with the notation of Theorem 3.1, it holds that*

$$\beta_n \Big| \, \mathrm{CTE}_\alpha^{T_n}(X,Y) - \widehat{\mathrm{CTE}}_\alpha^{T_n}(X,Y) \, \Big| \underset{n\to\infty}{\to} 0, \quad a.s., \tag{4}$$

where $\beta_n = \min\{\, p_n^{\frac{\epsilon}{2(1+\epsilon)}}\,, \sqrt{n}\,\}$*, with* $\epsilon > 0$.

The convergence in (4) must be interpreted as a componentwise convergence. In the case of a bounded density function $f_{(X,Y)}$ we obtain $\beta_n = \min\{\sqrt{p_n}, \sqrt{n}\,\}$.

Remark 4. Starting from Theorem 4.1, it could be interesting to consider the convergence $\big|\,\mathrm{CTE}_\alpha(X,Y) - \widehat{\mathrm{CTE}}_\alpha^{T_n}(X,Y)\,\big|$. We remark that in this case the speed of convergence will also depend on the convergence rate to zero of $\mathbb{P}[(X,Y) \in L(\alpha) \setminus L(\alpha)^{T_n}] \leq \mathbb{P}[X \geq T_n \text{ or } Y \geq T_n]$, for $n \to \infty$.

5. Illustrations

5.1. Estimation of the level sets

In this section we confront our estimator of level sets of the distribution function with various simulated samples. We consider two distribution functions which satisfy assumptions of Theorem 3.1: independent copula with exponential marginals and Survival Clayton copula with Burr marginals.

The plug-in estimation of level sets is constructed using the empirical estimator F_n. We take $T_n = n^{0.45}$. This choice is compatible with assumptions of Theorem 3.1. We consider a random grid of 10000 points in $[0,T_n]^2$. We provide a Monte Carlo approximation for $\lambda(L(\alpha)^{T_n} \triangle L_n(\alpha)^{T_n})$ (averaged on 100 iterations), for different values of α and n. The results are gathered in Tables $1-2$.

α	n= 500	n= 1000	n= 2000
0.10	0.331	0.326	0.223
0.24	0.519	0.391	0.249
0.38	0.591	0.469	0.396
0.52	1.057	0.906	0.881
0.66	1.222	0.989	0.904
0.80	1.541	1.367	1.334

Table 1. Distribution with independent and exponentially distributed marginals with parameter 1 and 2 respectively. Approximated $\lambda(L(\alpha)^{T_n} \triangle L_n(\alpha)^{T_n})$.

α	n= 500	n= 1000	n= 2000
0.10	0.697	0.633	0.536
0.24	0.893	0.872	0.809
0.38	0.971	0.911	0.879
0.52	1.001	0.982	1.229
0.66	1.569	1.522	1.413
0.80	2.377	2.269	2.175

Table 2. Distribution with Survival Clayton copula with parameter 1 and Burr$(2,1)$ marginals. Approximated $\lambda(L(\alpha)^{T_n} \triangle L_n(\alpha)^{T_n})$.

As expected, the greater n is, the better the estimations are. We remark that values in Tables $1-2$ can be considered good approximations of $\lambda(L(\alpha)^{T_n} \triangle L_n(\alpha)^{T_n})$ with respect to the square $[0,T_n]^2$. Moreover we note that for big values of α we need more data to get a good estimation of the level sets. This may come from the fact that when α grows the gradient of the distribution function decreases to zero and the constant A grows significantly (see proof of Theorem 3.1 in Section 6).

5.2. **Estimation of** $\mathrm{CTE}_\alpha(X,Y)$

5.2.1. *Simulated data*

In order to evaluate the performance of our estimator we present here some simulated cases of estimation of $\mathrm{CTE}_\alpha(X,Y)$, for different values of level α. To compare the estimated results with the theoretical ones we consider cases for which we can calculate (with Maple) the explicit value of the theoretical $\mathrm{CTE}_\alpha(X,Y)$. However our estimator can be applied to much more general cases: starting from the good performance of the level sets estimation (as in cases of Section 5.1) we can expect a good estimation of $\mathrm{CTE}_\alpha(X,Y)$.

We remark that distributions presented in Tables 3 - 5 satisfy assumptions of Theorem 4.1. Furthermore in these cases we know that $\mathbb{P}[X \geq T_n \text{ or } Y \geq T_n]$ decays to zero with a greater convergence rate than β_n in Theorem 4.1. Then, following Remark 4 (in Section 4), in Tables 3 - 5 we compare $\widehat{\mathrm{CTE}}_\alpha^{T_n}(X,Y)$ with the theoretical $\mathrm{CTE}_\alpha(X,Y)$.

In the following we denote $\overline{\widehat{\mathrm{CTE}}_\alpha^{T_n}}(X,Y) = \big(\overline{\widehat{\mathrm{CTE}}_\alpha^{T_n,1}}(X,Y), \overline{\widehat{\mathrm{CTE}}_\alpha^{T_n,2}}(X,Y)\big)$ the mean (coordinate by coordinate) of $\widehat{\mathrm{CTE}}_\alpha^{T_n}(X,Y)$ on 100 simulations. We denote $\widehat{\sigma} = (\widehat{\sigma}_1, \widehat{\sigma}_2)$ the empirical standard deviation (coordinate by coordinate) with

$$\widehat{\sigma}_1 = \sqrt{\frac{1}{99}\sum_{j=1}^{100}\left(\widehat{\mathrm{CTE}}_\alpha^{T_n,1}(X,Y)_j - \overline{\widehat{\mathrm{CTE}}_\alpha^{T_n,1}}(X,Y)\right)^2}$$

relatives to the first coordinate (resp. $\widehat{\sigma}_2$ relatives to the second one).

We denote RMSE$=$ $(\mathrm{RMSE}_1, \mathrm{RMSE}_2)$ the relative mean square error (coordinate by coordinate) with

$$\mathrm{RMSE}_1 = \sqrt{\frac{1}{100}\sum_{j=1}^{100}\left(\frac{\widehat{\mathrm{CTE}}_\alpha^{T_n,1}(X,Y)_j - \mathrm{CTE}_\alpha^{T_n,1}(X,Y)}{\mathrm{CTE}_\alpha^{T_n,1}(X,Y)}\right)^2}$$

relatives to the first coordinate of $\mathrm{CTE}_\alpha^{T_n}(X,Y)$ (resp. RMSE_2 relatives to the second one).

The plug-in estimation of level sets is constructed using the empirical estimator F_n of the bivariate distribution function with $n = 1000$. We choose $T_n = n^{0.45}$. This choice is compatible with assumptions of Theorem 4.1. Results are gathered in Tables 3 - 5.

α	$\mathrm{CTE}_\alpha(X,Y)$	$\overline{\widehat{\mathrm{CTE}}_\alpha^{T_n}}(X,Y)$	$\widehat{\sigma}$	RMSE
0.10	(0.627, 0.627)	(0.603, 0.656)	(0.031, 0.031)	(0.062, 0.068)
0.24	(0.761, 0.761)	(0.774, 0.731)	(0.061, 0.071)	(0.082, 0.130)
0.38	(0.896, 0.896)	(0.927, 0.871)	(0.072, 0.076)	(0.087, 0.119)
0.52	(1.051, 1.051)	(1.086, 1.128)	(0.094, 0.082)	(0.095, 0.107)
0.66	(1.246, 1.246)	(1.281, 1.322)	(0.127, 0.101)	(0.102, 0.101)
0.80	(1.531, 1.531)	(1.545, 1.611)	(0.157, 0.161)	(0.105, 0.117)

TABLE 3. (X,Y) with independent and exponentially distributed components with parameter 2.

α	$\mathrm{CTE}_\alpha(X,Y)$	$\widehat{\mathrm{CTE}}_\alpha^{T_n}(X,Y)$	$\widehat{\sigma}$	RMSE
0.10	(1.255, 0.627)	(1.233, 0.624)	(0.061, 0.023)	(0.051, 0.054)
0.24	(1.521, 0.761)	(1.514, 0.803)	(0.074, 0.039)	(0.048, 0.075)
0.38	(1.792, 0.896)	(1.793, 0.948)	(0.096, 0.055)	(0.053, 0.084)
0.52	(2.102, 1.051)	(2.087, 1.111)	(0.118, 0.076)	(0.056, 0.092)
0.66	(2.492, 1.246)	(2.477, 1.311)	(0.169, 0.108)	(0.068, 0.101)
0.80	(3.061, 1.531)	(3.056, 1.602)	(0.313, 0.153)	(0.102, 0.111)

TABLE 4. (X,Y) with independent and exponentially distributed components with parameter 1 and 2 respectively.

α	$\mathrm{CTE}_\alpha(X,Y)$	$\widehat{\mathrm{CTE}}_\alpha^{T_n}(X,Y)$	$\widehat{\sigma}$	RMSE
0.10	(1.188, 1.229)	(1.189, 1.238)	(0.061, 0.035)	(0.039, 0.029)
0.24	(1.448, 1.366)	(1.462, 1.365)	(0.065, 0.037)	(0.046, 0.031)
0.38	(1.727, 1.505)	(1.751, 1.536)	(0.082, 0.046)	(0.049, 0.037)
0.52	(2.049, 1.666)	(2.063, 1.713)	(0.091, 0.061)	(0.051, 0.045)
0.66	(2.454, 1.875)	(2.457, 1.951)	(0.117, 0.104)	(0.057, 0.068)
0.80	(3.039, 2.202)	(3.037, 2.322)	(0.192, 0.165)	(0.063, 0.108)

TABLE 5. (X,Y) with Clayton Copula with parameter 1, F_X exponential distribution with parameter 1, F_Y Burr$(4,1)$ distribution.

In Table 6 below, we show that for high levels (here $\alpha = 0.9$), one needs to use large samples (here $n > 2500$) to get reasonable estimates of CTE_α. We consider for the purpose (X, Y) independent and exponentially distributed with respective parameters 1 and 2. The theoretical value is $\mathrm{CTE}_{0.9}(X,Y) = (3.78, 1.89)$. In this case we need between 2500 and 5000 data to get the same performances as for lower level (see Table 4).

n	500	1000	1500	2000	2500	5000
$\widehat{\sigma}$	(0.919, 0.419)	(0.568, 0.319)	(0.511, 0.294)	(0.382, 0.239)	(0.348, 0.223)	(0.307, 0.151)
RMSE	(0.242, 0.221)	(0.151, 0.172)	(0.133, 0.165)	(0.101, 0.144)	(0.093, 0.129)	(0.092, 0.108)

TABLE 6. Evolution of $\widehat{\sigma}$ and RMSE in terms of sample size n for $\alpha = 0.9$; (X,Y) independent and exponentially distributed components with parameter 1 and 2 respectively.

The bad properties of the estimate for sample sizes less than 2500 can be explained by the fact that for high levels, the constant A is large (see proof of Theorem 3.1 in Section 6), but also by the fact that for $\alpha = 0.9$ the empirical estimate F_n of F is not the one to choose.

5.2.2. *Real data*

We consider here the estimation of CTE_α in a real case: **Loss-ALAE data** in the log scale (for details see Frees and Valdez [19]). Each claim consists of an indemnity payment (the loss, X) and an allocated loss adjustment expense (ALAE, Y). Examples of ALAE are the fees paid to outside attorneys, experts, and investigators used to defend claims.

The data size is $n = 1500$. Let $T_n = n^{0.4}$. Again our estimator is constructed using the empirical estimator F_n. Results about the estimation of $\mathrm{CTE}_\alpha^{T_n}$ are gathered in Table 7. In Figure 1 we represent data, estimated level sets and estimated bivariate Conditional Tail Expectation for several values of α.

α	0.10	0.24	0.38
$\widehat{\mathrm{CTE}}_\alpha^{T_n}$	(9.937, 9.252)	(10.361, 9.566)	(10.731, 9.728)
α	0.52	0.66	0.80
$\widehat{\mathrm{CTE}}_\alpha^{T_n}$	(11.096, 10.011)	(11.518, 10.315)	(12.057, 10.758)

Table 7. $\widehat{\mathrm{CTE}}_\alpha^{T_n}$ for Loss-ALAE data in log scale, with different values of level α.

In this real setting the estimation of CTE_α can be used in order to quantify the mean of the Loss (resp. ALAE) in the log scale conditionally to the fact that Loss and ALAE data belong jointly to the specific risk's area $L(\alpha)$.

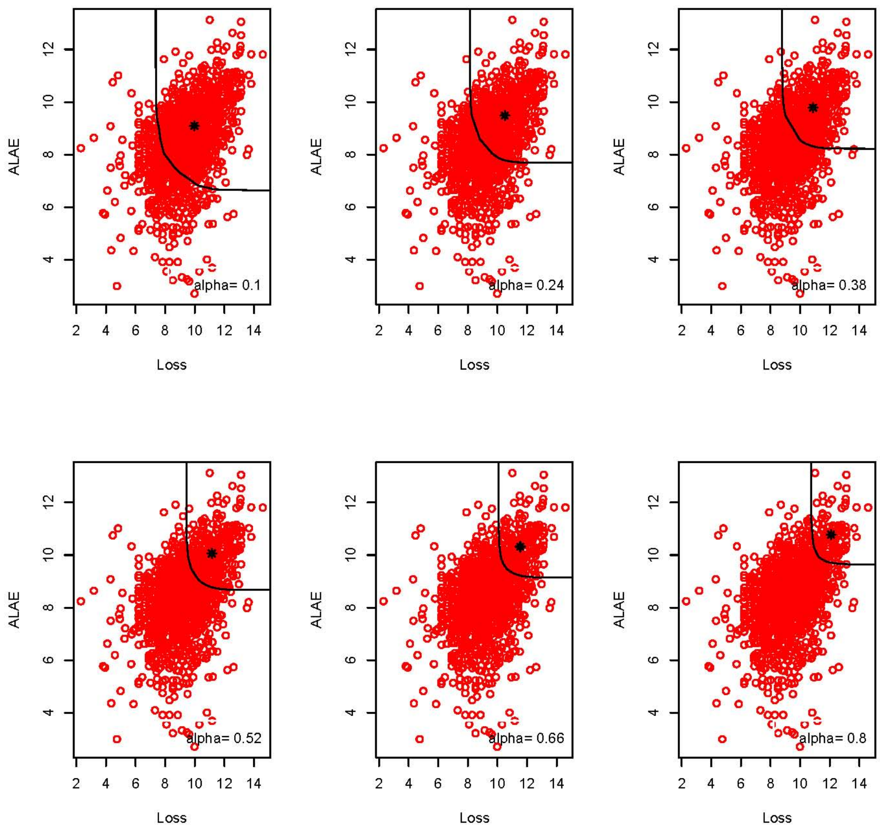

FIGURE 1. Loss-ALAE data in log scale: boundary of estimated level sets (line) and $\widehat{\mathrm{CTE}}_{\alpha}^{T_n}$ (star) for different values of α.

CONCLUSION

In this paper we have provided convergence results for the plug-in estimator of the level sets of an unknown distribution function on $\mathbb{R}_+^2$ in terms of Hausdorff distance and volume of the symmetric difference. In this setting we have proposed and estimated a new bivariate risk measure: $\mathrm{CTE}_{\alpha}(X, Y)$. A future work comparing this bivariate Conditional Tail Expectation with existing risk measures in terms of classical properties (monotonicity, translation invariance, homogeneity, ...), dependence structure, behavior with respect to different risk scenarios, is in preparation.

6. PROOFS

Proof of Proposition 1.1

Take $T > 0$ such that for all $t : \mid t - c \mid \leq r, \; \{F = t\}^T \neq \emptyset$, (from assumptions of Proposition 1.1 we know that such an T there exists).

Let $x \in \{z \in [0,T]^2 : | F(z) - c | \leq r\}$. Define for $\overline{\lambda} \in \mathbb{R}$

$$y_{\overline{\lambda}} \equiv y_{\overline{\lambda},x} = x + \overline{\lambda} \frac{(\nabla F)_x}{\|(\nabla F)_x\|},$$

so $\|y_{\overline{\lambda}} - x\| = | \overline{\lambda} |$. From the differentiability properties of F and using Taylor's formula we have, for $| \overline{\lambda} | < \zeta$

$$F(y_{\overline{\lambda}}) = F(x) + (\nabla F)_x^T (y_{\overline{\lambda}} - x) + \frac{1}{2}(y_{\overline{\lambda}} - x)^T (HF)_{\overline{x}} (y_{\overline{\lambda}} - x),$$

with $\overline{x}$ a point on the segment between x and $y_{\overline{\lambda}}$. So

$$F(y_{\overline{\lambda}}) = F(x) + \overline{\lambda} \|(\nabla F)_x\| + \frac{\overline{\lambda}^2}{2\|(\nabla F)_x\|^2} (\nabla F)_x^T (HF)_{\overline{x}} (\nabla F)_x.$$

By the Cauchy Schwarz Inequality we obtain

$$F(y_{\overline{\lambda}}) \geq F(x) + \overline{\lambda} \|(\nabla F)_x\| - \frac{\overline{\lambda}^2}{2\|(\nabla F)_x\|} \|(HF)_{\overline{x}} (\nabla F)_x\|$$

and

$$F(y_{\overline{\lambda}}) \leq F(x) + \overline{\lambda} \|(\nabla F)_x\| + \frac{\overline{\lambda}^2}{2\|(\nabla F)_x\|} \|(HF)_{\overline{x}} (\nabla F)_x\|.$$

Since $\|(HF)_{\overline{x}} (\nabla F)_x\| \leq \|(HF)_{\overline{x}}\| \|(\nabla F)_x\|$, we have

$$F(x) + \overline{\lambda} \|(\nabla F)_x\| - \frac{\overline{\lambda}^2}{2} \|(HF)_{\overline{x}}\| \leq F(y_{\overline{\lambda}}) \leq F(x) + \overline{\lambda} \|(\nabla F)_x\| + \frac{\overline{\lambda}^2}{2} \|(HF)_{\overline{x}}\|.$$

Since $\overline{x} \in E$ and $M_H < \infty$ we obtain

$$F(x) + \overline{\lambda} \|(\nabla F)_x\| - \frac{\overline{\lambda}^2}{2} M_H \leq F(y_{\overline{\lambda}}) \leq F(x) + \overline{\lambda} \|(\nabla F)_x\| + \frac{\overline{\lambda}^2}{2} M_H. \tag{5}$$

For $0 < \overline{\lambda} < \zeta$, we have from the left side of (5)

$$F(y_{\overline{\lambda}}) \geq F(x) + \overline{\lambda} \|(\nabla F)_x\| - \frac{\overline{\lambda}^2}{2} M_H \geq F(x) + \overline{\lambda} m^{\triangledown} - \frac{\overline{\lambda}^2}{2} M_H.$$

We assume now that $M_H > 0$ (the case $M_H = 0$ is trivial).

For $x \in \{z \in [0,T]^2 : | F(z) - c | \leq r\}$ and $0 < \overline{\lambda} < \zeta \wedge \frac{m^{\triangledown}}{M_H}$ we have then

$$F(y_{\overline{\lambda}}) \geq F(x) + \frac{\overline{\lambda}}{2} m^{\triangledown}, \tag{6}$$

Similarly using the right side of (5) we obtain for $0 < \overline{\lambda} < \zeta \wedge \frac{m^{\triangledown}}{M_H}$ and $x \in \{z \in [0,T]^2 : | F(z) - c | \leq r\}$,

$$F(y_{-\overline{\lambda}}) \leq F(x) - \frac{\overline{\lambda}}{2} m^{\triangledown}. \tag{7}$$

Define

$$\gamma = \left(\frac{m^{\triangledown}}{4} \left(\zeta \wedge \frac{m^{\triangledown}}{M_H} \right) \right) \wedge r > 0.$$

Suppose that $t = c+\varepsilon$, $0 < \varepsilon \leq \gamma$. Let $x \in [0,T]^2$ such that $F(x) = t = c+\varepsilon$ then $x \in \{z \in [0,T]^2 : \mid F(z) - c \mid \leq r\}$. Take now

$$0 < \overline{\lambda} = \frac{2\varepsilon}{m^\nabla} < \zeta \wedge \frac{m^\nabla}{M_H}.$$

We obtain from (7)

$$F(y_{-\overline{\lambda}}) \leq F(x) - \frac{\overline{\lambda}}{2} m^\nabla = c + \varepsilon - \varepsilon = c.$$

From the continuity property of F, we deduce that there exists y between x and $y_{-\overline{\lambda}}$ such that $F(y) = c$ and we have

$$\|x - y\| \leq \|x - y_{-\overline{\lambda}}\| = \mid \overline{\lambda} \mid = \frac{2\varepsilon}{m^\nabla} = \frac{2}{m^\nabla} \mid t - c \mid .$$

So we have proved that

$$\sup_{x \in \{F=t\}^T} d(x, \{F = c\}^T) \leq \frac{2}{m^\nabla} \mid t - c \mid .$$

Similarly, take $x \in [0,T]^2$ such that $F(x) = c$ and use (6) to obtain

$$\sup_{x \in \{F=c\}^T} d(x, \{F = t\}^T) \leq \frac{2}{m^\nabla} \mid t - c \mid .$$

The proof in case $t < c$ is completely analogous. So F satisfies Assumption **H** (see Section 1) with $A = \frac{2}{m^\nabla}$. □

Proof of Theorem 2.1

Under assumptions of Theorem 2.1, we can always take $T_1 > 0$ such that for all $t : \mid t - c \mid \leq r$, $\partial L(t)^{T_1} \neq \emptyset$. Then for each n, for all $t : \mid t - c \mid \leq r$, $\partial L(t)^{T_n}$ is a non-empty (and compact) set on $\mathbb{R}^2_+$.
In each $[0,T_n]^2$, from Proposition 1.1, Assumption **H** (Section 1) is satisfied with

$$\gamma = \left(\frac{m^\nabla}{4} \left(\zeta \wedge \frac{m^\nabla}{M_H} \right) \right) \wedge r > 0 \quad \text{ and } \quad A = \frac{2}{m^\nabla}.$$

First we have to find a bound for $\sup_{x \in \partial L(c)^{T_n}} d(x, \partial L_n(c)^{T_n})$.

Take $x \in \partial L(c)^{T_n}$ and define $\varepsilon_n = 2 \, \|F - F_n\|_\infty^{T_n}$. Using $\|F - F_n\|_\infty \to 0$, *a.s.*, for $n \to \infty$, then $\varepsilon_n \to 0$, *a.s.*, for $n \to \infty$. So with probability one there exists n_0 such that $\forall \, n \geq n_0$, $\varepsilon_n \leq \gamma$.

Since for all $t : \mid t - c \mid \leq r$ $\partial L(t)^{T_n} \neq \emptyset$, from Assumption **H**, there exist $u_n \equiv u_x^{\varepsilon_n}$ and $l_n \equiv l_x^{\varepsilon_n}$ in $[0,T_n]^2$ such that

$$F(u_n) = c + \varepsilon_n; \quad d(x, u_n) \leq A \, \varepsilon_n,$$
$$F(l_n) = c - \varepsilon_n; \quad d(x, l_n) \leq A \, \varepsilon_n.$$

Suppose now $\|F - F_n\|_\infty^{T_n} > 0$ (the other case is a trivial one). In this case

$$F_n(u_n) = c + \varepsilon_n + F_n(u_n) - F(u_n) \geq c + \varepsilon_n - \|F - F_n\|_\infty^{T_n} = c + 2\|F - F_n\|_\infty^{T_n} - \|F - F_n\|_\infty^{T_n} > c,$$

and in a similar way we can prove that $F_n(l_n) < c$.

As $F_n(l_n) < c$ and $F_n(u_n) > c$, with u_n and l_n in $[0,T_n]^2$, there exists $z_n \in \partial L_n(c)^{T_n} \cap B(u_n, d(u_n, l_n))$ with

$$d(z_n,x) \leq d(z_n,u_n)+d(u_n,x) \leq d(u_n,l_n)+d(u_n,x) \leq d(u_n,x)+d(x,l_n)+d(u_n,x) \leq 3\,A\,\varepsilon_n = 6\,A\|F-F_n\|_\infty^{T_n}.$$

Hence, for $n \geq n_0$

$$\sup_{x\in\partial L(c)^{T_n}} d(x,\partial L_n(c)^{T_n}) \leq\, 6\,A\,\|F-F_n\|_\infty^{T_n}.$$

Let us now bound $\sup_{x\in\partial L_n(c)^{T_n}} d(x,\partial L(c)^{T_n})$.

Take $x \in \partial L_n(c)^{T_n}$. From the *a.s.* continuity of F_n we obtain $F_n(x) = c$, *a.s.*, so

$$\mid F(x) - c \mid \leq \mid F(x) - F_n(x) \mid \leq\, \|F-F_n\|_\infty^{T_n} \,\leq\, \varepsilon_n, \quad a.s.$$

Remember that $\forall\, n \geq n_0$, $\varepsilon_n \leq \gamma$, *a.s.* Then from Assumption **H** $d(x,\partial L(c)^{T_n}) \leq\, A\,\mid F(x)-c \mid \leq\, A\,\|F-F_n\|_\infty^{T_n}$. We can conclude that with probability one, for $n \geq n_0$

$$\sup_{x\in\partial L_n(c)^{T_n}} d(x,\partial L(c)^{T_n}) \leq\, A\,\|F-F_n\|_\infty^{T_n}.$$

We obtain for $n \geq n_0$, $d_H(\partial L(c)^{T_n},\partial L_n(c)^{T_n}) \leq\, 6\,A\,\|F-F_n\|_\infty^{T_n}$, then

$$d_H(\partial L(c)^{T_n},\partial L_n(c)^{T_n}) = O(\|F-F_n\|_\infty^{T_n}), \quad a.s.$$

Hence the result. $\square$

Proof of Theorem 3.1

Under assumptions of Theorem 3.1, we can always take $T_1 > 0$ such that for all $\,t : |\,t-c\,| \leq r,\ \partial L(t)^{T_1} \neq \emptyset$. Then for each n, for all $\,t : |\,t-c\,| \leq r,\ \partial L(t)^{T_n}$ is a non-empty (and compact) set on $\mathbb{R}_+^2$.
We consider a positive sequence ε_n such that $\varepsilon_n \underset{n\to\infty}{\to} 0$. For each $n \geq 1$ the random sets $L(c)^{T_n} \triangle L_n(c)^{T_n}$, $Q_{\varepsilon_n} = \{x \in [0,T_n]^2 : \mid F - F_n \mid \leq \varepsilon_n\}$ and $\widetilde{Q}_{\varepsilon_n} = \{x \in [0,T_n]^2 : \mid F-F_n \mid > \varepsilon_n\}$ are measurable and

$$\lambda(L(c)^{T_n} \;\triangle\; L_n(c)^{T_n}) \;=\; \lambda(L(c)^{T_n} \;\triangle\; L_n(c)^{T_n} \;\cap\; Q_{\varepsilon_n}) \;+\; \lambda(L(c)^{T_n} \;\triangle\; L_n(c)^{T_n} \;\cap\; \widetilde{Q}_{\varepsilon_n}).$$

Since $L(c)^{T_n} \triangle L_n(c)^{T_n} \cap Q_{\varepsilon_n} \subset \{x \in [0,T_n]^2 : c-\varepsilon_n \leq F < c+\varepsilon_n\}$ we obtain

$$\lambda(L(c)^{T_n} \triangle L_n(c)^{T_n}) \leq \lambda(\{x \in [0,T_n]^2 : c-\varepsilon_n \leq F < c+\varepsilon_n\}) + \lambda(\widetilde{Q}_{\varepsilon_n}).$$

From Assumption **H** (Section 1) and Proposition 1.1, if $2\,\varepsilon_n \leq \gamma$ then

$$d_H(\partial L(c+\varepsilon_n)^{T_n},\partial L(c-\varepsilon_n)^{T_n}) \leq 2\,\varepsilon_n\,A.$$

So we can write

$$\lambda(\{x \in [0,T_n]^2 : c-\varepsilon_n \leq F < c+\varepsilon_n\}) \leq\, 2\,\varepsilon_n\,A\,2\,T_n.$$

If we now choose

$$\varepsilon_n = o\left(\frac{1}{p_n\,T_n}\right) \tag{8}$$

we obtain that for n large enough $2\,\varepsilon_n \leq \gamma$ and

$$p_n\,\lambda(\{x \in [0,T_n]^2 : c-\varepsilon_n \leq F < c+\varepsilon_n\}) \underset{n\to\infty}{\to} 0,\ a.s.$$

Let us now prove that $p_n\,\lambda(\widetilde{Q}_{\varepsilon_n}) \underset{n\to\infty}{\to} 0, \quad a.s.$

From Assumption **A1** (Section 3) we have $v_n\,\|F-F_n\|_\infty \underset{n\to\infty}{\to} 0,\ a.s.$

Take ε_n such that

$$\varepsilon_n = \frac{1}{v_n}. \tag{9}$$

Then there exist n_0 such that $v_n \, \|F - F_n\|_\infty \leq 1, \;\; a.s. \; \forall \, n \geq n_0$. So for all $\; n \geq n_0, \;\; \lambda(\widetilde{Q}_{\varepsilon_n}) = 0 \;$ and obviously $p_n \, \lambda(\widetilde{Q}_{\varepsilon_n}) = 0$. As $p_n = o(v_n)$ we can choose ε_n that satisfies (8) and (9). Hence the result. $\square$

Proof of Theorem 4.1

We only prove the result for the first coordinate of $\mathrm{CTE}_\alpha(X,Y)$ (the proof is similar for the second one).

We introduce these two preliminary results (Lemma 6.1 and 6.2):

Lemma 6.1. *Under Assumption* **A2**, *Assumptions of Theorem 3.1 and with the notation of Theorem 3.1, it holds that*

$$p_n^{\frac{\epsilon}{2(1+\epsilon)}} \left| \mathbb{E}[\, X \,|\, (X,Y) \in L(\alpha)^{T_n} \,] - \mathbb{E}[\, X \,|\, (X,Y) \in L_n(\alpha)^{T_n} \,] \right| \underset{n \to \infty}{\to} 0, \quad a.s.,$$

with $\epsilon > 0$.

Proof of Lemma 6.1

From Assumption **A2** and Theorem 3.1 we obtain

$$p_n^{\frac{\epsilon}{2(1+\epsilon)}} \left| \mathbb{P}[(X,Y) \in L(\alpha)^{T_n} \; \triangle \; L_n(\alpha)^{T_n}] \right| \; \leq \; p_n^{\frac{\epsilon}{2(1+\epsilon)}} \, d_\lambda\big(L(\alpha)^{T_n}, L_n(\alpha)^{T_n}\big)^{\frac{\epsilon}{1+\epsilon}} \;\; \| \;\; f \;\; \|_{1+\epsilon} \underset{n \to \infty}{\to} 0, \quad a.s. \tag{10}$$

As a straightforward consequence we find

$$p_n^{\frac{\epsilon}{2(1+\epsilon)}} \left| \mathbb{P}[(X,Y) \in L(\alpha)^{T_n}] - \mathbb{P}[(X,Y) \in L_n(\alpha)^{T_n}] \right| \underset{n \to \infty}{\to} 0, \quad a.s.$$

Using Assumption **A2** we also obtain

$$\begin{aligned} p_n^{\frac{\epsilon}{2(1+\epsilon)}} & \left| \int_{L(\alpha)^{T_n}} x \, f_{X,Y}(x,y) \lambda(\mathrm{d}x \, \mathrm{d}y) - \int_{L_n(\alpha)^{T_n}} x \, f_{X,Y}(x,y) \lambda(\mathrm{d}x \, \mathrm{d}y) \right| \\ & \leq \; p_n^{\frac{\epsilon}{2(1+\epsilon)}} \;\; \mathbb{E}[X^2]^{\frac{1}{2}} \;\, d_\lambda\big(L(\alpha)^{T_n}, L_n(\alpha)^{T_n}\big)^{\frac{\epsilon}{2(1+\epsilon)}} \; \| \, f \, \|_{1+\epsilon}^{\frac{1}{2}} \underset{n \to \infty}{\to} 0, \quad a.s. \end{aligned} \tag{11}$$

Then

$$\begin{aligned} p_n^{\frac{\epsilon}{2(1+\epsilon)}} & \left| \mathbb{E}[X | (X,Y) \in L(\alpha)^{T_n}] - \mathbb{E}[X | (X,Y) \in L_n(\alpha)^{T_n}] \right| \\ &= p_n^{\frac{\epsilon}{2(1+\epsilon)}} \Bigg| \int_{L(\alpha)^{T_n}} x \, f_{X,Y}(x,y) \lambda(\mathrm{d}x \, \mathrm{d}y) \; \mathbb{P}[(X,Y) \in L(\alpha)^{T_n}]^{-1} \\ & \quad - \int_{L_n(\alpha)^{T_n}} x \, f_{X,Y}(x,y) \lambda(\mathrm{d}x \, \mathrm{d}y) \; \mathbb{P}[(X,Y) \in L_n(\alpha)^{T_n}]^{-1} \Bigg| \\ &\leq \frac{p_n^{\frac{\epsilon}{2(1+\epsilon)}}}{\mathbb{P}[(X,Y) \in L(\alpha)^{T_n}] \, \mathbb{P}[(X,Y) \in L_n(\alpha)^{T_n}]} \Bigg(\mathbb{P}[(X,Y) \in L(\alpha)^{T_n}] \\ & \left| \int_{L(\alpha)^{T_n}} x \, f_{X,Y}(x,y) \lambda(\mathrm{d}x \, \mathrm{d}y) - \int_{L_n(\alpha)^{T_n}} x \, f_{X,Y}(x,y) \lambda(\mathrm{d}x \, \mathrm{d}y) \right| \\ & \quad + \int_{L(\alpha)^{T_n}} x \, f_{X,Y}(x,y) \lambda(\mathrm{d}x \, \mathrm{d}y) \left| \mathbb{P}[(X,Y) \in L(\alpha)^{T_n}] - \mathbb{P}[(X,Y) \in L_n(\alpha)^{T_n}] \right| \Bigg). \end{aligned} \tag{12}$$

Using (10)-(11) we obtain that (12) converges to zero $a.s.$, for $n \to \infty$. Hence the result. $\square$

Lemma 6.2. *Under Assumption* **A2**, *Assumptions of Theorem 3.1 and with the notation of Theorem 3.1, it holds that*

$$\sqrt{n}\left|\mathbb{E}[\,X \mid (X,Y) \in L_n(\alpha)^{T_n}\,] - \frac{\sum_{i=1}^n X_i 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}}{\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}}\right| \underset{n\to\infty}{\rightarrow} 0, \quad a.s.$$

Proof of Lemma 6.2

We can write

$$\sqrt{n}\left|\mathbb{E}[X|(X,Y) \in L_n(\alpha)^{T_n}] - \frac{\sum_{i=1}^n X_i 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}}{\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}}\right|$$

$$= \sqrt{n}\left|\frac{\int_{L_n(\alpha)^{T_n}} x\, f_{X,Y}(x,y)\lambda(\mathrm{d}x\,\mathrm{d}y)}{\mathbb{P}[(X,Y)\in L_n(\alpha)^{T_n}]} - \frac{\sum_{i=1}^n X_i 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}}{\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}}\right|.$$

Under assumptions of Lemma 6.2, from the central limit theorem for triangular arrays (e.g. Theorem 27.2 in Billingsley [7]) we obtain

$$\sqrt{n}\left|\mathbb{P}[(X,Y)\in L_n(\alpha)^{T_n}] - \frac{1}{n}\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}\right| \underset{n\to\infty}{\rightarrow} 0,\ a.s.,$$

$$\sqrt{n}\left|\int_{L_n(\alpha)^{T_n}} x\, f_{X,Y}(x,y)\lambda(\mathrm{d}x\,\mathrm{d}y) - \frac{1}{n}\sum_{i=1}^n X_i 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}\right| \underset{n\to\infty}{\rightarrow} 0,\ a.s.$$

Hence the result. □

Then to prove Theorem 4.1 we can write (4) as

$$\beta_n\left|\mathbb{E}[X|(X,Y)\in L(\alpha)^{T_n}] - \frac{\sum_{i=1}^n X_i 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}}{\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}}\right|$$

$$\leq \beta_n\left|\mathbb{E}[X|(X,Y)\in L(\alpha)^{T_n}] - \mathbb{E}[X|(X,Y)\in L_n(\alpha)^{T_n}]\right| + \beta_n\left|\mathbb{E}[X|(X,Y)\in L_n(\alpha)^{T_n}] - \frac{\sum_{i=1}^n X_i 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}}{\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}}\right|.$$

The result is a straightforward application of Lemma 6.1 and Lemma 6.2. □

This work has been partially supported by the French research national agency (ANR) under the reference ANR-08BLAN-0314-01. The authors thank Alberto Rodríguez-Casal for his fruitful comments. They thank also Stéphane Loisel and Norberg Ragnar for helpful discussions.

References

[1] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Math. Finance*, 9(3):203–228, 1999.

[2] A. Baíllo. Total error in a plug-in estimator of level sets. *Statist. Probab. Lett.*, 65(4):411–417, 2003.

[3] A. Baíllo, J. Cuesta-Albertos, and A. Cuevas. Convergence rates in nonparametric estimation of level sets. *Statistics and Probability Letters*, 53:27–35, 2001.

[4] A. Baíllo, A. Cuevas, and A. Justel. Set estimation and nonparametric detection. *Canad. J. Statist.*, 28(4):765–782, 2000.

[5] F. Belzunce, A. Castaño, A. Olvera-Cervantes, and A. Suárez-Llorens. Quantile curves and dependence structure for bivariate distributions. *Computational Statistics & Data Analysis*, 51(10):5112–5129, 2007.

[6] G. Biau, B. Cadre, and B. Pelletier. A graph-based estimator of the number of clusters. *ESAIM. Probability and Statistics*, 11:272–280, 2007.

[7] P. Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. A Wiley-Interscience Publication.

[8] B. Cadre. Kernel estimation of density level sets. *J. Multivariate Anal.*, 97(4):999–1023, 2006.

[9] J. Cai and H. Li. Conditional tail expectations for multivariate phase-type distributions. *Journal of Applied Probability*, 42(3):810–825, 2005.

[10] L. Cavalier. Nonparametric estimation of regression level sets. *Statistics. A Journal of Theoretical and Applied Statistics*, 29(2):131–160, 1997.

[11] Y. P. Chaubey and P. K. Sen. Smooth estimation of multivariate survival and density functions. *J. Statist. Plann. Inference*, 103(1-2):361–376, 2002. C. R. Rao 80th birthday felicitation volume, Part I.

[12] A. Cuevas and R. Fraiman. A plug-in approach to support estimation. *The Annals of Statistics*, 25(6):2300–2312, 1997.

[13] A. Cuevas, W. González-Manteiga, and A. Rodríguez-Casal. Plug-in estimation of general level sets. *Australian & New Zealand Journal of Statistics*, 48(1):7–19, 2006.

[14] A. Cuevas and A. Rodríguez-Casal. On boundary estimation. *Advances in Applied Probability*, 36(2):340–354, 2004.

[15] S. Dedu and R. Ciumara. Restricted optimal retention in stop-loss reinsurance under VaR and CTE risk measures. *Proc. Rom. Acad. Ser. A Math. Phys. Tech. Sci. Inf. Sci.*, 11(3):213–217, 2010.

[16] M. Denuit, J. Dhaene, M. Goovaerts, and R. Kaas. *Actuarial Theory for Dependent Risks*. Wiley, 2005.

[17] P. Embrechts and G. Puccetti. Bounds for functions of multivariate risks. *Journal of Multivariate Analysis*, 97(2):526–547, 2006.

[18] J. M. Fernández-Ponce and A. Suárez-Lloréns. Central regions for bivariate distributions. *Austrian Journal of Statistic*, 31(2–3):141–156, 2002.

[19] E. W. Frees and E. A. Valdez. Understanding relationships using copulas. *North American Actuarial Journal*, 2(1):1–25, 1998.

[20] J. A. Hartigan. Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.*, 82(397):267–270, 1987.

[21] V. I. Koltchinskii. M-estimation, convexity and quantiles. *Ann. Statist.*, 25(2):435–477, 1997.

[22] T. Laloë. *Sur Quelques Problèmes d'Apprentissage Supervisé et Non Supervisé*. PhD thesis, University Montpellier II, 2009.

[23] J.-C. Massé and R. Theodorescu. Halfplane trimming for bivariate distributions. *J. Multivariate Anal.*, 48(2):188–202, 1994.

[24] D. W. Müller and G. Sawitzki. Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.*, 86(415):738–746, 1991.

[25] G. Nappo and F. Spizzichino. Kendall distributions and level sets in bivariate exchangeable survival models. *Information Sciences*, 179:2878–2890, August 2009.

[26] W. Polonik. Measuring mass concentrations and estimating density contour clusters—an excess mass approach. *The Annals of Statistics*, 23(3):855–881, 1995.

[27] W. Polonik. Minimum volume sets and generalized quantile processes. *Stochastic Process. Appl.*, 69(1):1–24, 1997.

[28] P. Rigollet and R. Vert. Optimal rates for plug-in estimators of density level sets. *Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability*, 15(4):1154–1178, 2009.

[29] A. Rodríguez-Casal. *Estimacíon de conjuntos y sus fronteras. Un enfoque geometrico*. PhD thesis, University of Santiago de Compostela, 2003.

[30] C. Rossi. Sulle curve di livello di una superficie di ripartizione in due variabili; on level curves of two dimensional distribution function. *Giornale dell'Istituto Italiano degli Attuari*, 36:87–108, 1973.

[31] C. Rossi. Proprietà geometriche delle superficie di ripartizione. *Rend. Mat. (6)*, 9(4):725–736 (1977), 1976.

[32] R. Serfling. Quantile functions for multivariate analysis: approaches and applications. *Statist. Neerlandica*, 56(2):214–232, 2002. Special issue: Frontier research in theoretical statistics, 2000 (Eindhoven).

[33] L. Tibiletti. On a new notion of multidimensional quantile. *Metron. International Journal of Statistics*, 51(3-4):77–83, 1993.

[34] A. B. Tsybakov. On nonparametric estimation of density level sets. *The Annals of Statistics*, 25(3):948–969, 1997.