Plug-in estimation of level sets in a non-compact setting with applications in multivariate risk theory

Elena Di Bernardino[a]

joint work with Thomas Laloë[b] , Véronique Maume-Deschamps[c] and Clémentine Prieur[d]

Tunis, 43éme Journées de Statistique de la SFDS, 23-27 May 2011

[a, c] Université Lyon 1, ISFA, Laboratoire SAF
[b] Université de Nice Sophia-Antipolis, Laboratoire J-A Dieudonné
[d] Université Joseph Fourier, Grenoble

Framework

Goal: In this paper, we consider the problem of estimating the level sets of a bivariate distribution function F. More precisely our goal is to build a consistent estimator of

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

Idea: We consider a *plug-in* approach 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.

Tools and general aspects

- We do not assume any compactness property for the level sets we estimate. This requires special attention in the statement of the problem.
- We state consistency results with respect to two proximity criteria between sets: the Hausdorff distance and the volume of the symmetric difference.

Our results are motivated by *applications in bivariate risk theory*. In this sense we also present simulated and real data examples which illustrate our theoretical results.

Key words: Level sets, distribution function, plug-in estimation, Hausdorff distance, Conditional Tail Expectation.

Literature and background

Estimation of the level sets of a density function: Polonik (1995), Tsybakov (1997), Baíllo *et al.* (2001)

Estimation of the level sets of a regression function in a compact setting: Cavalier (1997), Biau *et al.* (2007), Laloë (2009)

Estimation of general compact level sets: Cuevas *et al.* (2006)

An alternative approach, based on the geometric properties of the compact support sets: Cuevas and Fraiman (1997), Cuevas and Rodríguez-Casal (2004).

Contents

Contents

Notation and preliminaries

Let ${\mathbb{R}_+^2}^* := \mathbb{R}_+^2 \setminus (0,0)$, $\mathcal{F}$ the set of continuous distribution functions $f : \mathbb{R}_+^2 \to [0,1]$ 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 $F_n(.) = F_n(X_1, X_2, \ldots, X_n, .)$ an estimator of F.

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

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

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

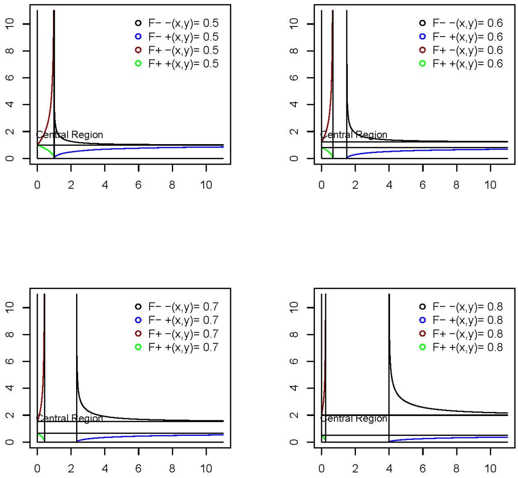
In addition, given $T > 0$, we set

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

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

For any $A \subset \mathbb{R}_+^2$ we note ∂A its boundary.

From a parametric formulation of $\partial L(c)$ (see Belzunce *et al.* 2007).
Case: Survival Clayton Copula with marginals (Burr(1), Burr(2))



Notation and preliminaries

For $r > 0$ and $\lambda > 0$, define

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

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

- $B(x, \xi)$ is the closed ball centered on x and with positive radius ξ,
- $(\nabla F)_x$ is the gradient of F evaluated at x,
- $\|(HF)_x\|$ is the matrix norm induced by Euclidean distance of the Hessian matrix in x.

Hausdorff distance between A_1 and A_2 is defined by:

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

where $B(S, \varepsilon) = \bigcup_{x \in S} B(x, \varepsilon)$; A_1 and A_2 are compact sets in $(\mathbb{R}_+^2, d)$.

Assumption H

Finally, we introduce the following assumption (e.g. see Tsybakov, 1997; Cuevas *et al.*, 2006):

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

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

Assumption **H** is satisfied under mild conditions (differentiability properties of F).

About Assumption **H**

H is satisfied under mild assumptions (modification of Proposition 3.1 in the PhD of Rodríguez-Casal, 2003).

Proposition

*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 $m^{\nabla} > 0$ and $M_H < \infty$. Then F satisfies Assumption **H**, with $A = \frac{2}{m^{\nabla}}$.*

Remark: Under assumptions of Proposition F is continuous and $m^{\nabla} > 0$, there is no plateau in the graph of F for each level t such that $|\, t - c \,| \leq r$. Furthermore from Theorem 1 in Rossi (1976) we know that $\{F = t\}$ is a curve in the quadrant $\mathbb{R}_+^2$. Finally under assumptions of Proposition we obtain $\partial L(c)^T = \{F = c\}^T = \{F = c\} \cap [0, T]^2$.

Contents

Consistency in terms of the Hausdorff distance

From now on we note, for $n \in \mathbb{N}^*$, and for $T > 0$,

$$\|F - F_n\|_\infty = \sup_{x \in \mathbb{R}_+^2} | F(x) - F_n(x) |, \quad \|F - F_n\|_\infty^T = \sup_{x \in [0,T]^2} | F(x) - F_n(x) | .$$

Theorem (Consistency Hausdorff distance)

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 $t : | t - c | \leq r$, $\partial L(t)^{T_1} \neq \emptyset$. 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

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

Consistency in terms of the Hausdorff distance

Theorem above 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: We provide an asymptotic result for a fixed level c. In particular following the proof of Theorem we remark that, for n large enough,

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

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.

L_1 consistency

Consistency criterion: the consistency of the volume (in the Lebesgue measure sense) of the symmetric difference between $L(c)^{T_n}$ and $L_n(c)^{T_n}$.

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

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

L_1 consistency

Theorem (Consistency volume)

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 $t : \mid t - c \mid \leq r$, $\partial L(t)^{T_1} \neq \emptyset$ 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 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 does not require continuity assumption on F_n.

Contents

Bivariate Value-at-Risk

The Value-at-Risk (VaR) measure is defined as

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

with $Q_X(\alpha) = F_X^{-1}(\alpha)$ (univariate quantile function of the continuous loss distribution function F_X).

Definition (Generalization of the VaR measure in dimension 2)

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. $\mathrm{VaR}_\alpha(F) = \partial L(\alpha)$.

References: Embrechts and Puccetti (2006), Tibiletti (1993) and Nappo and Spizzichino (2006). In Tibiletti (1993) with $\alpha = \frac{1}{2}$: natural extension of the bi-dimensional median.

Bivariate Value-at-Risk

Our (plug-in) estimator of the bivariate Value-at-Risk:

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

Consistency result for the $\mathrm{VaR}_\alpha(F_n)$ on the quadrant $\mathbb{R}^2_+$ with respect to the Hausdorff distance i.e.

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

VaR_α does not give any information about the thickness of the upper tail of the distribution function. Shortcoming of the VaR: we are not only concerned with the frequency of the default but also with the severity of loss in case of default.

In order to overcome this problem: Conditional Tail Expectation (CTE).

Bivariate Conditional Tail Expectation

For a continuous loss distribution function F_X the CTE at level α is defined by

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

where $\mathrm{VaR}_\alpha(X)$ is the univariate VaR defined above (e.g. for properties of the $CTE_\alpha(X)$ see Denuit *et al.*, 2005).

Several bivariate generalizations of the classical univariate CTE:

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

$$\mathbb{E}[\,(X,Y)\,|\, \max\{X,Y\} > Q_{\max\{X,Y\}}(\alpha)\,],$$

e.g see Cai and Li, (2005).

Conditioning events are the total risk or some univariate extreme risk in the portfolio (aggregate variable).

A new bivariate Conditional Tail Expectation

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 $\mathbb{E}(X^2) < \infty$, $\mathbb{E}(Y^2) < \infty$.

Remark: Our bivariate Conditional Tail Expectation is a natural extension of the univariate one.

A new bivariate Conditional Tail Expectation

Definition (Generalization of the CTE in dimension 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

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

2. the estimated bivariate α-Conditional Tail Expectation

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

A new bivariate Conditional Tail Expectation

Remarks:

- If X and Y are identically distributed with a symmetric copula then $\mathbb{E}[X \mid (X,Y) \in L_{(X,Y)}(\alpha)] = \mathbb{E}[Y \mid (X,Y) \in L_{(X,Y)}(\alpha)]$.
- 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 ($L(\alpha)$).

A new bivariate Conditional Tail Expectation

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

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

$$\widehat{\mathrm{CTE}}_\alpha^T(X,Y) = \left(\begin{array}{c} \frac{\sum_{i=1}^n X_i 1_{\{(X_i,Y_i)\in L_n(\alpha)^T\}}}{\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)^T\}}} \\ \frac{\sum_{i=1}^n Y_i 1_{\{(X_i,Y_i)\in L_n(\alpha)^T\}}}{\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)^T\}}} \end{array} \right),$$

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

A new bivariate Conditional Tail Expectation

Theorem (Consistency of $\widehat{CTE}_\alpha(X, Y)$)

Under Assumption **A2**, *Assumptions of Theorem Consistency volume and with the same notation, it holds that*

$$\beta_n \Big| \mathrm{CTE}_\alpha^{T_n}(X, Y) - \widehat{\mathrm{CTE}}_\alpha^{T_n}(X, Y) \Big| \underset{n \to \infty}{\rightarrow} 0, \quad a.s., \tag{2}$$

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

The convergence in (2) must be interpreted as a componentwise convergence. In the case of a bounded density function $f_{(X,Y)}$ we obtain $\beta_n = \min\{\sqrt{p_n}, \sqrt{n}\}$.

A new bivariate Conditional Tail Expectation

Remark: Starting from Theorem *Consistency of* $\widehat{CTE}_\alpha(X, Y)$, it could be interesting to consider the convergence

$$\left| \mathrm{CTE}_\alpha(X, Y) - \widehat{\mathrm{CTE}}_\alpha^{T_n}(X, Y) \right|.$$

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 $n \to \infty$.

Contents

Estimation of the level sets

F_n (empirical estimator); $T_n = n^{0.45}$; 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.

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

Estimation of the level sets

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: 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. 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.

Estimation of $CTE_\alpha(X, Y)$ on simulated data

In the following we denote $\overline{\widehat{\mathrm{CTE}}_\alpha^{T_n}}(X,Y) = \left(\overline{\widehat{\mathrm{CTE}}_\alpha^{T_n,1}}(X,Y), \overline{\widehat{\mathrm{CTE}}_\alpha^{T_n,2}}(X,Y)\right)$ 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 F_n empirical estimator, $n = 1000$.

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: (X, Y) independent; $X \sim Exp(2)$, $Y \sim Exp(2)$.

Estimation of $CTE_\alpha(X, Y)$ on simulated data

α	$\text{CTE}_\alpha(X, Y)$	$\overline{\widehat{\text{CTE}_\alpha}}^{T_n}(X, Y)$	$\widehat{\sigma}$	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: (X, Y) independent; $X \sim Exp(2)$, $Y \sim Exp(1)$.

Estimation of $CTE_\alpha(X, Y)$ on simulated data

α	$\mathrm{CTE}_\alpha(X, Y)$	$\overline{\widehat{\mathrm{CTE}_\alpha}}^{T_n}(X, Y)$	$\widehat{\sigma}$	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: (X, Y) Clayton Copula with parameter 1, $X \sim Exp(1)$, $Y \sim Burr(4, 1)$.

Estimation of $CTE_\alpha(X, Y)$ on simulated data

In Table 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.

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: Evolution of $\widehat{\sigma}$ and RMSE in terms of sample size n for $\alpha = 0.9$; (X, Y) independent, $X \sim Exp(1)$, $Y \sim Exp(2)$.

The theoretical value is $\text{CTE}_{0.9}(X, Y) = (3.78, 1.89)$. In this case we need between 2500 and 5000 data to get the same performances as for lower level.

Estimation of $CTE_\alpha(X, Y)$ on real data

We consider here the estimation of CTE_α in a real case: **Loss-ALAE data** in the log scale; $n = 1500$, $T_n = n^{0.4}$, F_n (empirical estimator).

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

Table: $\widehat{\text{CTE}}_\alpha^{T_n}$ 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 $L(\alpha)$.

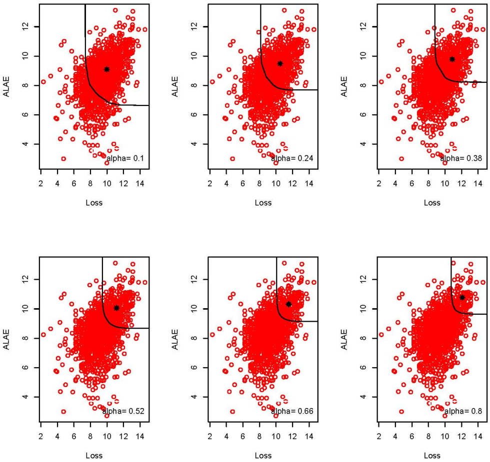


Figure: Loss-ALAE data in log scale: boundary of estimated level sets (line) and $\widehat{\mathrm{CTE}}_\alpha^{T_n}$ (star) for different values of α.

Contents

Conclusions and perspectives

We have provided:

- Convergence results for the plug-in estimator of the levels sets of an unknown distribution function (in terms of d_H and d_λ).
- Possible applications in multivariate risk theory (VaR and CTE).
- Illustrations on simulated and real data sets.

Further developments:

- Behavior of our estimator for high values of the level (using suitable F_n - Extreme Value Theory).
- Analysis of the CTE as risk measure: study of the classical properties (monotonicity, translation invariance, positive homogeneity, ...)
- Relation between our CTE and the dependence structure (definition of some statistic order; orthant order of the components, ...)

Thank for your attention.

- A. Baíllo, J.A. Cuesta-Albertos, A. Cuevas, *Convergence rates in nonparametric estimation of level sets*. Statistics and Probability Letters 53, (2001), 27-35.
- F. Belzunce, A. Castaño, A. Olvera-Cervantes, A. Suárez-Llorens, *Quantile curves and dependence structure for bivariate distributions*. Computational Statistics & Data Analysis 51, (2007), 5112-5129.
- G. Biau, B. Cadre, B. Pelletier, *A graph-based estimator of the number of clusters*. ESAIM, Probability and Statistics 11, (2007), 272-280.
- J. Cai, H. Li, *Conditional tail expectations for multivariate phase-type distributions*. Journal of Applied Probability 42(3), (2005), 810-825.
- L. Cavalier, *Nonparametric estimation of regression level sets*. Statistics 29, (1997), 131-160.
- A. Cuevas, R. Fraiman, *A plug-in approach to support estimation*. Annals of Statistics 25, (1997), 2300-2312.
- A. Cuevas, W. González-Manteiga, A. Rodríguez-Casal, *Plug-in estimation of general level sets*. Australian and New Zealand Journal of Statistics 48, (2006), 7-19.

- A. Cuevas, A. Rodríguez-Casal, *On boundary estimation*. Advances in Applied Probability 36, (2004), 340-354.
- M. Denuit, J. Dhaene, M. Goovaerts, R. Kaas. *Actuarial Theory for Dependent Risks*. Wiley, (2005).
- P. Embrechts, G. Puccetti, *Bounds for functions of multivariate risks*. Journal of Multivariate Analysis 97(2), (2006), 526-547.
- J.R. Fernéandez-Ponce, A. Suárez-Lloréns, *Central Regions for Bivariate Distributions*. Austrian Journal of Statistic 31(2-3), (2002), 141-156.
- E.W. Frees, E.A. Valdez, *Understanding relationships using copulas*. North American Actuarial Journal 2, (1998), 1-15.
- T. Laloë, *Sur Quelques Problèmes d'Apprentissage Supervisé et Non Supervisé.*, Ph.D. Thesis, University of Montpellier II, (2009).
- G. Nappo, F. Spizzichino, *Relations between Kendall Distributions and level sets in bivariate exchangeable survival models*. Information Sciences. 179 (17), (2009), 2878-2890.

- W. Polonik, *Measuring mass concentration and estimating density contour clusters - an excess mass approach*. Annals of Statistics 23, (1995), 855-881.
- A. Rodríguez-Casal, *Estimacíon de conjuntos y sus fronteras. Un enfoque geometrico*. Ph.D. Thesis, University of Santiago de Compostela, (2003).
- C. Rossi, *On surface level curves of distribution in two variables*. Giornale dell'Istituto Italiano degli Attuari, Anno XXXVI- N.1-2, (1973), 87-108.
- L. Tibiletti, *Connectedness, arcwise-connectedness and convexity for levels sets of multidimensional distribution functions*. Ratio Mathematica 8, (1994), 21-27.
- L. Tibiletti, *Quasi-concavity property of multivariate distribution functions*. Ratio Mathematica 9, (1995), 27-36.
- L. Tibiletti, *On a new notion of multidimensional quantile*. Metron, vol. LI n.3-4, (1993), 77-83.
- L. Tibiletti, *On quasi-concavity of n-dimensional distribution functions*. Atti del XV convegno A.M.A.S.E.S., (1991), 503-515.
- A.B. Tsybakov, *On nonparametric estimation of density level sets*. Annals of Statistics 25, (1997), 948-969.