A proposal of a bivariate Conditional Tail Expectation

Elena Di Bernardino[a]

joint works with Areski Cousin[b], Thomas Laloë[c] , Véronique Maume-Deschamps[d] and Clémentine Prieur[e]

Aussois, 4eme Rencontres des Jeunes Statisticiens, 5-9/09/11

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

Given an univariate continuous and strictly monotonic loss distribution function F_X,

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

Shortcoming of VaR measure:

- Var is not a coherent risk measure (in the "Artzner sense", see Artzner, 1999)
- Var does not give any information about thickness of the upper tail of the loss

To overcome problems of VaR $\rightarrow$ Conditional Tail Expectation (CTE):

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

(e.g. for properties of the $CTE_\alpha(X)$ see Denuit *et al.*, 2005).

A new bivariate Conditional Tail Expectation

Definition (Generalization of the CTE in dimension 2)

Consider a random vector (X, Y) satisfying regularity properties, with associate distribution function F. Let $L(\alpha) = \{F(x, y) \geq \alpha\}$. For $\alpha \in (0, 1)$, we define

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

Distributional approach to risk model:

Our $\mathrm{CTE}_\alpha(X, Y)$ does not use an aggregate variable (sum, min, max ...) in order to analyze the bivariate risk's issue. Conversely $\mathrm{CTE}_\alpha(X, Y)$ deals with the simultaneous joint damages considering the bivariate dependence structure of data in a specific risk's area (α-level set : $L(\alpha)$).

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:

$$\mathrm{CTE}_\alpha(X,Y)^{\mathrm{sum}} = \mathbb{E}[\,(X,Y)\,|\,X+Y > Q_{X+Y}(\alpha)\,],$$

$$\mathrm{CTE}_\alpha(X,Y)^{\mathrm{min}} = \mathbb{E}[\,(X,Y)\,|\,\min\{X,Y\} > Q_{\min\{X,Y\}}(\alpha)\,],$$

$$\mathrm{CTE}_\alpha(X,Y)^{\mathrm{max}} = \mathbb{E}[\,(X,Y)\,|\,\max\{X,Y\} > Q_{\max\{X,Y\}}(\alpha)\,].$$

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

Two different developments of research :

1. Problem of estimating the level sets $L(c) = \{F(x) \geq c\}$, with $c \in (0, 1)$, of an unknown distribution function F on $\mathbb{R}^2_+$ with a *plug-in* approach; in order to provide a consistent estimator for $\mathrm{CTE}_\alpha(X, Y)$ [Article *"Plug-in estimation of level sets in a non-compact setting with applications in multivariate risk theory"*, with Laloë L., Maume-Deschamps V. and Prieur C., submitted].

2. Analysis of the $\mathrm{CTE}_\alpha(X, Y)$ as risk measure: study of the classical properties (monotonicity, translation invariance, positive homogeneity, ...), behavior with respect to different risk scenarios and stochastic ordering of marginals risks [Article *"Some proposals about bivariate risk measures"*, with Cousin A.].

Literature and background

Estimation of the level sets of a density function: Polonik (1995), Tsybakov (1997), Baíllo *et al.* (2001), Biau *et al.* (2007).

Estimation of the level sets of a regression function in a compact setting: Cavalier (1997), 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).

Plug-in estimation of level sets

Goal: 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

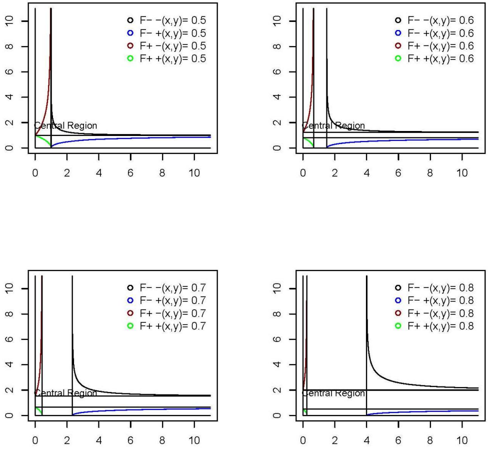
$$L(c) := \{F(x) \geq c\}, \quad \text{for } c \in (0,1).$$

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.

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



We state consistency results with respect to two proximity criteria between sets: the Hausdorff distance and the volume of the symmetric difference ("physical proximity" between sets).

$\rightarrow$ Problem of compactness property for the level sets we estimate.

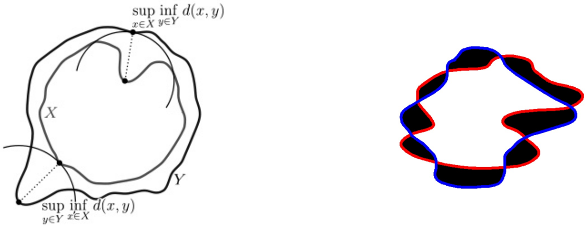


Figure: (left) Hausdorff distance between sets X and Y; (right) $\lambda(X \triangle Y)$, where λ stands for the Lebesgue measure on $\mathbb{R}^2$ and $\triangle$ for the symmetric difference.

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*:

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

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

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

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

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 $(\mathbb{R}_+^2, d)$.

Notation and preliminaries

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.

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 $A = \frac{2}{m^\nabla}$.*

Consistency result in terms of the Hausdorff distance

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

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

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 $t : \mid t - c \mid \leq r$, $\partial L(t)^{T_1} \neq \emptyset$. 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 for n large enough,

$$d_H(\partial L(c)^{T_n}, \partial L_n(c)^{T_n}) \leq 6\, A \, \|F - F_n\|_\infty^{T_n}, \quad a.s., \text{ where } A = \frac{2}{m^\nabla}.$$

Consistency result in terms of the Hausdorff distance

- $d_H(\partial L(c)^{T_n}, \partial L_n(c)^{T_n})$ converges to zero and the quality of our plug-in estimator is obviously related to the quality of the estimator F_n.
- Note that in the case c is close to one the constant $A = \frac{2}{m^{\nabla}}$ could be large. In this case, we will need a large number of data to get a reasonable estimation.
- In order to overcome the problem of "F_n is continuous with probability one" it can be considered a smooth version of estimator.

L_1 consistency

Let us introduce the following assumption:

A1 There exist positive increasing sequences $(v_n)_{n\in\mathbb{N}^*}$ and $(T_n)_{n\in\mathbb{N}^*}$ such that

$$v_n \int_{[0,T_n]^2} \mid F - F_n \mid^p \ \lambda(\mathrm{d}x) \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} 0, \quad \text{for some } 1 \leq p < \infty.$$

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 $M_H < \infty$. Let $(v_n)_{n\in\mathbb{N}^*}$ and $(T_n)_{n\in\mathbb{N}^*}$ positive increasing sequences such that Assumption **A1** *is satisfied and that for all $t : \mid t - c \mid \leq r$, $\partial L(t)^{T_1} \neq \emptyset$. Then it holds that*

$$p_n \, d_\lambda(L(c)^{T_n}, L_n(c)^{T_n}) \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} 0,$$

with p_n an increasing positive sequence such that $p_n = o\left(v_n^{\frac{1}{p+1}} / T_n^{\frac{p}{p+1}} \right)$.

L_1 consistency

- *Theorem Consistency volume* provides a convergence rate, which is closely related to the choice of the sequence T_n. A sequence T_n whose divergence rate is large implies a convergence rate p_n quite slow.
- *Theorem Consistency volume* does not require continuity assumption on F_n.
- Analysis in the case: F_n the bivariate empirical distribution function.
- Problem: optimal criterion for the choice of T_n related with the p in Assumption A1.

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

Definition

Consider a random vector (X, Y) with associate distribution function $F \in \mathcal{F}$. For $\alpha \in (0, 1)$, we define the estimated bivariate α-Conditional Tail Expectation

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

We introduce truncated versions of the theoretical and estimated CTE_α:

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

using $L(\alpha)^T$ and $L_n(\alpha)^T$ i.e. the truncated versions $L(\alpha)$ and $L_n(\alpha)$.

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

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

Under regularity properties of (X, Y), Assumptions of Theorem Consistency volume and with the same notation, it holds that

$$\beta_n \left| \mathrm{CTE}_\alpha^{T_n}(X, Y) - \widehat{\mathrm{CTE}}_\alpha^{T_n}(X, Y) \right| \underset{n \to \infty}{\overset{\mathbb{P}}{\to}} 0, \tag{2}$$

where $\beta_n = \min\{ p_n^{\frac{r}{2(1+r)}}, a_n \}$, with $r > 0$ such that the density $f_{(X,\,Y)} \in L^{1+r}(\lambda)$ and $a_n = o\left(\sqrt{n}\right)$.

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

On the choice of T_n

In the case of the bivariate empirical d.f. F_n: $\beta_n = o\left(n^{\frac{r}{6(1+r)}} / T_n^{\frac{2r}{3(1+r)}}\right)$.

In the case of a bounded density function $f_{(X,Y)}$, $\beta_n = o\left(n^{\frac{1}{6}} / T_n^{\frac{2}{3}}\right)$.

Obviously it could be interesting to consider the convergence:

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

$\rightarrow$ the speed of convergence will also depend on the convergence rate to zero of

$$\left|\ \mathrm{CTE}_\alpha(X,Y) - \mathrm{CTE}_\alpha^{T_n}(X,Y)\right|, \text{ then of } \mathbb{P}[(X,Y) \in L(\alpha) \setminus L(\alpha)^{T_n}].$$

We remark that $(\mathbb{P}[X \geq T_n \text{ or } Y \geq T_n])^{-1}$ is increasing in T_n, whereas the speed of convergence is decreasing in T_n. This kind of compromise provides an illustration on how to choose T_n, apart from satisfying the assumptions of consistency results above.

Perspectives

- Study of $\mathrm{CTE}_\alpha(X, Y)$ as risk measure (stochastic order for random variables and random vectors)
- $\mathrm{CTE}_\alpha(X, Y)$ and Kendall distribution function or bivariate probability integral transformation (BIPIT) $F(X, Y)$.
- In this work we provide asymptotic results for a fixed level c...
- Problem of constant A for $c \sim 1$: Plug in method with

$$\widehat{L}_n(c) := \{(x, y) \in \mathbb{R}_+^2 : \ \widehat{F}^*(x, y) \geq c\}, \qquad \text{for } c \sim 1,$$

with $\widehat{F}^*$ a "tail estimator" of F (Bivariate extreme value theory; first part of my PhD thesis).

- Artzner, P. and Delbaen, F. and Eber, J. M. and Heath, D., *Coherent measures of risk*. Math. Finance 9(3), (1999), 203–228.
- 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).
- 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).
- A.B. Tsybakov, *On nonparametric estimation of density level sets*. Annals of Statistics 25, (1997), 948-969.