

# ESTIMATION OF THE COVARIATE CONDITIONAL TAIL EXPECTATION : A DEPTH-BASED LEVEL SET APPROACH

Elisabeth Armaut<sup>1</sup>, Roland Diel<sup>2</sup> & Thomas Laloë<sup>3</sup>

<sup>1</sup> *Université de Nice Côte d'Azur, armaut@unice.fr*

<sup>2</sup> *Université de Nice Côte d'Azur, Roland.Diel@unice.fr*

<sup>3</sup> *Université de Nice Côte d'Azur, Thomas.Laloe@unice.fr*

**Résumé.** De nos jours, dans pratiquement tous les domaines tels que la finance, la médecine, l'écologie, l'industrie..., il est impossible d'éviter des risques! Par exemple, en finance, le "risque" signifie souvent la possibilité de perdre de l'argent. En hydrologie, le risque peut par ailleurs représenter la quantité d'eau dépassant le niveau de remplissage maximum d'un barrage. La *Conditionnal Covariate Tail Expectation* (ou CCTE) est une mesure de risque qui quantifie un coût moyen associé à  $d \geq 1$  facteurs de risque non nécessairement homogènes. Dans notre cadre d'étude, la zone de risque est représentée par un ensemble de niveau inférieur associé à une fonction de profondeur statistique multivariée. Nous proposons un estimateur consistant de la CCTE avec une vitesse de convergence : cet estimateur fait intervenir une estimation de l'ensemble de niveau associé à la profondeur en question via une méthode *plug-in*. Une étude sur simulation vient compléter l'étude des performances de notre estimateur.

**Mots-clés.** Estimation *plug-in*, fonction de profondeur multivariée, théorie du risque.

**Abstract.** Nowadays, in almost all fields such as finance, medicine, ecology, industry..., it is not possible to avoid risks. For instance, in finance, risk often means that there is potential for money loss. In hydrology, risk could represent the amount of water which exceeds the maximum storage level of a dam. The *Conditionnal Covariate Tail Expectation* (CCTE) is a risk measure that quantifies an expected cost associated to  $d \geq 1$  risk factors which are heterogeneous in nature. In our setting, the risk region in the problem at hand is represented by a depth-based lower level set. We provide a consistent estimator of the CCTE with a rate of convergence : this estimator involves a *plug-in* approach when estimating the lower level set. A simulation study complements the performances of our estimator.

**Key-Words.** *Plug-in* estimation, multivariate depth function, risk theory.

# 1 Introduction

La théorie du risque est une branche de la statistique dont le but principal recherché est de pouvoir gérer la part d'incertude de certains événements, ce qui permet la mise en place de mesures de prévention. Elle s'applique dans divers domaines, par exemple, en hydrologie pour prévoir les crues, en finance pour protéger la valeur du portefeuille après un investissement, etc...

Pour un phénomène donné, modéliser la structure de dépendance de données multivariées permet d'obtenir des résultats significatifs et représentatifs dans l'analyse du risque. Ainsi, l'étude de risques multivariés amène à l'étude de "régions" quantiles. Celle des risques univariés via des quantiles univariés a été largement traitée dans la littérature. Quant aux risques multivariés, l'étude de quantiles multivariés a gagné beaucoup d'attention ces dernières décennies, notamment les quantiles basés sur une loi de probabilité multivariée (cf. Belzunce et al. (2007), Dehaan and Huang (1995), Cousin and Di Bernardino (2013)), ou encore les quantiles basés sur une fonction de profondeur (cf. Zuo and Serfling (2000)).

Par ailleurs, il est pertinent d'analyser le comportement d'un coût modélisé par une variable aléatoire réelle  $Y$  (par exemple, la somme d'argent gagnée ou perdue dans un investissement sur une certaine période), et ce par rapport à  $d \geq 1$  facteurs de risque différents  $\mathbf{X} \in \mathbb{R}^d$ . Intuitivement, la variable de coût  $Y$  va dépendre des facteurs de risques du phénomène étudié. Plus précisément, la *Covariate Conditional Tail Expectation* est définie par Laloë et al. (2015)), comme suit :

$$\text{CCTE}_{F,\alpha}(Y, \mathbf{X}) := \mathbb{E}[Y | \mathbf{X} \in \mathcal{L}_{F_{\mathbf{X}}}(\alpha)], \alpha \in (0, 1), \quad (1.1)$$

où,

$$\mathcal{L}_{F_{\mathbf{X}}}(\alpha) := \{x \in \mathbb{R}^d : F_{\mathbf{X}}(x) \geq \alpha\}, \quad (1.2)$$

est l'ensemble de niveau (supérieur) associé à la fonction de répartition  $F_{\mathbf{X}} := F$  du v.a  $\mathbf{X} \in \mathbb{R}^d$ .

## 2 *Depth-based Covariate Conditionnal Tail Expectation*

L'utilisation d'une mesure de risque basée sur les ensembles de niveau d'une fonction de répartition nous restreint à considérer uniquement des orientations particulières du risque (cf. Figure 1, graphe gauche). Par exemple, on peut considérer des températures très basses ou très élevées, mais pas les deux à la fois. Afin de contourner ce problème de dépendance en une direction canonique, au lieu de considérer des ensembles de niveau de la forme  $\mathcal{L}_{F_{\mathbf{X}}}(\alpha)$ , des ensembles de niveau de la forme  $\mathcal{L}_{F_{R\mathbf{X}}}(\alpha)$  ont été étudiés dans la littérature, où  $R$  est une matrice de rotation dans  $\mathbb{R}^d$  (cf. Torres et al. (2020)). Une

telle matrice de rotation  $R$  permettrait de mettre en évidence d'autres zones de risque dans  $\mathbb{R}^d$ ; cependant, dans ce cas, la CCTE dépendrait encore d'une orientation via la rotation  $R$ . Pour cela, afin d'étudier une mesure de risque complètement indépendante de l'orientation, nous proposons une approche nouvelle en considérant la CCTE pour des ensembles de niveau associés à une fonction de profondeur statistique multivariée (au sens de la Définition 2.1 dans Zuo and Serfling (2000)), et qui est notée

$$D : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}_+,$$

avec  $\mathcal{P} := \mathcal{P}(\mathbb{R}^d)$  désignant l'ensemble des mesures de probabilités sur  $\mathbb{R}^d$ . En effet, une fonction de profondeur ordonne des données selon leur *degré de centralité*, et fournit un ordre statistique du *centre* vers l'extérieur (cf. Zuo and Serfling (2000)).

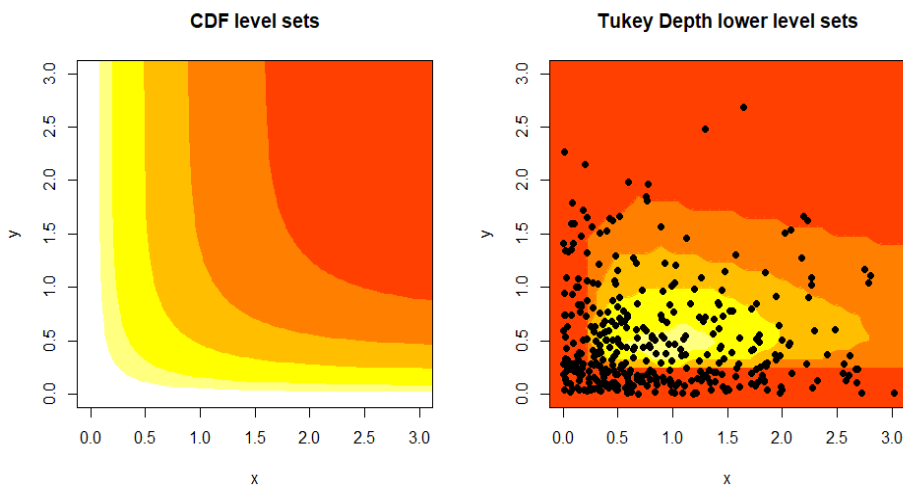


FIGURE 1 – Ensembles de niveaux (supérieurs) pour une fonction de répartition et ensembles de niveaux inférieurs pour une profondeur respectivement (de gauche à droite respectivement).

Afin d'étudier les zones de risque via une fonction de profondeur, il faut s'intéresser aux régions de faible profondeur sur lesquelles nous pouvons étudier le comportement de la covariable de coût  $Y$ . En d'autres termes, considérant un niveau  $\alpha > 0$ , une mesure de probabilité  $P$  sur  $\mathbb{R}^d$ , nous considérons l'ensemble de niveau

$$\mathcal{L}_D(\alpha) := \{x \in \mathbb{R}^d : D(x, P) \leq \alpha\}. \quad (2.1)$$

Étant donné une suite d'estimateurs consistants  $(\tilde{P}_n)_{n \geq 1}$  de  $P$ , nous proposons d'estimer  $\mathcal{L}_D(\alpha)$  par l'ensemble

$$\mathcal{L}_{n,D}(\alpha) := \mathcal{L}_n(\alpha) = \left\{x \in \mathbb{R}^d : D_n(x) := D(x, \tilde{P}_n) \leq \alpha\right\}, n \geq 1. \quad (2.2)$$

### 3 Estimation et vitesse de convergence

Soit  $\mathbf{X}$  un vecteur aléatoire de loi  $P \in \mathcal{P}(\mathbb{R}^d)$ ,  $Y$  une va réelle, et  $\alpha > 0$ . Dans notre cadre d'étude, on s'intéresse à l'estimation de la CCTE basée sur les ensembles de niveau d'une profondeur, définie par :

$$\text{CCTE}_{D,\alpha}(Y, \mathbf{X}) := \mathbb{E}[Y | \mathbf{X} \in \mathcal{L}_D(\alpha)]. \quad (3.1)$$

Soient  $n_1, n_2 \geq 1$ . Soient deux échantillons,

$$\begin{aligned} \tilde{S}_{n_1} &:= (\tilde{\mathbf{X}}_i)_{i=1,\dots,n_1} \text{ de même loi que } \mathbf{X}, \text{ et} \\ S_{n_2} &:= ((Y_i, \mathbf{X}_i))_{i=1,\dots,n_2} \text{ de même loi que } (Y, \mathbf{X}), \end{aligned} \quad (3.2)$$

tels que les  $(Y_i, \mathbf{X}_i)_i$  et les  $(\tilde{\mathbf{X}}_i)_i$  soient indépendants. Considérons  $\mathcal{L}_{n_1}(\alpha)$  un estimateur (calculé à partir de  $\tilde{S}_{n_1}$ ) de  $\mathcal{L}_D(\alpha)$ . On peut alors construire à partir de l'échantillon  $S_{n_2}$  un estimateur pour la CCTE (3.1) :

$$\widehat{\text{CCTE}}_{D,\alpha}^{n_1,n_2}(Y, \mathbf{X}) := \mathbb{E}_{S_{n_2}}[Y | \mathbf{X} \in \mathcal{L}_{n_1}(\alpha)] = \frac{\sum_{i=1}^{n_2} Y_i \mathbf{1}_{\mathbf{X}_i \in \mathcal{L}_{n_1}(\alpha)}}{\sum_{i=1}^{n_2} \mathbf{1}_{\mathbf{X}_i \in \mathcal{L}_{n_1}(\alpha)}}, \quad (3.3)$$

sous réserve que, pour  $n_1, n_2 \geq 1$ ,  $\mathbb{P}$ -p.s.  $\sum_{i=1}^{n_2} \mathbf{1}_{\mathbf{X}_i \in \mathcal{L}_{n_1}(\alpha)} > 0$ .

Notre premier résultat lie la vitesse de convergence de la  $\widehat{\text{CCTE}}$  à celle de  $\mathcal{L}_n(\alpha)$ . On suppose que :

**(H1)** : (i)  $\exists (v_{n_1})_{n_1 \geq 1}$  une suite de réels strictement positifs telle que  $(v_{n_1})_{n_1 \geq 1}$  s.t.

$$\lambda_d(\mathcal{L}_{n_1}(\alpha) \Delta \mathcal{L}_D(\alpha)) = \mathcal{O}_{P,n_1}(v_{n_1}^{-1}), \text{ and}$$

(ii)  $P$  est continue de densité  $f \in \mathbb{L}^p(\mathbb{R}^d, \lambda_d)$  pour un certain  $p \in (1, +\infty]$ .

**Théorème 1.** Soit  $\alpha > 0$ ,  $P \in \mathcal{P}(\mathbb{R}^d)$ . On suppose que

$$\mathbb{P}[\mathbf{X} \in \mathcal{L}_D(\alpha)] > 0.$$

Sous l'hypothèse **(H1)**, et en supposant qu'il existe un  $r \geq 2$  tel que  $Y$  est  $r$ -intégrable, on a :

$$|\widehat{\text{CCTE}}_{D,\alpha}^{n_1,n_2}(Y, \mathbf{X}) - \text{CCTE}_{D,\alpha}(Y, \mathbf{X})| = \mathcal{O}_{P,n_1,n_2} \left( n_2^{-\frac{1}{2}} \vee v_{n_1}^{-(1-\frac{1}{r})(1-\frac{1}{p})} \right).$$

Dans le cas particulier où la profondeur utilisée est la *profondeur de Mahalanobis* (définie ci-dessous), la CCTE est consistante (Théorème 2) avec une vitesse d'au plus  $\sqrt{n}$  (lorsque  $n_1 = n_2 = n$  par exemple).

La profondeur de *Mahalanobis* est définie dans Zuo et Serfling (2000) comme :

$$MHD(x, P) = \begin{cases} (1 + d_{\Sigma_{\mathbf{X}}}^2(x, \mu_{\mathbf{X}}))^{-1} & \text{if } \mathbb{E}_P[\|\mathbf{X}\|^2] < +\infty \\ 0 & \text{if } \mathbb{E}_P[\|\mathbf{X}\|^2] = +\infty, \end{cases}$$

où  $\mu_{\mathbf{X}} = \mathbb{E}_P[\mathbf{X}]$  est le vecteur moyen de  $\mathbf{X}$  et  $\Sigma_{\mathbf{X}}$  est sa matrice de covariance (supposée inversible), et

$$d_{\Sigma_{\mathbf{X}}}^2(x, \mu_{\mathbf{X}}) := \|x - \mu_{\mathbf{X}}\|_{\Sigma_{\mathbf{X}}}^2 := {}^t(x - \mu_{\mathbf{X}})\Sigma_{\mathbf{X}}^{-1}(x - \mu_{\mathbf{X}})$$

est la distance de *Mahalanobis*.

**Théorème 2.** Soit  $P \in \mathcal{P}$ ,  $D(\cdot, P) = MHD(\cdot, P)$  et  $\alpha \in (0, 1)$ . Sous les hypothèses du Théorème 1, on a

$$\left| \widehat{\text{CCTE}}_{D, \alpha}^{n_1, n_2}(Y, \mathbf{X}) - \text{CCTE}_{\alpha}(Y, \mathbf{X}) \right| = \mathcal{O}_{P, n_1, n_2} \left( n_2^{-\frac{1}{2}} \vee n_1^{-\frac{1}{2}(1-\frac{1}{r})(1-\frac{1}{p})} \right).$$

Enfin, nous proposerons une étude sur simulation pour illustrer nos résultats.

## Références

- 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.
- A. Cousin and E. Di Bernardino. On multivariate extensions of value-at-risk. *Journal of multivariate analysis*, 119 :32–46, 2013.
- L. Dehaan and X. Huang. Large quantile estimation in a multivariate setting. *Journal of Multivariate Analysis*, 53(2) :247–263, 1995.
- T. Laloë, R. Servien, and E. Di Bernardino. Estimating covariate functions associated to multivariate risks : a level set approach. *Metrika, Springer Verlag*, pages 497–526, 2015.
- R. Torres, E. Di Bernardino, H. Laniado, and R. Lillo. On the estimation of extreme directional multivariate quantiles. *Communications in Statistics-Theory and Methods*, 49(22) :5504–5534, 2020.
- Y. Zuo and R. Serfling. General notions of statistical depth function. *Annals of statistics*, pages 461–482, 2000.