

Une méthode de folding pour des extrêmes multivariés avec application à l'estimation de la mesure de probabilité spectrale

Armelle Guillo, Philippe Naveau, Alexandre You

► **To cite this version:**

Armelle Guillo, Philippe Naveau, Alexandre You. Une méthode de folding pour des extrêmes multivariés avec application à l'estimation de la mesure de probabilité spectrale. 42èmes Journées de Statistique, 2010, Marseille, France, France. 2010. <inria-00494782>

HAL Id: inria-00494782

<https://hal.inria.fr/inria-00494782>

Submitted on 24 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNE MÉTHODE DE FOLDING POUR DES EXTRÊMES MULTIVARIÉS AVEC APPLICATION À L'ESTIMATION DE LA MESURE DE PROBABILITÉ SPECTRALE

A FOLDING METHODOLOGY FOR MULTIVARIATE EXTREMES WITH APPLICATION TO THE ESTIMATION OF THE SPECTRAL PROBABILITY MEASURE

Armelle Guillou⁽¹⁾ & Philippe Naveau⁽²⁾ & Alexandre You⁽³⁾

⁽¹⁾ *Institut de Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et
CNRS, 7 rue René Descartes, 67084 Strasbourg Cedex, France*

⁽²⁾ *Laboratoire des Sciences du Climat et de l'Environnement, Orme des Merisiers /
Bat.701 C.E. Saclay, 91191 Gif-sur-Yvette, France*

⁽³⁾ *Université Paris VI, LSTA, Boîte 158, 175 rue du Chevaleret, 75013 Paris, France*

Mot clé principal: Extrêmes

Mot clé secondaire: Statistique Mathématique

Résumé : Nous nous plaçons dans le cas d'un échantillon indépendant et identiquement distribué de vecteurs aléatoires $\mathbf{X}_1, \dots, \mathbf{X}_n$ de loi F sur $R_+^d = [0, \infty)^d$. Soit $\|\cdot\|$ une norme sur R^d et $\mathfrak{N} := \{\mathbf{x} \in R^d : \|\mathbf{x}\| = 1\}$ la sphère unité de R^d relativement à cette norme. On note

$$T(\mathbf{X}) := (R, \Theta) = \left(\|\mathbf{X}\|, \frac{\mathbf{X}}{\|\mathbf{X}\|} \right)$$

les coordonnées polaires d'un vecteur aléatoire \mathbf{X} . On définit les espaces $E := \overline{R_+^d} \setminus \{\mathbf{0}\}$ et $\mathfrak{N}_+ := \mathfrak{N} \cap E$ où $\overline{R_+} := [0, \infty]$. $M_+(E)$ et $M_+((0, \infty] \times \mathfrak{N}_+)$ désignent respectivement l'ensemble des mesures de Radon sur E et sur $(0, \infty] \times \mathfrak{N}_+$. Notre principale hypothèse de travail consiste à supposer que la queue de la distribution jointe F est à variation régulière, c'est-à-dire qu'il existe une mesure de Radon ν sur E et une suite $b_n \rightarrow \infty$ telles que, pour tout vecteur aléatoire \mathbf{X} de loi F

$$nP \left(\frac{\mathbf{X}}{b_n} \in \cdot \right) \xrightarrow{v} \nu(\cdot), \quad n \rightarrow \infty \quad (1)$$

dans $M_+(E)$. De façon équivalente, il existe une mesure de probabilité $S(\cdot)$ sur les

boréliens de \mathfrak{N}_+ et une suite $b_n \rightarrow \infty$ telles que

$$nP\left(\left(\frac{R}{b_n}, \Theta\right) \in \cdot\right) \xrightarrow{v} c\nu_\alpha \times S(\cdot), \quad n \rightarrow \infty$$

dans $M_+((0, \infty] \times \mathfrak{N}_+)$ pour des constantes $c > 0$ et $\alpha > 0$, où $\nu_\alpha(r, \infty] = r^{-\alpha}$. S est la mesure de probabilité spectrale qui a été longuement étudiée dans la littérature (cf. e.g. Einmahl *et al.*, 1993, 1997, 2001; Einmahl et de Haan, 2009).

Nous commencerons notre exposé par introduire une nouvelle approche d'estimation dans le contexte de la théorie des valeurs extrêmes, appelée « folding ». Celle-ci a été développée par You *et al.* (2010) dans un cadre univarié. Elle consiste en une « transformation des données initiales » et a été initialement introduite dans un contexte de « perfect sampling » (Corcoran et Schneider, 2003). Cette nouvelle approche a permis d'améliorer de façon significative l'estimation de quantiles extrêmes en adaptant la méthode pic au delà d'un seuil (POT). Notre objectif dans cet exposé sera donc double : dans un premier temps, nous étendrons le concept de folding dans le cas multivarié ; dans un second temps, nous appliquerons cette extension du folding à un problème d'estimation classique en théorie des extrêmes multivariés, à savoir l'estimation de la mesure de probabilité spectrale S . Nous illustrerons, par le biais de simulations et sur des données réelles nos résultats.

Mots clés: Théorie des valeurs extrêmes, multivarié, folding, mesure spectrale.

Abstract: In this talk we consider a random vector $\mathbf{X} = (X^{(1)}, \dots, X^{(d)})$ with a distribution function F on $R_+^d = [0, \infty)^d$. Let $\|\cdot\|$ be a norm on R^d and $\mathfrak{N} := \{\mathbf{x} \in R^d : \|\mathbf{x}\| = 1\}$ the unit sphere of R^d with respect to this norm. Denote by

$$T(\mathbf{X}) := (R, \Theta) = \left(\|\mathbf{X}\|, \frac{\mathbf{X}}{\|\mathbf{X}\|} \right)$$

the polar coordinates of \mathbf{X} , where R is the *radius* of \mathbf{X} and Θ its *polar angle*. Define the spaces $E := \overline{R_+^d} \setminus \{\mathbf{0}\}$ and $\mathfrak{N}_+ := \mathfrak{N} \cap E$ where $\overline{R_+} := [0, \infty]$. $M_+(E)$ and $M_+((0, \infty] \times \mathfrak{N}_+)$ are the sets of all (non negative) Radon measures on E and $(0, \infty] \times \mathfrak{N}_+$, respectively. We assume that \mathbf{X} has a regularly varying tail $\overline{F} := 1 - F$, that is, there exist a Radon measure ν on E and $b_n \rightarrow \infty$ such that

$$nP\left(\frac{\mathbf{X}}{b_n} \in \cdot\right) \xrightarrow{v} \nu(\cdot), \quad n \rightarrow \infty \tag{2}$$

in $M_+(E)$, where \xrightarrow{v} denotes the vague convergence of measures. An equivalent statement in terms of the polar coordinates is the following: there exist a probability measure $S(\cdot)$ on Borel subsets of \mathfrak{N}_+ and $b_n \rightarrow \infty$ such that

$$nP\left(\left(\frac{R}{b_n}, \Theta\right) \in \cdot\right) \xrightarrow{v} c\nu_\alpha \times S(\cdot), \quad n \rightarrow \infty \quad (3)$$

in $M_+((0, \infty] \times \mathfrak{N}_+)$ for some $c > 0$ and $\alpha > 0$, where $\nu_\alpha(r, \infty] = r^{-\alpha}$. Hence, the radial component and the angular component are asymptotically independent. S is the so-called *spectral probability measure* which summarizes the dependency among extremes. In statistics of multivariate extremes, estimation of the spectral probability measure is a crucial problem which has focused much attention in the literature (see, e.g., Einmahl *et al.*, 1993, 1997, 2001; Einmahl and de Haan, 2009).

First, we will introduce a new approach, developed by You *et al.* (2010), for statistical estimation problems in extreme value theory. This method relies on the use of a *folding* transformation, defined by Corcoran and Schneider (2003) in the context of « perfect sampling » which allows to connect the center of the distribution and its tail. This transformation has been studied in the univariate independent and identically distributed case and an adaptation of the classical Peaks-Over-Thresholds approach has been proposed. Simulations tend to indicate that this method improves significantly the performances of extreme quantiles estimators in finite sample situations. In this talk, our aim is to extend the concept of the folding transformation in a multivariate extreme value framework with some application to statistical estimation problems. We will start by introducing a multidimensional version of the folding transformation and the statistical procedure which allows us to build the folded version of the sample. Then, we will propose a class of estimators of the spectral probability measure based on the folded sample and we will show their weak consistency. Finally, we will illustrate the performances of these estimators through numerical applications and simulations.

Keywords: Extreme value theory, multivariate, folding, spectral measure.

Bibliographie

- [1] Corcoran, J.N., Schneider, U. (2003). Shift and scale coupling methods for perfect simulation. *Probability in the Engineering and Informational Sciences*, **17**, 277-303.
- [2] Einmahl, J.H.J., de Haan, L. (2009). Estimating extreme bivariate quantile regions. *Discussion Paper 2009-29*, Tilburg University, Center for Economic Research.
- [3] Einmahl, J.H.J., de Haan, L., Huang, X. (1993). Estimating a multidimensional extreme-value distribution. *Journal of Multivariate Analysis*, **47**, 35-47.

- [4] Einmahl, J.H.J., de Haan, L., Piterbarg, V.I. (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *Annals of Statistics*, **29**, 1401-1423.
- [5] Einmahl, J.H.J., de Haan, L., Sinha, A.K. (1997). Estimating the spectral measure of an extreme value distribution. *Stochastic Processes and their Applications*, **70**, 143-171.
- [6] You, A., Schneider, U., Guillo, A., Naveau, P. (2010). Improving extreme quantile estimation via a folding procedure. *À paraître dans Journal of Statistical Planning and Inference*.