

Estimation of a discrete spectral measure and actuarial applications

Esterina Masiello ¹

Institut de Science Financière et d'Assurances – Université Lyon 1, France

Anne-Laure Fougères

Institut Camille Jordan – Université Lyon 1, France

Philippe Soulier

Modal'X – Université Paris-Ouest, France

Abstract: Let G be a bivariate extreme value distribution function. Various alternative ways to describe the dependence structure of G have been proposed in the literature. In this paper we focus on the spectral measure, introduced by de Haan and Resnick (1977). The estimation of this spectral measure is one of the main issues in multivariate extreme value theory. Working under the assumption that the spectral measure ν is a point measure, how can we estimate ν ? We propose a simple estimator and study some of its theoretical properties. Finally, we show how this estimator provides useful in the construction of an estimator of the multivariate ruin probability.

Keywords: Multivariate extremes, dependence structure, spectral measure, multivariate ruin probability

1. Introduction

Let $\|\cdot\|$ denote any norm on $\mathbb{R}^2$ and let $\mathbb{S}_1 = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| = 1\}$ denote the unit sphere for this norm. Let $\mathbf{Y}$ be a bivariate random vector whose distribution is *regularly varying* with tail index $\alpha > 0$, i.e. there exists a probability measure ν on the unit sphere $\mathbb{S}_1$ such that

$$\lim_{r \rightarrow \infty} \frac{P(\|\mathbf{Y}\| > yr, \mathbf{Y}/\|\mathbf{Y}\| \in B)}{P(\|\mathbf{Y}\| > r)} = y^{-\alpha} \nu(B) \quad (1)$$

for every $y > 0$ and Borel sets $B \subset \mathbb{S}_1$ with $\nu(\partial B) = 0$.

The probability measure ν is called the spectral measure of $\mathbf{Y}$. It describes in which directions we are likely to find extreme realizations of $\mathbf{Y}$. Let us just remark that if we put $y = 1$ in (1), then ν is the limiting distribution, as $r \rightarrow \infty$, of $\mathbf{Y}/\|\mathbf{Y}\|$ conditioned on the event $\{\|\mathbf{Y}\| > r\}$. The spectral measure was introduced by de Haan and Resnick (1977). For a general background see, for instance, the monographs by Beirlant et al. (2004) and de Haan and Ferreira (2006).

The form of the spectral measure is norm-dependent. For every choice of the norm, the spectral measure is a measure of dependence between the extreme values of the univariate components of $\mathbf{Y}$. Let the norm be the usual Euclidean norm.

The spectral measure will be defined on the interval $[0, \pi/2]$ and its density evaluated at $\theta \in [0, \pi/2]$ can be interpreted as a measure of the likelihood that $\arctan(y_2/y_1) = \theta$

¹Address of correspondence: esterina.masiello@univ-lyon1.fr

provided that at least one of the components was extreme. If most of the mass concentrates around $\pi/4$ then the occurrences of extreme movements in the two components are highly correlated, i.e. an extreme value in one component is almost surely matched by an extreme movement in the other. Most of the mass around 0 and $\pi/2$ corresponds to the situation when the occurrences of extreme on the two components are independent, i.e. extreme values in one component do not illicit extreme values from the other.

In this paper, we are interested in discrete spectral measures which arise whenever extreme values are generated by a finite number of independent heavy-tailed factors. An example is the following model

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{U}$$

where $\mathbf{X} = (X_1, \dots, X_k)^T$ is a $k \times 1$ random vector of *i.i.d.* random variables with common marginal Pareto-type tail, $\mathbf{U}$ is a $d \times 1$ column vector of noise variables having lighter tail than the $\mathbf{X}$ variables and $\mathbf{A}$ is a $d \times k$ matrix of coefficients. The intuitive interpretation is that the extreme observations of $\mathbf{Y}$ are dominated by a single component of $\mathbf{X}$. Therefore, in the limit, there are only a finite number of directions for extreme outcomes of $\mathbf{Y}$.

2. Estimation procedure and main result

Parametric estimators of the spectral measure have been proposed by Coles and Tawn (1991) and Joe et al. (1992). An alternative to the parametric approach is the nonparametric one: see Einmhal et al. (1997) and Einmhal et al. (2001) who work under the assumption that the density exists. Our main assumption is that the spectral measure ν is a point measure, i.e. there exist $\theta_1, \dots, \theta_K \in \mathbb{S}_1$ and real numbers $b_1, \dots, b_K$ such that $b_1 + \dots + b_K = 1$ and

$$\nu = \sum_{j=1}^K b_j \delta_{\theta_j} . \quad (2)$$

We want to estimate the points $\theta_1, \dots, \theta_K$ and the masses $b_1, \dots, b_K$ when K is known. We assume that $b_1 > b_i, i \geq 2$, i.e. θ_1 is the mode of the distribution ν .

Let h be an integer and $A_1, \dots, A_h$ be a regular partition of the unit sphere. For instance, if the norm is the L^2 norm, then L^2 is isomorphic to $[-\pi, \pi]$, and A_j can be an interval of length $2\pi/h$. Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be i.i.d. with the same distribution as $\mathbf{X}$ and k be an intermediate sequence, i.e. such that $k \rightarrow \infty$ and $n/k \rightarrow \infty$.

We define an estimator $\hat{\theta}$ of θ_1 as the midpoint which contains the greatest proportion of points among the k points with the largest radius.

Denote $R = \|\mathbf{X}\|$ and $\Theta = R^{-1}\mathbf{X}$ and let $R_{(1)} > \dots > R_{(n)}$ be the order statistics of $R_1, \dots, R_n$. Let $\hat{\nu}_n$ be the empirical estimator of the spectral measure defined by

$$\hat{\nu}_n(A) = \frac{1}{k} \sum_{i=1}^n 1_{\{\Theta_i \in A\}} 1_{\{R_i \geq R_{(k)}\}} . \quad (3)$$

It is known (Resnick (2007)) that $\hat{\nu}_n \Rightarrow \nu$. Define $\hat{j} = \arg \max_{1, \dots, K} \hat{\nu}_n(A_j)$, and let $\hat{\theta}$ be the midpoint of $A_{\hat{j}}$. Under weak conditions, $\hat{\theta}$ is shown to be a consistent estimator of θ_1 .

3. An application: the Poisson shock model

In the classical ruin problem, an insurance company gains capital from premium incomes and loses capital as a result of claims. Estimates are then given for the probability that the company ever incurs ruin, that is the probability that company's total capital becomes negative.

More precisely, the risk reserve process is modeled by $R(t) = u + tp - S(t)$, where u denotes the initial capital, p is the premium rate per unit time and $S(t)$ is a stochastic process modeling the amount of cumulative claims up to time t . The ruin probability is then given by $\psi(u) = P(R(t) < 0 \text{ for some } t \geq 0)$.

Ruin theory for the univariate risk model has been discussed extensively in the literature; for a summary of main results, see, for example, Asmussen (2000). An extension of classical risk theory towards a multidimensional model is possible while allowing $R(t)$, u , p and $S(t)$ to be vectors, with possible dependence between the components of $\mathbf{S}(t)$.

As above, vectors will be denoted by bold letters, capitals for random vectors and lower case for non-random vectors.

Really, we are interested in insurance claims which occur in different lines of business (for example, driving insurance, house insurance, health, etc.) of the same company or in different countries. We suppose that there are strong a priori reasons for believing that claims of different lines of business (or countries) are dependent. More precisely, we assume that all claims can be related to a series of underling and independent shock processes. In insurance these shocks might be natural catastrophes: hurricanes, for example. The most common tool to take this into account is the Poisson shock model, i.e. we assume that the different varieties of shocks arrive as independent Poisson processes, in which case the counting processes for the different loss types are also Poisson.

The mathematical framework is as follows. Let us consider an insurance company with d lines of business. Suppose $\{\mathbf{Z}_k, k \geq 1\}$ are *i.i.d.* $\mathbb{R}^d$ -valued random vectors representing claim sizes and Z is regularly varying, i.e. satisfies (1), with $\alpha > 1$. Let $N_t = \#\{n \geq 1 : T_n \leq t\}$ denote the total number of claims up to time t and $T_0 = 0$, $T_n = W_1 + \dots + W_n$ where $\{W_k, k \geq 1\}$ is a sequence of *i.i.d.* random variables representing the interarrival times.

We consider a d -dimensional risk process $\{\mathbf{R}_t, t \geq 0\}$ given by

$$\mathbf{R}_t = u\mathbf{b} + t\mathbf{p} - \sum_{k=1}^{N_t} \mathbf{Z}_k$$

where u is the initial capital and capital allocation to the different lines of business is determined according to $u\mathbf{b}$ with $\mathbf{b} \in (0, 1]^d$ and $b^{(1)} + \dots + b^{(d)} = 1$. $\mathbf{p} \in (0, \infty)^d$ is the premium income rate which is assumed to be constant.

The quantity of interest is the multivariate ruin probability which can be defined as

$$\psi_{d,F}(u) = P(\mathbf{R}_t \in F \text{ for some } t \geq 0)$$

for a given ruin set $F \subset \mathbb{R}^d$. Hence, $\psi(u)$ is the probability that the multivariate risk reserve process $\mathbf{R}_t$ enter some domain of $\mathbb{R}^d$, called insolvency region (or ruin set).

Let us define $\mathbf{c} = E(W)\mathbf{p} - E(\mathbf{Z})$ and assume that the net profit condition, $\mathbf{c} \in (0, \infty)^d$, holds, that is, each business line has a positive safety loading. Assume further that $E(W^\gamma) < \infty$ for some $\gamma > \alpha$. Under these hypothesis, Hult and Lindskog (2006) state

that it is possible to approximate the ruin probability of an insurance company for large initial capital u by

$$\psi_{d,F}(u) \approx \int_0^\infty \mu(v\mathbf{c} + \mathbf{b} - F)dv u P(|\mathbf{Z}| > u) \quad (4)$$

where μ is a measure on $\mathbb{R}^d \setminus \{0\}$ related to the spectral measure ν through

$$\mu(\mathbf{z} : |\mathbf{z}| > r, \mathbf{z}/|\mathbf{z}| \in S) = r^{-\alpha} \nu(S)$$

for $r > 0$ and Borel sets $S \subset \mathbb{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}$.

Once all the quantities on the right hand side of 4 are estimated, a simple estimator by plug-in method of the multivariate ruin probability is obtained, whose behavior is illustrated through a simulation study.

References

- Asmussen, S. (2000) *Ruin probabilities*. Advanced Series on Statistical Science and Applied Probability Vol. 2, World Scientific.
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J. (2004) *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, Chichester.
- Coles, S. and J., Tawn (1991) Modelling extreme multivariate events, *Journal of the Royal Statistical Society, Series B*, 53, 377392.
- Einmahl, J., de Haan, L. and Piterbarg, V. (2001) Nonparametric estimation of the spectral measure of an extreme value distribution, *Annals of Statistics*, 29, 14011423.
- Einmahl, J., de Haan, L. and Sinha, A. (1997) Estimating the spectral measure of an extreme value distribution, *Stoch. Proc. Appl.*, 70, 143171.
- de Haan, L. and Ferreira, A. (2006) *Extreme Value Theory: An Introduction*, New York: Springer.
- de Haan, L. and Resnick, S. (1977) Limit theory for multidimensional sample extremes, *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 40, 317337.
- Hult, H. and Lindskog F. (2006) Heavy-tailed insurance portfolios: buffer capital and ruin probabilities, *Technical Report No. 1441, School of ORIE, Cornell University*.
- Joe, H., Smith, R. L. and Weissman, I. (1992) Bivariate threshold methods for extremes. *J. Roy. Statist. Soc., Ser. B*, 54, 171183.