HAL open science

Étude statistique du coefficient de covariation symétrique signé

Bernédy Kodia, Bernard Garel

▶ To cite this version:

Bernédy Kodia, Bernard Garel. Étude statistique du coefficient de covariation symétrique signé. 42èmes Journées de Statistique, 2010, Marseille, France, France. inria-00494791

HAL Id: inria-00494791

https://inria.hal.science/inria-00494791

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.

Étude statistique du coefficient de covariation symétrique signé.

Bernédy KODIA & Bernard GAREL

Institut de Mathématiques de Toulouse, UPS et INP-ENSEEIHT.
E-mails : bernedy.kodia@enseeiht.fr, garel@enseeiht.fr.

Résumé

Le coefficient de covariation symétrique signé est une mesure permettant de capter la dépendance entre variables aléatoires α-stables symétriques. Dans le cas des vecteurs aléatoires sous-gaussiens, la matrice des coefficients de covariation symétriques signés coïncide avec la matrice de corrélation du vecteur gaussien sous-jacent. Deux estimateurs de ce coefficient sont proposés : le premier basé sur les moments fractionnaires d'ordre inférieur et le second sur les "screened ratio". Nous procédons à l'analyse de la robustesse de ces estimateurs, en utilisant des données normales contaminées.

Abstract

Signed symmetric covariation coefficient is a measure which allows to capture dependence between symmetric α-stable random variables. In the case of sub-Gaussian random vectors, the signed symmetric covariation coefficient matrix coincide with the correlation matrix of the underlying Gaussian vector. Two estimators of this coefficient are proposed: the first is based on fractional lower order moments and the second on screened ratio. We carry out a robustness study of these estimators, using contaminated normal data.

1 Introduction

Depuis les travaux pionniers de Mandelbrot (1960), les lois α-stables ont été l'objet d'un intérêt grandissant ces dernières décennies. Ces lois sont de plus en utilisées dans des domaines aussi divers que les finances, les télécommunications, la météorologie, la physique, la biologie, etc. Elles présentent de nombreuses propriétés très attractives pour les praticiens : elles prennent en compte l'asymétrie et les queues lourdes. Ce sont également les seules lois pour lesquelles le théorème limite centrale généralisé est encore applicable, faisant ainsi de la loi normale un élément de cette grande famille. Les lois de Cauchy et de Lévy sont deux autres lois bien connues de cette famille. L'inconvénient majeur des lois α-stables est que, à l'exception des trois lois mentionnées précédemment, on ne dispose pas de forme explicite de leur densité de probabilité mais seulement de leur fonction caractéristique. De nombreux algorithmes de simulation de variables aléatoires α-stables ont été implémentés, ce qui les rend plus abordables pour les praticiens. Les vecteurs α-stables sont plus difficiles à simuler. Un autre écueil est que les lois α-stables non-gaussiennes ne possèdent pas de moments de second ordre, d'où la non-existence de la matrice de corrélation.

Garel et Kodia [1] ont proposé le coefficient de covariation symétrique signé, comme mesure de dépendance permettant de capter la dépendance entre deux variables aléatoires α-stables

symétriques. Dans cet article, nous présentons une forme révisée de ce coefficient. Nous montrons que dans le cas des vecteurs aléatoires sous-gaussiens, la matrice des coefficients de covariation symétriques signés coïncide avec la matrice de corrélation du vecteur gaussien sous-jacent. Nous présentons deux estimateurs convergents de cette quantité. Nous procédons à l'analyse de la robustesse de ces estimateurs.

Dans la section 2 nous donnons quelques rappels de base sur les variables et vecteurs α-stables. Nous donnons la nouvelle écriture du coefficient de covariation symétrique signé dans la section 3. La section 4 est consacrée aux estimateurs de cette quantité et à leurs propriétés asymptotiques. Nous terminons cette partie par une analyse de la robustesse des estimateurs.

2 Les variables et vecteurs aléatoires α-stables

Une variable aléatoire α-stable X est caractérisée par quatre paramètres : $0 < \alpha \leq 2, \gamma \geq 0, -1 \leq \beta \leq 1$ et δ un réel. Le paramètre α est l'index de stabilité, β est le paramètre d'asymétrie, γ est un paramètre d'échelle et δ est un paramètre de position. Sa fonction caractéristique s'écrit :

$$E \exp i\theta X = \exp\Big\{ -\gamma^\alpha |\theta|^\alpha \left[1 + i\beta \mathrm{sign}(\theta) w(\theta,\alpha)\right] + i\delta\theta \Big\}, \tag{1}$$

où

$$w(\theta,\alpha) = \begin{cases} -\tan\frac{\pi\alpha}{2} & \text{si} \quad \alpha \neq 1, \\ \frac{2}{\pi}\ln|\theta| & \text{si} \quad \alpha = 1, \end{cases}$$

et $\mathrm{sign}(\theta) = 1$ si $\theta > 0$, $\mathrm{sign}(\theta) = -1$ si $\theta < 0$ et $\mathrm{sign}(\theta) = 0$ si $\theta = 0$. La variable alétoire X sera notée $S_\alpha(\gamma,\beta,\delta)$. Quand $\beta = \delta = 0$, la variable alétoire X est dite symétrique (c'est-à-dire que $-X$ et X ont même loi) et est notée $S\alpha S$.

La fonction caractéristique d'un vecteur aléatoire α-stable $\mathbf{X} = (X_1, X_2)$, pour $0 < \alpha < 2$, est donnée par

$$E \exp\left(i\langle \boldsymbol{\theta}, \mathbf{X}\rangle\right) = \exp\left\{ -\int_{S_2} |\langle \boldsymbol{\theta}, \mathbf{s}\rangle|^\alpha \left[1 + i\mathrm{sign}(\boldsymbol{\theta},\mathbf{s}) w(\langle \boldsymbol{\theta},\mathbf{s}\rangle,\alpha)\right] \boldsymbol{\Gamma}(d\mathbf{s}) + i\langle \boldsymbol{\theta}, \boldsymbol{\delta}\rangle \right\}, \tag{2}$$

où $\boldsymbol{\Gamma}$ est une mesure finie sur le cercle unité $S_2 = \{\mathbf{s} \in \mathbb{R}^2 : \|\mathbf{s}\| = 1\}$ et $\boldsymbol{\delta}$ est un vecteur de position. Ici $\langle \boldsymbol{\theta}, \mathbf{s}\rangle$ désigne le produit scalaire dans $\mathbb{R}^2$. La mesure $\boldsymbol{\Gamma}$ est appelée mesure spectrale du vecteur aléatoire α-stable $\mathbf{X}$ et le couple $(\boldsymbol{\Gamma}, \boldsymbol{\delta})$ est unique. Le vecteur est symétrique si et seulement si $\boldsymbol{\delta} = 0$ et $\boldsymbol{\Gamma}$ est une mesure symétrique sur S_2. Dans ce cas, sa fonction caractéristique est donnée par

$$E \exp\{i\langle \boldsymbol{\theta}, \mathbf{X}\rangle\} = \exp\left\{ -\int_{S_2} |\langle \boldsymbol{\theta}, \mathbf{s}\rangle|^\alpha \boldsymbol{\Gamma}(d\mathbf{s}) \right\}. \tag{3}$$

La mesure spectrale porte l'information essentielle sur le vecteur, en particulier sur la structure de dépendance entre ses composantes. Il s'ensuit que toutes les mesures de dépendance s'expriment en fonction de cette dernière.

Miller (1978) a introduit un nouveau concept de dépendance, la covariation, opérant entre variables aléatoires symétriques α-stables avec $\alpha > 1$.

Définition 2.1 *Soient X_1 et X_2 deux variables aléatoires conjointement $S\alpha S$ avec $\alpha > 1$ et soit $\mathbf{\Gamma}$ la mesure spectrale du vecteur aléatoire $\boldsymbol{X} = (X_1, X_2)$. La covariation de X_1 sur X_2 est définie par*

$$[X_1, X_2]_\alpha = \int_{S_2} s_1 s_2^{<\alpha-1>} \mathbf{\Gamma}(d\boldsymbol{s}), \tag{4}$$

où $a^{<p>} = |a|^p \text{sign}(a)$ est appelée puissance signée. Bien qu'elle soit linéaire en son premier argument, la puissance signée fait qu'en général la covariation ne soit pas linéaire en son deuxième argument et non symétrique en ses arguments. A partir de la covariation on définit une norme appelée norme de covariation comme suit

$$\|X_1\|_\alpha = ([X_1, X_1]_\alpha)^{1/\alpha} = \gamma_{X_1}, \tag{5}$$

où γ_{X_1} est le paramètre d'échelle de la variable aléatoire $S\alpha S$ X_1.

3 Le coefficient de covariation symétrique signé

Le coefficient de covariation symétrique signé est une mesure de dépendance construite à partir de la covariation. Cependant, à la différence de la covariation, ce coefficient est borné. Nous proposons ici une expression révisée du signe de ce coefficient, ce qui permet d'écarter des problèmes d'asymétrie apparente apparaissant dans la définition donnée dans Garel et Kodia [1].

3.1 Définition et premières propriétés

Définition 3.1 *Soit (X_1, X_2) un vecteur aléatoire $S\alpha S$ avec $\alpha > 1$. Le coefficient de covariation symétrique signé entre X_1 et X_2 est la quantité :*

$$\text{scov}(X_1, X_2) = \kappa_{(X_1,X_2)} \left| \frac{[X_1, X_2]_\alpha [X_2, X_1]_\alpha}{\|X_1\|_\alpha^\alpha \|X_2\|_\alpha^\alpha} \right|^{\frac{1}{2}}, \tag{6}$$

où

$$\kappa_{(X_1,X_2)} = \begin{cases} \text{sign}([X_1, X_2]_\alpha) & \text{si} \quad \text{sign}([X_1, X_2]_\alpha) = \text{sign}([X_2, X_1]_\alpha), \\ -1 & \text{si} \quad \text{sign}([X_1, X_2]_\alpha) = -\text{sign}([X_2, X_1]_\alpha). \end{cases} \tag{7}$$

Ce coefficient vérifie les propriétés suivantes : (i) $-1 \leq \text{scov}(X_1, X_2) \leq 1$ et si X_1, X_2 sont indépendantes, alors $\text{scov}(X_1, X_2) = 0$. (ii) $|\text{scov}(X_1, X_2)| = 1$ si et seulement si $X_2 = \lambda X_1$ pour tout $\lambda \in \mathbb{R}$. (iii) Pour $\alpha = 2$, $\text{scov}(X_1, X_2)$ coïncide avec le coefficient de corrélation habituel. (iv) Soient a et b deux réels non nuls, alors

$$\text{scov}(aX_1, bX_2) = \begin{cases} \text{sign}(ab)\text{scov}(X_1, X_2) & \text{si} \quad \text{sign}([X_1, X_2]_\alpha) = \text{sign}([X_2, X_1]_\alpha), \\ \text{scov}(X_1, X_2) & \text{si} \quad \text{sign}([X_1, X_2]_\alpha) = -\text{sign}([X_2, X_1]_\alpha) \end{cases} \tag{8}$$

Afin d'illustrer l'intérêt de ce nouveau coefficient, nous allons tout d'abord étudier le cas des vecteurs aléatoires sous-gaussiens.

3.2 Cas des vecteurs aléatoires sous-gaussiens

Soient $0 < \alpha < 2$, G_1, G_2 des variables conjointement normales de moyenne nulle et A une variable aléatoire positive, indépendante de (G_1, G_2), telle que $A \sim S_{\alpha/2}\big(\big(\cos\frac{\pi\alpha}{4}\big)^{2/\alpha}, 1, 0)$. Alors $\mathbf{X} = A^{1/2}\mathbf{G} = (A^{1/2}G_1, A^{1/2}G_2)$ est un vecteur aléatoire α-stable appelé sous-gaussien de vecteur gaussien sous-jacent $\mathbf{G} = (G_1, G_2)$. La fonction caractéristique de $\mathbf{X}$ est donnée par :

$$E\exp\Big\{i\sum_{k=1}^{2}\theta_k X_k\Big\} = \exp\Big\{-\Big|\frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}\theta_i\theta_j R_{ij}\Big|^{\alpha/2}\Big\}, \tag{9}$$

où les $R_{ij} = EG_iG_j$, $i, j = 1, 2$ sont les covariances du vecteur gaussien sous-jacent $\mathbf{G}$ (Samorodnitsky et Taqqu, 1994 page 78).

Proposition 3.2 *Soient $1 < \alpha < 2$ et $\boldsymbol{X} = (X_1, X_2)$ un vecteur sous gaussien de fonction caractéristique (9), alors la matrice des coefficients de covariation symétriques signés de $\boldsymbol{X}$ coïncide avec la matrice de corrélation du vecteur gaussien sous-jacent.*

Il s'agit là d'un résultat important car les vecteurs aléatoires sous-gaussiens sont complètement caractérisés par l'index de stabilité α et la matrice de corrélation du vecteur gaussien sous-jacent associé.

4 Estimation du coefficient de covariation symétrique signé

4.1 Les estimateurs du coefficient de covariation symétrique signé

Considérons, pour $1 \le p < \alpha$, la quantité

$$\mathrm{scov}^{FL}(X_1, X_2) = \hat{\kappa}_{(X_1,X_2)}\frac{\Big|\sum_{j,k=1}^{n} X_{1j}X_{2k}X_{1k}^{<p-1>}X_{2j}^{<p-1>}\Big|^{\frac{1}{2}}}{\Big(\sum_{j,k=1}^{n}|X_{1j}|^p|X_{2k}|^p\Big)^{1/2}} \tag{10}$$

où

$$\hat{\kappa}_{(X_1,X_2)} = \begin{cases} \mathrm{sign}\Big(\sum_{j=1}^{n}X_{1j}X_{2j}^{<p-1>}\Big) & \text{si} \quad \mathrm{sign}\Big(\sum_{j=1}^{n}X_{1j}X_{2j}^{<p-1>}\Big) = \mathrm{sign}\Big(\sum_{j=1}^{n}X_{2j}X_{1j}^{<p-1>}\Big) \\ -1 & \text{sinon.} \end{cases}$$

Les couples $(X_{11}, X_{21}), ..., (X_{1n}, X_{2n})$ sont des observations de (X_1, X_2) indépendantes. Soient

c_1 et c_2 deux constantes positives, considérons également la quantité

$$\mathrm{scov}^{SR}(X_1, X_2) = \tilde{\kappa}_{(X_1,X_2)} \frac{\Big| \sum_{j,k=1}^{n} X_{1j} X_{2k} X_{1k}^{-1} X_{2j}^{-1} \mathbb{I}_{]c_1,c_2[}(|X_{1k}|, |X_{2j}|) \Big|^{\frac{1}{2}}}{\Big(\sum_{j,k=1}^{n} \mathbb{I}_{]c_1,c_2[}(|X_{1k}|, |X_{2j}|) \Big)^{1/2}} \tag{11}$$

où

$$\tilde{\kappa}_{(X_1,X_2)} = \begin{cases} \mathrm{sign}\Big(\sum_{j=1}^{n} X_{1j} X_{2j}^{-1} \mathbb{I}_{]c_1,c_2[}(|X_{2j}|) \Big) & \text{si} \quad A_1 = A_2 \\ -1 & \text{sinon,} \end{cases}$$

avec $A_1 = \mathrm{sign}\Big(\sum_{j=1}^{n} X_{1j} X_{2j}^{-1} \mathbb{I}_{]c_1,c_2[}(|X_{2j}|) \Big)$ et $A_2 = \mathrm{sign}\Big(\sum_{j=1}^{n} X_{2j} X_{1j}^{-1} \mathbb{I}_{]c_1,c_2[}(|X_{1j}|) \Big)$

Proposition 4.1 *Les quantités* scov^{FL} *et* scov^{SR} *sont deux estimateurs convergents du coefficient de covariation symétrique signé.*

4.2 Le coefficient de covariation symétrique signé comme outil d'estimation robuste

Les deux estimateurs précédents fonctionnent avec des distributions ayant une variance infinie, donc présentant un nombre non négligeagle de valeurs extrêmes. Ces considérations nous ont fait penser que de tels estimateurs pourraient fournir des estimations à la fois consistantes et robustes dans le cas gaussien contaminé. Nous avons généré des données gaussiennes contaminées de la forme

$$p\mathcal{N}_n(0, \Sigma) + (1-p)\mathcal{N}_n(0, \Sigma + \kappa I), \tag{12}$$

où Σ est une matrice de covariance $n \times n$, I est la matrice identité et $\kappa > 0$ est une constante à spécifier. Nous avons généré des données de dimension 3, pour lesquelles la matrice de covariance Σ est donnée par

$$\Sigma = \begin{pmatrix} 17.95 & 5.85 & 3.93 \\ 5.85 & 5.36 & 1.83 \\ 3.93 & 1.83 & 1.05 \end{pmatrix}.$$

La matrice de corrélation associée est

$$\boldsymbol{r}_\Sigma = \begin{pmatrix} 1 & 0.596 & 0.905 \\ 0.596 & 1 & 0.772 \\ 0.905 & 0.772 & 1 \end{pmatrix}.$$

Nous posons $p = 0.9$ et $\kappa = 100$. Nous posons $c_1 = 0.3$ et $c_2 = 4$ dans l'estimateur (**??**). Nous calculons les moyennes des estimations de la matrice de corrélation du mélange sur 100 réplications

d'un échantillon de taille 100. Les valeurs entre parenthèses sont les écarts arithmétiques moyens à la moyenne figurant au dessus. Nous obtenons

$$\boldsymbol{\rho}_{\widehat{r}} = \begin{pmatrix} \underset{0.000}{1} & \underset{0.117}{0.311} & \underset{0.166}{0.247} \\ \underset{0.117}{0.311} & \underset{0.000}{1} & \underset{0.208}{0.125} \\ \underset{0.166}{0.247} & \underset{0.208}{0.125} & \underset{0.000}{1} \end{pmatrix}, \quad \boldsymbol{\rho}_{\widehat{\mathrm{SCOV}}^{\mathrm{FL}}} = \begin{pmatrix} \underset{0.000}{1} & \underset{0.086}{0.449} & \underset{0.139}{0.563} \\ \underset{0.086}{0.449} & \underset{0.000}{1} & \underset{0.153}{0.424} \\ \underset{0.139}{0.563} & \underset{0.153}{0.424} & \underset{0.000}{1} \end{pmatrix}$$

$$\text{et} \quad \boldsymbol{\rho}_{\widehat{\mathrm{SCOV}}^{\mathrm{SR}}} = \begin{pmatrix} \underset{0.000}{1} & \underset{0.216}{0.572} & \underset{0.229}{0.894} \\ \underset{0.216}{0.572} & \underset{0.000}{1} & \underset{0.179}{0.691} \\ \underset{0.229}{0.894} & \underset{0.179}{0.691} & \underset{0.000}{1} \end{pmatrix}.$$

Ici $\boldsymbol{\rho}_{\widehat{r}}$ est un estimateur obtenu à partir des estimateurs du maximum de vraisemblance de la covariance et de l'écart-type. Les résultats de $\boldsymbol{\rho}_{\widehat{\mathrm{SCOV}}^{\mathrm{SR}}}$ confirment l'intuition. Cet estimateur, qui dispose de constantes de calage c_1 et c_2, fournit un estimateur robuste de la matrice de corrélation d'un vecteur gaussien dont la qualité paraît excellente sur les cas traités.

Bibliographie

[1] Garel, B. et Kodia, B. (2009) Signed symmetric covariation coefficient for alpha-stable dependence modeling. *C. R. Acad. Sci. Paris. Ser. I 347-352*

[2] Miller, G. (1978) Properties of certain symmetric stable distributions. *Journal of Multivariate Analysis*, 8 (3) 346-360.

[3] Samorodnitsky, G et Taqqu M. S. (1994) Stable non-Gaussian random processes. *Stochastic Modeling. Chapman & Hall, New York-London.*

[4] Uchaikin, V. V. and Zolotarev, V. M. (1999) Chance and Stability. Stable Distributions and their Applications. *de Gruyter, Berlin-New York.*