Généralisation de la méthode des moments pondérés
Application à la loi des excès

Jean Diebolt[(1)], Armelle Guillou[(2)] et **Imen Rached**[(1)]

[(1)] *Université de Marne-la-Vallée*
Equipe d'Analyse et de Mathématiques Appliquées
5, boulevard Descartes, Batiment Copernic
Champs-sur-Marne
77454 Marne-la-Vallée Cedex 2

[(2)] *Université Paris VI*
Laboratoire de Statistique Théorique et Appliquée
Boîte 158
175 rue du Chevaleret
75013 Paris

Résumé
La méthode P.O.T. (Peaks-Over-Threshold) consiste à approximer la loi des excès au-delà d'un seuil par une loi de Pareto généralisée (GPD). Nous proposons une généralisation de la méthode des moments pondérés que nous utilisons pour estimer les paramètres de la loi GPD approximante. Nous étudions ensuite la loi limite des estimateurs proposés.

Abstract
The P.O.T. (Peaks-Over-Threshold) approach consists of approximating the distribution of excesses over thresholds by a generalized Pareto distribution (GPD). We propose a generalization of the probability-weighted moments method and we apply it in order to estimate the parameters of the approximating GPD. We study the limiting distribution of the estimators.
Mots clés
Loi de Pareto généralisée, loi des excès, moments pondérés généralisés.
Keywords
Generalized Pareto distribution, distribution of excesses, generalized probability-weighted moments.

1 Introduction

L'étude des événements rares tels que les phénomènes métérologiques extrêmes (tornades, tempêtes de verglas, inondations...) ou les crashes boursiers en finance, se ramènent à l'estimation des queues de distribution et des quantiles extrêmes. La méthode des excès, initialement introduite par Pickands (1975), est une approche permettant d'étudier ces phénomènes. Cette méthode s'appuie sur l'approximation de la loi des excès au-delà d'un seuil u (assez élevé), c'est-à-dire la loi de $X-u$ conditionnellement à $X>u$, par une loi de Pareto généralisée (Generalized Pareto Distribution ou GPD).

On définit les lois GPD par leur fonction de répartition :

$$G_{\gamma,\sigma}(x) = \begin{cases} 1 - (1 + \frac{\gamma}{\sigma}x)^{-1/\gamma} & \text{si } \gamma \neq 0, \sigma > 0 \\ 1 - \exp(-x/\sigma) & \text{si } \gamma = 0, \sigma > 0, \end{cases}$$

où $x \in [0, \infty[$ si $\gamma \geq 0$ et $[0, -\sigma/\gamma[$ si $\gamma < 0$.
Dans ce travail, on procède à l'estimation des paramètres d'échelle σ et de forme γ de la loi GPD approximante. Soit, pour chaque n, un seuil ν_n non aléatoire et $Y_1, ..., Y_{N_n}$ un échantillon d'excès au-dessus de ν_n obtenu à partir d'un échantillon $X_1, ..., X_n$ de loi F (défini par $Y_j = X_{i_j} - \nu_n > 0$, où les i_j sont les i, $1 \leq i \leq n$, tels que $X_{i_j} > \nu_n$), N_n désignant le nombre aléatoire d'excès au-dessus de ν_n.
La méthode P.O.T. consiste à estimer la fonction de survie $\bar{F}_{\nu_n}$ de la loi des excès au-delà d'un seuil ν_n par la fonction de survie d'une loi GPD approximante de paramètres γ_n et σ_n, paramètres estimés à partir des excès observés au-dessus de ν_n. Ces estimateurs se calculent à partir de la fonction de répartition empirique des excès, notée $\mathbb{A}_{n,\nu_n}$, définie par

$$\mathbb{A}_{n,\nu_n}(x) = \frac{1}{N_n} \sum_{j=1}^{N_n} \mathbf{I}_{\{\mathbf{Y_j} \leq \mathbf{x}\}}.$$

Plusieurs techniques ont été proposées pour approximer la loi des excès par une loi GPD. Hosking et Wallis (1987) ont montré par des essais numériques et des simulations que, pour un échantillon de taille inférieure à 500 environ, les méthodes des moments et des moments pondérés sont plus efficaces que les estimateurs du maximum de vraisemblance mais le problème essentiel de leur approche est le domaine de validité : $\gamma < 1/2$. Nous proposons une généralisation de la méthode des moments pondérés afin d'élargir le domaine à $\gamma < 3/2$ et nous appliquons cette nouvelle méthode pour estimer les paramètres γ_n et σ_n de la loi de Pareto approximante. Nous étudions ensuite le comportement asymptotique des estimateurs proposés.

2 Généralisation de la méthode des moments pondérés

Nous proposons à présent une généralisation de la méthode des moments pondérés utilisée par Hosking et Wallis (1987). Il s'agit, dans un premier temps, de définir des moments pondérés généralisés (GPWM) μ_{ω_1} et μ_{ω_2} dans le cas d'une loi GPD.

Définition 1 *Soit X une v.a. de fonction de répartition $G_{\gamma,\sigma}$ et $\omega : [0,1] \longrightarrow \mathbb{R}$ une fonction continue nulle en 0 et dérivable à droite en 0. On définit le moment pondéré généralisé μ_ω de X, par*

$$\mu_\omega = \mathbb{E}\left[X\omega(1 - G_{\gamma,\sigma}(X))\right], \qquad \textit{pour} \quad \gamma < 2.$$

Autrement dit, si on pose $W(x)=\int_0^x \omega(t)\,dt$

$$\mu_\omega = \sigma \int_0^1 W(x)x^{-\gamma-1}\,dx =: \sigma\phi_{\omega_1}(\gamma).$$

Le principe de la méthode des moments pondérés généralisés est d'exprimer les paramètres de la loi GPD(γ_n,σ_n) approximante en fonction de deux GPWM μ_{ω_1} et μ_{ω_2} et de déduire ensuite des estimateurs de γ_n et σ_n à partir des GPWM estimateurs. Dans la proposition suivante, nous donnons des conditions suffisantes à l'existence d'un difféomorphisme qui transforme deux GPWM $(\mu_{\omega_1},\mu_{\omega_2})$ en (γ,σ).

Proposition 1 *Soient* μ_{ω_1} *et* μ_{ω_2} *deux moments pondérés généralisés d'une loi GPD(*γ*,*σ*) définis comme précédement. On a,*
(i) Les fonctions ϕ_{ω_1} *et* ϕ_{ω_2} *sont de classe* C^1 *sur* $]-\infty,2[$.
(ii) Si, pour tout $\gamma<2$*, la fonction* $\rho'(\gamma)=\left(\dfrac{\phi_{\omega_1}}{\phi_{\omega_2}}\right)'(\gamma)$ *garde un signe constant, alors il existe un* C^1*-difféomorphisme* $T_{(\omega_1,\omega_2)}$ *tel que :*

$$T_{(\omega_1,\omega_2)}(\mu_{\omega_1},\mu_{\omega_2})=(\gamma,\sigma),$$

où γ *et* σ *sont donnés respectivement par :*

$$\gamma=\rho^{-1}\left(\frac{\mu_{\omega_1}}{\mu_{\omega_2}}\right) \qquad \text{et} \qquad \sigma=\frac{\mu_{\omega_2}}{\phi_{\omega_2}\circ\rho^{-1}\left(\frac{\mu_{\omega_1}}{\mu_{\omega_2}}\right)}.$$

Notons $A_{(\omega_1,\omega_2)}=DT_{(\omega_1,\omega_2)}(\mu_{\omega_1},\mu_{\omega_2})$.

3 Loi asymptotique des estimateurs des paramètres de la loi GPD approximante

Notons μ_ω^1 le moment pondéré généralisé d'une loi GPD$(\gamma,1)$. Nous proposons d'estimer μ_ω^1 par $\widehat{\mu}_{\omega,n}=\int_0^\infty W(1-\mathbb{A}_{n,\nu_n}(x))\,dx$. Nous supposons que F est deux fois dérivable et admet une réciproque F^{-1}. Soient les deux fonctions V et M définies par

$$V(t)=\bar{F}^{-1}(e^{-t}) \qquad \text{et} \qquad M(t)=\frac{V''(\ln t)}{V'(\ln t)}-\gamma$$

et $a_n=M(e^{V^{-1}(\nu_n)})$. Nous nous plaçons sous les conditions suivantes

$$\lim_{t\to+\infty} M(t)=0, \tag{1}$$

M est de signe constant à l'infini et il existe $\rho\le 0$ tel que M est à variation régulière d'ordre ρ. (2)

Théorème 1 *Supposons que ω_1 et ω_2 sont deux fonctions de classes C^1 et nulles en 0. Sous des hypothèses (1) et (2), avec $-1 < \gamma < 3/2$,*
on a conditionnellement à $N_n = k_n$, où k_n est une suite tendant vers $+\infty$ et telle que $\sqrt{k_n} a_n \to \lambda,\ \lambda \in \mathbb{R}$

$$\sqrt{k_n}\left(\frac{\widehat{\mu}_{\omega_1,n}}{\sigma_n} - \mu^1_{\omega_1}, \frac{\widehat{\mu}_{\omega_2,n}}{\sigma_n} - \mu^1_{\omega_2}\right) \xrightarrow{d} \mathcal{N}\left(\lambda C,\, \Gamma\right)\,,$$

où

$$C = \left(\int_0^\infty \omega_1(e^{-u})e^{-u}\int_0^u e^{\gamma t}\int_0^t e^{\rho z}\,dz\,dt\,du, \int_0^\infty \omega_2(e^{-u})e^{-u}\int_0^u e^{\gamma t}\int_0^t e^{\rho z}\,dz\,dt\,du\right)$$

et Γ est la matrice de variance-covariance de (X_1, X_2), avec

$$X_1 = \int_0^1 t^{-\gamma-1}\omega_1(t)\mathbb{B}(t)\,dt \qquad et \qquad X_2 = \int_0^1 t^{-\gamma-1}\omega_2(t)\mathbb{B}(t)\,dt,$$

où $\mathbb{B}$ est un pont brownien sur $[0,1]$.

Corollaire 1 *Sous les mêmes hypothèses que celles du Théorème 1 et si de plus*

$$n\left(1 - F(\nu_n)\right)a_n \to \infty \quad quand \quad n \to +\infty,$$

alors conditionnellement à $N_n = k_n$, on a

$$\sqrt{k_n}\left(\widehat{\gamma}_n - \gamma, \frac{\widehat{\sigma}_n}{\sigma_n} - 1\right) \xrightarrow{d} \mathcal{N}(\lambda C',\, \Sigma)\,,$$

où $\Sigma = A_{(\omega_1,\omega_2)}\Gamma A^t_{(\omega_1,\omega_2)}$ et $C' = A_{(\omega_1,\omega_2)}C$.

4 Exemples

La loi limite et le biais asymptotique dépendent tous deux des fonctions ω_1 et ω_2. Nous choisissons un exemple de telles fonctions qui réduit le biais et donne de bons estimateurs en terme d'erreur en moyenne quadratique (MSE).

Bibliographie

[1] Diebolt J., Guillou A., Worms R.(2003), Asymptotic behaviour of the probability-weighted moments and penultimate approximation, *ESAIM Probability and Statistics*, 7, 219-238.
[2] Hosking J.R.M (1986), The theory of probability weighted moments, *Research Report RC12210,IBM Thomas J. Watson Reasearch Center*, Yorktown Heights, NY.
[3] Pickands J. (1975), Statistical inference using extreme order statistics, *Ann. Statist*, 3, 119-131.