Université Paris Dauphine
Département MIDO
L3 - Statistique Mathématique

Année 2012-2013

Projet de statistique mathématique. Deuxième partie

Le projet compte pour 30% de la note finale de Statistique Mathématique.

Déroulement du projet : Les étudiants travaillent par groupes de 2 ou 3 étudiants du même groupe de TD. Un rapport écrit en version papier sur cette première partie doit être rendu lors de la semaine du 5 avril. Tout retard équivaut à un 0 pour cette première partie qui compte pour 7/20 de la note finale du projet.

Les étudiants doivent utiliser le logiciel R [1]. Les principales fonctions utilisées doivent apparaître en annexe dans le rapport.

Dans la première partie du projet, chaque binôme ou trinôme a traité une série de données réelles provenant des volumes à l'Ask dans un carnet d'ordre. Les données ainsi obtenues $x_1, \ldots, x_n$ sont les réalisations d'observations $X_1, \ldots, X_n$ supposées iid.

1 Objectif de la seconde partie

L'objectif de la seconde partie du projet est de modéliser les données réelles et d'estimer numériquement les paramètres inconnus dans le modèle GEV par un Z- et un M-estimateur calculer théoriquement et approcher numériquement.

2 Modélisation des données et estimation des paramètres

On rappelle la distribution des valeurs extrêmes généralisées (GEV) (generalized extreme values en anglais) de paramètres $\mu \in \mathbb{R}$, $\sigma > 0$ et $\xi \in \mathbb{R}$ de fonction de répartition

$$G_{(\mu,\sigma,\xi)}(x) = \exp\Big(- \Big(1 + \frac{x-\mu}{\sigma}\xi\Big)_+^{-1/\xi}\Big), \qquad x \in \mathbb{R} \text{ et } a_+ = \max\{a, 0\}.$$

$\boxed{\textbf{Question 1}}$: Calculer (si possible) les 3 premiers moments $m_1 = \mathbb{E}_\theta(X)$, $m_2 = \mathbb{E}_\theta(X^2)$ et $m_3 = \mathbb{E}_\theta(X^3)$ (on pourra faire des restrictions sur l'ensemble des paramètres à estimer). La méthode des moments est-elle applicable dans le modèle GEV ? Calculer les moments pondérés, c.f.

1. voir ci-besoin introduction à R : http ://www.ceremade.dauphine.fr/%7Exian/Noise/R.pdf ou le site officiel http ://www.r-project.org/

Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments
J. R. M. Hosking, J. R. Wallis and E. F. Wood, Technometrics Vol. 27, No. 3 (Aug., 1985), pp. 251-261
afin d'appliquer la méthode des moments généralisés.

Question 2 : Poser un système qui lie les paramètres en fonction des moments (généralisés ou non). Le résoudre analytiquement conditionnellement à la valeur de l'estimateur $\hat{\xi}_n^{MM}$.

Question 3 : Approcher numériquement le Z-estimateur avec la fonction "nleqslv" sous R (qu'on charge en tapant library(nleqslv)) en explicitant ce que fait cette fonction R.

Question 4 : Calculer la vraisemblance des observations, puis trouver analytiquement l'estimateur du maximum de vraisemblance dans le cas $\xi = 0$. Le modèle est il régulier ?

Question 5 : Approcher numériquement l'estimateur du maximum de vraisemblance en utilisant sous R la fonction "constrOptim" ou la fonction "stableFit" du package "fBasics" (que l'on installe en tapant install.packages('fBasics') et que l'on charge en tapant library(fBasics)) en explicitant ce que font ces fonctions R.

Question 6 Comparer les 4 estimateurs obtenus (Z- et M-estimateurs, valeurs analytique et numérique) en traçant les 4 densités correspondant aux valeurs des paramètres obtenus et l'histogramme des données sur le même graphique. Quel semble être le meilleur estimateur ?