

Projet R - Cours de statistique

M2 TIDE 2018-2019

Première partie : Loi de Pareto

La v.a. X suit la loi de *Pareto* de paramètre 1 et $p > 0$ si elle a pour densité la fonction

$$f(x) = \frac{p}{x^{p+1}}, \quad x \geq 1.$$

Exercice 1. Loi des grands nombres

Dans cet exercice on illustre la loi des grands nombres en utilisant la simulation de loi de Pareto.

- Générer une réalisation x_1, \dots, x_n de la suite de v.a. i.i.d. X_1, \dots, X_n avec $n = 50000$ où X_1 suit une loi de Pareto de paramètre $p = 3$.
- Montrer que $\mathbb{E}(X_1) = \frac{p}{p-1}$ pour $p > 1$ et $Var(X_1) = \frac{p}{(p-1)^2(p-2)}$ pour $p > 2$.
- Calculer $\bar{x}_m = \frac{1}{m} \sum_{i=1}^m x_i$ pour $m = 1, \dots, n$.
- Tracer \bar{x}_m en fonction de m . Conclusion ?
- Procéder de même avec $p = 1$. Conclusion ?

Exercice 2. Théorème central limite

Dans cet exercice on illustre le théorème central limite en utilisant la simulation de loi de Pareto.

- Générer R réalisations $x_{1,r}, \dots, x_{n,r}$ de la suite de v.a. i.i.d. X_1, \dots, X_n où X_1 suit une loi de Pareto de paramètre $p = 3$, ($r = 1, \dots, R$). On a $\mathbb{E}[X_1] = \frac{p}{p-1}$ et $Var[X_1] = \frac{p}{(p-1)^2(p-2)}$. On prendra $n = 1000$ et $R = 5000$.
- Calculer $\bar{x}_{n,r} = \frac{1}{n} \sum_{i=1}^n x_{i,r}$ pour chaque r .
- Tracer (1) un histogramme de $\sqrt{n} \frac{\bar{x}_{n,r} - \mathbb{E}[X_1]}{\sqrt{Var[X_1]}}$ sur le même graphique que la fonction de densité de la loi $\mathcal{N}(0, 1)$ et (2) la fonction de répartition empirique de cette suite sur le même graphique que la fonction de répartition de la loi $\mathcal{N}(0, 1)$. Essayer avec d'autres valeurs de n . Conclusion ?
- Refaire les questions a) et b) avec $p = 1$ et tracer les histogrammes de $\bar{x}_{n,r}$ pour ces deux valeurs de p dans des graphes côte à côte se partageant la même fenêtre graphique. Conclusion ?

Deuxième partie : ... des plans sur les comètes ...

Les jeux des données dont nous disposons représentent des perturbations planétaires qui sont appliquées aux comètes du nuage d'Oort. Il est tout à fait possible de regarder ces données comme des objets dans un espace $K \times M$. K est donnée par le rectangle $[0, 15] \times [-1, 1]$. $(q, \cos i)$, un point dans K représente la distance périhélique et l'angle d'inclinaison de la trajectoire elliptique d'une comète quelconque. M représente l'espace des per-

turbations Δz , qui prennent des valeurs dans \mathbb{R} .

En très simple, la dynamique des comètes est étudiée de la manière suivante. Premièrement, les perturbations planétaires sont calculées à l'aide d'un intégrateur numérique. Ces sont nos données. Ensuite, la perturbation Δz qui se trouve en $(q, \cos i)$ est appliquée au demi-grand axe de la trajectoire de la comète ayant les mêmes périhélie et angle d'inclinaison que la perturbation considérée. Il en résulte que plus on a des perturbations calculées, plus précis sera l'étude de la dynamique des comètes.

Cependant le calcul des perturbations est vraiment très, très coûteux ... D'où l'intérêt des statistiques ...

Téléchargez les deux jeux des données disponibles à partir de la page suivante :

<http://samm.univ-paris1.fr/Shuyan-LIU-Enseignement>

L'objectif de cette partie du projet est de caractériser statistiquement ces perturbations ...

Exercice 3.

- a) Faites une analyse exploratoire de ces deux jeux de données. Réfléchissez au choix de vos statistiques. Argumentez ce choix.
- b) Est-ce que vous observez des structures particulières ?
- c) Avez-vous des indications concernant la modélisation par une loi de probabilité de ces perturbations ?
- d) Est-ce que vous pouvez proposer un test qui rejette certaines lois concernant ces perturbations. Exemple : loi normale, exponentielle bilatérale, Cauchy ou autre ? Si oui, il faudrait expliquer comment vous estimez les paramètres des lois testées ...

A l'attention de l'étudiant : Envoyez moi **avant le 7 mai 2019 minuit**, un compte-rendu contenant les solutions aux trois exercices de ce projet. Après cette date tout compte-rendu sera irrecevable. Vous pouvez travailler en binôme. Les solutions doivent être soigneusement rédigées. Veillez à inclure les graphiques demandés ainsi que les codes (ou fragments de code) qui vous semblent les plus appropriés et qui argumentent au mieux vos réponses.

Vous pouvez me contacter directement dans mon bureau (C.20.07) ou bien par email (Shuyan.Liu@univ-paris1.fr).