Génie Mathématique **Année 2021-2022**

Projet GM3 – Semestre 2

Deuxième vague

Probabilités et Statistique

• **Titre : Simulation et Estimation dans le Cadre de la Loi de Pareto**

• **Consigne :** Dès l'affectation du sujet, prendre un premier rendez-vous avec M. Bruno PORTIER pour la présentation du projet ; ensuite, prendre rendez-vous par mail avec M. Bobbia.

• **Contact :**

Bruno Portier – LMI

Tél. : 02 32 95 65 39

Mail : bruno.portier@insa-rouen.fr

Page Web : `http://lmi2.insa-rouen.fr/~bportier/`

Michel Bobbia – Atmo Normandie

Mail : michel.bobbia@atmonormandie.fr

• **Description :** Ce projet porte sur l'étude de la loi de Pareto, la simulation d'un échantillon de valeurs indépendantes d'une variable aléatoire de loi de Pareto et l'étude d'un estimateur du paramètre de cette loi.

1 Etude Mathématique

1.1 Etude probabiliste

1. Soient α et θ deux réels strictement positifs. Déterminer le réel C pour que la fonction

$$x \mapsto \begin{cases} \dfrac{C}{x^{\alpha+1}} & \text{si } x \geq \theta \\ 0 & \text{sinon.} \end{cases}$$

   soit une densité de probabilité.

Dans toute la suite, on notera cette densité f.

Définition 1 *On dit que la variable aléatoire X suit une loi de Pareto de paramètres α et θ si et seulement si sa densité est f. On notera alors $X \sim \mathcal{P}(\alpha, \theta)$.*

Dans ce qui suit, X désignera une variable aléatoire de loi $\mathcal{P}(\alpha, \theta)$.

2. Dans quelle contexte cette loi est-elle utilisée ? Donner des exemples de modélisation avec une telle loi.
3. A quelle condition la variable X possède-t-elle une espérance ? une variance ? On se placera sous cette dernière condition dans la suite.
4. Calculer l'espérance et la variance de X.
5. Calculer la fonction de répartition F de X.
6. Déterminer la fonction réciproque de F.

1.2 Simulation d'une valeur d'une variable de loi de Laplace

L'objet de ce paragraphe est de proposer une méthode de construction d'une valeur d'une variable aléatoire de loi $\mathcal{P}(\alpha, \theta)$.

7. Expliquer de manière formelle comment simuler une valeur d'une variable aléatoire de loi $\mathcal{P}(\alpha, \theta)$.

1.3 Etude statistique

On considère n données réelles $x_1, x_2, \ldots, x_n$. On fait l'hypothèse que ces données sont les réalisations de n variables aléatoires X_1, X_2, $\cdots$, X_n , indépendantes et de même loi $\mathcal{P}(\alpha, \theta = 1)$. On suppose que le paramètre α est inconnu et on s'intéresse au problème de son estimation par la méthode du maximum de vraisemblance.

8. Ecrire la log-vraisemblance de l'échantillon et trouver l'estimation du maximum de vraisemblance de α. En déduire alors l'estimateur du maximum de vraisemblance du paramètre α. On le notera $\widehat{\alpha}_n$.
9. Calculer $\mathbb{E}\left[\log(X_1)\right]$ et $\mathbb{V}\mathrm{ar}(\log(X_1))$.
10. Proposer un estimateur sans biais et convergent du paramètre $\beta = \dfrac{1}{\alpha}$. On notera cet estimateur B_n et b_n sa réalisation.
11. Enoncer le TLC que satisfait cet estimateur.
12. Proposer un estimateur du paramètre α. On le notera A_n.
13. En déduire la convergence de A_n vers α et énoncer le TLC que satisfait A_n.

2 Etude par simulation sous *R*

Dans un premier temps, l'objectif est de simuler n valeurs indépendantes d'une variable aléatoire de loi $\mathcal{P}(\alpha, \theta = 1)$ et d'apprécier la qualité de simulation de la loi. On étudiera ensuite les performances de l'estimateur du maximum de vraisemblance du paramètre α.

Le travail à réaliser sous *R* est le suivant : écrire un programme permettant de

1. fixer la taille n de l'échantillon à simuler ;

2. fixer la valeur du paramètre α de la loi de Pareto ;
3. simuler un échantillon de n valeurs indépendantes, notées $x_1, x_2, \ldots, x_n$, d'une variable aléatoire de loi $\mathcal{P}(\alpha, \theta = 1)$;
4. construire l'histogramme en fréquences (20 classes) de ces valeurs et superposer à cet histogramme, la courbe représentative de la densité de la loi $\mathcal{P}(\alpha, \theta = 1)$;
5. calculer les valeurs de la suite $(a_k)_{1 \leq k \leq n}$, a_k étant l'estimation du paramètre α obtenue à partir des données $(x_1, \cdots, x_k)$;
6. visualiser à l'aide de deux graphiques le comportement de l'estimateur, l'un, illustrant son comportement transitoire (les 30 ou 40 premières valeurs de la suite (a_k)) et l'autre, illustrant sa convergence vers la valeur de α choisie.

On effectuera ce travail pour différentes valeurs de n et de α. On pourra aussi illustrer l'aspect fluctuation d'échantillonnage en superposant plusieurs trajectoires de l'estimateur.

On commentera bien évidemment les résultats obtenus.

Dans un second temps, le travail consiste à étudier par simulations le comportement à distance finie de l'estimateur du paramètre α de la loi de Pareto.

Pour cela, on simulera 400 échantillons de taille 1000, tous issus d'une Loi de Pareto de paramètre α bien choisi. On stockera ces valeurs dans la matrice `data` de format 1000x400.

Pour chaque colonne de données, on estimera le paramètre α à partir des données $(x_1, \cdots, x_n)$ pour les tailles d'échantillon $n = 50, 100, 200, 500, 1000$.

Pour analyser le comportement de l'estimateur, on tracera les boîtes à moustaches des 400 estimations pour chaque taille d'échantillon retenue. On superposera sur chaque boîte à moustaches la moyenne des 400 estimations. Analysez et commentez les résultats obtenus. Que sommes-nous amener à illustrer avec un tel graphique ?

- **Logiciels :** *R*

- **Rédaction du rapport :** Le document qui sera remis à la fin du projet devra contenir :

  1. une table des matières ;
  2. une introduction précisant notamment l'objet du problème posé, les résultats obtenus et le plan du document ;
  3. une première partie **manuscrite** (rédigée sur feuilles blanches) présentant le travail mathématique réalisé ;
  4. une deuxième partie, présentant et commentant les résultats et les graphiques obtenus lors du travail de simulations ;
  5. une conclusion ;
  6. une bibliographie ;
  7. une annexe contenant le listing du programme R.