Projet R - Cours de statistique

M1 MIAGE 2018-2019

Exercice 1. 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

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

a) Créer une fonctions `dpareto(x, p)` en R qui calcule $f(x)$ avec le paramètre `p` donné.

b) En utilisant la fonction `dpareto(x, p)` tracer dans le même graphe les courbes de la densité de Pareto avec le paramètre p de valeurs différentes. Différencier les courbes par couleur ou par type de ligne. Ajouter une légende indiquant les valeurs différentes de p.

c) La fonction de répartition de loi de Pareto est la suivante

$$F(x) = \begin{cases} 0, & x < 1, \\ 1 - \frac{1}{x^p}, & x \geq 1. \end{cases}$$

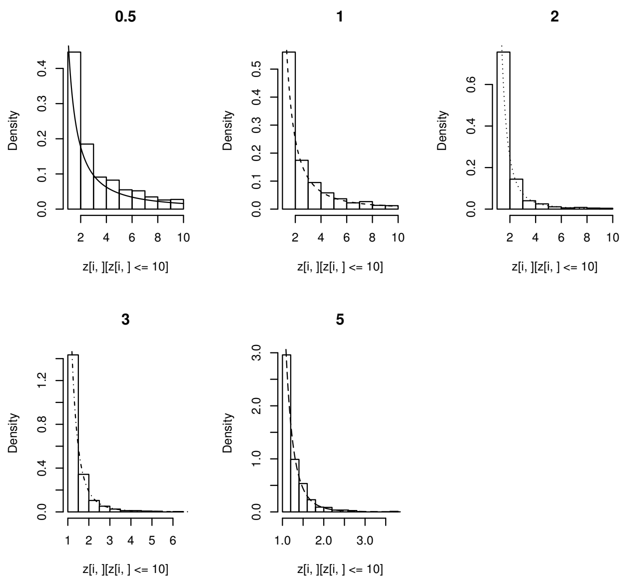
Créer une fonctions `ppareto(x, p)` en R qui calcule $F(x)$ avec le paramètre `p` donné.

d) En utilisant la fonction `ppareto(x, p)` tracer dans le même graphe les courbes $(\log x, \log(1 - F(x)))$ avec différentes valeurs de p que vous avez pris pour la question b). Différencier les courbes par couleur ou par type de ligne. Ajouter une légende indiquant les valeurs différentes de p. Pourquoi les courbes obtenues sont des lignes droites? Pouvez vous commenter la figure obtenue en reliant la figure obtenue dans la question b) ?

e) Que fait le programme ci dessous ? Quelle est la relation entre la fonction `qpareto(y, p)` définie ici et `ppareto(x, p)` définie dans la question c) ? Relancer ce programme en modifiant la valeur de p et tracer pour chaque p les images dans des graphes différents se partageant la même fenêtre graphique comme par exemple la figure dans la page suivante.

```
qpareto <- function(y, p){
    if((y > 0)&&(y < 1)&&(p > 0)){(1/(1-y))^(1/p)}
    else{stop("y<=0 ou y>=1 ou p<=0")}
}
rpareto <- function(n, p){qpareto(runif(n), p)}

p = 1/2; n = 1000; x = seq(1, 10, length.out = n)
y = dpareto(x, p); z = rpareto(n, p)
hist(z[z <= 10], freq = F, main = as.character(p));
lines(x, y, lty = 1, col = "red", lwd = 2)
```



Exercice 2. Loi des grands nombres
Dans cet exercice on illustre la loi des grands nombres en utilisant la simulation de loi de Pareto.
a) Générer une réalisation $x_1, \ldots, x_n$ de la suite de v.a. i.i.d. $X_1, \ldots, X_n$ avec $n = 50000$ où X_1 suit une loi de Pareto de paramètre $p = 3$.
b) 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 $p > 2$.
c) Calculer $\bar{x}_m = \frac{1}{m} \sum_{i=1}^{m} x_i$ pour $m = 1, \ldots, n$.
d) Tracer $\bar{x}_m$ en fonction de m. Conclusion ?
e) Procéder de même avec $p = 1$. Conclusion ?

Exercice 3. Théorème central limite
Dans cet exercice on illustre le théorème central limite en utilisant la simulation de loi de Pareto.
a) Générer R réalisations $x_{1,r}, \ldots, x_{n,r}$ de la suite de v.a. i.i.d. $X_1, \ldots, X_n$ où X_1 suit une loi de Pareto de paramètre $p = 3$, $(r = 1, \ldots, R)$. 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 $R = 5000$.
b) Calculer $\bar{x}_{n,r} = \frac{1}{n} \sum_{i=1}^{n} x_{i,r}$ pour chaque r.
c) 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 ?
d) 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 ?

A l'attention de l'étudiant : Envoyez moi **avant le 12 avril 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).