DM15

À RENDRE LE MARDI 25 MARS

Estimation

Dans cet exercice, toutes les variables aléatoires sont supposées définies sur un même espace probabilisé.

Partie A : Loi de Pareto

Soient a et b deux réels strictement positifs. On définit la fonction f sur $\mathbb{R}$ par :

$$f : x \mapsto \begin{cases} 0 & \text{si } x < b \\ a\,\dfrac{b^a}{x^{a+1}} & \text{si } x \geqslant b \end{cases}$$

1. Montrer que f est une densité de probabilité.

Une variable aléatoire suit la loi de Pareto de paramètres a et b lorsqu'elle admet pour densité la fonction f.
Dans toute la suite de l'exercice, on considère une variable aléatoire X suivant la loi de Pareto de paramètres a et b.

2. Déterminer la fonction de répartition de X.

3. (a) Soit U une variable aléatoire suivant la loi uniforme sur $[0,1[$.
Montrer que la variable aléatoire $bU^{-\frac{1}{a}}$ suit la loi de Pareto de paramètres a et b.

   (b) En déduire une fonction `Python` d'en-tête def pareto(a,b) qui prend en arguments deux réels a et b strictement positifs et qui renvoie une simulation de la variable aléatoire X.

   (c) On considère la fonction `Python` ci-dessous.
Que contient la liste `L` renvoyée par la fonction `mystere` ?

```python
import numpy as np

def mystere(a,b):
    L= []
    for p in range(2,7):
        S=0
        for k in range(10**p) :
            S=S+pareto(a,b)
        L.append(S/10**p)
    return L
```

   (d) On exécute la fonction précédente avec différentes valeurs de a et de b.
Comment interpréter les résultats obtenus ?

```
--> mystere(2,1)
 ans =
     1.9306917   1.9411352   1.9840089   1.9977684   2.0012415
--> mystere(3,2)
 ans =
     3.1050951   3.0142956   2.9849407   2.9931656   2.9991517
--> mystere(1,4)
 ans =
     21.053151   249.58609   51.230522   137.64549   40.243918
```

4. (a) Montrer que X admet une espérance si et seulement si $a > 1$ et que, dans ce cas :

$$\mathrm{E}(X) = \frac{a\,b}{a-1}$$

   (b) Montrer que X admet une variance si et seulement si $a > 2$ et que, dans ce cas :

$$V(X) = \frac{a\,b^2}{(a-1)^2\,(a-2)}$$

Partie B : Estimation du paramètre b

On suppose **dans cette partie uniquement** que $a = 3$ et on chercher Ã déterminer un estimateur performant de b.
Ainsi, la variable aléatoire X admet pour densité la fonction f définie par :

$$f : x \mapsto \begin{cases} 0 & \text{si } x < b \\ \dfrac{3\,b^3}{x^4} & \text{si } x \geqslant b \end{cases}$$

Soient $n \in \mathbb{N}^*$ et $X_1, \ldots, X_n$ des variables aléatoires indépendantes, toutes de même loi que X.
On définit :

$$Y_n = \min(X_1, \ldots, X_n) \qquad \text{et} \qquad Z_n = \frac{X_1 + \cdots + X_n}{n}$$

On admet que Y_n et Z_n sont encore des variables aléatoires définies sur le même espace probabilisé.

5. (a) Calculer, pour tout x de $[b, +\infty[$, $P(Y_n > x)$.
   (b) En déduire que Y_n suit une loi de Pareto dont on précisera les paramètres.
   (c) Montrer que $Y_n' = \dfrac{3n-1}{3n}\,Y_n$ est un estimateur sans biais de b.
   Calculer le risque quadratique de cet estimateur.

6. (a) Déterminer l'espérance et la variance de Z_n.
   (b) En déduire un estimateur noté Z_n' sans biais de b de la forme $\alpha\, Z_n$ où α est un réel à préciser.
   Calculer le risque quadratique de cet estimateur.

7. Entre Y_n' et Z_n', quel estimateur choisir ? Justifier.

Partie C : Estimation du paramètre a

On suppose **dans cette partie uniquement** que $b = 1$ et on cherche Ã construire un intervalle de confiance pour a.
Ainsi, la variable aléatoire X admet pour densité la fonction f définie par :

$$f : x \mapsto \begin{cases} 0 & \text{si } x < 1 \\ \dfrac{a}{x^{a+1}} & \text{si } x \geqslant 1 \end{cases}$$

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes, toutes de même loi que X.

8. Soit $n \in \mathbb{N}^*$. On pose : $W_n = \ln(X_n)$.
Montrer que la variable aléatoire W_n suit une loi exponentielle dont on précisera le paramètre.
En déduire l'espérance et la variance de W_n.

9. On définit, pour tout n de $\mathbb{N}^*$:

$$M_n \;=\; \frac{\ln(X_1) + \cdots + \ln(X_n)}{n} \qquad \text{et} \qquad T_n = \sqrt{n}\,(a\,M_n - 1)$$

(a) Justifier que la suite de variables aléatoires $(T_n)_{n\in\mathbb{N}^*}$ converge en loi vers une variable aléatoire suivant la loi normale centrée réduite.

(b) En déduire que l'intervalle $\left[\dfrac{\sqrt{n}-2}{\sqrt{n}\,M_n}\ ;\ \dfrac{\sqrt{n}+2}{\sqrt{n}\,M_n}\right]$ est un intervalle de confiance asymptotique pour a au niveau de confiance 95%.
On admettra que $\Phi(2) \geqslant 0,975$*, où* Φ *désigne la fonction de répartition d'une variable aléatoire suivant la loi normale centrée réduite.*