TD10

ESTIMATION, CONVERGENCE EN LOI.

Exercice 1.- EDHEC 2018 Exercice 3.

On admet que toutes les variables aléatoires considérées dans cet exercice sont définies sur le même espace probabilisé $(\Omega, \mathcal{A}, P)$ que l'on ne cherchera pas à déterminer.

Soit a un réel strictement positif et f la fonction définie par

$$f(x) = \begin{cases} \frac{x}{a} e^{-x^2/2a} & \text{si } x \geqslant 0 \\ 0 & \text{si } x \leqslant 0 \end{cases}$$

1. Montrer que la fonction f est une densité.

Dans toute la suite de l'exercice, on considère une variable aléatoire X de densité f.

2. Déterminer la fonction de répartition F_X de X.

3. On considère la variable aléatoire Y définie par $Y = \frac{X^2}{2a}$.

- **a.** Montrer que Y suit la loi exponentielle de paramètre 1.
- **b.** Rappeler la méthode du cours qui permet de simuler une variable exponentielle de paramètre a, c'est-à-dire, écrire un script Python demandant la valeur de a à l'utilisateur et permettant de simuler la variable aléatoire X.

4.
- **a.** Vérifier que la fonction g, qui à tout réel x associe $x^2 e^{-x^2/2a}$ est paire.
- **b.** Rappeler l'expression intégrale ainsi que la valeur du moment d'ordre 2 d'une variable aléatoire Z suivant la loi normale de paramètre 0 et a.
- **c.** En déduire que X possède une espérance et la déterminer.

5. Rappeler l'espérance de Y puis montrer que X possède un moment d'ordre 2 et le calculer.

6. En déduire que la variance est donnée par

$$V(X) = \frac{(4-\pi)a}{2}$$

On suppose désormais que la paramètre a est inconnu et on souhaite l'estimer.

7. Soit n un entier naturel supérieur ou égal à 1. On considère un échantillon $(X_1, X_2, \cdots, X_n)$ composé de variables aléatoires indépendantes ayant toutes la même loi que X.

On note S_n la variable aléatoire définie par $S_n = \frac{1}{2n} \sum_{k=1}^{n} X_k^2$.

- **a.** Montrer que S_n est un estimateur sans biais de a.
- **b.** Montrer que X^2 possède une variance et que $V(X^2) = 4a^2$.
- **c.** Déterminer la variance $V_a(S_n)$ de S_n en tant qu'estimateur de a.

8. On suppose que a est inférieur à 1.

a. Écrire l'inégalité de Bienaymé-Tchebychev pour la variable aléatoire S_n et en déduire :

$$\forall \varepsilon > 0, \quad P(|S_n - a| \leqslant \varepsilon) \geqslant 1 - \frac{1}{n\varepsilon^2}.$$

b. Déterminer une valeur de n pour laquelle $\left[S_n - \frac{1}{10}, S_n + \frac{1}{10}\right]$ est un intervalle de confiance pour a avec un niveau de confiance au moins égal à 95%.

Exercice 2.- EML 2020 Exercice 3.

Dans cet exercice, toutes les variables aléatoires sont supposées définies sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$.

Partie A : Loi de Pareto.
Soit a et b deux réels strictement positifs. On définit la fonction f sur $\mathbb{R}$ par

$$f(x) = \begin{cases} 0 & \text{si } x < b \\ a\frac{b^a}{x^{a+1}} & \text{si } x \geqslant b \end{cases}$$

1. Montrer que f est une densité de probabilité.

On dit qu'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 suivant la loi de Pareto de paramètre 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 paramètre 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 ?

```
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.

```
--> print(mystere(2,1))

  [ 1.9306917   1.9411352   1.9840089   1.9977684   2.0012415   ]

--> print(mystere(3,2))

  [ 3.1050951   3.0142956   2.9849407   2.9931656   2.9991517  ]

--> print(mystere(1,4))

  [ 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 :

$$\mathbb{E}(X) = \frac{ab}{a-1}$$

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

$$\mathbb{V}(X) = \frac{ab^2}{(a-1)(a-2)}$$

Partie B : Estimation du paramètre b.

On suppose **dans cette partie uniquement** que $a = 3$ et on cherche à déterminer un estimateur performant de b.

Ainsi la variable aléatoire X admet pour densité la fonction f définie par :

$$f(x) = \begin{cases} 0 & \text{si } x < b \\ \frac{3b^3}{x^4} & \text{si } x \geqslant b \end{cases}$$

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

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

On admet que Y_n et Z_n sont encore des variables aléatoires définies sur $(\Omega, \mathcal{A}, P)$.

5. Calculer, pour tout x de $[b, +\infty[$, $P([Y_n > x])$.
6. En déduire que Y_n suit une loi de Pareto dont on précisera les paramètres.
7. Montrer que $Y_n' = \frac{3n-1}{3n} Y_n$ est un estimateur sans biais de b.
8. **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 αZ_n où α est un réel à préciser.
9. Calculer $V_b(Y_n')$ et $V_b(Z_n')$. Entre Y_n' et Z_n', quel estimateur choisir ?

Partie C : Estimation du paramètre a.

On suppose dans cette partie uniquement que $b = 1$ et on cherche à construire un intervalle de confiance pour a.

Ainsi la variable aléatoire X admet pour densité la fonction f définie par :

$$f(x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{a}{x^{a+1}} & \text{si } x \geqslant 1 \end{cases}$$

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

10. 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.
11. On définit, pour tout $n \in \mathbb{N}^3$:

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

   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 $\left[\frac{\sqrt{n}-2}{\sqrt{n}M_n}, \frac{\sqrt{n}+2}{\sqrt{n}M_n}\right]$ 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.

Exercice 3.- inspiré d'**ORAL ESCP** maths appro.

Soit $n \in \mathbb{N}^*$. On considère un n-échantillon $(X_1, X_2, \cdots, X_n)$ de la loi exponentielle $\mathcal{E}(\lambda)$ de paramètre $\lambda > 0$. On pose $L_n = \min(X_1, \cdots, X_n)$ et $M_n = \max(X_1, \cdots, X_n)$.

1. Déterminer les fonctions de répartition des variables L_n et M_n.

2. Quelle est la loi de la variable $Y_n = n\lambda L_n$?

3. On pose maintenant $Z_n = \lambda M_n - \ln(n)$.

- **a.** Déterminer la fonction de répartition F_n de Z_n.
- **b.** Pour $t \in \mathbb{R}$, déterminer la limite $F(t)$ de $F_n(t)$ lorsque $n \to +\infty$.
- **c.** En déduire soigneusement que Z_n converge en loi vers une variable aléatoire à densité que l'on notera Z. (*On dit que Z suit la loi de Gumbel.*)

La durée de vie d'une ampoule est modélisée par une variable aléatoire $X \hookrightarrow \mathcal{E}(\lambda)$ où $\lambda > 0$ est inconnu. On cherche à estimer la durée de vie moyenne $\mu = E(X) = \frac{1}{\lambda}$ et on dispose d'un échantillon de n ampoules (dont les durées de vie sont supposées indépendantes).

4. Dans cette question, on suppose que la seule information dont on dispose est la durée de vie de l'ampoule qui *a grillé* le plus tôt.

- **a.** À l'aide de la question **2**, proposer un estimateur $\hat{L}_n$ de μ construit à partir de L_n et dont l'espérance vaut μ.
- **b.** Quelle est sa variance ?
- **c.** Montrer que pour tout $\varepsilon > 0$,
$$P\left(\left|\bar{L}_n - \mu\right| > \varepsilon\right) = 1 - e^{-1+\lambda\varepsilon} + e^{-1-\lambda\varepsilon}.$$
- **d.** L'estimateur $\bar{L}_n$ est-il convergent ?
- **e.** Soit $\alpha \in]0,1[$. Montrer que si $Y \hookrightarrow \mathcal{E}(1)$, alors
$$P\left(Y < -\ln\left(1 - \frac{\alpha}{2}\right)\right) = P\left(Y > -\ln\left(\frac{\alpha}{2}\right)\right) = \frac{\alpha}{2}.$$
- **f.** Montrer alors que l'intervalle
$$I_{\alpha,n} = \left[\frac{\bar{L}_n}{-\ln\left(\frac{\alpha}{2}\right)}, \frac{\bar{L}_n}{-\ln\left(1 - \frac{\alpha}{2}\right)}\right]$$
est un intervalle de confiance su seuil $1 - \alpha$ pour μ.

5. Proposer une stratégie utilisant Z_n et Z pour obtenir un autre intervalle de confiance (asymptotique) pour μ.