Devoir maison 13

à rendre pour le mercredi 22 mars

Dans les questions 1 à 4, λ désigne un paramètre réel strictement positif, inconnu. Pour n élément de $\mathbb{N}^*$, on considère un n-échantillon $(X_1, X_2, ..., X_n)$ de variables aléatoires à valeurs strictement positives, indépendantes, de même loi exponentielle de paramètre λ.

On pose pour tout n de $\mathbb{N}^*$: $S_n = \sum\limits_{k=1}^{n} X_k$ et $J_n = \lambda S_n$.

1. Calculer pour tout n de $\mathbb{N}^*$, $E(S_n)$, $V(S_n)$, $E(J_n)$ et $V(J_n)$.
2. On admet qu'une densité f_{J_n} de J_n est donnée par $f_{J_n}(x) = \begin{cases} \dfrac{e^{-x}x^{n-1}}{(n-1)!} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$.
   (a) A l'aide du théorème de transfert, établir pour tout n supérieur ou égal à 3, l'existence de $E\left(\dfrac{1}{J_n}\right)$ et de $E\left(\dfrac{1}{J_n^2}\right)$, et donner leur valeurs respectives.
   (b) On pose pour tout n supérieur ou égal à 3 : $\widehat{\lambda_n} = \dfrac{n}{S_n}$. Justifier que $\widehat{\lambda_n}$ est un estimateur de λ. Est-il sans biais ?
   (c) Soit T_n un estimateur d'un paramètre θ. On définit le risque quadratique de T_n par $E\left((T_n - \theta)^2\right)$.
      i. Montrer que $E\left((T_n - \theta)^2\right) = V(T_n)$ lorsque T_n est un estimateur sans biais.
      ii. Calculer la limite, lorsque n tend vers $+\infty$, du risque quadratique associé à $\widehat{\lambda_n}$ en λ.
3. Dans cette question, on veut déterminer un intervalle de confiance du paramètre λ au risque α. On note Φ la fonction de répartition de la loi normale centrée réduite, et u_α le réel strictement positif tel que $\Phi(u_\alpha) = 1 - \frac{\alpha}{2}$.
   (a) Énoncer le théorème de la limite centrée. En déduire que la variable aléatoire N_n définie par $N_n = \lambda\dfrac{S_n}{\sqrt{n}} - \sqrt{n}$ converge en loi vers la loi normale centrée réduite.
   (b) En déduire que pour n assez grand, on a approximativement : $P([-u_\alpha \leq N_n \leq u_\alpha]) = 1 - \alpha$.
   (c) On note λ_0 la valeur observée de $\widehat{\lambda_n}$ sur le n-échantillon. Montrer que pour n assez grand, l'intervalle $\left[\left(1 - \dfrac{u_\alpha}{\sqrt{n}}\right)\lambda_0, \left(1 + \dfrac{u_\alpha}{\sqrt{n}}\right)\lambda_0\right]$ est un intervalle de confiance de λ au risque α.
4. Avec le n-échantillon $(X_1, X_2, \ldots, X_n)$, on construit un nouvel intervalle de confiance de λ au risque β ($\beta \neq \alpha$), tel que la longueur de cet intervalle soit k ($k > 1$) fois plus petite que celle obtenue avec le risque α. Établir l'égalité $\beta = 2\Phi\left(\dfrac{1}{k}\Phi^{-1}(\alpha/2)\right)$. En déduire que $\beta > \alpha$. Ce dernier résultat était-il prévisible ?

Dans les questions 5 à 7, on suppose que $\lambda = 1$.

5. On pose pour tout n de $\mathbb{N}^*$: $T_n = \max(X_1, X_2, \ldots, X_n)$.
   Pour tout n de $\mathbb{N}^*$, pour tout réel x positif ou nul, on pose : $g_n(x) = \displaystyle\int_0^x F_{T_n}(t)\,\mathrm{d}t$ et $h_n(x) = \displaystyle\int_0^x t f_{T_n}(t)\,\mathrm{d}t$.
   (a) Exprimer $h_n(x)$ en fonction de $F_{T_n}(x)$ et $g_n(x)$.
   (b) Déterminer pour tout réel t, l'expression de $F_{T_n}(t)$ en fonction de t.
   Établir pour tout n supérieur ou égal à 2, la relation : $g_{n-1}(x) - g_n(x) = \dfrac{1}{n}F_{T_n}(x)$.
   (c) En déduire que pour tout n de $\mathbb{N}^*$, pour tout réel x positif ou nul, l'expression de $g_n(x)$ en fonction de $x, F_{T_1}(X), F_{T_2}(x), \ldots, F_{T_n}(x)$.
   (d) Montrer que $F_{T_n}(x) - 1$ est équivalent à $-ne^{-x}$, lorsque x tend vers $+\infty$.
   (e) Déduire des questions c) et d) l'existence de $E(T_n)$ et montrer que $E(T_n) = \sum\limits_{k=1}^{n} \dfrac{1}{k}$.
6. On veut étudier dans cette question la convergence en loi de la suite de variables aléatoires $(G_n)_{n\geq1}$ définie par : pour tout n de $\mathbb{N}^*$, $G_n = T_n - E(T_n)$.
   On pose pour tout n de $\mathbb{N}^*$: $\gamma_n = -\ln n + E(T_n)$.

(a) Montrer que les suites de termes généraux $\gamma_n = \sum_{k=1}^{n} \frac{1}{k} - \ln(n)$ et $\delta_n = \gamma_n - \frac{1}{n}$ convergent vers une même limite γ vérifiant

$$\forall n \in \mathbb{N}^*, \quad \gamma_n - \frac{1}{n} \leq \gamma \leq \gamma_n$$

On pourra utiliser sans démonstration le fait que, pour tout $x \in]-1, +\infty[$, $\ln(1+x) \leq x$.

(b) Montrer que pour tout x réel et n assez grand, on a : $F_{G_n}(x) = \left(1 - \frac{1}{n} e^{-(x+\gamma_n)}\right)^n$.

(c) En déduire que pour tout x réel, on a : $\lim_{n \to +\infty} F_{G_n}(x) = e^{-e^{-(x+\gamma)}}$.

(d) Montrer que la fonction $F_G : \mathbb{R} \longrightarrow \mathbb{R}$ définie par $F_G(x) = e^{-e^{-(x+\gamma)}}$ est la fonction de répartition d'une variable aléatoire G à densité. Conclure. La loi de G s'appelle la loi de Gumbel.

7. (a) Écrire une fonction `GAMMA()` qui renvoie une valeur approchée de γ à 10^{-3} près.

(b) i. Soit X une variable aléatoire à densité de fonction de répartition F_X strictement croissante. Déterminer la loi de la variable aléatoire Y définie par $Y = F_X(X)$.

ii. En déduire une fonction `GUMBEL` qui permet de simuler la variable aléatoire G. On utilisera la fonction `GAMMA()` pour avoir une valeur approchée de la constante γ.

(c) On considère le programme suivant :

```python
from random import random
from numpy import log

def EXPO():
    return -log(random())

def AGRO(n):
    dodo=EXPO()
    for k in range(n-1):
        boulot=EXPO()
        if boulot>dodo:
            dodo=boulot
    return dodo

def VETO(n):
    bio=AGRO(n)
    for k in range(n):
        bio=bio-1/(k+1)
    return bio
```

Expliquer pourquoi peut-on considérer que l'instruction `VETO(100)` simule également la variable aléatoire G.