

Problème

Toutes les variables aléatoires qui interviennent dans ce problème sont supposées définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathcal{P})$. Sous réserve d'existence, on note $E(X)$ et $V(X)$ respectivement l'espérance et la variance d'une variable aléatoire X , et $Cov(X, Y)$ la covariance de deux variables aléatoires X et Y .

Dans les parties I et III, la fonction de répartition et une densité d'une variable aléatoire X à densité sont notées respectivement F_X et f_X .

On admet que les formules donnant l'espérance et la variance d'une somme de variables aléatoires discrètes, ainsi que la définition et les propriétés de la covariance et du coefficient de corrélation linéaire de deux variables aléatoires discrètes, s'appliquent au cas de variables aléatoires à densité.

Pour n entier supérieur ou égal à 2, on dit que les variables aléatoires à densité X_1, X_2, \dots, X_n sont indépendantes si pour tout n -uplet (x_1, x_2, \dots, x_n) de réels, les événements $[X_1 \leq x_1], [X_2 \leq x_2], \dots, [X_n \leq x_n]$ sont indépendants.

L'objet du problème est double. D'une part, montrer certaines analogies entre les lois géométriques et exponentielles, d'autre part mettre en évidence quelques propriétés asymptotiques de variables aléatoires issues de la loi exponentielle.

La partie II est indépendante de la partie I. La partie III est indépendante de la partie II et largement indépendante de la partie I.

Partie I. Loi exponentielle

1. a. Rappeler la valeur de $\int_0^{+\infty} e^{-t} dt$. Etablir pour tout n de \mathbb{N}^* la convergence de l'intégrale

$$\int_0^{+\infty} t^n e^{-t} dt.$$

On pose alors $I_0 = \int_0^{+\infty} e^{-t} dt$ et, pour tout n de \mathbb{N}^* $I_n = \int_0^{+\infty} t^n e^{-t} dt$.

- b. Soit n un entier de \mathbb{N}^* . A l'aide d'une intégration par parties, établir une relation de récurrence entre I_n et I_{n-1} . En déduire la valeur de I_n en fonction de n .

Soit λ un réel strictement positif. Soit X_1 et X_2 deux variables indépendantes de même loi exponentielle de paramètre λ (d'espérance $1/\lambda$).

on pose : $Y = X_1 - X_2$, $T = \max(X_1, X_2)$ et $Z = \min(X_1, X_2)$.

2. Justifier les relations $T + Z = X_1 + X_2$ et $T - Z = |X_1 - X_2| = |Y|$.
3. a. Rappeler sans démonstration les valeurs respectives de $V(X_1)$ et de $P([X_1 \leq x])$, pour tout réel x .
b. Calculer $E(X_1 + X_2)$, $V(X_1 + X_2)$, $E(Y)$, $V(Y)$.
4. Déterminer pour tout réel z , $F_Z(z)$ et $f_Z(z)$. Reconnaître la loi de Z et en déduire $E(Z)$ et $V(Z)$.

5. a. Montrer que pour tout réel t , on a : $F_T(t) = \begin{cases} (1 - e^{-\lambda t})^2 & \text{si } t \geq 0 \\ 0 & \text{si } t < 0 \end{cases}$. Exprimer pour tout réel t , $f_T(t)$.

- b. Justifier l'existence de $E(T)$ et $V(T)$. Montrer que $E(T) = \frac{3}{2\lambda}$ et $V(T) = \frac{5}{4\lambda^2}$.
(on pourra utiliser des changements de variables affine).
6. On note r le coefficient de corrélation linéaire de Z et T . Montrer que $r = 1/\sqrt{5}$.
7. a. Préciser $Y(\Omega)$ et $|Y|(\Omega)$.
b. Déterminer une densité de la variable aléatoire $-X_2$.
c. Montrer que pour tout réel y , l'intégrale $\int_{-\infty}^{+\infty} f_{X_1}(t) f_{-X_2}(y-t) dt$ est convergente et qu'elle vaut $\frac{\lambda}{2} e^{-\lambda|y|}$.
(on distinguera les deux cas : $y \geq 0$ et $y < 0$)
d. Etablir que la fonction $y \mapsto \frac{\lambda}{2} e^{-\lambda|y|}$ est une densité de probabilité sur \mathbb{R} ; on admet que c'est une densité de la variable aléatoire Y .
e. Déterminer pour tout y réel, $f_{|Y|}(y)$. Reconnaître la loi de $|Y| = T - Z$.

Partie II. Loi géométrique

Soit p un réel de $]0, 1[$ et $q = 1 - p$. Soit X_1 et X_2 deux variables indépendantes de même loi géométrique de paramètre p (d'espérance $1/p$).

on pose : $Y = X_1 - X_2$, $T = \max(X_1, X_2)$ et $Z = \min(X_1, X_2)$. On rappelle que $T + Z = X_1 + X_2$ et $T - Z = |X_1 - X_2| = |Y|$.

1. a. Rappeler sans démonstration les valeurs respectives de $V(X_1)$ et de $P([X_1 \leq k])$, pour tout k de $X_1(\Omega)$.
b. Calculer $E(X_1 + X_2)$, $V(X_1 + X_2)$, $E(X_1 - X_2)$, $V(X_1 - X_2)$.
c. Etablir la relation : $P([X_1 = X_2]) = \frac{p}{1+q}$
2. a. Montrer que Z suit la loi géométrique de paramètre $1 - q^2$. En déduire $E(Z)$, $V(Z)$ et $E(T)$.
b. Soit k un entier de \mathbb{N}^* . Justifier l'égalité : $[Z = k] \cup [T = k] = [X_1 = k] \cup [X_2 = k]$.
En déduire la relation suivante : $P(T = k) = 2P(X_1 = k) - P(Z = k)$.
c. Etablir la formule : $V(T) = \frac{q(2q^2 + q + 2)}{(1 - q^2)^2}$.
3. a. Préciser $(T - Z)(\Omega)$. Exprimer pour tout j de \mathbb{N}^* , l'événement $[Z = j] \cap [Z = T]$ en fonction des événements $[X_1 = j]$ et $[X_2 = j]$. En déduire pour tout j de \mathbb{N}^* , l'expression de $P([Z = j] \cap [Z = T])$
b. Montrer que pour tout couple (j, l) de $(\mathbb{N}^*)^2$, on a : $P([Z = j] \cap [T - Z = l]) = 2p^2 q^{2j+l-2}$
c. Montrer que pour tout k de \mathbb{Z} , $P([X_1 - X_2 = k]) = \frac{pq^{|k|}}{1+q}$ (on distinguera trois cas : $k = 0$, $k > 0$ et $k < 0$).
d. En déduire la loi de la variable aléatoire $|X_1 - X_2|$.
e. Etablir à l'aide des questions précédentes que les variables Z et $T - Z$ sont indépendantes.
4. a. A l'aide du résultat de la précédente (3.e., calculer $Cov(Z, T)$. Les variables Z et T sont-elles indépendantes ?
b. Calculer en fonction de q , le coefficient de corrélation linéaire ρ de Z et T .
c. Déterminer la loi de probabilité du couple (Z, T) .

- d. Déterminer pour tout j de \mathbb{N}^* , la loi de probabilité conditionnelle de T sachant l'événement $[Z = j]$.
- e. Soit j un élément de \mathbb{N}^* . On suppose qu'il existe une variable aléatoire D_j à valeur dans \mathbb{N}^* , dont la loi de probabilité est la loi conditionnelle de T sachant l'événement $[Z = j]$. Calculer $E(D_j)$.

Partie III. Convergences

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, \dots, 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}^* : $S_n = \sum_{k=1}^n X_k$ 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 $f_{J_n}(x) = \begin{cases} \frac{e^{-x} x^{n-1}}{(n-1)!} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$.

- a. A l'aide du théorème de transfert, établir pour tout n supérieur ou égal à 3, l'existence de $E\left(\frac{1}{J_n}\right)$ et de $E\left(\frac{1}{J_n^2}\right)$, et donner leur valeurs respectives.

- b. On pose pour tout n supérieur ou égal à 3 : $\widehat{\lambda}_n = \frac{n}{S_n}$. Justifier que $\widehat{\lambda}_n$ est un estimateur de λ . Est-il sans biais ? Calculer la limite, lorsque n tend vers $+\infty$, du risque quadratique associé à $\widehat{\lambda}_n$ en λ .

3. **[Cubes]** 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. Enoncer le théorème de la limite centrée. En déduire que la variable aléatoire N_n définie par $N_n = \lambda \frac{S_n}{\sqrt{n}} - \sqrt{n}$ 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. Montrer que pour n assez grand, l'intervalle $\left[\left(1 - \frac{u_\alpha}{\sqrt{n}}\right) \widehat{\lambda}_n, \left(1 + \frac{u_\alpha}{\sqrt{n}}\right) \widehat{\lambda}_n\right]$ est un intervalle de confiance de λ au risque α . On note λ_0 la réalisation de $\widehat{\lambda}_n$ sur le n -échantillon.

4. Avec le n -échantillon (X_1, X_2, \dots, 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 . **[Hors-programme maintenant] - mais la question a. est faisable.**

- a. Justifier l'existence de la fonction réciproque Φ^{-1} de Φ . Quel est le domaine de définition de Φ^{-1} ?

- b. Etablir l'égalité $\beta = 2\Phi\left(\frac{1}{k}\Phi^{-1}(\alpha/2)\right)$. 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}^* : $T_n = \max(X_1, X_2, \dots, X_n)$.

Pour tout n de \mathbb{N}^* , pour tout réel x positif ou nul, on pose :

$$g_n(x) = \int_0^x F_{T_n}(t) dt \text{ et } h_n(x) = \int_0^x t f_{T_n}(t) dt$$

a. Exprimer $h_n(x)$ en fonction de $F_n(x)$ et $g_n(x)$.

b. Déterminer pour tout réel t , l'expression de $F_{T_n}(t)$ en fonction de t .

Etablir pour tout n supérieur ou égal à 2, la relation : $g_{n-1}(x) - g_n(x) = \frac{1}{n} F_{T_n}(x)$

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 $x, F_{T_1}(x), F_{T_2}(x), \dots, F_{T_n}(x)$.

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 $E(T_n) = \sum_{k=1}^n \frac{1}{k}$.

6. On veut étudier dans cette question la convergence en loi de la suite de variables aléatoires $(G_n)_{n \geq 1}$ définie par : pour tout n de \mathbb{N}^* , $G_n = T_n - E(T_n)$.

On pose pour tout n de \mathbb{N}^* : $\gamma_n = -\ln n + E(T_n)$ et on admet sans démonstration que la suite $(\gamma_n)_{n \geq 1}$ est convergente ; on note γ sa limite.

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

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

c. Montrer que la fonction $F_G : \mathbb{R} \rightarrow \mathbb{R}$ définie par $F_G(x) = e^{-e^{-(x+\gamma)}}$ est la fonction de répartition d'une variable aléatoire G à densité. Conclure **[Cubes]**.

7. a. 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 $Y = F_X(X)$.

b. On suppose qu'on importe les bibliothèques suivantes sous Python :

```
import numpy as np
import numpy.random as rd
```

Sachant que `rd.rand()` permet de simuler la loi uniforme sur $]0, 1[$, et la valeur de γ est donnée par la commande : `np.euler_gamma`, écrire une fonction Python d'en-tête `def Gumbel()` permettant de simuler la variable aléatoire G .