

CONSIGNES À LIRE !

La qualité de la rédaction, le soin porté à la copie, la lisibilité, l'orthographe, la rigueur du vocabulaire ainsi que la clarté des raisonnements sont des critères importants d'évaluation.

Quelques précisions :

- la copie doit être prise de sorte que la marge se situe à droite de chaque page,
- la première page de la copie doit rester vierge et sera réservée aux appréciations,
- toutes les pages de la copie devront être numérotées et rangées dans l'ordre de lecture,
- les résultats finaux doivent être clairement mis en évidence (soulignés ou encadrés),
- les questions d'un même exercice doivent être présentées dans l'ordre du sujet.
- Pour toutes les questions **Python** du sujet, on supposera avoir importé les différents modules nécessaires de la sorte :

```
import numpy as np
import numpy.random as rd
import numpy.linalg as al
import matplotlib.pyplot as plt
import pandas as pd
```

L'usage de tout matériel électronique est interdit. Aucun document n'est autorisé.

Toutes les variables aléatoires de ce problème sont supposées définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

Pour les programmes en Python, on suppose importées :

- la bibliothèque **numpy** sous l'alias **np**,
- la bibliothèque **numpy.random** sous l'alias **rd**,
- la bibliothèque **matplotlib.pyplot** sous l'alias **plt**.

Définition. Soient $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires et X une variable aléatoire, toutes définies sur $(\Omega, \mathcal{A}, \mathbb{P})$. On dit que la suite $(X_n)_{n \in \mathbb{N}^*}$ **converge en probabilité** vers X lorsque :

$$\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0$$

On notera $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$.

On rappelle le résultat suivant :

Inégalité de Markov. Si X est une variable aléatoire à valeurs positives admettant une espérance, alors :

$$\forall \varepsilon > 0, \mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}(X)}{\varepsilon}$$

PARTIE 1

Soient f une densité de probabilité sur $\mathbb{R}$, X une variable aléatoire réelle ayant f pour densité et F sa fonction de répartition. On note $\mathcal{E}$ l'ensemble des réels en lesquels la fonction F n'est pas de classe $\mathcal{C}^1$.

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires réelles mutuellement indépendantes, de même loi que X .

Pour tout entier $n \geq 1$, on ordonne, pour chaque $\omega \in \Omega$, le n -uplet $(X_1(\omega), X_2(\omega), \dots, X_n(\omega))$ en :

$$X_{(n,1)}(\omega) \leq X_{(n,2)}(\omega) \leq \dots \leq X_{(n,n)}(\omega),$$

et on admet que les applications $X_{(n,k)}$ ainsi définies sont des variables aléatoires réelles sur $(\Omega, \mathcal{A}, \mathbb{P})$.

Pour tous k, n entiers tels que $1 \leq k \leq n$, la variable aléatoire $X_{(n,k)}$ donne la k -ième valeur obtenue en classant dans l'ordre croissant les valeurs prises par les n premières X_i .

On remarque en particulier que $X_{(n,1)} = \min_{1 \leq i \leq n} (X_i)$ et $X_{(n,n)} = \max_{1 \leq i \leq n} (X_i)$.

1. Soit $n \in \mathbb{N}^*$.

1.a. Exprimer, à l'aide de F , la fonction de répartition $F_{(n,n)}$ de la variable aléatoire $X_{(n,n)}$. En déduire que $X_{(n,n)}$ admet comme densité la fonction $f_{(n,n)}$ définie par :

$$\forall x \in \mathbb{R}, f_{(n,n)}(x) = nf(x)F(x)^{n-1}.$$

1.b. Par une méthode similaire, montrer que $X_{(n,1)}$ admet comme densité la fonction $f_{(n,1)}$ définie par :

$$\forall x \in \mathbb{R}, f_{(n,1)}(x) = nf(x)(1 - F(x))^{n-1}.$$

2. Soit $x \in \mathbb{R}$ et soient n et k deux entiers tels que $1 \leq k \leq n$.

Pour tout $i \in \mathbb{N}^*$, on note $Y_{i,x}$ la variable aléatoire indicatrice de l'événement $[X_i \leq x]$. C'est-à-dire la variable aléatoire définie par :

$$\forall \omega \in \Omega, Y_{i,x}(\omega) = \begin{cases} 1 & \text{si } X_i(\omega) \leq x \\ 0 & \text{sinon} \end{cases}$$

2.a. Déterminer la loi de $Y_{i,x}$. On veillera à donner un paramètre indépendant de i .

2.b. Montrer l'égalité d'événements :

$$[X_{(n,k)} \leq x] = \left[\sum_{i=1}^n Y_{i,x} \geq k \right].$$

2.c. En déduire que la fonction de répartition $F_{(n,k)}$ de la variable aléatoire $X_{(n,k)}$ est donnée par :

$$\forall x \in \mathbb{R}, F_{(n,k)}(x) = \sum_{\ell=k}^n \binom{n}{\ell} F(x)^\ell (1 - F(x))^{n-\ell}.$$

3. Pour tout triplet d'entiers (n, k, ℓ) tel que $0 \leq k \leq n$ et $0 \leq \ell \leq n$, on introduit les fonctions polynomiales $b_{n,\ell}$ et $a_{n,k}$ définies respectivement par, pour tout $z \in \mathbb{R}$:

$$b_{n,\ell}(z) = \binom{n}{\ell} z^\ell (1 - z)^{n-\ell} \quad \text{et} \quad a_{n,k}(z) = \sum_{\ell=k}^n b_{n,\ell}(z).$$

On convient également que pour tout $n \in \mathbb{N}^*$, la fonction $b_{n-1,n}$ est identiquement nulle.

3.a. Soient $n \in \mathbb{N}^*$ et $\ell \in \llbracket 1; n \rrbracket$. Montrer que pour tout $z \in \mathbb{R}$:

$$b'_{n,\ell}(z) = n(b_{n-1,\ell-1}(z) - b_{n-1,\ell}(z)).$$

3.b. Soient $n \in \mathbb{N}^*$ et $k \in \llbracket 1; n \rrbracket$. D  duire de la question pr  c  dente une expression de $a'_{n,k}$ en fonction de l'une des fonctions $b_{m,\ell}$ avec m et ℓ    d  terminer en fonction de n et k .

3.c. Soient $n \in \mathbb{N}^*$ et $k \in \llbracket 1; n \rrbracket$. Montrer que la variable al  atoire $X_{(n,k)}$ admet comme densit   la fonction $f_{(n,k)}$ d  finie par :

$$\forall x \in \mathbb{R}, f_{(n,k)}(x) = n \binom{n-1}{k-1} f(x) F(x)^{k-1} (1 - F(x))^{n-k}.$$

4. Soit x un r  el tel que $0 < F(x) < 1$.

4.a. Soient $\alpha \in]F(x); 1[$ et $\varepsilon > 0$ tel que $\alpha - F(x) > \varepsilon$.

Montrer que pour tout entier n suffisamment grand, on a l'inclusion d'  v  nements :

$$\left[\sum_{i=1}^n Y_{i,x} \geq \lfloor \alpha n \rfloor \right] \subset \left[\left| \frac{1}{n} \sum_{i=1}^n Y_{i,x} - F(x) \right| \geq \varepsilon \right],$$

et en d  duire que :

$$\lim_{n \rightarrow +\infty} F_{(n, \lfloor \alpha n \rfloor)}(x) = 0.$$

4.b. Soient $\alpha \in]0; F(x)[$ et $\varepsilon > 0$ tel que $F(x) - \alpha > \varepsilon$.

Montrer que pour tout $n \in \mathbb{N}^*$:

$$\left[\sum_{i=1}^n Y_{i,x} \geq \lfloor \alpha n \rfloor \right] \supset \left[\left| \frac{1}{n} \sum_{i=1}^n Y_{i,x} - F(x) \right| < \varepsilon \right],$$

et en d  duire que :

$$\lim_{n \rightarrow +\infty} F_{(n, \lfloor \alpha n \rfloor)}(x) = 1.$$

5. On suppose pour cette question qu'il existe un intervalle ouvert $I =]a; b[$, avec   ventuellement $a = -\infty$ ou $b = +\infty$, tel que f est continue sur I , $f(x) > 0$ pour tout $x \in I$ et $f(x) = 0$ pour tout $x \in \mathbb{R} \setminus I$.

5.a. Montrer que la fonction de r  partition associ  e F est bijective de I dans $]0; 1[$.

5.b. On note F^{-1} la bijection r  ciproque de la fonction F restreinte    I et on fixe $\alpha \in]0; 1[$.

Montrer que la suite de variables al  atoires $(X_{(n, \lfloor \alpha n \rfloor)})_{n \in \mathbb{N}^*}$ converge en loi vers une variable al  atoire certaine   gale    $F^{-1}(\alpha)$.

5.c. En d  duire que la suite de variables al  atoires $(X_{(n, \lfloor \alpha n \rfloor)})_{n \in \mathbb{N}^*}$ converge en probabilit   vers une variable al  atoire certaine   gale    $F^{-1}(\alpha)$.

6. Dans cette question, on souhaite   crire une fonction **Python** d'en-t  te **def tri(L)** : permettant de trier dans l'ordre croissant les valeurs d'une liste **L** sans utiliser de fonction de tri pr  d  finie dans **Python**.

Pour cela, on utilise un algorithme de tri appel   *tri par insertion* et d  fini de la fa  on suivante :

- On agit sur les coefficients de la liste **L**, que l'on suppose num  rot  s de 0    $s - 1$.
- Pour chaque k compris entre 1 et $s - 1$, avant l'  tape k les coefficients num  rot  s 0    $k - 1$ de **L** ont d  j   t   class  s dans l'ordre croissant, les coefficients suivants   tant inchang  s.
L'  tape k consiste alors    ins  rer la valeur x du coefficient num  rot   k de **L** en bonne position parmi les $k + 1$ premiers coefficients de **L** afin qu'   l'issue de l'  tape k les coefficients num  rot  s 0    k de **L** soient class  s dans l'ordre croissant, les coefficients suivants   tant inchang  s.

Par exemple, en partant de la liste **L**=[38,28,35,24,31,15], les   tapes de l'algorithme de tri par insertion donnent successivement :

```
[28, 38, 35, 24, 31, 15]
[28, 35, 38, 24, 31, 15]
[24, 28, 35, 38, 31, 15]
[24, 28, 31, 35, 38, 15]
[15, 24, 28, 31, 35, 38]
```

Compl  ter la fonction **tri** ci-dessous afin que son ex  cution mette en   uvre l'algorithme de tri par insertion d  crit ci-dessus puis renvoie la liste **L** dans laquelle les valeurs sont rang  es dans l'ordre croissant.

```
1 def tri(L):
2     s=...
3     for k in ...
4         x=L[k]
5         i=...
6         while x<L[i] and i>=0:
7             ...
8             ...
9         L[i+1]=...
10    return L
```

7. On   crit le programme **Python** suivant    la suite de la fonction **tri** :

```

1 def echantillonX(n):
2     ...
3
4 def X(n,k):
5     L=echantillonX(n)
6     return ...
7
8 n=5000
9 A=np.linspace(0.01,0.99,99)
10 B=[]
11 for a in A:
12     k=int(np.floor(n*a)) #k=partie entière de n*a
13     B.append(X(n,k))
14 plt.plot(A,B,"r+")
15 plt.show()

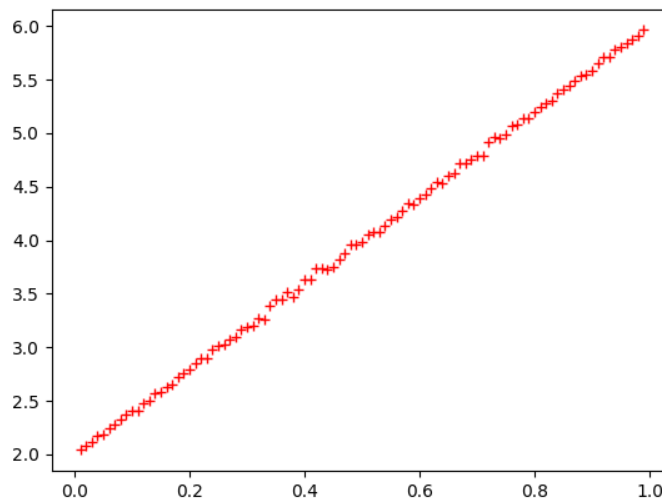
```

La fonction **echantillonX**, que l'on suppose déjà écrite pour les trois questions suivantes, renvoie, sous la forme d'une liste, une réalisation du n -uplet de variables aléatoires $(X_1, X_2, \dots, X_n)$.

On suppose que la densité de X satisfait les hypothèses de la question 5.

7.a. Compléter la fonction **X** afin qu'elle renvoie une réalisation de la variable aléatoire $X_{(n,k)}$ lorsqu'on lui fournit en entrée deux entiers n et k tels que $1 \leq k \leq n$. On pourra utiliser la fonction **tri**.

7.b. En exécutant le programme, on obtient l'affichage suivant :



À quoi correspondent les points affichés sur cette représentation graphique ?

7.c. Démontrer que pour tout $x \in]2; 6]$, $F(x) = \frac{x-2}{4}$. En déduire la loi commune des X_i simulées par la fonction **echantillonX**.

7.d. Compléter alors la fonction **echantillonX** par une ou plusieurs lignes de code pour que le script produise une représentation graphique de valeurs simulées proche de la représentation graphique ci-dessus.

PARTIE 2

On s'intéresse dans cette partie au cas où $(X_i)_{i \in \mathbb{N}^*}$ est une famille de variables aléatoires mutuellement indépendantes suivant toutes la même loi exponentielle $\mathcal{E}(\lambda)$, avec $\lambda \in]0; +\infty[$.

On admet le résultat suivant :

Si Y et Z sont deux variables aléatoires à densité, indépendantes, de densités respectives f_Y et f_Z telles qu'au moins l'une de ces deux densités est bornée, alors la variable aléatoire $Y + Z$ est à densité et admet pour densité la fonction h définie sur $\mathbb{R}$ par :

$$\forall x \in \mathbb{R}, h(x) = \int_{-\infty}^{+\infty} f_Y(x-t)f_Z(t)dt$$

8. Soit $n \in \mathbb{N}^*$ fixé.

8.a. On note $(T_1, \dots, T_n)$ une famille de variables aléatoires mutuellement indépendantes telles que :

pour tout $i \in \llbracket 1; n \rrbracket$, T_i suit la loi exponentielle $\mathcal{E}((n-i+1)\lambda)$.

Pour tout $k \in \llbracket 1; n \rrbracket$, on note de plus $S_k = T_1 + \dots + T_k$.

Montrer par récurrence sur $k \in \llbracket 1; n \rrbracket$ que la variable aléatoire S_k est une variable aléatoire à densité de densité la fonction $g_{n,k}$ définie par :

$$\forall x \in \mathbb{R}, g_{n,k}(x) = \begin{cases} n \binom{n-1}{k-1} \lambda e^{-\lambda(n-k+1)x} (1 - e^{-\lambda x})^{k-1} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

8.b. Soit $k \in \llbracket 1; n \rrbracket$. Vérifier que S_k suit la même loi que $X_{(n,k)}$.

9. Soit (n, k) un couple d'entiers tel que $1 \leq k \leq n$.

Montrer que $X_{(n,k)}$ admet une espérance et une variance données par :

$$\mathbb{E}(X_{(n,k)}) = \frac{1}{\lambda} \sum_{i=1}^k \frac{1}{n-i+1} \quad \text{et} \quad \mathbb{V}(X_{(n,k)}) = \frac{1}{\lambda^2} \sum_{i=1}^k \frac{1}{(n-i+1)^2}.$$

10.10.a. Montrer que pour tout entier $j \geq 1$ et tout $x \in [j; j+1]$, on a :

$$0 \leq \frac{1}{j} - \frac{1}{x} \leq \frac{2}{x^2}.$$

10.b. En déduire l'encadrement, pour $(n, k) \in \mathbb{N}^2$ tel que $1 \leq k \leq n$:

$$\frac{1}{\lambda} \ln \left(\frac{n+1}{n-k+1} \right) \leq \mathbb{E}(X_{(n,k)}) \leq \frac{1}{\lambda} \ln \left(\frac{n+1}{n-k+1} \right) + \frac{2}{\lambda} \left(\frac{1}{n-k+1} - \frac{1}{n+1} \right).$$

10.c. Montrer que pour tout $\alpha \in]0; 1[$:

$$\lim_{n \rightarrow +\infty} \mathbb{E}(X_{(n, \lfloor \alpha n \rfloor)}) = -\frac{1}{\lambda} \ln(1 - \alpha).$$

11. Montrer que pour tout $\alpha \in]0; 1[$:

$$\lim_{n \rightarrow +\infty} \mathbb{V}(X_{(n, \lfloor \alpha n \rfloor)}) = 0.$$

12. Soit $\alpha \in]0; 1[$.

12.a. Soient $n \in \mathbb{N}^*$ et $\varepsilon > 0$. Montrer que :

$$\mathbb{P} \left(\left| X_{(n, \lfloor \alpha n \rfloor)} + \frac{1}{\lambda} \ln(1 - \alpha) \right| \geq \varepsilon \right) \leq \frac{1}{\varepsilon^2} \mathbb{E} \left(\left(X_{(n, \lfloor \alpha n \rfloor)} + \frac{1}{\lambda} \ln(1 - \alpha) \right)^2 \right).$$

12.b. En déduire que la suite de variables aléatoires $(X_{(n, \lfloor \alpha n \rfloor)})_{n \in \mathbb{N}^*}$ converge en probabilité vers une variable aléatoire certaine égale à $-\frac{1}{\lambda} \ln(1 - \alpha)$.

13. Quelle question de la partie 1 permet de retrouver le résultat précédent ?

PARTIE 3

14. On admet **pour cette question** le théorème suivant :

Théorème :

Pour tout $n \in \mathbb{N}^*$, il existe une famille $(T_{n,1}, \dots, T_{n,n})$ de variables aléatoires mutuellement indépendantes telles que :

- pour tout $i \in \llbracket 1; n \rrbracket$, la variable aléatoire $T_{n,i}$ suit la loi exponentielle $\mathcal{E}((n-i+1)\lambda)$;
- pour tout $k \in \llbracket 1; n \rrbracket$, on a $X_{(n,k)} = T_{n,1} + \dots + T_{n,k}$.

En quoi ce théorème donne-t-il un résultat plus fort que ce qui a été prouvé en **8.b.** ?

Le but de la fin du problème est de démontrer ce théorème dans le cas où $n = 2$.

Soit $q \in]0; 1[$ et soit $p = 1 - q$. On dit qu'une variable aléatoire Z à valeurs dans $\mathbb{N}$ suit la loi $\mathcal{L}(q)$ si :

$$\forall k \in \mathbb{N}, \mathbb{P}([Z = k]) = pq^k.$$

15. On suppose que Z suit la loi $\mathcal{L}(q)$. Donner la valeur de l'espérance et de la variance de Z . Déterminer $\mathbb{P}([Z \geq \ell])$ pour tout $\ell \in \mathbb{N}$.

16. Soit $n \in \mathbb{N}^*$ et soit $(Z_1, \dots, Z_n)$ une famille de variables aléatoires mutuellement indépendantes de loi $\mathcal{L}(q)$.

On pose $U_n = \min(Z_1, \dots, Z_n)$.

16.a. Calculer $\mathbb{P}(U_n \geq \ell)$ pour tout $\ell \in \mathbb{N}$.

16.b. En déduire que U_n suit une loi $\mathcal{L}(r)$ pour un certain paramètre r à déterminer.

17. Soient Z_1, Z_2 deux variables aléatoires indépendantes et de même loi $\mathcal{L}(q)$. Soient k_1, k_2 deux entiers naturels. Montrer que :

$$\mathbb{P}([Z_1 \geq k_1] \cap [Z_2 - Z_1 \geq k_2]) = \frac{1}{1+q} q^{2k_1} q^{k_2}.$$

18. Soient X_1, X_2 deux variables aléatoires indépendantes suivant la loi exponentielle $\mathcal{E}(\lambda)$. Soient x, y deux réels positifs ou nuls.

18.a. Soit m un entier naturel non nul. On pose $Z_1 = \lfloor mX_1 \rfloor$ et $Z_2 = \lfloor mX_2 \rfloor$. Montrer que les variables aléatoires Z_1 et Z_2 suivent toutes les deux la loi $\mathcal{L}(q)$, où q est un paramètre que l'on précisera.

18.b. On conserve les notations de la question précédente et on pose de plus $k_1 = \lfloor mx \rfloor$ et $k_2 = \lfloor my \rfloor$. Montrer les inclusions entre événements :

$$[Z_1 \geq k_1 + 1] \subset [X_1 > x] \subset [Z_1 \geq k_1],$$

$$[Z_2 - Z_1 \geq k_2 + 2] \subset [X_2 - X_1 > y] \subset [Z_2 - Z_1 \geq k_2].$$

18.c. En déduire l'encadrement :

$$\frac{1}{1 + e^{-\lambda/m}} e^{-2\lambda \frac{\lfloor mx \rfloor + 1}{m}} e^{-\lambda \frac{\lfloor my \rfloor + 2}{m}} \leq \mathbb{P}([X_1 > x] \cap [X_2 - X_1 > y]) \leq \frac{1}{1 + e^{-\lambda/m}} e^{-2\lambda \frac{\lfloor mx \rfloor}{m}} e^{-\lambda \frac{\lfloor my \rfloor}{m}}.$$

18.d. Montrer que :

$$\mathbb{P}([X_1 > x] \cap [X_2 - X_1 > y]) = \frac{1}{2} e^{-2\lambda x} e^{-\lambda y}.$$

19. Soient X_1, X_2 deux variables aléatoires indépendantes suivant la loi exponentielle $\mathcal{E}(\lambda)$. On note $X_{(1)}$ et $X_{(2)}$ les variables aléatoires notées précédemment $X_{(2,1)}$ et $X_{(2,2)}$.

19.a. Montrer que $\mathbb{P}([X_1 = X_2]) = 0$. En déduire que les variables aléatoires $X_{(1)}$ et $X_{(2)} - X_{(1)}$ sont presque-sûrement à valeurs strictement positives.

19.b. Montrer que pour tous réels positifs x et y :

$$\mathbb{P}([X_{(1)} > x] \cap [X_{(2)} - X_{(1)} > y]) = 2\mathbb{P}([X_1 > x] \cap [X_2 - X_1 > y]).$$

19.c. Montrer que les variables aléatoires $T_1 = X_{(1)}$ et $T_2 = X_{(2)} - X_{(1)}$ sont indépendantes, de loi respective $\mathcal{E}(2\lambda)$ et $\mathcal{E}(\lambda)$. Conclure quant à l'objectif de cette partie.

★★★★★★★ FIN ★★★★★★★