DS10

Concours blanc type Maths 2 du 23/01/2024

Durée : 4h

La qualité de la rédaction, la clarté et la précision des raisonnements interviendront pour une part importante dans l'appréciation des copies. Les résultats doivent être encadrés.
La calculatrice n'est pas autorisée.

Dans ce problème, toutes les variables aléatoires sont supposées définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$*. Si* X *est une variable aléatoire, on note respectivement* $\mathrm{E}(X)$ *et* $\mathrm{Var}(X)$ *son espérance et sa variance, sous réserve d'existence.*

Le but de ce problème est de mettre en évidence quelques résultats asymptotiques liés au modèle du collectionneur de vignettes. Dans chaque paquet de céréales se trouve une vignette et il y a en tout des vignettes de n types différents, où n est un entier supérieur ou égal à 1 . Chacun des n types de vignettes se retrouve avec la même fréquence dans les paquets de céréales. Une collection est alors complète lorsqu'elle comporte n vignettes de types différents.

On modélise le nombre total de paquets de céréales qu'il est nécessaire d'acheter pour obtenir la collection complète de n vignettes de types différents par la variable aléatoire notée C_n.

On pose par convention $C_0 = 0$ et pour tout $i \in [\![1, n]\!]$, on note C_i le nombre d'achats de paquets de céréales nécessaires pour obtenir i vignettes de types différents.

De même, pour tout $i \in [\![1, n]\!]$, on pose $X_i = C_i - C_{i-1}$, qui représente le nombre d'achats supplémentaires de paquets de céréales qu'il est nécessaire d'effectuer pour obtenir une nouvelle vignette d'un type différent des $(i-1)$ vignettes de types différents déjà obtenues. Par convention, on pose $X_1 = C_1 = 1$.

On suppose que les variables aléatoires $X_1, \dots, X_n$ sont mutuellement indépendantes. Enfin, on pose :

$$V_n = \frac{C_n}{n} - \ln(n).$$

Questions préliminaires

1. (a) Montrer que :

$$C_n = X_1 + \cdots + X_n = \sum_{i=1}^{n} X_i.$$

   (b) Justifier que pour tout $i \in [\![1, n]\!]$, la variable aléatoire X_i suit la loi géométrique $\mathcal{G}\left(\dfrac{n-i+1}{n}\right)$.

Partie I

Pour tout entier $n \geq 1$, on pose :

$$H_n = \sum_{k=1}^{n} \frac{1}{k} \quad , \quad S_n = \sum_{k=1}^{n} \frac{1}{k^2}.$$

Nous allons démontrer la convergence de la suite $(S_n)_{n\geq 1}$ *et déterminer une valeur approchée de sa limite* S*.*

2. (a) Montrer que pour tout entier $k \geq 2$, on a l'encadrement :

$$\frac{1}{k} - \frac{1}{k+1} \leq \frac{1}{k^2} \leq \frac{1}{k-1} - \frac{1}{k}.$$

(b) Montrer que pour tout entier $n \geq 2$, on a :
$$\frac{3}{2} - \frac{1}{n+1} \leq S_n \leq 2 - \frac{1}{n}.$$

(c) Montrer que la suite $(S_n)_{n\geq 1}$ est convergente et donner un encadrement de sa limite S.

(d) Montrer que pour tout entier $n \geq 1$, on a l'encadrement :
$$\frac{1}{n+1} \leq S - S_n \leq \frac{1}{n}.$$

(e) En déduire un programme `Python` qui permet d'obtenir une valeur approchée de S à 10^{-7}-près.

3. Pour tout entier $n \geq 1$, on pose $u_n = H_n - \ln(n)$.

(a) Montrer que pour tout entier $n \geq 1$, on a l'encadrement :
$$\frac{1}{n+1} \leq \ln\left(\frac{n+1}{n}\right) \leq \frac{1}{n}.$$

(b) Montrer que pour tout entier $n \geq 1$, on a : $0 \leq u_n \leq 1$.

(c) Montrer que la suite $(u_n)_{n\geq 1}$ converge vers une certaine limite $\gamma \in [0,1]$.

4. Justifier l'existence de l'espérance de C_n et montrer que $E(C_n) = nH_n$.

5. Justifier l'existence de la variance de C_n et exprimer $\mathrm{Var}(C_n)$ en fonction de n, S_n et H_n.

6. (a) Montrer que pour tout réel $a > 0$, on a :
$$\mathbb{P}(|C_n - nH_n| \geq an) \leq \frac{S}{a^2}.$$

(b) Montrer que pour tout réel $c > 1$, on a :
$$\mathbb{P}\left(\left|\frac{C_n}{n} - \ln(n)\right| \geq c\right) \leq \frac{S}{(c-1)^2}.$$

(c) Pour cette question, on suppose que $n = 10^6$.
On donne l'approximation $\ln(n) \approx 13.816$.
Montrer que :
$$\mathbb{P}\left(\frac{C_n}{n} \in [7.81, 19.82]\right) \geq 0.92.$$

Partie II

7. Soient $T_1, \ldots, T_n, n$ variables aléatoires mutuellement indépendantes de même loi exponentielle $\mathcal{E}(1)$. On pose :
$$M_n = \max_{1\leq i\leq n} T_i.$$
Montrer que M_n suit la loi de densité f_n donnée par :
$$\forall x \in \mathbb{R}, \quad f_n(x) = \begin{cases} ne^{-x}\left(1-e^{-x}\right)^{n-1} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

8. (a) Soit Z une variable aléatoire de loi exponentielle $\mathcal{E}(n+1)$ et indépendante de M_n. On note g la densité de la loi de Z qui est nulle sur $]-\infty, 0]$ et continue sur $]0, +\infty[$. Montrer que pour tout $x > 0$ et tout $t \in]0, x[$, on a :
$$f_n(t)g(x-t) = (n+1)e^{-(n+1)x} ne^t \left(e^t - 1\right)^{n-1}.$$

(b) Soit $(Z_i)_{i\geq 1}$ une suite de variables aléatoires mutuellement indépendantes telle que pour tout entier $i \geq 1$, Z_i est de loi exponentielle $\mathcal{E}(i)$. Montrer que pour tout entier $n \geq 1$, la variable aléatoire $\displaystyle\sum_{i=1}^{n} Z_i$ suit la loi de densité f_n.

9. On définit la fonction $f : \mathbb{R} \to \mathbb{R}_+$ par :

$$\forall x \in \mathbb{R}, \quad f(x) = \mathrm{e}^{-x}\mathrm{e}^{-\mathrm{e}^{-x}}.$$

Montrer que f est une densité de probabilité sur $\mathbb{R}$. *La loi de densité f est appelée loi de Gumbel.*

10. (a) Soit $(Z_i)_{i\geq 1}$ la suite de variables aléatoires introduite précédemment. On pose $W_n = \displaystyle\sum_{i=1}^{n} Z_i - \ln(n)$.
Montrer que la fonction de répartition F_{W_n} de W_n est donnée par :

$$\forall x \in \mathbb{R}, \quad F_{W_n}(x) = \begin{cases} \left(1 - \frac{e^{-x}}{n}\right)^n & \text{si } x > -\ln(n), \\ 0 & \text{si } x \leq -\ln(n). \end{cases}$$

(b) Montrer que la suite de variables aléatoires $(W_n)_{n\geq 1}$ converge en loi vers une variable aléatoire de loi de Gumbel.

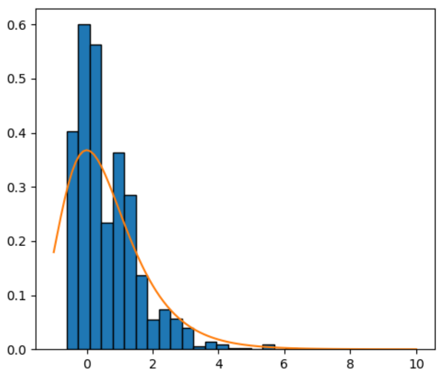
11. (a) On rappelle qu'en `Python` l'instruction `rd.geometric(p)` permet la simulation d'une variable aléatoire suivant la loi géométrique de paramètre p.
Écrire une fonction d'en-tête `def simulV(n)` qui pour un entier `n` fourni en entrée, renvoie une simulation de la variable aléatoire V_n définie en introduction de ce problème.

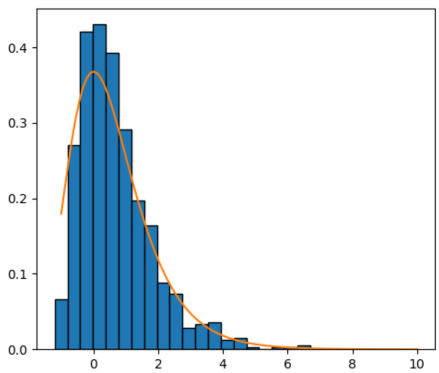
(b) À la suite de la fonction `simulV`, écrire un programme `Python` qui construit un vecteur-ligne V contenant 1000 simulations indépendantes de la variable aléatoire V_n pour un certain entier n entré par l'utilisateur.

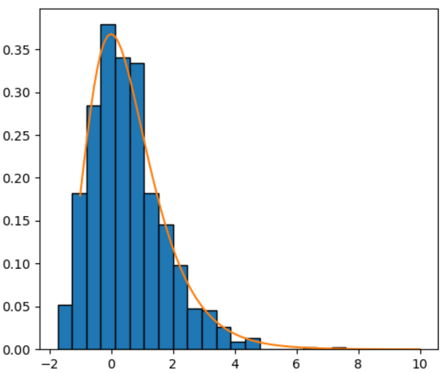
(c) On complète ce programme par le code suivant :

```python
plt.hist(V,20,density='True',edgecolor='k')

def f(x):
    y = np.exp(-x)*np.exp(-np.exp(-x))
    return y

absc = np.linspace(-1,10,100)
plt.plot(absc,f(absc))
plt.show()
```

On obtient les sorties graphiques suivantes en exécutant le programme pour $n = 5$, $n = 10$ puis $n = 50$:







Que peut-on observer sur ces figures ? Quelle conjecture peut-on en déduire pour la suite $(V_n)_{n\geq 1}$?

12. Soit $p \in]0,1[$ et soit U une variable aléatoire de loi exponentielle $\mathcal{E}(p)$. On pose $\alpha = -\frac{p}{\ln(1-p)}$.
Montrer que la variable aléatoire $V = \lfloor \alpha U \rfloor + 1$ est de loi géométrique $\mathcal{G}(p)$.
On rappelle que $\lfloor t \rfloor$ désigne la partie entière du réel t.

13. (a) Montrer que la variable aléatoire $V - \alpha U$ est à valeurs dans $[0,1]$. En déduire que $\mathrm{Var}(V - \alpha U) \leq 1$.
(b) Soient X et Y deux variables aléatoires admettant un moment d'ordre 2. Justifier l'existence de la covariance $\mathrm{Cov}(X,Y)$ puis montrer que :
$$\mathrm{Var}(X+Y) \leq 2(\mathrm{Var}(X) + \mathrm{Var}(Y)).$$
On pourra calculer $\mathrm{Var}(X+Y)$ et $\mathrm{Var}(X-Y)$.
(c) En déduire que :
$$\mathrm{Var}(V-U) \leq 2 + 2\frac{(1-\alpha)^2}{p^2}.$$

14. Dans cette question, on suppose que $n \geq 2$.
Soient Y_1, …, Y_n, n variables aléatoires mutuellement indépendantes telles que pour tout $i \in [\![1,n]\!]$, Y_i suit la loi exponentielle $\mathcal{E}\left(\dfrac{n-i+1}{n}\right)$.
On pose $\alpha_1 = 0$, $X_1' = 1$ et pour tout $i \in [\![2,n]\!]$, $\alpha_i = -\dfrac{n-i+1}{n} \times \dfrac{1}{\ln\left(\frac{i-1}{n}\right)}$ et $X_i' = \lfloor \alpha_i Y_i \rfloor + 1$.
De plus, on pose : $W_n' = \dfrac{Y_1 + \cdots + Y_n}{n} - \ln(n)$ et $V_n' = \dfrac{X_1' + \cdots + X_n'}{n} - \ln(n)$.
(a) Montrer que :
$$\mathrm{Var}\left(V_n' - W_n'\right) \leq \frac{1}{n^2} + \frac{2}{n^2}\sum_{i=2}^{n}\left[\left(\frac{n}{n-i+1}\right)^2 (1-\alpha_i)^2 + 1\right].$$
(b) i. Montrer que la fonction $\phi :]0,1[\to \mathbb{R}$ définie par :
$$\forall z \in]0,1[, \quad \phi(z) = \frac{1}{(1-z)^2}\left(1 + \frac{1-z}{\ln(z)}\right)^2$$
peut être prolongée en une fonction continue sur le segment $[0,1]$.
ii. En déduire qu'il existe un réel $A > 0$ tel que :
$$\lim_{n\to+\infty} \frac{1}{n}\sum_{i=2}^{n}\left(\frac{n}{n-i+1}\right)^2 (1-\alpha_i)^2 = A.$$
(c) Montrer que la suite de variables aléatoires $(V_n' - W_n')_{n\geq 2}$ converge en probabilité vers 0.

On admet le résultat suivant (théorème de Slutsky) : soient (X_n) et (Y_n) des suites de variables aléatoires. Supposons que (X_n) converge en loi vers une variable X, et que (Y_n) converge en probabilité vers une constante $c \in \mathbb{R}$. Alors $(X_n + Y_n)$ converge en loi vers $X + c$.

(d) Montrer que la suite de variables aléatoires $(V_n')_{n\geq 2}$ converge en loi vers une variable aléatoire de loi de Gumbel.
(e) On donne les valeurs numériques suivantes : $\mathrm{e}^{-\mathrm{e}^{-x}} \approx 0.96$ si $x = 3.20$ et $\mathrm{e}^{-\mathrm{e}^{-x}} \approx 0.04$ si $x = -1.17$. Lorsque $n = 10^6$, montrer que :
$$\mathbb{P}\left(\frac{C_n}{n} \in [12.65, 17.02]\right) \approx 0.92.$$
Comparer avec le résultat de la première partie et commenter.

Partie III

Dans cette partie, on suppose que $n \geq 2$. On suppose de plus que les vignettes sont numérotées de 1 à n. Pour tout $i \in [\![1, n]\!]$ et tout entier $m \geq 1$, on note $A_{i,m}$ la variable aléatoire donnant le nombre de vignettes numérotées i obtenues dans les m premiers paquets de céréales achetés.

15. (a) Justifier que pour tout $i \in [\![1, n]\!]$, la variable aléatoire $A_{i,m}$ suit la loi binomiale $\mathcal{B}\left(m, \frac{1}{n}\right)$.
    (b) Calculer la covariance $\text{Cov}\left(A_{1,m}, m - A_{1,m}\right)$.
    (c) À l'aide d'un raisonnement par l'absurde, en déduire que les variables aléatoires $A_{1,m}, \ldots, A_{n,m}$ ne sont pas mutuellement indépendantes.

16. (a) Montrer que si $(E_i)_{i\geq 1}$ est une famille d'événements, alors pour tout entier $r \geq 1$:
$$\mathbb{P}\left(\bigcup_{i=1}^{r} E_i\right) \leq \sum_{i=1}^{r} \mathbb{P}\left(E_i\right).$$
    (b) Montrer que :
$$\mathbb{P}\left(C_n > m\right) \leq n\left(1 - \frac{1}{n}\right)^m \leq n\mathrm{e}^{-\frac{m}{n}}.$$
    (c) Soit c un réel strictement positif. Montrer qu'on a la majoration :
$$\mathbb{P}\left(C_n > cn\ln(n)\right) \leq n^{1-c}.$$
    (d) Montrer que pour tout réel $x > -\ln(n)$, on a :
$$\mathbb{P}\left(V_n > x\right) \leq \mathrm{e}^{-x}.$$

Dans la suite de cette partie, on introduit un modèle légèrement différent : le nombre N de paquets achetés est décrit par une variable aléatoire N de loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda > 0$. On cherche à calculer la probabilité de compléter, à partir des N vignettes obtenues, la collection de vignettes. On suppose toujours que les vignettes sont numérotées de 1 à n. On note $\widetilde{A}_i$ la variable aléatoire donnant le nombre de vignettes numérotées i obtenues dans les N paquets de céréales achetés.

17. (a) Soit $p \in \mathbb{N}^*$. Justifier que pour tout $i \in [\![1, n]\!]$, la loi conditionnelle de $\widetilde{A}_i$ sachant l'événement $[N = p]$ est la loi binomiale $\mathcal{B}\left(p, \frac{1}{n}\right)$ et en déduire que la variable aléatoire $\tilde{A}_i$ suit une loi de Poisson dont on déterminera le paramètre.
    (b) Montrer que pour tout n-uplet $(k_1, \ldots, k_n)$ avec $k_i \in \mathbb{N}$ et $k_1 + \cdots + k_n = p$, on a:
$$\mathbb{P}_{[N=p]}\left(\left[\tilde{A}_1 = k_1\right] \cap \cdots \cap \left[\tilde{A}_n = k_n\right]\right) = \frac{p!}{k_1! \cdots k_n!} \times \frac{1}{n^p}.$$
    (c) Montrer que pour tout n-uplet $(k_1, \ldots, k_n)$ avec $k_i \in \mathbb{N}$, on a :
$$\mathbb{P}\left(\left[\tilde{A}_1 = k_1\right] \cap \cdots \cap \left[\tilde{A}_n = k_n\right]\right) = \prod_{i=1}^{n} \frac{\mathrm{e}^{-\frac{\lambda}{n}}\left(\frac{\lambda}{n}\right)^{k_i}}{k_i!}.$$
    (d) Montrer que les variables aléatoires $\tilde{A}_1, \ldots, \widetilde{A}_n$ sont mutuellement indépendantes. Commenter en comparant avec le résultat de la question 15.(c).

18. Soit D_n l'événement « à l'issue des N achats de paquets de céréales, la collection de vignettes est complète ».

    Montrer que :
$$\mathbb{P}\left(D_n\right) = \left(1 - \mathrm{e}^{-\frac{\lambda}{n}}\right)^n.$$

    On admet le résultat suivant : pour tout $p \in \mathbb{N}^*$, on a $\mathbb{P}_{[N=p]}\left(D_n\right) = \mathbb{P}\left(C_n \leq p\right)$.

19. (a) Montrer que :

$$\mathbb{P}(D_n) = \sum_{p=0}^{\infty} \mathbb{P}(C_n \leq p)\,\mathbb{P}(N = p).$$

(b) On suppose maintenant que $\lambda > 1$.
Soit $a \in]\sqrt{\lambda}, \lambda[$. On pose $k_1 = \lfloor \lambda - a \rfloor$ et $k_2 = \lfloor \lambda + a \rfloor + 1$.
Montrer que :

$$\sum_{p=k_1}^{k_2} \mathbb{P}(N = p) \geq 1 - \frac{\lambda}{a^2}.$$

(c) Montrer l'encadrement :

$$\mathbb{P}(C_n \leq k_1)\left(1 - \frac{\lambda}{a^2}\right) \leq \mathbb{P}(D_n) \leq \mathbb{P}(C_n \leq k_2) + \frac{\lambda}{a^2}.$$

20. (a) Soit $(c_n)_{n\geq 1}$ une suite convergente de nombres réels.
Montrer que pour tout entier n suffisamment grand, on a l'encadrement :

$$\mathbb{P}\left(V_n \leq c_n - \frac{1}{n^{1/3}} - \frac{1}{n}\right)\left(1 - \frac{\ln(n) + c_n}{n^{1/3}}\right) \leq \left(1 - \frac{e^{-c_n}}{n}\right)^n \leq \mathbb{P}\left(V_n \leq c_n + \frac{1}{n^{1/3}} + \frac{1}{n}\right) + \frac{\ln(n) + c_n}{n^{1/3}}.$$

On pourra appliquer la question précédente avec $\lambda = n\ln(n) + c_n n$ *et* $a = n^{2/3}$.

(b) Retrouver alors la convergence en loi de la suite $(V_n)_{n\geq 1}$ vers une variable aléatoire de loi de Gumbel.