



PRÉPASIMPLE

prepasimple.fr

HEC — 2022

Sujet de mathématiques approfondies

Mathématiques approfondies — Prépa ECG

Annales de concours · Sujet

Dans ce problème, toutes les variables aléatoires sont définies sur un même espace probabilisé. Pour une variable aléatoire X , on note $E(X)$ et $V(X)$ son espérance et sa variance, sous réserve d'existence.

Le problème étudie quelques résultats asymptotiques du modèle du collectionneur de vignettes. Chaque paquet de céréales contient une vignette parmi n types, où $n \geq 1$. Les types des vignettes successivement obtenues sont indépendants et uniformément distribués sur $\{1, \dots, n\}$. Une collection est complète lorsqu'elle contient les n types.

On note C_n le nombre total de paquets qu'il faut acheter pour obtenir la collection complète. On pose $C_0 = 0$ et, pour $i \in \{1, \dots, n\}$, on note C_i le nombre d'achats nécessaires pour obtenir i types différents. Ces notations sont utilisées à n fixé.

Pour $i \in \{1, \dots, n\}$, on pose $X_i = C_i - C_{i-1}$: c'est le nombre d'achats supplémentaires nécessaires pour obtenir un type différent des $i - 1$ types déjà obtenus. Par convention, $X_1 = C_1 = 1$.

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

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

Questions préliminaires

1. (a) Montrer que :

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

1. (b) Justifier que, pour tout $i \in \{1, \dots, n\}$, X_i suit la loi géométrique $\mathcal{G}\left(\frac{n-i+1}{n}\right)$ sur les entiers strictement positifs. Pour un paramètre égal à 1, cette loi est la loi certaine de valeur 1.

Partie I

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

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

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

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

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

2. (b) Montrer que, pour tout entier $n \geq 2$:

$$\frac{3}{2} - \frac{1}{n+1} \leq S_n \leq 2 - \frac{1}{n}.$$

2. (c) Montrer que $(S_n)_{n \geq 1}$ converge et encadrer sa limite S .

2. (d) Montrer que, pour tout entier $n \geq 1$:

$$\frac{1}{n+1} \leq S - S_n \leq \frac{1}{n}.$$

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

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

3. (a) Montrer que, pour tout entier $n \geq 1$:

$$\frac{1}{n+1} \leq \ln\left(\frac{n+1}{n}\right) \leq \frac{1}{n}.$$

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

3. (c) Montrer que $(u_n)_{n \geq 1}$ converge vers une 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 $V(C_n)$ en fonction de n , S_n et H_n .

6. (a) Montrer que, pour tout réel $a > 0$:

$$P(|C_n - nH_n| \geq an) \leq \frac{S}{a^2}.$$

6. (b) Montrer que, pour tout réel $c > 1$:

$$P\left(\left|\frac{C_n}{n} - \ln(n)\right| \geq c\right) \leq \frac{S}{(c-1)^2}.$$

6. (c) On suppose $n = 10^6$ et on donne l'approximation $\ln(n) \approx 13.816$. Montrer que :

$$P\left(\frac{C_n}{n} \in [7.81, 19.92]\right) \geq 0.92.$$

Partie II

7. Soient $T_1, \dots, T_n$ des 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 admet la densité :

$$\forall x \in \mathbb{R}, \quad f_n(x) = \begin{cases} ne^{-x}(1 - e^{-x})^{n-1} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

8. (a) Soit Z de loi exponentielle $\mathcal{E}(n+1)$, indépendante de M_n . On note g la densité de Z , nulle sur $]-\infty, 0]$ et continue sur $]0, +\infty[$. Montrer que, pour $x > 0$ et $t \in]0, x[$:

$$f_n(t)g(x-t) = (n+1)e^{-(n+1)x}ne^t(e^t - 1)^{n-1}.$$

8. (b) Soit $(Z_i)_{i \geq 1}$ une suite de variables aléatoires mutuellement indépendantes, avec $Z_i \sim \mathcal{E}(i)$. Montrer que, pour tout entier $n \geq 1$,

$$\sum_{i=1}^n Z_i$$

admet la densité f_n .

9. On définit $f : \mathbb{R} \rightarrow \mathbb{R}_+$ par :

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

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

10. (a) Avec la suite $(Z_i)_{i \geq 1}$ précédente, on pose

$$W_n = \sum_{i=1}^n Z_i - \ln(n).$$

Montrer que sa fonction de répartition est :

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

10. (b) Montrer que $(W_n)_{n \geq 1}$ converge en loi vers une variable aléatoire de loi de Gumbel.
11. (a) En Python, après `import numpy as np`, l'instruction `np.random.geometric(p)` simule une loi géométrique de paramètre p sur les entiers strictement positifs. Écrire une fonction `simulV(n)` qui, pour un entier $n \geq 1$, renvoie une simulation de V_n .
11. (b) À la suite de cette fonction, écrire un programme Python qui construit un tableau `V` contenant 1000 simulations indépendantes de V_n , pour un entier n entré par l'utilisateur. L'instruction `int(input("n = "))` lit cet entier au clavier.
11. (c) On complète le programme par le code suivant. Les paramètres `bins=20` et `density=True` demandent 20 classes et une normalisation de l'histogramme en densité; `facecolor="white"` et `edgecolor="black"` règlent les couleurs des rectangles.

Python

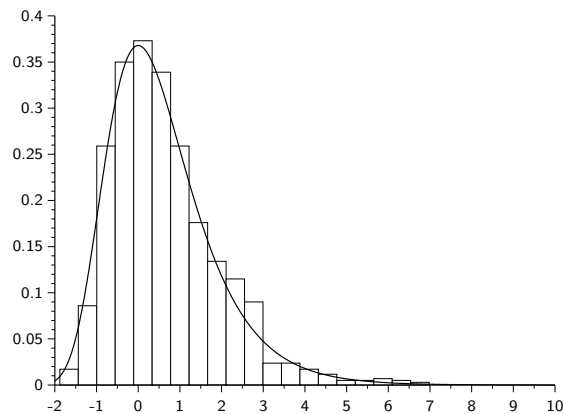
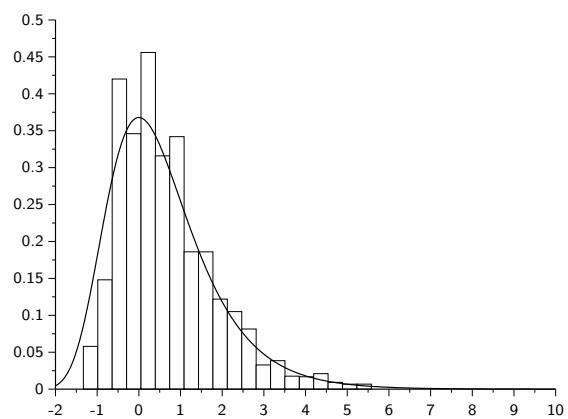
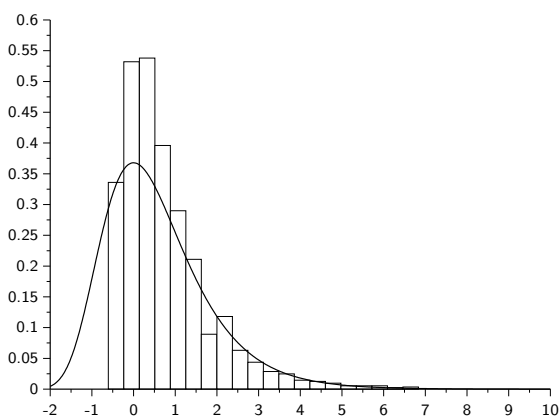
```
import matplotlib.pyplot as plt

plt.hist(V, bins=20, density=True, facecolor="white", edgecolor="black")

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

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

Les trois sorties graphiques ci-dessous correspondent, dans l'ordre, à $n = 5$, $n = 10$ puis $n = 50$. La courbe superposée est celle de f .



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

12. Soient $p \in]0, 1[$ et $U \sim \mathcal{E}(p)$. On pose :

$$\alpha = -\frac{p}{\ln(1-p)}.$$

Montrer que $V = \lfloor \alpha U \rfloor + 1$ suit la loi géométrique $\mathcal{G}(p)$. La notation $\lfloor t \rfloor$ désigne la partie entière de t .

13. (a) Montrer que $V - \alpha U$ est à valeurs dans $[0, 1]$. En déduire que $V(V - \alpha U) \leq 1$.

13. (b) Soient X et Y deux variables aléatoires admettant un moment d'ordre 2. Justifier l'existence de $\text{Cov}(X, Y)$ puis montrer que :

$$V(X + Y) \leq 2(V(X) + V(Y)).$$

On pourra calculer $V(X + Y)$ et $V(X - Y)$.

13. (c) En déduire que :

$$V(V - U) \leq 2 + 2\frac{(1-\alpha)^2}{p^2}.$$

14. On suppose $n \geq 2$. Soient $Y_1, \dots, Y_n$ mutuellement indépendantes, avec, pour $i \in \{1, \dots, n\}$:

$$Y_i \sim \mathcal{E}\left(\frac{n-i+1}{n}\right).$$

On pose $\alpha_1 = 0$, $X'_1 = 1$ et, pour $i \in \{2, \dots, n\}$:

$$\alpha_i = -\frac{n-i+1}{n} \times \frac{1}{\ln\left(\frac{i-1}{n}\right)}, \quad X'_i = \lfloor \alpha_i Y_i \rfloor + 1.$$

On pose aussi :

$$W'_n = \frac{Y_1 + \dots + Y_n}{n} - \ln(n), \quad V'_n = \frac{X'_1 + \dots + X'_n}{n} - \ln(n).$$

14. (a) Montrer que :

$$V(V'_n - W'_n) \leq \frac{1}{n^2} + \frac{2}{n^2} \sum_{i=2}^n \left[\left(\frac{n}{n-i+1} \right)^2 (1 - \alpha_i)^2 + 1 \right].$$

14. (b) i. Montrer que la fonction $\phi :]0, 1[\rightarrow \mathbb{R}$ définie par :

$$\forall z \in]0, 1[, \quad \phi(z) = \frac{1}{(1-z)^2} \left(1 + \frac{1-z}{\ln(z)} \right)^2$$

se prolonge en une fonction continue sur $[0, 1]$.

14. (b) ii. En déduire qu'il existe un réel $A > 0$ tel que :

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=2}^n \left(\frac{n}{n-i+1} \right)^2 (1 - \alpha_i)^2 = A.$$

14. (c) Montrer que $(V'_n - W'_n)_{n \geq 2}$ converge en probabilité vers 0.

14. (d) Montrer que $(V_n)_{n \geq 2}$ converge en loi vers une variable aléatoire de loi de Gumbel.

14. (e) On donne $e^{-e^{-x}} \approx 0.96$ pour $x = 3.20$ et $e^{-e^{-x}} \approx 0.04$ pour $x = -1.17$. Lorsque $n = 10^6$, montrer que :

$$P\left(\frac{C_n}{n} \in [12.65, 17.02]\right) \approx 0.92.$$

Comparer avec le résultat de la première partie et commenter.

Partie III

On suppose $n \geq 2$ et les vignettes numérotées de 1 à n . Pour $i \in \{1, \dots, n\}$ et $m \geq 1$ entier, $A_{i,m}$ désigne le nombre de vignettes de type i obtenues dans les m premiers achats.

15. (a) Justifier que, pour tout $i \in \{1, \dots, n\}$:

$$A_{i,m} \sim \mathcal{B}\left(m, \frac{1}{n}\right).$$

15. (b) Calculer $\text{Cov}(A_{1,m}, m - A_{1,m})$.

15. (c) Par un raisonnement par l'absurde, en déduire que $A_{1,m}, \dots, A_{n,m}$ ne sont pas mutuellement indépendantes.

16. (a) Si $(E_i)_{i \geq 1}$ est une famille d'événements, montrer que, pour tout entier $r \geq 1$:

$$P\left(\bigcup_{i=1}^r E_i\right) \leq \sum_{i=1}^r P(E_i).$$

16. (b) Montrer que, pour tout entier $m \geq 1$:

$$P(C_n > m) \leq n \left(1 - \frac{1}{n}\right)^m \leq ne^{-m/n}.$$

16. (c) Soit $c > 0$. En tenant compte du caractère entier de C_n , montrer que :

$$P(C_n > cn \ln(n)) \leq ne^{-[cn \ln(n)]/n} \leq e^{1/n} n^{1-c}.$$

16. (d) Montrer que, pour tout réel $x > -\ln(n)$:

$$P(V_n > x) \leq e^{1/n} e^{-x}.$$

Dans la suite, le nombre N de paquets achetés suit la loi de Poisson $\mathcal{P}(\lambda)$, avec $\lambda > 0$, et est indépendant des types des vignettes successivement obtenues. Les types restent indépendants et uniformes sur $\{1, \dots, n\}$. On note $\tilde{A}_i$ le nombre de vignettes de type i obtenues lors des N achats.

17. (a) Soit $p \in \mathbb{N}^*$. Justifier que, pour tout $i \in \{1, \dots, n\}$, la loi conditionnelle de $\tilde{A}_i$ sachant $[N = p]$ est $\mathcal{B}\left(p, \frac{1}{n}\right)$. En déduire que $\tilde{A}_i$ suit une loi de Poisson dont on précisera le paramètre.

17. (b) Pour tout $(k_1, \dots, k_n) \in \mathbb{N}^n$ tel que $k_1 + \dots + k_n = p$, montrer que :

$$P_{[N=p]}([\tilde{A}_1 = k_1] \cap \dots \cap [\tilde{A}_n = k_n]) = \frac{p!}{k_1! \dots k_n!} \frac{1}{n^p}.$$

17. (c) Pour tout $(k_1, \dots, k_n) \in \mathbb{N}^n$, montrer que :

$$P([\tilde{A}_1 = k_1] \cap \dots \cap [\tilde{A}_n = k_n]) = \prod_{i=1}^n \frac{e^{-\lambda/n} (\lambda/n)^{k_i}}{k_i!}.$$

17. (d) Montrer que $\tilde{A}_1, \dots, \tilde{A}_n$ sont mutuellement indépendantes. Comparer avec la question 15. (c).

18. Soit D_n l'événement «à l'issue des N achats, la collection est complète». Montrer que :

$$P(D_n) = (1 - e^{-\lambda/n})^n.$$

On admet que, pour tout $p \in \mathbb{N}^*$:

$$P_{[N=p]}(D_n) = P(C_n \leq p).$$

19. (a) Montrer que :

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

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

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

19. (c) Montrer que :

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

20. (a) Soit $(c_n)_{n \geq 1}$ une suite réelle convergente. Montrer que, pour tout entier n suffisamment grand :

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

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

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

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