

HEC Paris – ESCP BS

Mathématiques II

Mathématiques approfondies

Sujet adapté – 2022

Les parties du sujet contenant des éléments n'étant plus au programme depuis 2022 ont été adaptées pour convenir aux nouveaux programmes de la filière Mathématiques approfondies en vigueur depuis 2023.

Table des matières

1	Partie I – Premiers résultats asymptotiques	2
	<i>Thèmes abordés : suites ; séries ; probabilités discrètes ; estimations ; Python.</i>	2
2	Partie II – Convergence vers la loi de Gumbel	3
	<i>Thèmes abordés : variables à densité ; convergences de variables ; Python.</i>	3
3	Partie III – Poissonnisation du modèle	4
	<i>Thèmes abordés : probabilités discrètes ; convergences de variables.</i>	4

Problème – Le collectionneur de vignettes

Dans ce problème, toutes les variables aléatoires sont supposées définies sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$. Si X est une variable aléatoire, on note respectivement $E(X)$ et $\text{Var}(X)$ 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 n vignettes de types différents, où $n \geq 1$. Chacun des n types se retrouve avec la même fréquence. Une collection est complète lorsqu'elle comporte n vignettes de types différents.

On modélise le nombre total de paquets nécessaires à l'obtention de la collection complète par C_n . On pose $C_0 = 0$ et, pour tout $i \in \{1, \dots, n\}$, C_i désigne le nombre d'achats nécessaires pour obtenir i types différents. On pose aussi $X_i = C_i - C_{i-1}$; par convention, $X_1 = C_1 = 1$. Les variables $X_1, \dots, X_n$ sont mutuellement indépendantes et

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

Questions préliminaires

1. (a) Montrer que $C_n = X_1 + \dots + X_n = \sum_{i=1}^n X_i$.
- (b) Justifier que, pour tout $i \in \{1, \dots, n\}$, X_i suit la loi géométrique $\mathcal{G}\left(\frac{n-i+1}{n}\right)$.

1 Partie I – Premiers résultats asymptotiques

Thèmes abordés : suites ; séries ; probabilités discrètes ; estimations ; Python.

Pour tout entier $n \geq 1$, on pose $H_n = \sum_{k=1}^n \frac{1}{k}$ et $S_n = \sum_{k=1}^n \frac{1}{k^2}$. Nous allons démontrer la convergence de $(S_n)_{n \geq 1}$ et déterminer une valeur approchée de sa limite S .

2. (a) Montrer que, pour tout $k \geq 2$, $\frac{1}{k} - \frac{1}{k+1} \leq \frac{1}{k^2} \leq \frac{1}{k-1} - \frac{1}{k}$.
- (b) Montrer que, pour tout $n \geq 2$, $\frac{3}{2} - \frac{1}{n+1} \leq S_n \leq 2 - \frac{1}{n}$.
- (c) Montrer que (S_n) est convergente et donner un encadrement de S .
- (d) Montrer que, pour tout $n \geq 1$, $\frac{1}{n+1} \leq S - S_n \leq \frac{1}{n}$.
- (e) En déduire un programme Python qui permet d'obtenir une valeur approchée de S à 10^{-7} près.
3. Pour tout $n \geq 1$, on pose $u_n = H_n - \ln(n)$.
 - (a) Montrer que $\frac{1}{n+1} \leq \ln\left(\frac{n+1}{n}\right) \leq \frac{1}{n}$.
 - (b) Montrer que $0 \leq u_n \leq 1$.
 - (c) Montrer que (u_n) converge vers une limite $\gamma \in [0, 1]$.
4. Justifier l'existence de $E(C_n)$ et montrer que $E(C_n) = nH_n$.
5. Justifier l'existence de $\text{Var}(C_n)$ et exprimer $\text{Var}(C_n)$ en fonction de n , S_n et H_n .
6. (a) Montrer que, pour tout $a > 0$, $P(|C_n - nH_n| \geq an) \leq \frac{S}{a^2}$.
- (b) Montrer que, pour tout $c > 1$, $P\left(\left|\frac{C_n}{n} - \ln(n)\right| \geq c\right) \leq \frac{S}{(c-1)^2}$.
- (c) Pour $n = 10^6$, avec $\ln(n) \simeq 13,816$, montrer que $P\left(\frac{C_n}{n} \in [7,81, 19,92]\right) \geq 0,92$.

2 Partie II – Convergence vers la loi de Gumbel

Thèmes abordés : variables à densité ; convergences de variables ; Python.

7. Soient $T_1, \dots, T_n$ indépendantes de même loi exponentielle $\mathcal{E}(1)$ et $M_n = \max_{1 \leq i \leq n} T_i$. Montrer que M_n suit la densité

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

8. (a) Soit $Z \sim \mathcal{E}(n+1)$ indépendante de M_n , de densité g . Montrer que, pour $x > 0$ et $t \in]0, x[$,

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

- (b) Soit $(Z_i)_{i \geq 1}$ indépendante, avec $Z_i \sim \mathcal{E}(i)$. Montrer que $\sum_{i=1}^n Z_i$ suit la densité f_n .

9. On définit $f : \mathbb{R} \rightarrow \mathbb{R}_+$ par $f(x) = e^{-x}e^{-e^{-x}}$. Montrer que f est une densité sur $\mathbb{R}$. Cette loi est appelée loi de Gumbel.

10. (a) Avec les Z_i précédentes, poser $W_n = \sum_{i=1}^n Z_i - \ln(n)$. Montrer que

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

- (b) Montrer que (W_n) converge en loi vers une variable de loi de Gumbel.

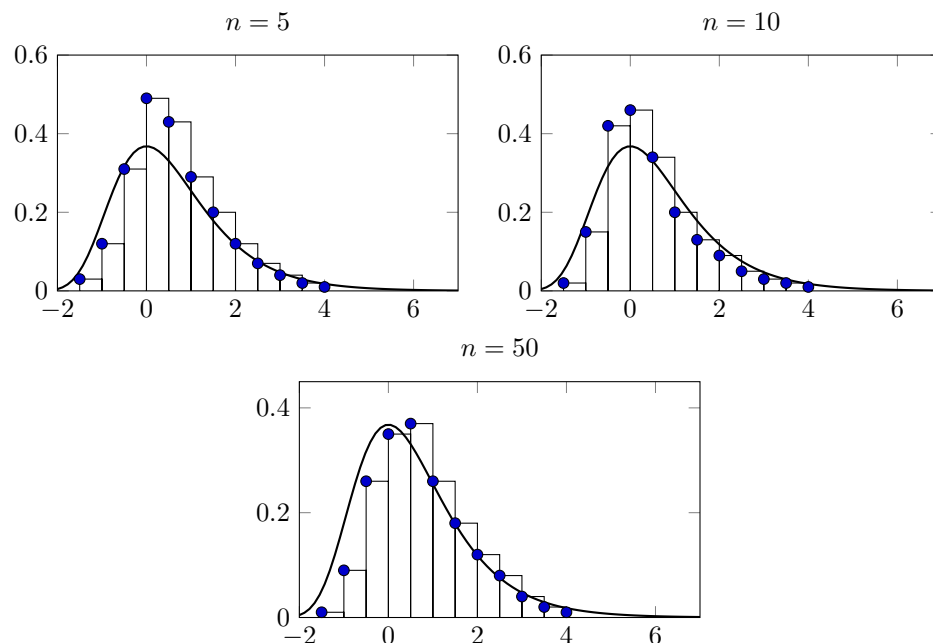
11. (a) En Python, `rng.geometric(p)` simule une loi géométrique de paramètre p . Écrire une fonction `def simulV(n):` qui renvoie une simulation de V_n .

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

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

```
plt.hist(V, bins=20, density=True)
def f(x):
    return np.exp(-x) * np.exp(-np.exp(-x))
absc = np.linspace(-1, 10, 100)
plt.plot(absc, f(absc))
```

On obtient les sorties ci-dessous pour $n = 5$, 10 puis 50.



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

12. Soit $p \in]0, 1[$, $U \sim \mathcal{E}(p)$ et $\alpha = -\frac{p}{\ln(1-p)}$. Montrer que $V = \lfloor \alpha U \rfloor + 1$ suit $\mathcal{G}(p)$. On rappelle que $\lfloor t \rfloor$ désigne la partie entière de t .
13. (a) Montrer que $V - \alpha U \in [0, 1]$, puis que $\text{Var}(V - \alpha U) \leq 1$.
 (b) Si X, Y admettent un moment d'ordre 2, justifier l'existence de $\text{Cov}(X, Y)$ et montrer $\text{Var}(X+Y) \leq 2(\text{Var}(X) + \text{Var}(Y))$. On pourra calculer $\text{Var}(X+Y)$ et $\text{Var}(X-Y)$.
 (c) En déduire $\text{Var}(V - U) \leq 2 + 2\frac{(1-\alpha)^2}{p^2}$.
14. Pour $n \geq 2$, soient Y_i indépendantes et $Y_i \sim \mathcal{E}\left(\frac{n-i+1}{n}\right)$. Posons $\alpha_1 = 0$, $X'_1 = 1$ et, pour $i \in \{2, \dots, n\}$,

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

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

(a) Montrer que

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

- (b) i. Montrer que $\phi :]0, 1[\rightarrow \mathbb{R}$, $\phi(z) = \frac{1}{(1-z)^2} \left(1 + \frac{(1-z)^2}{\ln(z)} \right)$, se prolonge continûment sur $[0, 1]$.
 ii. En déduire qu'il existe $A > 0$ tel que $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=2}^n \left(\frac{n}{n-i+1} \right)^2 (1 - \alpha_i)^2 = A$.
- (c) Montrer que $(V'_n - W'_n)$ converge en probabilité vers 0.
- (d) Montrer que (V'_n) converge en loi vers une variable de Gumbel.
- (e) Sachant que $e^{-e^{-x}} \simeq .96$ pour $x = 3.20$ et $\simeq .04$ pour $x = -1.17$, montrer lorsque $n = 10^6$ que $P\left(\frac{C_n}{n} \in [12.65, 17.02]\right) \simeq .92$. Comparer avec la première partie et commenter.

3 Partie III – Poissonnisation du modèle

Thèmes abordés : probabilités discrètes ; convergences de variables.

Dans cette partie, $n \geq 2$ et les vignettes sont numérotées de 1 à n . Pour $i \in \{1, \dots, n\}$ et $m \geq 1$, $A_{i,m}$ désigne le nombre de vignettes i obtenues dans les m premiers paquets.

15. (a) Justifier que $A_{i,m} \sim \mathcal{B}(m, 1/n)$.
 (b) Calculer $\text{Cov}(A_{1,m}, m - A_{1,m})$.
 (c) Par l'absurde, déduire que $A_{1,m}, \dots, A_{n,m}$ ne sont pas mutuellement indépendantes.
16. (a) Montrer que $P(\bigcup_{i=1}^r E_i) \leq \sum_{i=1}^r P(E_i)$.
 (b) Montrer que $P(C_n > m) \leq n(1 - 1/n)^m \leq ne^{-m/n}$.
 (c) Pour $c > 0$, montrer $P(C_n > cn \ln n) \leq n^{1-c}$.
 (d) Montrer que, pour $x > -\ln n$, $P(V_n > x) \leq e^{-x}$.

Dans la suite, le nombre N de paquets est une variable de Poisson $\mathcal{P}(\lambda)$, $\lambda > 0$. On note $\tilde{A}_i$ le nombre de vignettes i obtenues dans les N paquets.

17. (a) Pour $p \in \mathbb{N}^*$, justifier que, conditionnellement à $[N = p]$, $\tilde{A}_i$ suit $\mathcal{B}(p, 1/n)$, puis déterminer la loi de $\tilde{A}_i$.
 (b) Pour $k_i \in \mathbb{N}$ et $\sum k_i = p$, montrer

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

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

- (d) Montrer que $\tilde{A}_1, \dots, \tilde{A}_n$ sont indépendantes et commenter par rapport à 15.c).
18. Soit D_n l'événement « la collection est complète à l'issue des N achats ». Montrer $P(D_n) = (1 - e^{-\lambda/n})^n$.
On admet que $P_{[N=p]}(D_n) = P(C_n \leq p)$.
19. (a) Montrer $P(D_n) = \sum_{p=0}^{\infty} P(C_n \leq p)P(N = p)$.
(b) Pour $\lambda > 1$, $a \in]\sqrt{\lambda}, \lambda[$, $k_1 = \lfloor \lambda - a \rfloor$, $k_2 = \lfloor \lambda + a \rfloor + 1$, montrer $\sum_{p=k_1}^{k_2} P(N = p) \geq 1 - \frac{\lambda}{a^2}$.
(c) Montrer $P(C_n \leq k_1)(1 - \frac{\lambda}{a^2}) \leq P(D_n) \leq P(C_n \leq k_2) + \frac{\lambda}{a^2}$.
20. (a) Soit (c_n) convergente. Montrer que, pour n assez grand,

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

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

- (b) Retrouver la convergence en loi de (V_n) vers une loi de Gumbel.