HEC Paris – ESCP Europe – EAP

Mathématiques II

Mathématiques approfondies

Sujet adapté – 2009

Table des matières

1 Préliminaire **2**

Thèmes abordés : probabilités discrètes. 2

2 Infimum et supremum **2**

Thèmes abordés : probabilités discrètes ; suites ; Python. 2

3 Couple (Inf, Sup) **3**

Thèmes abordés : couples de variables discrètes ; suites. 3

4 Prévision **4**

Thèmes abordés : probabilités discrètes ; fonctions de plusieurs variables. 4

5 Estimation **5**

Thèmes abordés : probabilités discrètes ; estimations. 5

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.

1 Préliminaire

Thèmes abordés : probabilités discrètes.

Dans tout le problème, N désigne un entier supérieur ou égal à 1. On note $\mathbb{E}(X)$ et $\mathbb{V}(X)$ respectivement l'espérance et la variance, lorsqu'elles existent, de toute variable aléatoire réelle X définie sur un espace probabilisé.

Soit $(U_n)_{n\geqslant 1}$ une suite de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, mutuellement indépendantes et de même loi uniforme discrète sur $[\![1, N]\!]$. Pour tout $n \in \mathbb{N}^*$, on pose

$$T_n = \sup(U_1, U_2, \ldots, U_n) \qquad \text{et} \qquad Z_n = \inf(U_1, U_2, \ldots, U_n).$$

On admet que T_n et Z_n sont des variables aléatoires définies sur $(\Omega, \mathcal{A}, \mathbb{P})$. Ainsi, pour tout $\omega \in \Omega$,

$$T_n(\omega) = \max\big(U_1(\omega), U_2(\omega), \ldots, U_n(\omega)\big) \quad \text{et} \quad Z_n(\omega) = \min\big(U_1(\omega), U_2(\omega), \ldots, U_n(\omega)\big).$$

Si $C \in \mathcal{A}$, on note $\mathbf{1}_C$ la variable aléatoire indicatrice de l'événement C, définie par

$$\mathbf{1}_C(\omega) = \begin{cases} 1 & \text{si } \omega \in C, \\ 0 & \text{si } \omega \notin C. \end{cases}$$

Pour tout $n \in \mathbb{N}^*$, on pose

$$d_n(N) = \begin{cases} \displaystyle\sum_{k=1}^{N-1} \left(\frac{k}{N}\right)^n & \text{si } N \geqslant 2, \\ 0 & \text{si } N = 1. \end{cases}$$

1. Soit Y une variable aléatoire définie sur $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans $[\![1, N]\!]$. Établir les deux relations suivantes :

$$\mathbb{E}(Y) = \sum_{k=0}^{N-1} \mathbb{P}(Y > k) \qquad \text{et} \qquad \mathbb{E}(Y^2) = \sum_{k=0}^{N-1} (2k+1)\mathbb{P}(Y > k).$$

2 Infimum et supremum

Thèmes abordés : probabilités discrètes ; suites ; Python.

2. Rappeler, sans démonstration, les valeurs respectives de $\mathbb{E}(U_1)$ et de $\mathbb{V}(U_1)$.
3. a) Calculer, pour tout $k \in [\![1, N]\!]$, $\mathbb{P}(T_n \leqslant k)$.
   b) En déduire la loi de probabilité de T_n.
4. a) Montrer que la suite $(d_n(N))_{n\geqslant 1}$ est convergente et calculer sa limite.
   b) Exprimer $\mathbb{E}(T_n)$ en fonction de N et de $d_n(N)$. En déduire la valeur de $\lim\limits_{n\to+\infty} \mathbb{E}(T_n)$.

c) Établir la formule suivante :

$$\mathbb{V}(T_n) = (2N-1)d_n(N) - 2Nd_{n+1}(N) - d_n^2(N).$$

En déduire la valeur de $\lim\limits_{n\to+\infty} \mathbb{V}(T_n)$.

d) Montrer que, si $N \geqslant 2$, on a $\lim\limits_{n\to+\infty} \dfrac{d_{n+1}(N)}{d_n(N)} = 1 - \dfrac{1}{N}$; en déduire que, lorsque n tend vers $+\infty$, $\mathbb{V}(T_n) \sim d_n(N)$.

5. Déterminer la loi de Z_n. Calculer $\mathbb{E}(Z_n)$ et $\mathbb{V}(Z_n)$.

6. La fonction Python `randint(a,b)` du module `random` permet de simuler un entier aléatoire uniforme entre a et b, bornes incluses. Écrire une fonction Python `simulmax(n)` qui simule la variable aléatoire T_n.

```
from random import randint

def simulmax(n):
    """Simule T_n pour une loi uniforme sur {1, ..., N}."""
    return max(randint(1, N) for _ in range(n))
```

3 Couple (Inf, Sup)

Thèmes abordés : couples de variables discrètes ; suites.

7. Pour tout $n \in \mathbb{N}^*$ et tout couple $(k,\ell) \in \mathbb{N}^2$, on pose

$$\phi_n(k,\ell) = \mathbb{P}\big([T_n \leqslant k] \cap [Z_n \leqslant \ell]\big).$$

a) Montrer que, pour tout $(k,\ell) \in [\![1,N]\!]^2$,

$$\phi_n(k,\ell) = \begin{cases} \left(\dfrac{k}{N}\right)^n & \text{si } k \leqslant \ell, \\ \left(\dfrac{k}{N}\right)^n - \left(\dfrac{k-\ell}{N}\right)^n & \text{si } k > \ell. \end{cases}$$

b) Établir, pour tout $(k,\ell) \in [\![1,N]\!]^2$, la formule suivante :

$$\mathbb{P}\big([T_n = k] \cap [Z_n = \ell]\big) = \phi_n(k,\ell) + \phi_n(k-1,\ell-1) - \phi_n(k-1,\ell) - \phi_n(k,\ell-1).$$

c) En déduire, en distinguant les trois cas $k < \ell$, $k = \ell$ et $k > \ell$, l'expression de $\mathbb{P}\big([T_n = k] \cap [Z_n = \ell]\big)$ en fonction de k et ℓ.

8. On donne, pour tout couple $(m,n) \in (\mathbb{N}^*)^2$, les deux relations suivantes :

$$\sum_{j=1}^{m} \big[(j+1)^n - 2j^n + (j-1)^n\big] = (m+1)^n - m^n - 1,$$

$$\sum_{j=1}^{m} j\big[(j+1)^n - 2j^n + (j-1)^n\big] = m(m+1)^n - (m+1)m^n.$$

a) En déduire, pour tout $n \in \mathbb{N}^*$, la formule suivante : $\mathbb{E}(T_n Z_n) = N(1 + d_{n+1}(N))$.

b) On note, pour tout $n \in \mathbb{N}^*$, ρ_n le coefficient de corrélation linéaire entre T_n et Z_n. Calculer $\lim\limits_{n\to+\infty} \rho_n$ lorsque $N \geqslant 2$.

9. a) Pour tout $n \in \mathbb{N}^*$ et tout $(k,\ell) \in [\![1,N]\!]^2$, calculer la probabilité conditionnelle $\mathbb{P}_{[T_n=k]}([Z_n = \ell])$.

b) En déduire, pour tout $n \in \mathbb{N}^*$ et pour tout $k \in [\![1,N]\!]$, l'expression de l'espérance conditionnelle $\mathbb{E}(Z_n \mid [T_n = k])$ de Z_n sachant $[T_n = k]$.

4 Prévision

Thèmes abordés : probabilités discrètes ; fonctions de plusieurs variables.

Pour $n \in \mathbb{N}^*$, on dispose d'un $(n+1)$-échantillon indépendant et identiquement distribué $(U_1, U_2, \ldots, U_{n+1})$ de la loi uniforme sur $[\![1,N]\!]$. On pose

$$T_n = \sup(U_1, U_2, \ldots, U_n) \qquad \text{et} \qquad T_{n+1} = \sup(U_1, U_2, \ldots, U_{n+1}) = \sup(T_n, U_{n+1}).$$

Pour tout $t = (t_1, t_2, \ldots, t_N) \in \mathbb{R}^N$, on pose $W_t(T_n) = \sum\limits_{k=1}^{N} t_k \mathbf{1}_{[T_n=k]}$.

Dans cette partie, on se propose de déterminer la valeur de t pour laquelle les deux conditions suivantes sont vérifiées :

$$\text{i) } \mathbb{E}(W_t(T_n)) = \mathbb{E}(T_{n+1}), \qquad \text{ii) } \mathbb{E}\big((T_{n+1} - W_t(T_n))^2\big) \text{ est minimale.}$$

10. On suppose, dans cette question, que les réels $t_1, \ldots, t_N$ sont deux à deux distincts. Montrer, pour tout $k \in [\![1,N]\!]$, la relation : $\mathbb{P}(W_t(T_n) = t_k) = \mathbb{P}(T_n = k)$.

11. Établir, pour tout $k \in [\![1,N]\!]$, la formule suivante :

$$\mathbb{E}\big(T_{n+1}\mathbf{1}_{[T_n=k]}\big) = \mathbb{E}(T_{n+1} \mid [T_n = k]) \times \mathbb{P}(T_n = k).$$

12. a) Calculer, pour tout $(k,j) \in [\![1,N]\!]^2$, $\mathbb{P}([T_n = k] \cap [T_{n+1} = j])$.

b) En déduire, pour tout $(k,j) \in [\![1,N]\!]^2$, la probabilité conditionnelle $\mathbb{P}_{[T_n=k]}([T_{n+1} = j])$.

c) Déterminer, pour tout $k \in [\![1,N]\!]$, l'expression de l'espérance conditionnelle $\mathbb{E}(T_{n+1} \mid [T_n = k])$ de T_{n+1} sachant $[T_n = k]$.

d) En appliquant la formule de l'espérance totale, déduire de la question précédente la relation suivante :

$$\mathbb{E}(T_{n+1}) = \frac{N+1}{2} + \frac{1}{2N}\big(\mathbb{E}(T_n^2) - \mathbb{E}(T_n)\big).$$

13. Établir l'égalité suivante : $(W_t(T_n))^2 = \sum\limits_{k=1}^{N} t_k^2 \mathbf{1}_{[T_n=k]}$.

14. Soit g la fonction définie sur $\mathbb{R}^N$ à valeurs réelles par

$$g(t_1, t_2, \ldots, t_N) = \mathbb{E}\big((T_{n+1} - W_t(T_n))^2\big).$$

a) À l'aide des résultats des questions 11, 12 et 13, expliciter g en fonction des variables $t_1, t_2, \ldots, t_N$.

b) Montrer que g admet un minimum global sur $\mathbb{R}^N$ atteint en un point $\theta = (\theta_1, \theta_2, \ldots, \theta_N)$ que l'on déterminera en fonction de $\mathbb{E}(T_{n+1} \mid [T_n = 1]), \mathbb{E}(T_{n+1} \mid [T_n = 2]), \ldots, \mathbb{E}(T_{n+1} \mid [T_n = N])$.

15. Établir les deux relations suivantes :

$$\mathbb{E}(W_\theta(T_n)) = \mathbb{E}(T_{n+1}) \qquad \text{et} \qquad \mathbb{V}(W_\theta(T_n)) \leqslant \mathbb{V}(T_{n+1}).$$

16. a) Établir, pour tout $i \in \mathbb{N}^*$, l'égalité suivante : $\displaystyle\sum_{k=1}^{N} k^i \mathbf{1}_{[T_n=k]} = (T_n)^i$.

b) En déduire la relation suivante : $W_\theta(T_n) = \dfrac{N+1}{2} + \dfrac{1}{2N}(T_n^2 - T_n)$.

5 Estimation

Thèmes abordés : probabilités discrètes ; estimations.

Soit U une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, de loi uniforme discrète sur $[\![1, N]\!]$. On suppose que le paramètre N est inconnu. Cette partie a pour objet la détermination d'un estimateur ponctuel de N, sans biais et de variance minimale.

Pour n entier supérieur ou égal à 1, soit $(U_1, U_2, \ldots, U_n)$ un n-échantillon i.i.d. de la loi de U.

17. Soit ε un réel strictement positif. On pose

$$A_n(\varepsilon) = [|T_n - N| \geqslant \varepsilon] \qquad \text{et} \qquad B_n(\varepsilon) = [|T_n - \mathbb{E}(T_n)| + |d_n(N)| \geqslant \varepsilon].$$

a) Peut-on dire que $T_n + d_n(N)$ est un estimateur sans biais de N ?

b) Montrer que la suite $(T_n)_{n\geqslant 1}$ est une suite d'estimateurs asymptotiquement sans biais du paramètre N.

c) Montrer que $A_n(\varepsilon) \subset B_n(\varepsilon)$ et qu'il existe un entier naturel n_0 tel que, pour tout $n > n_0$, on a :

$$B_n(\varepsilon) \subset [|T_n - \mathbb{E}(T_n)| \geqslant \varepsilon/2].$$

d) En déduire que la suite d'estimateurs $(T_n)_{n\geqslant 1}$ est convergente.

18. a) Calculer, pour tout n-uplet $(u_1, u_2, \ldots, u_n) \in [\![1, N]\!]^n$, $\mathbb{P}\left(\bigcap_{i=1}^{n} [U_i = u_i]\right)$.

b) En déduire que, pour tout $k \in [\![1, N]\!]$, la loi conditionnelle du vecteur aléatoire $(U_1, U_2, \ldots, U_n)$ sachant $[T_n = k]$ est donnée par :

$$\mathbb{P}_{[T_n=k]}\left(\bigcap_{i=1}^{n} [U_i = u_i]\right) = \begin{cases} \dfrac{1}{k^n - (k-1)^n} & \text{si, pour tout } i \in [\![1, n]\!],\ 1 \leqslant u_i \leqslant N \text{ et } \max_{1\leqslant i\leqslant n}(u_i) = k, \\ 0 & \text{sinon.} \end{cases}$$

On remarquera que cette loi conditionnelle ne dépend pas du paramètre N.

19. Pour $n \in \mathbb{N}^*$, on pose $S_n = T_n + Z_n - 1$ et, pour tout $k \in [\![1, N]\!]$,

$$\psi_n(k) = \frac{k^{n+1} - (k-1)^{n+1}}{k^n - (k-1)^n}.$$

a) Montrer que S_n est un estimateur sans biais de N.

b) Établir, pour tout $k \in [\![1, N]\!]$, l'égalité : $\psi_n(k) = \mathbb{E}(S_n \mid [T_n = k])$.

c) En déduire que $\psi_n(T_n)$ est un estimateur sans biais de N.

d) On pose, pour tout $k \in [\![1, N]\!]$, $\varphi_n(k) = \mathbb{E}(S_n^2 \mid [T_n = k])$. Établir, pour tout $k \in [\![1, N]\!]$, l'inégalité : $\psi_n^2(k) \leqslant \varphi_n(k)$ (on pourra utiliser la fonction définie sur $\mathbb{R}$ par $\lambda \mapsto \mathbb{E}((S_n - \lambda)^2 \mid [T_n = k])$). En déduire que $\mathbb{V}(\psi_n(T_n)) \leqslant \mathbb{V}(S_n)$.

e) Calculer $\mathbb{V}(S_n)$. En déduire que $\psi_n(T_n)$ est un estimateur convergent de N.

20. Soit, pour $n \in \mathbb{N}^*$, un estimateur sans biais R_n du paramètre N. On pose, pour tout $k \in [\![1, N]\!]$, $f_n(k) = \mathbb{E}(R_n \mid [T_n = k])$.

a) En utilisant une méthode analogue à celle de la question 19.d, montrer que $\mathbb{V}(f_n(T_n)) \leqslant \mathbb{V}(R_n)$.

b) Soit F une fonction réelle. Montrer que, pour n fixé dans $\mathbb{N}^*$, la condition « pour tout $N \in \mathbb{N}^*$, $\mathbb{E}(F(T_n)) = N$ » est vérifiée si et seulement si, pour tout $k \in [\![1, N]\!]$, on a : $F(k) = \psi_n(k)$.

c) En déduire que, dans l'ensemble des estimateurs sans biais de N, l'estimateur $\psi_n(T_n)$ est optimal, dans le sens où $\mathbb{V}(\psi_n(T_n))$ est minimale.

La partie IV constitue une démonstration du théorème de Lehmann–Scheffé dans le cas particulier d'une loi uniforme sur $[\![1, N]\!]$, avec N inconnu.