



PRÉPASIMPLE

prepasimple.fr

HEC — 2009

Sujet de mathématiques approfondies

Mathématiques approfondies — Prépa ECG

Annales de concours · Sujet

Dans tout le problème, N désigne un entier supérieur ou égal à 1.

On note $E(X)$ et $V(X)$ respectivement l'espérance et la variance, lorsqu'elles existent, de toute variable aléatoire réelle X définie sur un espace probabilisé.

Soit $(U_n)_{n \geq 1}$ une suite de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, mutuellement indépendantes et de même loi uniforme discrète sur $\llbracket 1, N \rrbracket$.

Pour tout $n \in \mathbb{N}^*$, on pose $T_n = \sup(U_1, \dots, U_n)$ et $Z_n = \inf(U_1, \dots, U_n)$. On admet que T_n et Z_n sont des variables aléatoires définies sur $(\Omega, \mathcal{A}, P)$. Ainsi, pour tout $\omega \in \Omega$:

$$T_n(\omega) = \max(U_1(\omega), \dots, U_n(\omega)), \quad Z_n(\omega) = \min(U_1(\omega), \dots, U_n(\omega)).$$

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 :

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

Préliminaire

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

$$E(Y) = \sum_{k=0}^{N-1} P(Y > k), \quad E(Y^2) = \sum_{k=0}^{N-1} (2k+1)P(Y > k).$$

Partie I. Inf et Sup

2. Rappeler, sans démonstration, les valeurs respectives de $E(U_1)$ et de $V(U_1)$.

3. (a) Calculer, pour tout $k \in \llbracket 1, N \rrbracket$, $P(T_n \leq k)$.

3. (b) En déduire la loi de probabilité de T_n .

4. (a) Montrer que la suite $(d_n(N))_{n \geq 1}$ est convergente et calculer sa limite.

4. (b) Exprimer $E(T_n)$ en fonction de N et de $d_n(N)$. En déduire la valeur de $\lim_{n \rightarrow +\infty} E(T_n)$.

4. (c) Établir la formule suivante :

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

En déduire la valeur de $\lim_{n \rightarrow +\infty} V(T_n)$.

4. (d) Montrer que, si $N \geq 2$, on a :

$$\lim_{n \rightarrow +\infty} \frac{d_{n+1}(N)}{d_n(N)} = 1 - \frac{1}{N}.$$

En déduire que, lorsque n tend vers $+\infty$, $V(T_n) \sim d_n(N)$.

5. Déterminer la loi de Z_n . Calculer $E(Z_n)$ et $V(Z_n)$.

6. La commande Python `rd.randint(1, N + 1)`, après `import numpy.random as rd`, simule une variable aléatoire suivant la loi uniforme sur $\llbracket 1, N \rrbracket$; sa borne supérieure est exclue.

Écrire une fonction Python `simulmax(n, N)` qui simule la variable aléatoire T_n .

Partie II. Couple (Inf, Sup)

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

$$\phi_n(k, \ell) = P([T_n \leq k] \cap [Z_n \leq \ell]).$$

7. (a) Montrer, pour tout $(k, \ell) \in \llbracket 1, N \rrbracket^2$, la relation suivante :

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

7. (b) Établir, pour tout $(k, \ell) \in \llbracket 1, N \rrbracket^2$, la formule suivante :

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

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

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

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

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

8. (a) En déduire, pour tout $n \in \mathbb{N}^*$, la formule suivante :

$$E(T_n Z_n) = N(1 + d_{n+1}(N)).$$

8. (b) Pour tout $n \in \mathbb{N}^*$, on note ρ_n le coefficient de corrélation linéaire entre T_n et Z_n .
Calculer $\lim_{n \rightarrow +\infty} \rho_n$ lorsque $N \geq 2$.
9. (a) Pour tout $n \in \mathbb{N}^*$ et tout couple $(k, \ell) \in \llbracket 1, N \rrbracket^2$, calculer la probabilité conditionnelle $P_{[T_n=k]}(Z_n = \ell)$.
9. (b) En déduire, pour tout $n \in \mathbb{N}^*$ et tout $k \in \llbracket 1, N \rrbracket$, l'expression de l'espérance conditionnelle $E(Z_n \mid T_n = k)$.

Partie III. Prévision

Pour $n \in \mathbb{N}^*$, on dispose d'un $(n+1)$ -échantillon indépendant et identiquement distribué $(U_1, \dots, U_{n+1})$ de la loi uniforme sur $\llbracket 1, N \rrbracket$.

On pose $T_n = \sup(U_1, \dots, U_n)$ et $T_{n+1} = \sup(U_1, \dots, U_{n+1}) = \sup(T_n, U_{n+1})$.

Pour tout $t = (t_1, \dots, t_N) \in \mathbb{R}^N$, on pose :

$$W_t(T_n) = \sum_{k=1}^N t_k \mathbf{1}_{[T_n=k]}.$$

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

- $E(W_t(T_n)) = E(T_{n+1})$;
- $E((T_{n+1} - W_t(T_n))^2)$ est minimale.

10. Montrer, pour tout $k \in \llbracket 1, N \rrbracket$, la relation :

$$P(W_t(T_n) = t_k) = \sum_{\substack{1 \leq \ell \leq N \\ t_\ell = t_k}} P(T_n = \ell).$$

En déduire que, lorsque les t_k sont deux à deux distincts, $P(W_t(T_n) = t_k) = P(T_n = k)$.

Établir également, pour tout $t \in \mathbb{R}^N$, la formule :

$$E(W_t(T_n)) = \sum_{k=1}^N t_k P(T_n = k).$$

11. Établir, pour tout $k \in \llbracket 1, N \rrbracket$, la formule suivante :

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

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

12. (b) En déduire, pour tout couple $(k, j) \in \llbracket 1, N \rrbracket^2$, la probabilité conditionnelle $P_{[T_n=k]}(T_{n+1} = j)$.

12. (c) Déterminer, pour tout $k \in \llbracket 1, N \rrbracket$, l'expression de l'espérance conditionnelle $E(T_{n+1} \mid T_n = k)$.

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

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

13. Établir l'égalité suivante :

$$(W_t(T_n))^2 = \sum_{k=1}^N t_k^2 \mathbf{1}_{[T_n=k]}.$$

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

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

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

14. (b) Montrer que g admet un minimum global sur $\mathbb{R}^N$, atteint en un point $\theta = (\theta_1, \dots, \theta_N)$ que l'on déterminera en fonction des $E(T_{n+1} \mid T_n = k)$ pour $k \in \llbracket 1, N \rrbracket$.

15. Établir les deux relations suivantes :

$$E(W_\theta(T_n)) = E(T_{n+1}), \quad V(W_\theta(T_n)) \leq V(T_{n+1}).$$

16. (a) Établir, pour tout $i \in \mathbb{N}^*$, l'égalité suivante :

$$\sum_{k=1}^N k^i \mathbf{1}_{[T_n=k]} = T_n^i.$$

16. (b) En déduire la relation suivante :

$$W_\theta(T_n) = \frac{N+1}{2} + \frac{1}{2N}(T_n^2 - T_n).$$

Partie IV. Estimation

Soit U une variable aléatoire de loi uniforme discrète sur $\llbracket 1, N \rrbracket$. 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 $n \geq 1$, soit $(U_1, \dots, U_n)$ un n -échantillon indépendant et identiquement distribué de la loi de U . Les estimateurs considérés sont des fonctions réelles de cet échantillon, dont l'expression ne dépend pas du paramètre inconnu N .

17. Soit $\varepsilon > 0$. On pose :

$$A_n(\varepsilon) = [|T_n - N| \geq \varepsilon], \quad B_n(\varepsilon) = [|T_n - E(T_n)| + |d_n(N)| \geq \varepsilon].$$

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

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

17. (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 :

$$B_n(\varepsilon) \subset [|T_n - E(T_n)| \geq \varepsilon/2].$$

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

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

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

$$P_{[T_n=k]} \left(\bigcap_{i=1}^n [U_i = u_i] \right) = \begin{cases} \frac{1}{k^n - (k-1)^n} & \text{si tous les } u_i \in \llbracket 1, k \rrbracket \text{ et } \max_{1 \leq i \leq n} u_i = k, \\ 0 & \text{sinon.} \end{cases}$$

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 $k \in \mathbb{N}^*$:

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

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

19. (b) Établir, pour tout $k \in \llbracket 1, N \rrbracket$, l'égalité :

$$\psi_n(k) = E(S_n \mid T_n = k).$$

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

19. (d) Pour tout $k \in \llbracket 1, N \rrbracket$, on pose $\varphi_n(k) = E(S_n^2 \mid T_n = k)$.

Établir, pour tout $k \in \llbracket 1, N \rrbracket$, l'inégalité $\psi_n(k)^2 \leq \varphi_n(k)$. On pourra utiliser la fonction définie sur $\mathbb{R}$ par $\lambda \mapsto E((S_n - \lambda)^2 \mid T_n = k)$.

En déduire que $V(\psi_n(T_n)) \leq V(S_n)$.

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

20. Pour $n \in \mathbb{N}^*$, soit R_n un estimateur sans biais du paramètre N . Pour tout $k \in \llbracket 1, N \rrbracket$, on pose $f_n(k) = E(R_n \mid T_n = k)$.

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

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

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