



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 $\llbracket 1, N \rrbracket$.

On pose, pour tout n de $\mathbb{N}^*$: $T_n = \sup(U_1, U_2, \dots, U_n)$ et $Z_n = \inf(U_1, U_2, \dots, U_n)$. On admet que T_n et Z_n sont des variables aléatoires définies sur $(\Omega, \mathcal{A}, P)$. Ainsi, pour tout ω de Ω , on a :

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

On rappelle que si C désigne un élément de $\mathcal{A}$, on note $\mathbf{1}_C$ la variable aléatoire indicatrice de l'événement C , définie sur $(\Omega, \mathcal{A}, P)$ par :

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

On pose, pour tout n de $\mathbb{N}^*$:

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

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 :

$$E(Y) = \sum_{k=0}^{N-1} P([Y > k]) \quad \text{et} \quad E(Y^2) = \sum_{k=0}^{N-1} (2k+1) P([Y > k])$$

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 de $\llbracket 1, N \rrbracket$, $P([T_n \leq k])$.

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.

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

c) Établir la formule suivante : $V(T_n) = (2N-1)d_n(N) - 2Nd_{n+1}(N) - d_n^2(N)$.

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

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 $+\infty$, on a : $V(T_n) \sim d_n(N)$.

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

6) On rappelle qu'après l'instruction `import numpy.random as rd`, l'appel `rd.randint(a, b)`, où a et b sont deux entiers tels que $a < b$, renvoie un entier choisi au hasard selon la loi uniforme sur $\llbracket a, b-1 \rrbracket$. Écrire une fonction Python d'en-tête `def simulmax(n, N)` : qui simule la variable aléatoire T_n .

Partie II : Couple (Inf, Sup)

7) On pose, pour tout n de $\mathbb{N}^*$ et pour tout couple (k, ℓ) de $\mathbb{N}^2$:

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

a) Montrer, pour tout (k, ℓ) de $\llbracket 1, N \rrbracket^2$, la relation suivante :

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

b) Établir, pour tout (k, ℓ) de $\llbracket 1, N \rrbracket^2$, la formule suivante :

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

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 ℓ .

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

• i)

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

• ii)

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

- a) En déduire, pour tout n de $\mathbb{N}^*$, la formule suivante : $E(T_n Z_n) = N(1 + d_{n+1}(N))$.
- b) On note, pour tout n de $\mathbb{N}^*$, 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 de $\mathbb{N}^*$ et pour tout couple (k, ℓ) de $\llbracket 1, N \rrbracket^2$, calculer la probabilité conditionnelle $P_{[T_n=k]}([Z_n = \ell])$.
- b) En déduire, pour tout n de $\mathbb{N}^*$ et pour tout k de $\llbracket 1, N \rrbracket$, l'expression de l'espérance conditionnelle $E(Z_n/[T_n = k])$ de Z_n sachant $[T_n = k]$.

Partie III : Prévision

Pour n entier de $\mathbb{N}^*$, on dispose d'un $(n+1)$ -échantillon indépendant identiquement distribué (i.i.d.) $(U_1, U_2, \dots, U_{n+1})$ de la loi uniforme sur $\llbracket 1, N \rrbracket$.

On pose : $T_n = \sup(U_1, U_2, \dots, U_n)$ et $T_{n+1} = \sup(U_1, U_2, \dots, U_{n+1}) = \sup(T_n, U_{n+1})$.

Pour tout $t = (t_1, t_2, \dots, t_N)$ de $\mathbb{R}^N$, on pose :

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

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

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

- 10) On suppose dans cette question que les réels $t_1, t_2, \dots, t_N$ sont deux à deux distincts.
Montrer, pour tout k de $\llbracket 1, N \rrbracket$, la relation : $P([W_t(T_n) = t_k]) = P([T_n = k])$.

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

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

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

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

c) Déterminer, pour tout k de $\llbracket 1, N \rrbracket$, l'expression de l'espérance conditionnelle $E(T_{n+1}/[T_n = k])$ 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 :

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

13) Établir l'égalité suivante :

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

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

$$g(t_1, t_2, \dots, t_N) = E[(T_{n+1} - W_t(T_n))^2]$$

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

- b) Montrer que g admet un minimum global sur $\mathbb{R}^N$ atteint en un point $\theta = (\theta_1, \theta_2, \dots, \theta_N)$ que l'on déterminera en fonction de

$$E(T_{n+1}/[T_n = 1]), E(T_{n+1}/[T_n = 2]), \dots, E(T_{n+1}/[T_n = N]).$$

- 15) Établir les deux relations suivantes :

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

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

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

- 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 définie sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, 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 entier supérieur ou égal à 1, soit $(U_1, U_2, \dots, U_n)$ un n -échantillon i.i.d. de la loi de U .

- 17) Soit ε un réel strictement positif. On pose :

$$A_n(\varepsilon) = [|T_n - N| \geq \varepsilon] \quad \text{et} \quad B_n(\varepsilon) = [|T_n - E(T_n)| + |d_n(N)| \geq \varepsilon]$$

- a) Peut-on dire que $T_n + d_n(N)$ est un estimateur sans biais de N ?
- b) Montrer que la suite $(T_n)_{n \geq 1}$ est une suite d'estimateurs asymptotiquement sans biais du paramètre N , c'est-à-dire que $\lim_{n \rightarrow +\infty} E(T_n) = N$.
- 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 : $B_n(\varepsilon) \subset [|T_n - E(T_n)| \geq \varepsilon/2]$.
- d) En déduire que la suite d'estimateurs $(T_n)_{n \geq 1}$ est convergente.

- 18) a) Calculer, pour tout n -uplet $(u_1, u_2, \dots, u_n)$ de $\llbracket 1, N \rrbracket^n$, $P\left(\bigcap_{i=1}^n [U_i = u_i]\right)$.
- b) En déduire que, pour tout k de $\llbracket 1, N \rrbracket$, la loi conditionnelle du vecteur aléatoire $(U_1, U_2, \dots, U_n)$ sachant $[T_n = k]$ est donnée par :

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

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

- 19) On pose, pour n entier de $\mathbb{N}^*$: $S_n = T_n + Z_n - 1$ et, pour tout k de $\llbracket 1, N \rrbracket$:

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

- a) Montrer que S_n est un estimateur sans biais de N .
- b) Établir, pour tout k de $\llbracket 1, N \rrbracket$, l'égalité : $\psi_n(k) = E(S_n/[T_n = k])$.

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

d) On pose, pour tout k de $\llbracket 1, N \rrbracket$: $\varphi_n(k) = E(S_n^2/[T_n = k])$.

Établir, pour tout k de $\llbracket 1, N \rrbracket$, l'inégalité : $\psi_n^2(k) \leq \varphi_n(k)$ (on pourra utiliser la fonction définie sur $\mathbb{R}$ par : $\lambda \mapsto E((S_n - \lambda)^2/[T_n = k])$). En déduire que $V(\psi_n(T_n)) \leq V(S_n)$.

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

20) Soit, pour n entier de $\mathbb{N}^*$, un estimateur sans biais R_n du paramètre N .

On pose, pour tout k de $\llbracket 1, N \rrbracket$: $f_n(k) = E(R_n/[T_n = k])$.

a) En utilisant une méthode analogue à celle de la question 19) d), montrer que :

$$V(f_n(T_n)) \leq V(R_n).$$

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