

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 deux variables aléatoires définies sur $(\Omega, \mathcal{A}, P)$. Ainsi, pour tout $\omega \in \Omega$, on a :

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

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

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

$$\text{On pose, pour tout } n \text{ de } \mathbb{N}^* : d_n(N) = \begin{cases} \sum_{k=1}^{N-1} \left(\frac{k}{N}\right)^n & \text{si } N \geq 2 \\ 0 & \text{si } N = 1 \end{cases}$$

Préliminaire

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

(a) Justifier que, pour tout k de $\llbracket 0; N-1 \rrbracket$, $P(Y > k) = \sum_{i=k+1}^N P(Y = i)$.

(b) 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 de $d_n(N)$. 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) Montre 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 que la fonction Scilab **grand(1,1,'uin',1,N)** permet de simuler une variable aléatoire suivant la loi uniforme sur $\llbracket 1, N \rrbracket$. Écrire une fonction Scilab d'en-tête **function T=simulmax(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 $(k, \ell) \in \llbracket 1, N \rrbracket^2$, la relation suivante :

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

(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 :

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

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

(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. Préviation

Pour n entier de $\mathbb{N}^*$, on dispose d'un $n+1$ variables $(U_1, U_2, \dots, U_{n+1})$ indépendantes et toutes de la loi uniforme sur $\llbracket 1, N \rrbracket$. On pose :

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

Pour tout $t = (t_1, t_2, \dots, t_N)$ de $\mathbb{R}^N$, on pose : $W_t(T_n) = \sum_{k=1}^N t_k \times \mathbb{1}_{[T_n=k]}$.

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. Montrer, pour tout k de $[[1, N]]$, la relation : $P([W_t(T_n) = t_k]) = P([T_n = k])$.

11. Etablir, pour tout k de $[[1, N]]$, la formule suivante :

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

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

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

(c) Déterminer, pour tout k de $[[1, N]]$, 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. Etablir l'égalité suivante : $(W_t(T_n))^2 = \sum_{k=1}^N t_k^2 \times \mathbb{1}_{[T_n=k]}$.

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

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

Pour tout k de $[[1; N]]$, on note $\theta_k = E(T_{n+1} / [T_n = k])$, et $\theta = (\theta_1, \dots, \theta_N) \in \mathbb{R}^N$.

(a) A l'aide des résultats des questions 11, 12 et 13, montrer que :

$$g(t) = \sum_{k=1}^N (t_k^2 - 2\theta_k t_k + E(T_{n+1}^2)) P(T_n = k), \text{ puis que :}$$

$$g(t) = \sum_{k=1}^N (t_k - \theta_k)^2 P(T_n = k) + \sum_{k=1}^N (E(T_{n+1}^2) - \theta_k^2) P(T_n = k).$$

(b) Montrer que g admet un minimum global sur $\mathbb{R}^N$ atteint au point θ .

15. Etablir 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) Etablir, pour tout i de $\mathbb{N}^*$, l'égalité suivante : $\sum_{k=1}^N k^i \times \mathbb{1}_{[T_n=k]} = (T_n)^i$.

(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 $[[1, N]]$. On suppose que le paramètre N est inconnu.

Pour n entier supérieur ou égal à 1, soit $(U_1, U_2, \dots, U_n)$ n variables indépendantes toutes de même loi que U .

Dans cette partie, on s'intéresse aux variables aléatoires définies à l'aide de T_n qui permettent d'estimer N .

17. (a) Montrer que la suite $(E(T_n))_{n \geq 1}$ converge vers N .

(b) Montrer que la suite $(V(T_n))_{n \geq 1}$ converge vers 0.

(c) Pourquoi peut-on dire que, lorsque n devient grand, T_n fournit une estimation de N ?

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

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

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

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

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

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

(a) Montrer que, pour tout n de $\mathbb{N}^*$, $E(S_n) = N$.

(b) Etablir, pour tout k de $[[1, N]]$, l'égalité : $\psi_n(k) = E(S_n / [T_n = k])$.

(c) En déduire que, pour tout n de $\mathbb{N}^*$, $E(\psi_n(T_n)) = N$.

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

Établir, pour tout k de $[[1, N]]$, 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) En déduire que $V(\psi_n(T_n))$ tend vers 0 lorsque n tend vers $+\infty$.

20. Soit, pour n entier de $\mathbb{N}^*$, une variable aléatoire R_n d'espérance N .

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

(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. 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 $[[1, N]]$, on a : $F(k) = \psi_n(k)$.

(c) En déduire que, parmi les variables d'espérance N , $\psi_n(T_n)$ est optimale, dans le sens où $V(\psi_n(T_n))$ est minimale.

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