 Samedi 27 mars 2010.

Devoir Surveillé n° 8 –4 heuresx
Épreuve de type « Parisiennes »

Les élèves ont le choix entre l'épreuve de type « Parisiennes » et l'épreuve de type « Ecricome ». Ils ne devront traiter qu'une des deux épreuves.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté, la précision et la concision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

L'usage de tout document et de tout matériel électronique est interdit. Notamment, les téléphones portables doivent être éteints *et* rangés.

Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il le signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre

Problème – ***(HEC-ESSEC-EAP 2009, Maths 2)***

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 $(U_n)_{n\geqslant 1}$ une suite de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, mutuellement indépendantes, de même loi uniforme discrète sur $[\![1, N]\!]$.
On pose, pour tout n de $\mathbb{N}^*$: $T_n = \sup(U_1, U_2, \ldots, U_n)$ et $Z_n = \inf(U_1, U_2, \ldots, U_n)$. On admet que T_n et Z_n sont des variables aléatoires définies sur $(\Omega, \mathcal{T}, P)$. Ainsi, pour tout ω de Ω, on a :

$$T_n(\omega) = \max(U_1(\omega), U_2(\omega), \ldots, U_n(\omega)) \qquad et \qquad Z_n(\omega) = \min(U_1(\omega), U_2(\omega), \ldots, U_n(\omega)).$$

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 :

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

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

Préliminaire

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

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

Partie I – Inf et Sup

1. Rappeler, sans démonstration, les valeurs respectives de $E(U_1)$ et de $V(U_1)$.
2. (a) Calculer, pour tout k de $[\![1, N]\!]$, $P([T_n \leqslant k])$.
   (b) En déduire la loi de probabilité de T_n.
3. (a) Montrer que la suite $(d_n(N))_{n \geqslant 1}$ est convergente et calculer sa limite.
   (b) Exprimer $E(T_n)$ en fonction de N et $d_n(N)$. En déduire la valeur de $\lim\limits_{n \to +\infty} E(T_n)$.
   (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 $\lim\limits_{n \to +\infty} V(T_n)$.
   (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 $+\infty$, on a : $V(T_n) \underset{+\infty}{\sim} d_n(N)$.
4. Déterminer la loi de Z_n. Calculer $E(Z_n)$ et $V(Z_n)$.
5. On rappelle que la fonction Pascal `random(N)` permet de simuler une variable aléatoire suivant la loi uniforme sur $[\![0, N-1]\!]$. Écrire une fonction Pascal d'en-tête `simulmax(n :integer) :integer` qui simule la variable aléatoire T_n.

Partie II – Couple (Inf , Sup)

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

$$\varphi_n(k, \ell) = P([T_n \leqslant k] \cap [Z_n \leqslant \ell]).$$

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

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

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

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

   (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.
2. On donne, pour tout couple (m, n) de $(\mathbb{N}^*)^2$, les deux relations suivantes :
   (i) $\sum\limits_{j=1}^{m} [(j+1)^n - 2j^n + (j-1)^n] = (m+1)^n - m^n - 1$;
   (ii) $\sum\limits_{j=1}^{m} j[(j+1)^n - 2j^n + (j-1)^n] = m(m+1)^n - (m+1)m^n$.
   (a) En déduire, pour tout n de $\mathbb{N}^*$, la formule suivante : $E(T_nZ_n) = N(1 + d_{n+1}(N))$.

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

3. (a) Pour tout n de $\mathbb{N}^*$, et pour tout couple (k,ℓ) de $[\![1,N]\!]^2$, 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 $[\![1,N]\!]$, l'expression de l'espérance conditionnelle $E(Z_n \mid [T_n=k])$ 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.)* $(U_1, U_2, \ldots, U_{n+1})$ *de la loi uniforme sur* $[\![1,N]\!]$.

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

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

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.*

1. Montrer, pour tout k de $[\![1,N]\!]$, la relation : $P([W_t(T_n) = t_k]) = P([T_k = k])$.
2. Établir, pour tout k de $[\![1,N]\!]$, la formule suivante :

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

3. (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 $E(T_{n+1} \mid [T_n=k])$ 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 :

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

4. Établir l'égalité suivante : $(W_t(T_n))^2 = \sum\limits_{k=1}^{N} t_k^2 \times \mathbf{1}_{[T_n=k]}$.
5. Soit g la fonction définie sur $\mathbb{R}^N$ à valeurs réelles par :

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

   (a) À l'aide des résultats des questions III-2,3,4, 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 $E(T_{n+1} \mid [T_n=1]), E(T_{n+1} \mid [T_n=2]), \ldots, E(T_{n+1} \mid [T_n=N])$.

6. Établir les deux relations suivantes :

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

7. (a) Établir, pour tout i de $\mathbb{N}^*$, l'égalité suivante : $\sum\limits_{k=1}^{N} k^i \times \mathbf{1}_{[T_n=k]} = (T_n)^i$.

   (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 $[\![1, N]\!]$. 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.

1. Soit ε un réel strictement positif. On pose :

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

   (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 :

$$B_n(\varepsilon) \subset \left[|T_n - E(T_n)| \geqslant \frac{\varepsilon}{2}\right].$$

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

2. (a) Calculer, pour tout n-uplet $(u_1, u_2, \ldots, u_n)$ de $[\![1, N]\!]^2$, $P\left(\bigcap\limits_{i=1}^{n} [U_i = u_i]\right)$.

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

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

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

3. On pose, pour n entier 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 S_n est un estimateur sans biais de N.

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

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

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

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

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

   (a) En utilisant une méthode analogue à celle de la question IV 3(d), montrer que : $V(f_n(T_n)) \leqslant V(R_n)$.

   (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 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.

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