PROBABILITÉS ET STATISTIQUES

COMMENTAIRES

Bien que les trois parties du problème puissent, pour l'essentiel, être traitées séparément, il est conseillé de les traiter dans l'ordre de l'énoncé.

NOTATIONS

Dans tout le problème, $P(\Theta)$ désigne la probabilité d'un événement Θ. Si (Ω, E, P) est un espace de probabilité, on dira que l'espace de probabilité (Ω', E', P') est un espace de probabilité *agrandi* à partir de (Ω, E, P), si $\Omega \subset \Omega'$, $E \subset E'$, et si la restriction $P'_{|E}$ de la probabilité P' à E coïncide avec P. On notera alors $P' = P$ par abus de langage.

On désignera respectivement par $E(X)$ et $V(X)$ l'espérance et la variance d'une variable aléatoire X. La loi de probabilité (ou la distribution) de X sera désignée par $L(X)$.

Le logarithme népérien de x est noté $\ln x$. On pose de même $\ln \ln x = \ln(\ln x)$.

PREMIÈRE PARTIE

Soient U et V deux variables aléatoires à valeurs dans $\mathbb{N} = \{0, 1, 2, \dots\}$. On définit la *distance en variation* entre les lois $L(U)$ de U et $L(V)$ de V par :

$$d(L(U), L(V)) = \sup_{D \subset \mathbb{N}} |P(U \in D) - P(V \in D)|$$

On note A, B, C la partition de $\mathbb{N}$ définie par :

$$A = \{k : k \in \mathbb{N},\ P(U = k) < P(V = k)\}$$
$$B = \{k : k \in \mathbb{N},\ P(U = k) = P(V = k)\}$$
$$C = \{k : k \in \mathbb{N},\ P(U = k) > P(V = k)\}$$

1° Établir les identités :

$$d(L(U), L(V)) = \frac{1}{2} \sum_{k=0}^{+\infty} |P(U = k) - P(V = k)| = 1 - \sum_{k=0}^{+\infty} \min\{P(U = k), P(V = k)\}$$

2° On suppose que U et V sont définies simultanément sur le même espace de probabilités. Montrer que :

$$d(L(U), L(V)) \leqslant P(U \neq V)$$

3° On pose :

$$p_{kk} = \min\{P(U = k), P(V = k)\}, \qquad k \in \mathbb{N},$$
$$p_{kl} = 0 \quad \text{si} \quad k \neq l, \quad \text{avec} \quad (k, l) \in (A \times \mathbb{N}) \cup (B \times \mathbb{N}) \cup (\mathbb{N} \times B) \cup (\mathbb{N} \times C)$$

a. Soient deux mesures de probabilités P′ sur $\Omega' = \{1, \dots, I\}$, et P″ sur $\Omega'' = \{1, \dots, J\}$. On pose :

$$P_i = P'(i), \quad Q_j = P''(j), \quad \text{et} \quad R_{ij} = P_i Q_j ,$$
$$1 \leqslant i \leqslant I, \quad 1 \leqslant j \leqslant J.$$

Vérifier que $R(i, j) = R_{ij}$ définit une mesure de probabilité sur l'ensemble produit $\Omega' \times \Omega''$, de marg P′ et P″.

b. Montrer qu'il est possible de définir p_{kl} pour $k \neq l$ avec $(k, l) \in (C \times A)$, de manière que :

— pour tout $k \in \mathbb{N}$, $\displaystyle\sum_{j=0}^{+\infty} p_{kj} = P(U = k)$,

— pour tout $l \in \mathbb{N}$, $\displaystyle\sum_{i=0}^{+\infty} p_{il} = P(V = l)$,

— pour tout $k \in \mathbb{N}$ et pour tout $l \in \mathbb{N}$, $p_{kl} \geqslant 0$.

c. En déduire qu'il existe toujours un espace de probabilité sur lequel U et V, de lois L (U) et L (V) données sont définies simultanément, de manière que :

$$d(L(U), L(V)) = P(U \neq V)$$

On dira d'une loi jointe de U et V satisfaisant l'égalité ci-dessus qu'elle définit un *couplage maximal de* U et V

4° On considère dans cette question le cas particulier où :

— la loi L (U) de U est une loi de Bernoulli :

$$P(U = 1) = 1 - P(U = 0) = p \in (0, 1)$$

où en général on désigne par (α, β) l'intervalle ouvert d'extrémités α et β.

— la loi L (V) de V est une loi de Poisson d'espérance $E(V) = p \in (0, 1)$:

$$P(V = r) = \frac{p^r}{r!} e^{-p}, \qquad r = 0, 1, 2, \dots$$

a. Évaluer $d(L(U), L(V))$ en fonction de p. On exprimera le résultat par une expression ne faisant intervenir ni sommation infinie, ni valeur absolue, et exprimée en fonction de p et e^{-p}.

b. Montrer qu'il est possible de définir simultanément U et V sur le même espace de probabilités, jointement à une variable aléatoire W, de manière que U et W soient indépendants et que V = WU. Déterminer la loi de W et prouver qu'une telle construction établit un couplage maximal de U et V.

5° On considère maintenant une suite $X_1, X_2, \dots, X_n$ de variables aléatoires indépendantes de Bernoulli, telles que, pour $i = 1, 2, \dots, n$,

$$P(X_i = 1) = 1 - P(X_i = 0) = p_i \in (0, 1)$$

On pose $S_n = X_1 + \dots + X_n$.

Parallèlement, on considère une suite $Y_1, Y_2, \dots, Y_n$ de variables aléatoires indépendantes de Poisson, telles que, pour $i = 1, 2, \dots, n$,

$$P(Y_i = r) = \frac{p_i^r}{r!} e^{-p_i}, \qquad r = 0, 1, 2, \dots$$

On pose $T_n = Y_1 + \dots + Y_n$.

a. Quelle est la distribution de T_n ?

b. Établir les inégalités (indication : on pourra utiliser la construction du 4°, *b*) :

$$d\left(\mathrm{L}(\mathrm{S}_n),\ \mathrm{L}(\mathrm{T}_n)\right) \leqslant 1 - \prod_{i=1}^{n} \left(1 - d\left(\mathrm{L}(\mathrm{X}_i),\ \mathrm{L}(\mathrm{Y}_i)\right)\right) \leqslant \sum_{i=1}^{n} p_i^2 .$$

DEUXIÈME PARTIE

Dans toute cette partie, on considère une suite infinie de variables aléatoires de Bernoulli indépendantes $\mathrm{X}_1, \mathrm{X}_2, \ldots$, telles que :

$$\mathrm{P}(\mathrm{X}_i = 1) = 1 - \mathrm{P}(\mathrm{X}_i = 0) = \frac{1}{i}, \qquad i = 1, 2, \ldots$$

On notera $\iota = \sqrt{-1}$ le nombre complexe de module 1 et d'argument $\frac{\pi}{2}$.

On pose $\mathrm{S}_0 = 0$, $\mathrm{S}_n = \mathrm{X}_1 + \ldots + \mathrm{X}_n$, $\quad n = 1, 2, \ldots,$

$$\mathrm{N}(m, n) = \mathrm{S}_n - \mathrm{S}_m, \quad 0 \leqslant m \leqslant n < \infty$$

On admettra que la limite :

$$\lim_{n \to \infty} \left(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} - \ln n\right) = \gamma$$

existe et est finie, ainsi que la formule :

$$\sum_{n=1}^{+\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

1° Montrer que :

$$\lim_{\mathrm{N} \to \infty} \left\{ \sup_{\mathrm{N} \leqslant m \leqslant n} \left| \mathrm{E}(\mathrm{N}(m, n)) - \ln \frac{n}{m} \right| \right\} = \lim_{\mathrm{N} \to \infty} \left\{ \sup_{\mathrm{N} \leqslant m \leqslant n} \left| \mathrm{V}(\mathrm{N}(m, n)) - \ln \frac{n}{m} \right| \right\} = 0$$

2° On désigne par $[u]$ la partie entière de u ($[u] \leqslant u < [u] + 1$).

On pose :

$$\mathrm{M}_\mathrm{T}(s, t) = \mathrm{N}([\mathrm{T}\, e^s], [\mathrm{T}\, e^t]), \quad 0 \leqslant s \leqslant t, \quad \mathrm{T} > 0$$

a. Évaluer la fonction caractéristique $\mathrm{E}(\exp(\iota\, u\, \mathrm{N}(m, n)))$ de $\mathrm{N}(m, n)$. En déduire :

$$\lim_{\mathrm{T} \to \infty} \mathrm{E}(\exp(\iota\, u\, \mathrm{M}_\mathrm{T}(s, t)))$$

b. Montrer que, pour tout k-uple $0 \leqslant s_1 < s_2 < \ldots < s_k$ fixé, la loi limite jointe de $\mathrm{M}_\mathrm{T}(s_1, s_2), \mathrm{M}_\mathrm{T}(s_2, s_3), \ldots, \mathrm{M}_\mathrm{T}(s_{k-1}, s_k)$ lorsque T tend vers l'infini est un produit de lois de Poisson indépendantes, dont on précisera les paramètres.

c. Montrer que (indication : on pourra utiliser les fonctions caractéristiques) :

$$\lim_{n \to \infty} \mathrm{P}\left(\frac{\mathrm{S}_n - \ln n}{\sqrt{\ln n}} < t\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-v^2/2}\, dv$$

d. Évaluer :

$$\lim_{n\to\infty} \{ \mathrm{E}(\mathrm{S}_n) - \ln n \}, \quad \text{et} \quad \lim_{n\to\infty} \{ \mathrm{V}(\mathrm{S}_n) - \ln n \}$$

3° *a.* Montrer, en se servant des résultats obtenus dans la première partie, qu'il existe une suite de variables aléatoires $Y_1, Y_2, \dots$ construites sur le même espace de probabilité que $X_1, X_2, \dots$, éventuellement agrandi, et telles que :

— pour tout $i = 1, 2, \dots$, Y_i suit une loi de Poisson d'espérance $1/i$;

— pour tout $i = 1, 2, \dots$, on a :

$$d(\mathrm{L}(X_i), \mathrm{L}(Y_i)) = \mathrm{P}(X_i \neq Y_i) \leqslant \frac{1}{i^2}$$

b. On pose $T_n = Y_1 + \dots + Y_n$. Quelle est la loi de T_n ?

Montrer que la limite :

$$\lim_{n\to\infty} \{ S_n - T_n \}$$

existe et est finie presque sûrement.

4° *a.* On considère trois variables aléatoires ξ, η et ζ, telles que :

— pour $r = 0, 1, 2, \dots$, $\mathrm{P}(\zeta = r) = \dfrac{\alpha^r}{r!} e^{-\alpha}$, où $\alpha > 0$ est un nombre fixé;

— on a $\xi \geqslant 0$, $\eta \geqslant 0$, $\xi + \eta = \zeta$, la loi conditionnelle de ξ et η sachant ζ étant donnée par :

$$\mathrm{P}(\xi = m, \eta = r - m \mid \zeta = r) = \binom{r}{m} \theta^m (1-\theta)^{r-m}, \quad m = 0, 1, \dots, r$$

où $\theta \in (0, 1)$ est un paramètre fixé.

Déterminer la loi jointe de ξ et η.

b. En déduire qu'il est possible de construire sur le même espace de probabilité les suites $X_1, X_2, \dots$, et $Y_1, Y_2, \dots$, ainsi qu'une suite de variables aléatoires indépendantes $Z_1, Z_2, \dots$, de même loi de Poisson d'espérance 1, de telle manière que :

— pour tout $n = 1, 2, \dots$, si $l(n) = \left[1 + \dfrac{1}{2} + \dots + \dfrac{1}{n}\right]$, alors :

$$\sum_{i=1}^{l(n)} Z_i \leqslant T_n \leqslant \sum_{i=1}^{l(n)+1} Z_i$$

— pour tout $n = 1, 2, \dots$, $T_n - \displaystyle\sum_{i=1}^{l(n)} Z_i$ suit une loi de Poisson d'espérance $1 + \dfrac{1}{2} + \dots + \dfrac{1}{n} - l(n)$, et est une variable aléatoire indépendante de $Y_{n+1}, Y_{n+2}, \dots$

c. Montrer que :

$$\lim_{n\to\infty} \frac{S_n}{\ln n} = 1 \quad \text{presque sûrement.}$$

TROISIÈME PARTIE

Dans cette partie, on considère une suite $\omega_1, \omega_2, \dots, \omega_n$ de variables aléatoires indépendantes et uniformément distribuées sur l'intervalle (0, 1).

1° Montrer qu'avec probabilité un, pour tout $n = 2, 3, \dots,$ $\omega_1, \dots, \omega_n$ sont distincts. En déduire que la statistique ordonnée :

$$\omega_{1,n} < \omega_{2,n} < \dots < \omega_{n,n}$$

obtenue en rangeant $\omega_1, \dots, \omega_n$ par ordre croissant, est définie de manière unique presque sûrement.

2° Soit $t \in (0, 1)$ fixé, et soit $D_n(t)$ le nombre de variables parmi $\omega_1, \dots, \omega_n$, inférieures ou égales à t.

a. Pour $i = 0, 1, \dots, n$, déterminer $P(D_n(t) = i)$.

b. En déduire, pour $i = 1, \dots, n$, $P(\omega_{i,n} \leqslant t)$.

c. Montrer que cette probabilité peut s'écrire sous la forme :

$$P(\omega_{i,n} \leqslant t) = \frac{1}{\beta(u, v)} \int_0^t x^{u-1} (1 - x)^{v-1}\, dx \qquad t \in (0, 1)$$

où :

$$\beta(u, v) = \int_0^1 x^{u-1} (1 - x)^{v-1}\, dx$$

et où u et v sont des paramètres dont on précisera les valeurs, en fonction de i et n. On admettra que, pour j, k entiers,

$$\beta(j, k) = (k - 1)!\,(j - 1)!\,/\,(k + j - 1)!.$$

3° *a.* Montrer que, pour tout $n = 1, 2, \dots,$ l'identité :

$$\omega_i = \omega_{r_{i,n},n} \qquad , \qquad i = 1, \dots, n$$

définit presque sûrement une permutation $r = \{r_{1,n}, \dots, r_{n,n}\}$ de $\{1, \dots, n\}$.

b. Vérifier que, pour $r = \{r_{1,n}, \dots, r_{n,n}\}$ donné, la correspondance entre $\{\omega_1, \dots, \omega_n\}$ et $\{\omega_{1,n}, \dots, \omega_{n,n}\}$ est biunivoque et de jacobien unité si $\omega_{1,n} < \dots < \omega_{n,n}$.

En déduire que la densité jointe de $\{\omega_{1,n}, \dots, \omega_{n,n}\}$ est $n!$ sur l'ensemble $\{(x_1, \dots, x_n) : 0 < x_1 < \dots < x_n < 1\}$.

c. En déduire que $r = \{r_{1,n}, \dots, r_{n,n}\}$ est indépendant de $\{\omega_{1,n}, \dots, \omega_{n,n}\}$ et équidistribué sur l'ensemble de toutes les permutations de $\{1, \dots, n\}$.

4° On pose désormais $R(n) = r_{n,n}$, c'est-à-dire $R(1) = 1$ et, pour $n \geqslant 2$:

$$R(n) = \left|\begin{array}{ll} 1 & \text{si } \omega_n < \omega_{1,n-1}, \\ i & \text{si } \omega_{i-1,n-1} < \omega_n < \omega_{i,n-1}, \qquad i = 2, 3, \dots, n-1, \\ n & \text{si } \omega_n > \omega_{n-1,n-1}. \end{array}\right.$$

a. Montrer que la suite $\{R(n), n \geqslant 1\}$ est définie presque sûrement.

b. Déterminer la probabilité conditionnelle $P(R(n) \leqslant i \mid \omega_{i,n-1} = t)$.

En déduire $P(R(n) \leqslant i)$, puis $P(R(n) = i)$, $i = 1, \dots, n$. Montrer que :

$$P(R(n) = n) = \frac{1}{n}, \qquad n = 1, 2, \dots$$

c. Montrer que la suite $\{ R(n), n \geqslant 1 \}$ est une suite de variables aléatoires indépendantes.

5° Pour $n = 2, 3, \ldots$, on pose $X_n = 1$ si $\omega_n \geqslant \max \{ \omega_1, \ldots, \omega_{n-1} \}$ et $X_n = 0$ autrement. On pose $X_1 =$ ‹
Lorsque $X_n = 1$, on dit que ω_n est un *record* (sous-entendu de la suite $\omega_1, \omega_2, \ldots$), et que l'indice n correspor dant est un *temps de record.*

a. Soit $N(n, kn) = \sum_{i=n+1}^{kn} X_i$ le nombre de records observés dans l'intervalle $i \in \{ n+1, \ldots, kn \}$. En se servan des résultats de la première et de la deuxième partie, évaluer la loi limite de $N(n, kn)$ lorsque n tend ver l'infini. En particulier, déterminer :

$$\lim_{n \to \infty} P(N(n, kn) = r), \qquad r = 0, 1, 2, \ldots$$

NOTA BENE : k est ici un nombre entier fixé.

b. Établir les inégalités, pour $0 < \theta < N$,

$$\frac{\theta^N}{N!} e^{-\theta} \leqslant \sum_{r=N}^{+\infty} \frac{\theta^r}{r!} e^{-\theta} \leqslant \frac{\theta^N}{N!} e^{-\theta} \left\{ \frac{1}{1 - \frac{\theta}{N}} \right\}$$

c. En déduire que, pour tout $\varepsilon > 0$, il existe presque sûrement un indice n_ε tel que $n \geqslant n_\varepsilon$ implique (pour $k \geqslant 2$, fixé à l'avance) :

$$N(n, kn) \leqslant \frac{(1+\varepsilon) \ln n}{\ln \ln n}.$$

Indications : On rappelle la formule de Stirling :

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2 \pi n}\, (1 + o(1)), \quad n \to \infty$$

On pourra faire usage des résultats du 3° de la deuxième partie.

6° On utilise dans cette question la suite $\{ Z_n, n \geqslant 1 \}$ construite au 4°, *b* de la deuxième partie, ainsi que les notations introduites dans cette question. On supposera par la suite que $k \geqslant 3$ est fixé.

a. On donne les évaluations numériques suivantes :

$$0{,}577 < \gamma < 0{,}578, \quad 1{,}098 < \ln 3 < 1{,}099, \quad 0{,}693 < \ln 2 < 0{,}694$$

Montrer que, pour tout $k \geqslant 3$,

$$1 \leqslant \liminf_{n \to \infty} \{ l(kn) - l(n) \} \leqslant \limsup_{n \to \infty} \{ l(kn) - l(n) \} \leqslant [\ln k] + 1$$

b. On pose, pour $K \geqslant 1$, $\tau_n = \sum_{i=n+1}^{n+K} Z_i$, $n = 0, 1, 2, \ldots$

Montrer que, indépendamment de $K \geqslant 1$ fixé,

$$\limsup_{n \to \infty} \left\{ \frac{\ln \ln n}{\ln n} \right\} \tau_n = 1 \text{ presque sûrement.}$$

c. En déduire que, pour tout $k \geqslant 3$ fixé,

$$\limsup_{n \to \infty} \left\{ \frac{\ln \ln n}{\ln n} \right\} N(n, kn) = 1 \text{ presque sûrement.}$$

d. Comment pourrait-on montrer que le résultat ci-dessus reste valable pour $k = 2$?