

PROBABILITES

Durée 3 heures. Les documents, téléphones ou calculatrices ne sont pas autorisés.

Exercice I (barème approximatif: 5 points)

Soit (X, Y) un couple de variables aléatoires réelles, de densité

$$p(x, y) = \frac{1}{2\sqrt{x}} \exp\{-y\} \mathbf{1}_D(x, y)$$

avec $D = \{(x, y) : 0 < x < y^2, y > 0\}$.

1. On pose $U = \sqrt{X}, V = Y - \sqrt{X}$. Calculer la densité du couple (U, V) .
2. Les variables X et Y sont-elles indépendantes?
Les variables U et V sont-elles indépendantes?
3. Quelles sont les lois des variables X, Y, U, V ?

Exercice II (7 points)

Soient X_1, X_2, X_3, X_4 des variables gaussiennes centrées réduites indépendantes. On pose

$$Y = X_1 X_2 + X_3 X_4, \quad Z = \frac{X_1 X_2 + X_3 X_4}{\sqrt{X_2^2 + X_4^2}}$$

1. Calculer $\mathbf{E}(f(Z)|X_2, X_4)$ pour f borélienne positive. En déduire la loi conditionnelle de Z sachant (X_2, X_4) , puis la loi de Z .
2. Calculer $\mathbf{E}(\exp\{itX_1 X_2\}|X_2)$. En déduire la fonction caractéristique de la variable $X_1 X_2$.
3. Vérifier que la fonction caractéristique de $Y = X_1 X_2 + X_3 X_4$ est

$$\Phi_Y(t) = (1 + t^2)^{-1} = \frac{1}{2} \left((1 - it)^{-1} + (1 + it)^{-1} \right)$$

4. Calculer la fonction caractéristique de la loi exponentielle de densité $f(x) = e^{-x} \mathbf{1}_{\mathbb{R}_+^*}(x)$.
En déduire la densité de Y .

Problème (10 points)

Soit $\theta > 0$ un paramètre. Des individus numérotés $1, 2, \dots$ arrivent successivement à la cantine, qui abrite une infinité de tables infiniment longues. Le premier individu s'assied à une table. Puis, lorsque le $k + 1$ -ème individu se présente ($k \geq 1$), il choisit au hasard l'un des k individus déjà attablés avec la probabilité $1/(k + \theta)$ et s'assied à la même table, ou occupe une nouvelle table avec la probabilité $\theta/(k + \theta)$, indépendamment de ce qui s'est passé avant.

On note K_n la variable aléatoire égale au nombre de tables occupées lorsque n individus ont pris place, et

$$p_{n,i} = \mathbf{P}(K_n = i), \quad 1 \leq i \leq n.$$

Partie A:

1. Montrer que $p_{n+1,1} = \frac{n!}{(n+\theta)(n-1+\theta)\dots(1+\theta)}$. Trouver une relation entre $p_{n+1,i}$, $p_{n,i}$ et $p_{n,i-1}$, pour $2 \leq i \leq n$.
2. En déduire une relation de récurrence satisfaite par le polynôme P_n

$$P_n(x) = \sum_{i=1}^n p_{n,i} x^i, \quad x \in \mathbb{R}.$$

En conclure que

$$P_n(x) = \frac{R_n(\theta x)}{R_n(\theta)}, \quad R_n(x) = \prod_{i=0}^{n-1} (x + i).$$

Quelle est la fonction génératrice de la variable K_n ?

3. Calculer $\mathbf{E}(K_n)$ et $\text{Var}(K_n)$ à l'aide de cette formule. (Laisser le résultat sous la forme d'une somme.)

Partie B:

1. Montrer, à partir de sa définition, que K_n s'écrit

$$K_n = \sum_{i=1}^n X_i \tag{1}$$

avec X_1, \dots, X_n des variables aléatoires de Bernoulli indépendantes dont on précisera les paramètres.

2. Retrouver, à l'aide de (1), les expressions obtenues précédemment pour $\mathbf{E}(K_n)$ et $\text{Var}(K_n)$.
3. Montrer que

$$\int_0^n \frac{\theta}{\theta + x} dx \leq \mathbf{E}(K_n) \leq 1 + \int_0^{n-1} \frac{\theta}{\theta + x} dx,$$

et en déduire un équivalent de $\mathbf{E}(K_n)$.

4. Montrer que $\text{Var}(K_n) \leq \mathbf{E}(K_n)$. En déduire que $\frac{K_n}{\ln n}$ converge dans L^2 et en probabilité vers θ quand $n \rightarrow \infty$.
5. Calculer la fonction caractéristique Φ_{K_n} de K_n . Trouver une suite $c_n \rightarrow \infty$ telle que $\lim_{n \rightarrow \infty} \log \Phi_{K_n - \mathbf{E}K_n}(t/c_n) = -t^2/2$. En déduire la limite en loi de $(K_n - \mathbf{E}K_n)/\sqrt{\log n}$.

Corrigé Succint

Exercice I

1. En suivant l'énoncé, on considère l'application $\Phi : (x, y) \mapsto (u, v)$, avec $u = \sqrt{x}$ et $v = y - \sqrt{x}$, définie sur D et à valeurs dans $]0, +\infty[^2$. Cette application est bijective, d'inverse $\Phi^{-1}(u, v) = (x, y)$ donnée par

$$x = u^2, \quad y = v + u.$$

On constate que Φ est indéfiniment différentiable ainsi que Φ^{-1} , et on calcule le déterminant jacobien

$$\text{Jac}(\Phi^{-1})(u, v) = \det \begin{pmatrix} 2u & 0 \\ 1 & 1 \end{pmatrix} = 2u > 0$$

D'après la formule de changement de variables différentiable, le couple (U, V) a pour densité

$$\begin{aligned} p_{U,V}(u, v) &= p_{X,Y}(\Phi^{-1}(u, v)) \times |\text{Jac}(\Phi^{-1})(u, v)| \times \mathbf{1}_{]0, \infty[^2}(u, v) \\ &= e^{-u} \mathbf{1}_{]0, \infty[}(u) \times e^{-v} \mathbf{1}_{]0, \infty[}(v) \end{aligned}$$

On reconnaît le produit de deux densités exponentielles de paramètre 1.

2. Les variables U et V sont donc indépendantes (et de loi exponentielle). Par contre, les variables X et Y ne sont pas indépendantes car la fonction indicatrice de D ne peut pas se mettre sous forme produit. (On a $X < Y^2$.)
3. On a déjà remarqué que U et V suivent la loi exponentielle de paramètre 1. On calcule la densité de Y par la formule des marginales,

$$p_Y(y) = \int p_{X,Y}(x, y) dx = e^{-y} \mathbf{1}_{y>0} \int_0^{y^2} \frac{1}{2\sqrt{x}} dx = e^{-y} \mathbf{1}_{y>0} \left[\sqrt{x} \right]_0^{y^2} = ye^{-y} \mathbf{1}_{y>0},$$

c'est-à-dire une loi gamma $\gamma(s = 2, \theta = 1)$. De même, la densité de X est

$$p_X(x) = \int p_{X,Y}(x, y) dy = \frac{1}{2\sqrt{x}} \mathbf{1}_{x>0} \left[-e^{-y} \right]_{\sqrt{x}}^{+\infty} = \frac{1}{2\sqrt{x}} e^{-\sqrt{x}} \mathbf{1}_{x>0}$$

Remarque: X a même loi que le carré d'une variable de loi exponentielle (puisque l'on a $X = U^2$).

Exercice II

1. Par indépendance de (X_1, X_3) et (X_2, X_4) , et par la méthode de calcul vue en cours, l'espérance conditionnelle est donnée par

$$\mathbf{E}(f(Z) | X_2, X_4) = \Psi(X_2, X_4), \quad \text{avec } \Psi(x_2, x_4) = \mathbf{E}f\left(\frac{x_2 X_1 + x_4 X_3}{\sqrt{x_2^2 + x_4^2}}\right)$$

pour $(x_2, x_4) \in \mathbb{R}^2 \setminus \{(0, 0)\}$. Puisque (X_1, X_3) est un vecteur gaussien, la variable $(x_2 X_1 + x_4 X_3) / (\sqrt{x_2^2 + x_4^2})$ est gaussienne; on voit que cette variable est centrée réduite. Donc,

$$\mathbf{E}(f(Z)|X_2, X_4) = (2\pi)^{-1/2} \int_{\mathbb{R}} f(x) \exp\{-x^2/2\} dx$$

pour toute f borélienne positive, et la loi conditionnelle de Z sachant (X_2, X_4) est gaussienne centrée réduite. Il en est de même pour la loi (non conditionnelle) de Z .

2. En utilisant cette fois l'indépendance de X_1 et X_2 ainsi que la valeur de la fonction caractéristique de la loi gaussienne centrée réduite, on calcule de même

$$\mathbf{E}(\exp\{itX_1X_2\}|X_2) = \exp\{-t^2X_2^2/2\},$$

puis, par conditionnement,

$$\begin{aligned} \Phi_{X_1X_2}(t) &= \mathbf{E}(\exp\{itX_1X_2\}) \\ &= \mathbf{E}(\mathbf{E}(\exp\{itX_1X_2\}|X_2)) \\ &= \mathbf{E} \exp\{-t^2X_2^2/2\} \\ &= (2\pi)^{-1/2} \int_{\mathbb{R}} \exp\{-(1+t^2)x^2/2\} dx \\ &= (1+t^2)^{-1/2}, \end{aligned}$$

d'après la normalisation gaussienne.

3. Par indépendance de (X_1, X_2) et (X_3, X_4) , on a

$$\begin{aligned} \Phi_{X_1X_2+X_3X_4}(t) &= \Phi_{X_1X_2}(t)\Phi_{X_3X_4}(t) \\ &= ((1+t^2)^{-1/2})^2 \\ &= (1+t^2)^{-1} \\ &= \frac{1}{2}((1-it)^{-1} + (1+it)^{-1}) \end{aligned}$$

par calcul direct.

4. On calcule facilement

$$\int e^{itx} \times e^{-x} \mathbf{1}_{\mathbb{R}_+^*}(x) dx = \int_0^{\infty} e^{-(1-it)x} dx = \left[\frac{e^{-(1-it)x}}{-(1-it)} \right]_0^{\infty} = (1-it)^{-1}$$

On remarque alors que

$$(1+it)^{-1} = \int e^{-itx} \times e^{-x} \mathbf{1}_{\mathbb{R}_+^*}(x) dx = \int e^{itx} \times e^x \mathbf{1}_{\mathbb{R}_-^*}(x) dx,$$

et ainsi,

$$\frac{1}{2} \int_{\mathbb{R}} e^{itx} \times e^{-|x|} dx = \frac{1}{2} ((1-it)^{-1} + (1+it)^{-1}) = \Phi_Y(t)$$

pour tout $t \in \mathbb{R}$. Par injectivité de la transformation de Fourier, cela implique que Y suit la loi de densité $q(x) = \frac{1}{2}e^{-|x|}$ sur \mathbb{R} , qu'on appelle loi exponentielle symétrique.

Problème

Partie A:

1. Puisque la probabilité pour que le $k+1$ -ème individu s'assoie à une table déjà occupée vaut $k/(k+\theta)$, on a

$$p_{n+1,1} = \frac{1}{1+\theta} \times \frac{2}{2+\theta} \times \dots \times \frac{n}{n+\theta}.$$

De même,

$$\begin{aligned} p_{n+1,i} &= \mathbf{P}(K_{n+1} = i) \\ &= \mathbf{P}(K_{n+1} = i | K_n = i) \mathbf{P}(K_n = i) + \mathbf{P}(K_{n+1} = i | K_n = i-1) \mathbf{P}(K_n = i-1) \\ &= \frac{n}{n+\theta} p_{n,i} + \frac{\theta}{n+\theta} p_{n,i-1} \end{aligned}$$

2. On a

$$\begin{aligned} P_{n+1}(x) &= \sum_{i=1}^{n+1} p_{n+1,i} x^i \\ &= p_{n+1,1} x + \sum_{i=2}^n p_{n+1,i} x^i + p_{n+1,n+1} x^{n+1} \end{aligned}$$

On applique alors la formule de récurrence de la question 1), et on utilise l'égalité $p_{n+1,n+1} = \theta^n / ((1+\theta) \dots (n+\theta))$, pour obtenir

$$P_{n+1}(x) = \frac{n+\theta x}{n+\theta} P_n(x).$$

Puisque $P_1(x) = x$, on en déduit que $P_n(x) = R_n(\theta x) / R_n(\theta)$. Par ailleurs, la fonction génératrice de K_n est

$$\mathbf{E}x^{K_n} = \sum_{i=1}^n \mathbf{P}(K_n = i) x^i = P_n(x), \quad x \in [0, 1].$$

3. Puisque P_n est la fonction génératrice de K_n et que $P_n(1) = 1$, on a

$$\mathbf{E}K_n = P'_n(1) = \frac{P'_n(1)}{P_n(1)} = (\ln P_n)'(1) = \left(\sum_{i=0}^{n-1} \ln(\theta x + i) \right)'(1) = \sum_{i=0}^{n-1} \frac{\theta}{\theta + i} \quad (2)$$

Semblablement, $\text{Var}(K_n) = P''_n(1) + P'_n(1) - P'_n(1)^2$, et en utilisant les dérivées logarithmiques,

$$P'_n(x) = \left(\sum_{i=0}^{n-1} \frac{\theta}{\theta x + i} \right) P_n(x),$$

donc

$$P''_n(1) = - \sum_{i=0}^{n-1} \left(\frac{\theta}{\theta + i} \right)^2 + \left(\sum_{i=0}^{n-1} \frac{\theta}{\theta + i} \right)^2,$$

et finalement

$$\text{Var}(K_n) = \sum_{i=0}^{n-1} \frac{\theta}{\theta+i} \left(1 - \frac{\theta}{\theta+i}\right) = \sum_{i=1}^{n-1} \frac{i\theta}{(\theta+i)^2} \quad (3)$$

Partie B:

1. Au convive numéro i on associe une variable X_i valant 1 ou 0 selon que i choisit une nouvelle table ou non. Ainsi, X_i est une variable de Bernoulli de paramètre

$$q_i := \frac{\theta}{\theta+i-1}.$$

On a $K_n = X_1 + \dots + X_n$, et les X_i sont indépendantes par hypothèse.

2. On a $\mathbf{E}K_n = \sum_{i=1}^n \mathbf{E}X_i$, et $\text{Var}(K_n) = \sum_{i=1}^n \text{Var}(X_i)$ par indépendance. Avec $\mathbf{E}X_i = q_i$ et $\text{Var}(X_i) = q_i(1 - q_i)$, on retrouve les résultats de (2) et (3).

3. Par monotonie de $x \mapsto \theta/(\theta+x)$, on a

$$\frac{\theta}{\theta+i} \leq \int_{i-1}^i \frac{\theta}{\theta+x} dx \leq \frac{\theta}{\theta+i-1},$$

ce qui donne l'encadrement voulu. Puisque

$$\int_0^n \frac{\theta}{\theta+x} dx = \theta [\ln(\theta+x)]_0^n = \theta \ln(1+n/\theta) \sim \theta \ln n$$

quand $n \rightarrow \infty$, on en déduit que

$$\mathbf{E}K_n \sim \theta \ln n, \quad n \rightarrow \infty.$$

4. Puisque $\text{Var}(X_i) = q_i(1 - q_i) \leq q_i = \mathbf{E}X_i$, on a $\text{Var}(K_n) \leq \mathbf{E}K_n$. Ainsi,

$$\begin{aligned} \left\| \frac{K_n}{\ln n} - \theta \right\|_2^2 &= \mathbf{E} \left(\frac{K_n}{\ln n} - \theta \right)^2 \\ &= \frac{\text{Var}(K_n)}{\ln^2 n} + \left(\frac{\mathbf{E}K_n}{\ln n} - \theta \right)^2 \\ &= \mathcal{O}(1/\ln n) + o(1) \end{aligned}$$

tend vers 0 quand n tend vers l'infini. Par conséquent, $K_n/\ln n \rightarrow \theta$ dans L^2 et donc aussi en probabilité.

5. Par indépendance des sommants dans la formule $K_n = \sum_{i=1}^n X_i$ et puisque les X_i sont de loi de Bernoulli, la fonction caractéristique de K_n est donnée par

$$\Phi_{K_n}(t) = \prod_{i=1}^n \Phi_{X_i}(t) = \prod_{i=1}^n [1 + q_i(e^{it} - 1)].$$

Puisque $c_n \rightarrow \infty$, on a les développements limités pour tout $t \in \mathbb{R}$,

$$\begin{aligned} 1 + q_i(e^{it/c_n} - 1) &= 1 + q_i \left(i \frac{t}{c_n} - \frac{t^2}{2c_n^2} + \mathcal{O}(c_n^{-3}) \right), \\ \log [1 + q_i(e^{it/c_n} - 1)] &= q_i \left(i \frac{t}{c_n} - \frac{t^2}{2c_n^2} (1 - q_i) + \mathcal{O}(c_n^{-3}) \right), \\ \log \Phi_{K_n - \mathbf{E}K_n}(t/c_n) &= -\frac{t^2}{2c_n^2} \sum_{i=1}^n q_i(1 - q_i) + \mathcal{O}(c_n^{-3}) \sum_{i=1}^n q_i. \end{aligned} \quad (4)$$

On est conduit à poser

$$c_n = \sqrt{\sum_{i=1}^n q_i(1 - q_i)} = \sqrt{\sum_{i=1}^{n-1} \frac{\theta i}{(\theta + i)^2}}.$$

Puisque $\sum_{i=1}^n q_i \sim \theta \ln n$ et $\sum_{i \geq 1} q_i^2 < \infty$, cette suite c_n convient. De plus, on a encore

$$c_n \sim \sqrt{\theta \ln n}, \quad (5)$$

quand $n \rightarrow \infty$. D'après (4) et (eq:333), il vient

$$\lim_{n \rightarrow \infty} \Phi_{\frac{K_n - \mathbf{E}K_n}{\sqrt{\log n}}}(t) = \exp -\theta t^2 / 2.$$

Comme la convergence des fonctions caractéristiques entraîne la convergence en loi des variables aléatoires, on obtient que

$$\frac{K_n - \mathbf{E}K_n}{\sqrt{\log n}} \longrightarrow Z \quad \text{en loi,}$$

avec Z une variable gaussienne $\mathcal{N}(0, \theta)$.

Commentaire: Ce modèle n'est pas seulement la description de l'occupation des tables à la cantine. En écologie et ou génétique, on s'intéresse au nombre et à l'abondance des espèces; la variable K_n apparaît naturellement pour le nombre d'espèces dans un échantillon de n individus.