Mathématiques

HEC-ESCP-EML 99 : corrigé de l'épreuve de Maths II (voie S)

Roger Cuculière,
Professeur en classe préparatoire, lycée Pasteur (Neuilly sur Seine), collaborateur de diverses revues (Quadrature, Tangente, Repères, Sciences et Vie Junior...), co-auteur de "Olympiades internationales de Mathématiques 1988-97", Ed. du choix.

Nota : l'énoncé du sujet n'est pas reproduit ici pour des raisons d'espace éditorial. Il est disponible dans les rapports de jurys transmis par les Écoles à tous les établissements.

On rappelle que la fonction Γ est définie, pour $x > 0$, par : $\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}\,dt$. Cette fonction vérifie : $\Gamma(x+1) = x\Gamma(x)$, pour tout $x > 0$. Pour $n \in \mathbb{N}^*$, on a : $\Gamma(n) = (n-1)!$.

Par définition, une variable aléatoire X suivant la loi $\Gamma(b,\tau)$, avec $b > 0$ et $\tau > 0$, a pour densité : $f(x) = \dfrac{e^{-x/b}x^{\tau-1}}{\Gamma(\tau)b^{\tau}}$ si $x > 0$ et $f(x) = 0$ si $x \leq 0$. On a : $E(X) = b\tau$ et $V(X) = b^2\tau$.

Partie 1

1. 1. Soit r un entier *strictement positif* (l'énoncé semble oublier qu'il existe des entiers négatifs). La loi du χ^2 à r degrés de liberté est la loi $\Gamma\left(2,\dfrac{r}{2}\right)$, qui admet donc pour densité la fonction : $f_r(x) = \dfrac{x^{(r/2)-1}e^{-x/2}}{\Gamma(r/2)2^{r/2}}$ si $x > 0$ et $f(x) = 0$ si $x \leq 0$. On a : $E(X) = r$ et $V(X) = 2r$.

1. 2. a. Pour $n = 1$, l'égalité proposée : $e^{\lambda} = \sum_{k=0}^{n-1}\dfrac{\lambda^k}{k!} + \int_0^{\lambda} e^{\lambda-t}\dfrac{t^{n-1}}{(n-1)!}dt$, s'écrit : $e^{\lambda} = 1 + \int_0^{\lambda} e^{\lambda-t}dt$, qui est vraie, puisque $\int_0^{\lambda} e^{\lambda-t}dt = e^{\lambda} = e^{\lambda}\int_0^{\lambda} e^{-t}dt = e^{\lambda}\left(-e^{-\lambda}+1\right)$.

Si l'on suppose que cette égalité proposée est vraie pour un certain $n \in \mathbb{N}^*$, on calcule l'intégrale $\int_0^{\lambda} e^{\lambda-t}\dfrac{t^{n-1}}{(n-1)!}dt$ au moyen d'une intégration par parties : $u = e^{\lambda-t}$, $v' = \dfrac{t^{n-1}}{(n-1)!}$, d'où : $u' = -e^{\lambda-t}$ et $v = \dfrac{t^n}{n!}$, et : $\int_0^{\lambda} e^{\lambda-t}\dfrac{t^{n-1}}{(n-1)!}dt = \dfrac{\lambda^n}{n!} + \int_0^{\lambda} e^{\lambda-t}\dfrac{t^n}{n!}dt$, et en reportant ceci dans l'égalité proposée, il vient :

$e^{\lambda} = \sum_{k=0}^{n} \frac{\lambda^k}{k!} + \int_0^{\lambda} e^{\lambda - t} \frac{t^n}{n!} dt$, ce qui prouve par récurrence ladite égalité.

1. 2. b. Si la variable aléatoire Y_λ suit la loi de Poisson $\boldsymbol{P}(\lambda)$, $\lambda > 0$, alors pour tout $k \in \mathbb{N}$, on a : $P(Y_\lambda = k) = e^{-\lambda} \frac{\lambda^k}{k!}$,

d'où, pour tout $n \in \mathbb{N}^*$: $P(Y_\lambda < n) = P(Y_\lambda \leq n-1) = e^{-\lambda} \sum_{k=0}^{n-1} \frac{\lambda^k}{k!}$.

Si la variable aléatoire X_{2n} suit la loi du χ^2 à $2n$ degrés de liberté, autrement dit la loi $\Gamma(2, n)$, alors pour $\lambda > 0$, on a :

$P(X_{2n} > 2\lambda) = 1 - P(X_{2n} \leq 2\lambda) = 1 - \int_0^{2\lambda} \frac{x^{n-1} e^{-x/2}}{2^n \Gamma(n)} dx$. Avec l'égalité : $\Gamma(n) = (n-1)!$ et le changement de variable : $t = \frac{x}{2}$,

il vient : $P(X_{2n} > 2\lambda) = 1 - \int_0^{\lambda} \frac{t^{n-1} e^{-t}}{(n-1)!} dt$.

En multipliant par $e^{-\lambda}$ les deux membres de l'égalité trouvée à la question 1. 2. a., on obtient :

$1 - \int_0^{\lambda} \frac{t^{n-1} e^{-t}}{(n-1)!} dt = e^{-\lambda} \sum_{k=0}^{n-1} \frac{\lambda^k}{k!}$, soit : $P(X_{2n} > 2\lambda) = P(Y_\lambda < n)$.

1. 2. c. La question 1. 2. b. assure que : $P(X_{2n} > x) = P(Y_{x/2} < n) = e^{-x/2} Q_{n-1}\left(\frac{x}{2}\right)$, où l'on définit : $Q_n(x) = \sum_{k=0}^{n} \frac{x^k}{k!}$. Le calcul de $Q_n(x)$ par le procédé de Thomas Horner se ferait naturellement en posant : $y_0 = \frac{1}{n!}$ et $y_{k+1} = y_k x + \frac{1}{(n-k-1)!}$, et donnerait : $y_n = Q_n(x)$. Mais il est très malcommode d'avoir à calculer successivement toutes ces factorielles, et cela comporte de grands risques d'erreurs d'arrondis. Alors, il est plus indiqué de poser : $z_k = (n-k)! y_k$. Cette suite z_k se définit par : $z_0 = 1$, $z_{k+1} = z_k \frac{x}{n-k} + 1$, et donne au bout du compte : $z_n = Q_n(x)$. Cet algorithme se traduit immédiatement par une fonction en langage Pascal :

```
(**************************************************************************)
FUNCTION PROB(Var n : integer ; x : real) : real ;

VAR z : real ;
    k : integer ;

BEGIN

  x := x/2 ; z :=1 ;
  for k:= 1 to n-1 do z := z*x/(n-k)+1 ;
  PROB := z*exp(-x)

END ;
(**************************************************************************)
```

1. 2. d. Pour $\lambda > 0$, on a : $F_6(2\lambda) = P(X_6 \leq 2\lambda) = 1 - P(X_6 > 2\lambda) = 1 - P(Y_\lambda < 3) = 1 - P(Y_\lambda \leq 2)$.

La table donnée par l'énoncé (et sur qui il faut lire Y_λ au lieu de X_λ) fournit alors les valeurs suivantes :

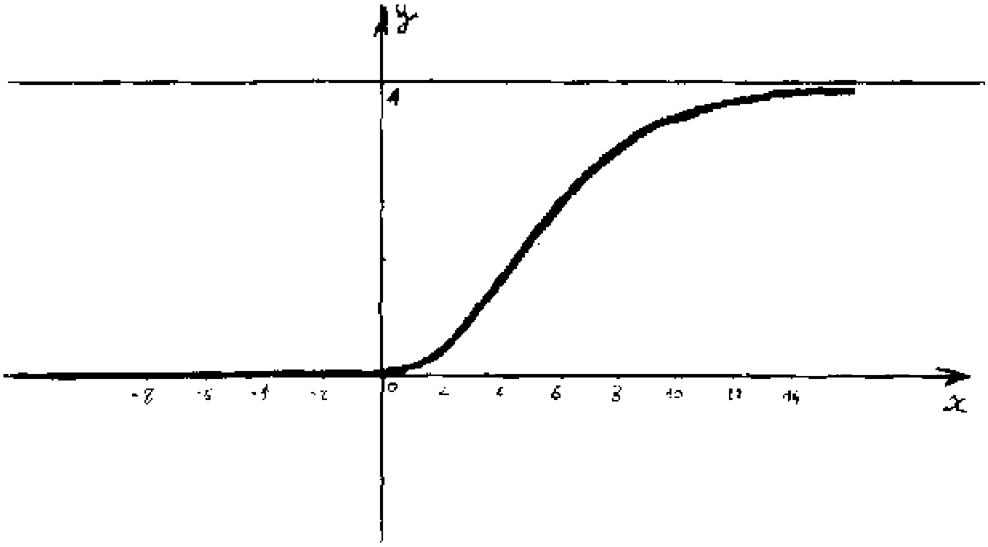
$F_6(2) = 1 - P(Y_1 \leq 2) \cong 0{,}0803$, $F_6(4) = 1 - P(Y_2 \leq 2) \cong 0{,}3233$, $F_6(6) = 1 - P(Y_3 \leq 2) \cong 0{,}5768$,

$F_6(8) = 1 - P(Y_4 \leq 2) \cong 0{,}7619$, $F_6(10) = 1 - P(Y_5 \leq 2) \cong 0{,}8753$, $F_6(12) = 1 - P(Y_6 \leq 2) \cong 0{,}9380$,

$F_6(14) = 1 - P(Y_7 \le 2) \cong 0{,}9704$.

A quoi il faut ajouter : $F_6(0) = P(X_6 \le 0) = 0$ de par la définition de la loi $\Gamma(b,\tau)$, et évidemment : $F_6(x) = 0$ pour tout $x \le 0$.

Graphe :



On sait que $F_6'(x) = f_6(x) = \dfrac{x^2 e^{-x/2}}{8\Gamma(3)}$ pour $x > 0$. Il en résulte que $F_6'(0) = 0$, et que le graphe de la fonction F_6 présente un point d'inflexion pour $x = 4$.

1. 3. a. Puisque la variable aléatoire X_1 suit une loi normale (gaussienne) centrée réduite, sa fonction de répartition est :

$P(X_1 \le x) = \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt = \Phi(x)$. La répartition de X_1^2 est donc, pour $x > 0$:

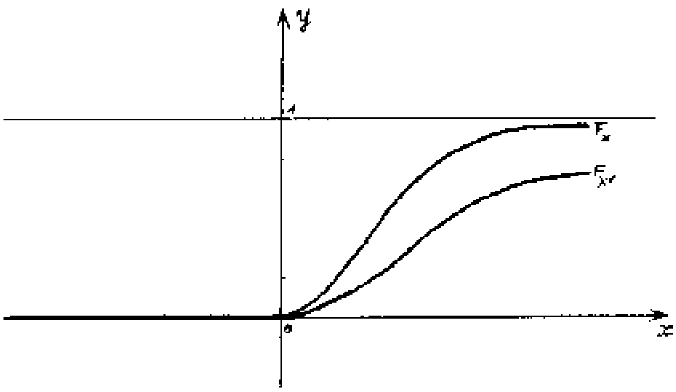
$F(x) = P\left(X_1^2 \le x\right) = P\left(-\sqrt{x} \le X_1 \le \sqrt{x}\right) = \Phi\left(\sqrt{x}\right) - \Phi\left(-\sqrt{x}\right)$, ce qui implique que, pour $x > 0$, la densité de X_1^2 est :

$f(x) = F'(x) = \dfrac{\Phi'\left(\sqrt{x}\right)}{2\sqrt{x}} + \dfrac{\Phi'\left(-\sqrt{x}\right)}{2\sqrt{x}} = \dfrac{1}{\sqrt{2\pi}} x^{-1/2} e^{-x/2}$, et bien sûr : $f(x) = 0$ si $x \le 0$.

Par ailleurs, la loi du χ^2 à 1 degré de liberté est la loi $\Gamma\left(2, \dfrac{1}{2}\right)$, qui admet pour densité la fonction : $f_1(x) = \dfrac{e^{-x/2} x^{-1/2}}{\Gamma(1/2)\sqrt{2}}$ pour $x > 0$, et $f(x) = 0$ si $x \le 0$. Il en résulte que : $\int_0^{+\infty} t^{-1/2} e^{-t/2} dt = \sqrt{2\pi} = \Gamma\left(\dfrac{1}{2}\right)\sqrt{2}$, et que la variable aléatoire X_1^2 suit la loi du χ^2 à 1 degré de liberté, c'est-à-dire la loi $\Gamma\left(2, \dfrac{1}{2}\right)$.

1. 3. b. Les variables X_1^2 , X_2^2 , ..., X_k^2 sont indépendantes et suivent chacune la loi $\Gamma\left(2, \dfrac{1}{2}\right)$. En vertu de la stabilité par la somme de la loi Γ, la somme $X_1^2 + X_2^2 + \ldots + X_k^2$ suit la loi $\Gamma\left(2, \dfrac{k}{2}\right)$, autrement dit la loi du χ^2 à k degrés de liberté.

1. 3. c. Soient deux entiers r et r' tels que : $0 < r < r'$, et soient X_1 , X_2 , ... , X_r , ... , $X_{r'}$ des variables indépendantes gaussiennes centrées réduites. La variable aléatoire $X = X_1^2 + X_2^2 + \ldots + X_r^2$ suit la loi du χ^2 à r degrés de liberté, et la variable aléatoire $X' = X_1^2 + X_2^2 + \ldots + X_r^2 + \ldots + X_{r'}^2$ suit la loi du χ^2 à r' degrés de liberté. On a toujours : $X' \ge X$, et l'événement $(X' \le x)$ implique donc l'événement $(X \le x)$. Il en résulte : $F_{X'}(x) = P(X' \le x) \le P(X \le x) = F_X(x)$. Le graphe de la fonction $x \mapsto F_{X'}(x)$ est situé au-dessous du graphe de la fonction $x \mapsto F_X(x)$, et même strictement au-dessous pour $x > 0$.



Partie 2

2. A. Etude des variables X_i.

2. A. 1. La loi de la variable aléatoire X_i est la loi binomiale $\boldsymbol{B}(n, p_i)$, son espérance est donc : $E(X_i) = np_i$, et sa variance : $V(X_i) = np_i(1-p_i)$.

2. A. 2. La loi de la variable aléatoire $X_i + X_j$, avec $i \neq j$, est la loi binomiale $\boldsymbol{B}(n, p_i + p_j)$, son espérance est donc : $n(p_i + p_j)$ et sa variance : $n(p_i + p_j)(1 - p_i - p_j)$.

On en déduit immédiatement : $\operatorname{cov}(X_i, X_j) = \frac{1}{2}\left(V(X_i + X_j) - V(X_i) - V(X_j)\right) =$

$$= \frac{1}{2}\left(n(p_i + p_j)(1 - p_i - p_j) - np_i(1 - p_i) - np_j(1 - p_j)\right) = -np_i p_j .$$

2. B. On suppose dans cette partie que $s = 2$.

2. B. 1. Dans ce cas, on a : $X_1 + X_2 = n$ et $p_1 + p_2 = 1$, ce qui conduit à :

$X_2 - np_2 = n - X_1 - n(1 - p_1) = -X_1 + np_1$, d'où :

$$U_n = \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_2 - np_2)^2}{np_2} = (X_1 - np_1)^2\left(\frac{1}{np_1} + \frac{1}{np_2}\right) = (X_1 - np_1)^2\,\frac{p_1 + p_2}{np_1 p_2} = \left(\frac{X_1 - np_1}{\sqrt{np_1 p_2}}\right)^2,$$ soit :

$U_n = Z_1^2$, où $Z_1 = \dfrac{X_1 - np_1}{\sqrt{np_1 p_2}} = \dfrac{X_1 - np_1}{\sqrt{np_1(1 - p_1)}}$.

2. B. 2. a. On se souvient que $E(X_1) = np_1$ et $V(X_1) = np_1(1 - p_1)$. D'après l'approximation usuelle de la loi binomiale par la loi normale, on peut donc approcher la loi de Z_1, lorsque n est grand, par la loi normale centrée réduite $\boldsymbol{N}(0,1)$.

2. B. 2. b. D'après le théorème de Moivre-Laplace, la variable aléatoire $Z_1 = \dfrac{X_1 - np_1}{\sqrt{np_1(1 - p_1)}}$ converge en loi, quand n tend vers $+\infty$, vers une variable aléatoire X qui suit la loi normale centrée réduite $\boldsymbol{N}(0,1)$. Il en résulte que $U_n = Z_1^2$ converge en loi vers X^2, qui suit la loi $\Gamma\left(2, \frac{1}{2}\right)$, autrement dit la loi du χ^2 à 1 degré de liberté (d'après la question 1. 3. a.). La suite (U_n) converge donc en loi vers la loi du χ^2 à 1 degré de liberté, quand n tend vers l'infini.

2. C. On suppose dans cette partie que $s = 3$ et que $p_1 = p_2 = \frac{1}{4}$ et $p_3 = \frac{1}{2}$.

2. C. 1. On a : $U_n = \dfrac{4}{n}\left[\left(X_1 - \dfrac{n}{4}\right)^2 + \left(X_2 - \dfrac{n}{4}\right)^2\right] + \dfrac{2}{n}\left(X_3 - \dfrac{n}{2}\right)^2$. Traitons le contenu du crochet :

$$\left(X_1-\frac{n}{4}\right)^2+\left(X_2-\frac{n}{4}\right)^2 = X_1^2+X_2^2-\frac{n}{2}X_1-\frac{n}{2}X_2+\frac{n^2}{8}$$

$$=\frac{1}{2}(X_1-X_2)^2+\frac{1}{2}\left(X_1^2+X_2^2-nX_1-nX_2+\frac{n^2}{4}+2X_1X_2\right)=\frac{1}{2}(X_1-X_2)^2+\frac{1}{2}\left(X_1+X_2-\frac{n}{2}\right)^2$$

$=\frac{1}{2}(X_1-X_2)^2+\frac{1}{2}\left(\frac{n}{2}-X_3\right)^2$. En reportant ceci dans l'expression précédente de U_n, il vient :

$U_n=\frac{2}{n}(X_1-X_2)^2+\frac{4}{n}\left(X_3-\frac{n}{2}\right)^2=Z_2^2+Z_1^2$. *CQFD.*

2. C. 2. Il est clair que : $E(X_1)=E(X_2)=\frac{n}{4}$, $E(X_3)=\frac{n}{2}$, $V(X_1)=V(X_2)=\frac{3n}{16}$, $V(X_3)=\frac{n}{4}$, et :

$\operatorname{cov}(X_1,X_2)=-\frac{n}{16}$, $\operatorname{cov}(X_1,X_3)=\operatorname{cov}(X_2,X_3)=-\frac{n}{8}$. On en déduit les espérances : $E(Z_1)=0$, $E(Z_2)=0$, les variances : $V(Z_1)=\frac{4}{n}V(X_3)=1$, $V(Z_2)=\frac{2}{n}\big(V(X_1)+V(X_2)-2\operatorname{cov}(X_1,X_2)\big)=1$, et la covariance :

$\operatorname{cov}(Z_1,Z_2)=\frac{2}{\sqrt{n}}\sqrt{\frac{2}{n}}\operatorname{cov}(X_3,X_1-X_2)=0$.

2. C. 3. On a : $Z_1=\frac{X_3-np_3}{\sqrt{np_3(1-p_3)}}$, comme à la question 2. B. 2. a. Lorsque n est grand, on peut donc approcher la loi de Z_3 par la loi normale centrée réduite $\boldsymbol{N}(0,1)$.

2. C. 4. a. Dans la somme : $\sum_{i=1}^{n}T_i$, le nombre de termes égaux à 1 est X_1, le nombre de termes égaux à –1 est X_2, et le nombre de termes égaux à 0 est X_3. Cette somme est donc égale à X_1-X_2. On a ainsi prouvé que :

$X_1-X_2=\sum_{i=1}^{n}T_i$.

2. C. 4. b. Les variables aléatoires T_i sont indépendantes et suivent la même loi, d'espérance $m=0$ et d'écart-type $\sigma=\frac{1}{\sqrt{2}}$.

L'espérance de la moyenne $\frac{1}{n}\sum_{i=1}^{n}T_i$ est : $m=0$, et son écart-type est : $\frac{\sigma}{\sqrt{n}}=\frac{1}{\sqrt{2n}}$. D'après le théorème de la limite centrée, $\dfrac{\frac{1}{n}\sum_{i=1}^{n}T_i-m}{\frac{\sigma}{\sqrt{n}}}$ converge en loi, quand n tend vers l'infini, vers une variable aléatoire qui suit la loi normale centrée réduite $\boldsymbol{N}(0,1)$. Or, $\dfrac{\frac{1}{n}\sum_{i=1}^{n}T_i-m}{\frac{\sigma}{\sqrt{n}}}=\sqrt{\frac{2}{n}}\sum_{i=1}^{n}T_i=\sqrt{\frac{2}{n}}(X_1-X_2)=Z_2$. Ceci prouve qu'on peut donc approcher la loi de Z_2, lorsque n est grand, par la loi normale centrée réduite $\boldsymbol{N}(0,1)$.

2. C. 5. L'énoncé dit que quand *n* est " grand ", on peut assimiler les variables aléatoires Z_1 et Z_2 à des variables indépendantes gaussiennes centrées réduites. Dès lors, d'après la question 1. 3. b., la variable aléatoire $U_n = Z_1^2 + Z_2^2$ suit la loi $\Gamma(2{,}1)$, autrement dit la loi du χ^2 à 2 degrés de liberté.

2. C. 6. a. La fonction **`random(4)`** du langage Pascal retourne un nombre aléatoire compris entre 0 et 3, avec équiprobabilité. Pour obtenir les couleurs 1, 2, 3 avec les fréquences respectives $\frac{1}{4}, \frac{1}{4}, \frac{1}{2}$, il suffit de considérer que l'on a : la couleur 1 quand on obtient le nombre 0, la couleur 2 quand on obtient le nombre 1, et la couleur 3 quand on obtient le nombre 2 ou le nombre 3. Voici une procédure Pascal qui réalise ceci :

```
(**********************************************************************)
PROCEDURE Tirage(Var C : Tableau);

VAR i, coul : integer ;

BEGIN

for i := 1 to 100 do
  begin
     coul := 1 + random(4) ;
     if coul = 4 then coul := 3 ;
     C[i] := coul
  end

END ;
(**********************************************************************)
```

2. C. 6. b. On associe au *i*-ème tirage la variable aléatoire T_i définie à la question 2. C. 4., et la somme des valeurs prises par T_i donne $X_1 - X_2$. Voici une fonction Pascal qui réalise ceci :

```
(**********************************************************************)
FUNCTION Difference(Var C : tableau) : integer ;

VAR  dif, i, T : integer ;

BEGIN

  Tirage(C) ;
  dif := 0 ;
  for i := 1 to 100 do
     begin
       if C[i] = 1 then T := C[i] else T := C[i] - 3 ;
       dif := dif + T
     end ;
  Difference := dif

END ;
(**********************************************************************)
```

2. D. On suppose désormais s entier quelconque supérieur ou égal à 2.

2. D. 1. D'une façon générale, notons $(A)_{ij}$ le coefficient de la *i*-ème ligne et *j*-ème colonne d'une matrice *A*.

La matrice de covariance *M* est donc définie par : $(M)_{ij} = \operatorname{cov}(Y_i, Y_j)$.

On se souvient que : $V(X_i) = np_i(1-p_i)$ et que : $\operatorname{cov}(X_i, X_j) = -np_ip_j$ pour $i \neq j$. Il en résulte : $V(Y_i) = 1 - p_i$, et :

$\operatorname{cov}(Y_i, Y_j) = -\sqrt{p_i p_j}$ pour $i \neq j$. Ceci s'écrit : $(M)_{ij} = \delta_{ij} - \sqrt{p_i p_j}$, où δ_{ij} est le *symbole de Kronecker*, défini par : $\delta_{ij} = 1$ si $i = j$ et $\delta_{ij} = 0$ si $i \neq j$. Comme la matrice (δ_{ij}) n'est autre que la matrice unité *I*, il en résulte : $M = I - N$, avec $(N)_{ij} = \sqrt{p_i p_j}$.

2. D. 2. • La simple formule du produit matriciel donne :

$$\left(N^2\right)_{ij} = \sum_{k=1}^{s} (N)_{ik} (N)_{kj} = \sum_{k=1}^{s} \sqrt{p_i p_k} \sqrt{p_k p_j} = \sqrt{p_i p_j} \sum_{k=1}^{s} p_k = \sqrt{p_i p_j} = (N)_{ij}$$, ce qui prouve que $N^2 = N$.

• La *j*-ème colonne de la matrice *N* est : $C_j = \begin{pmatrix} \sqrt{p_1 p_j} \\ \sqrt{p_2 p_j} \\ \ldots \\ \ldots \\ \sqrt{p_s p_j} \end{pmatrix} = \sqrt{p_j} \begin{pmatrix} \sqrt{p_1} \\ \sqrt{p_2} \\ \ldots \\ \ldots \\ \sqrt{p_s} \end{pmatrix} = \sqrt{p_j}\, C$, avec $C = \begin{pmatrix} \sqrt{p_1} \\ \sqrt{p_2} \\ \ldots \\ \ldots \\ \sqrt{p_s} \end{pmatrix}$. Tous ces vecteurs-colonnes appartiennent donc à la droite vectorielle engendrée par le vecteur *C*. Le rang de la famille de ces vecteurs-colonnes est donc inférieur ou égal à 1, et il n'est pas nul puisque $p_i > 0$. Ce rang est donc égal à 1, et c'est le rang de la matrice *N*.

2. D. 3. Conformément à la norme AFNOR en vigueur aujourd'hui, nous notons $A^{\boldsymbol{T}}$ la transposée d'une matrice *A*, en abandonnant sans regret la notation ancienne ${}^tA$.

La matrice *N*, qui est symétrique réelle, se diagonalise sur $\mathbb{R}$ avec une base orthonormale de vecteurs propres. Cela signifie qu'il existe une matrice diagonale *D* et une matrice de passage *Q* telles que : $D = Q^{-1}NQ$, et que cette matrice de passage vérifie : $Q^{\boldsymbol{T}} Q = QQ^{\boldsymbol{T}} = I$, autrement dit : $Q \in GL_n(\mathbb{R})$ et $Q^{\boldsymbol{T}} = Q^{-1}$.

L'égalité $N^2 = N$ démontrée plus haut signifie que la matrice *N* annule le polynôme $x^2 - x$. Il s'ensuit que les valeurs propres de cette matrice sont parmi les racines de ce polynôme, et donc ne peuvent être que 0 ou 1. Les coefficients diagonaux de la matrice *D* sont ces valeurs propres, et sont donc égaux à 0 ou à 1. Le nombre de ces coefficients égaux à 1 est la rang de la matrice *D*, qui est le rang de la matrice *N*, égal à 1. La matrice *D* a donc sur sa diagonale s - 1 coefficients égaux à 0, et un coefficient égal à 1. En ordonnant convenablement les vecteurs propres, on peut obtenir une matrice diagonale *D* dont les $s - 1$ premiers coefficients diagonaux sont égaux à 0, et le dernier est nul.

Il en résulte : $M = I - N = Q(I - D)Q^{-1} = Q(I - D)Q^{\boldsymbol{T}}$, et la matrice $I - D$ est bien la matrice J_{s-1} décrite par l'énoncé.

2. D. 4. • On a : $Z_i = \sum_{k=1}^{s} a_{ik} Y_k$, ce qui montre que chaque variable Z_i est combinaison linéaire des variables Y_1, Y_2, ..., Y_s, et comme ces dernières sont centrées, force est d'en déduire que les variables Z_i sont centrées elles aussi.

• On a : $\operatorname{cov}(Z_i, Z_j) = \operatorname{cov}\left(\sum_{k=1}^{s} a_{ik} Y_k, \sum_{h=1}^{s} a_{jh} Y_h\right) = \sum_{k=1}^{s} \sum_{h=1}^{s} a_{ik} a_{jh} \operatorname{cov}(Y_k, Y_h) = \sum_{k=1}^{s} \sum_{h=1}^{s} \left(Q^{\boldsymbol{T}}\right)_{ik} \left(Q^{\boldsymbol{T}}\right)_{jh} (M)_{kh} =$

$$= \sum_{h=1}^{s} \left(Q^{\boldsymbol{T}}\right)_{jh} \sum_{k=1}^{s} \left(Q^{\boldsymbol{T}}\right)_{ik} (M)_{kh} = \sum_{h=1}^{s} (Q)_{hj} \left(Q^{\boldsymbol{T}} M\right)_{ih} = \left(Q^{\boldsymbol{T}} MQ\right)_{ij} = \left(Q^{-1} MQ\right)_{ij} = (J_{s-1})_{ij}.$$

La matrice de covariance de Z_1, Z_2, ..., Z_s est donc la matrice J_{s-1}.

• La variance de Z_s est donc : $V(Z_s) = \operatorname{cov}(Z_s, Z_s) = (J_{s-1})_{ss} = 0$. Ainsi, la variable aléatoire Z_s est quasi-certaine : elle prend une seule valeur avec probabilité 1, et comme elle est centrée, cette valeur est nulle.

2. D. 5. On a : $Z_i^2 = \sum_{k=1}^{s} a_{ik} Y_k \sum_{h=1}^{s} a_{ih} Y_h = \sum_{k=1}^{s} \left(Q^{\boldsymbol{T}}\right)_{ik} Y_k \sum_{h=1}^{s} \left(Q^{\boldsymbol{T}}\right)_{ih} Y_h$, d'où :

$$\sum_{i=1}^{s} Z_i^2 = \sum_{k=1}^{s}\sum_{h=1}^{s} Y_k Y_h \sum_{i=1}^{s} \left(Q^{\mathsf{T}}\right)_{ik}\left(Q^{\mathsf{T}}\right)_{ih}$$. Effectuons :

$$\sum_{i=1}^{s} \left(Q^{\mathsf{T}}\right)_{ik}\left(Q^{\mathsf{T}}\right)_{ih} = \sum_{i=1}^{s} (Q)_{ki}\left(Q^{\mathsf{T}}\right)_{ih} = \left(QQ^{\mathsf{T}}\right)_{kh} = \left(QQ^{-1}\right)_{kh} = (I)_{kh} = \delta_{kh}.$$

Il en résulte : $\sum_{i=1}^{s} Z_i^2 = \sum_{k=1}^{s}\sum_{h=1}^{s} \delta_{kh} Y_k Y_h = \sum_{k=1}^{s} Y_k^2 = U_n$. Puisque $Z_s = 0$, cette égalité devient : $U_n = \sum_{i=1}^{s-1} Z_i^2$.

D'après la question 1. 3. b., la variable aléatoire U_n suit la loi $\Gamma\left(2, \frac{s-1}{2}\right)$, autrement dit la loi du χ^2 à s - 1 degrés de liberté.

Partie 3

Si l'on fait l'hypothèse que les naissances sont uniformément réparties au cours de l'année, alors la probabilité qu'une naissance ait lieu durant la période T_i est égale à la proportion p_i de journées de cette période dans l'année. On peut considérer n naissances comme n tirages indépendants dans un ensemble de jours répartis en quatre périodes T_1, T_2, T_3, T_4. Pour i = 1, 2, 3, 4, on désigne par X_i la variable aléatoire égale au nombre de ces naissances survenues durant la période T_i. Ces variables aléatoires X_i sont les mêmes que celles définies à la partie 2, concernant n tirages successifs avec remise dans une urne contenant des boules de $s = 4$ couleurs C_i en proportion p_i.

On pose de même : $U_n = \sum_{i=1}^{4} \frac{(X_i - np_i)^2}{np_i}$, et d'après la partie D, cette variable aléatoire suit la loi du χ^2 à s - 1 = 3 degrés de liberté. D'où : $P(U_n \le 11{,}3) = F_3(11{,}3) = 0{,}99$.

On dispose d'un échantillon de n = 8000 naissances, et l'on a les proportions, autrement dit les fréquences f_i des naissances survenues durant la période T_i. La valeur de X_i observée est nf_i, et la valeur de U_n observée est *ipso facto* :

$\sum_{i=1}^{4} \frac{(nf_i - np_i)^2}{np_i} = n\sum_{i=1}^{4} \frac{(f_i - p_i)^2}{p_i}$. Or, le calcul numérique de cette quantité pour n = 8000 donne 12,03, alors que nous avons vu que $P(U_n \le 11{,}3) = 0{,}99$, ce qui signifie que sous l'hypothèse faite, il serait pratiquement certain que $U_n \le 11{,}3$. Ceci conduit donc à rejeter l'hypothèse de répartition uniforme des naissances au cours de l'année.

Commentaire

Ce problème portait sur une loi de probabilité : la loi du χ^2, et en proposait une application, pour déboucher sur une question de statistique.

On peut regretter une ou deux maladresses : "entier non nul" au lieu de "entier strictement positif", et dans l'annexe X_λ au lieu de Y_λ.

Plusieurs questions concernaient les lois de probabilité classiques, ainsi que la convergence et l'approximation des variables aléatoires. Le Calcul intégral, l'Algèbre linéaire et l'Algèbre bilinéaire étaient aussi mis à contribution, ce qui montre bien l'unité des mathématiques. De plus, ce problème apparaît comme complémentaire du problème de Mathématiques I, et ces deux problèmes couvrent l'intégralité du programme de mathématiques des classes préparatoires HEC. Une bonne année pour les annales.

R. C.

e-mail : cuculier@imaginet.fr