

Conception : HEC Paris

OPTION SCIENTIFIQUE

MATHÉMATIQUES

Mercredi 27 avril 2016, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision 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.

Ils ne doivent faire usage d'aucun document. **L'utilisation de toute calculatrice et de tout matériel électronique est interdite.** Seule l'utilisation d'une règle graduée est autorisée.

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

Dans tout le problème :

- On note n et k deux entiers vérifiant $2 \leq k \leq n$ et E un $\mathbf{R}$ -espace vectoriel de dimension n muni d'un produit scalaire $\langle \cdot, \cdot \rangle_E$ qui en fait un espace euclidien.
- On note 0_E et $0_{\mathcal{L}(E)}$ respectivement, le vecteur nul et l'endomorphisme nul de E et $\mathcal{B} = (e_1, e_2, \dots, e_n)$ une base orthonormale de E . L'endomorphisme identité de E est noté id_E .
- Pour tout sous-espace vectoriel F de E , on note $F^\perp$ l'orthogonal de F et p_F le *projecteur orthogonal d'image F* , c'est-à-dire l'unique endomorphisme de E vérifiant : $\forall x \in F, p_F(x) = x$ et $\forall x \in F^\perp, p_F(x) = 0_E$.
- On note $\mathcal{M}_{n,m}(\mathbf{R})$ l'ensemble des matrices à n lignes et m colonnes ($m \geq 1$) à coefficients réels. La transposée d'une matrice $A \in \mathcal{M}_{n,m}(\mathbf{R})$ est notée tA .
- Pour tout $(\rho_1, \rho_2, \dots, \rho_k) \in \mathbf{R}^k$, on note $\text{Diag}(\rho_1, \rho_2, \dots, \rho_k)$ la matrice diagonale de $\mathcal{M}_k(\mathbf{R})$ dont les coefficients diagonaux sont, dans cet ordre, $\rho_1, \rho_2, \dots, \rho_k$.
- On note I_n la matrice identité de $\mathcal{M}_n(\mathbf{R})$.

On rappelle que la *somme* de k sous-espaces vectoriels $F_1, F_2, \dots, F_k$ de E est le sous-espace vectoriel de E ,

noté $\sum_{i=1}^k F_i$, défini par : $\sum_{i=1}^k F_i = \left\{ \sum_{i=1}^k x_i; (x_1, x_2, \dots, x_k) \in F_1 \times F_2 \times \dots \times F_k \right\}$.

On rappelle aussi que les sous-espaces vectoriels $F_1, F_2, \dots, F_k$ sont en *somme directe* si chaque vecteur de $\sum_{i=1}^k F_i$ n'admet qu'une seule décomposition de la forme précédente. Dans ce cas, et seulement dans ce cas, la somme

des sous-espaces vectoriels $F_1, F_2, \dots, F_k$ est notée $\bigoplus_{i=1}^k F_i$.

L'objet de ce problème est la mise en évidence de quelques propriétés algébriques dont les conséquences probabilistes fondent les tests statistiques qui permettent de mesurer l'influence effective d'une ou plusieurs variables explicatives sur une variable endogène.

La partie II est indépendante de la partie I.

Tournez la page S.V.P.

Partie I. Partitions de l'identité.

Soit k endomorphismes $u_1, u_2, \dots, u_k$ de E . On dit que $u_1, u_2, \dots, u_k$ constituent une *partition de l'identité* de E si : $u_1 + u_2 + \dots + u_k = \text{id}_E$.

1. *Exemple 1.* Dans cette question, $n = 3$ et $E = \mathbf{R}^3$. Soit $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$ et f l'endomorphisme de $\mathbf{R}^3$ de matrice A dans la base canonique de $\mathbf{R}^3$.

- a) Préciser le spectre de la matrice A et montrer que A n'est pas diagonalisable.
- b) Montrer que le polynôme $Q \in \mathbf{R}[X]$ tel que $Q(X) = X^3 + X^2$ est un polynôme annulateur de A .
- c) Existe-t-il un polynôme de degré 2 annulateur de A ?
- d) Trouver deux polynômes Q_1 et Q_2 de $\mathbf{R}[X]$ pour lesquels les deux endomorphismes $Q_1(f)$ et $Q_2(f)$ sont des projecteurs et constituent une partition de l'identité de $\mathbf{R}^3$.

2. *Exemple 2.* On considère dans cette question un endomorphisme f de E diagonalisable et possédant k valeurs propres distinctes $\lambda_1, \lambda_2, \dots, \lambda_k$.

Pour tout $i \in \llbracket 1, k \rrbracket$, on note :

- $L_i(X)$ le polynôme de $\mathbf{R}[X]$ défini par $L_i(X) = \prod_{\substack{j \in \llbracket 1, k \rrbracket \\ j \neq i}} (X - \lambda_j)$;

- $E_{\lambda_i}(f)$ le sous-espace propre de f associé à la valeur propre λ_i ;
- v_i l'endomorphisme de E défini par $v_i = L_i(f)$.

- a) Justifier l'égalité : $E = \bigoplus_{i=1}^k E_{\lambda_i}(f)$. En déduire que $\prod_{j=1}^k (X - \lambda_j)$ est un polynôme annulateur de f .

- b) Établir pour tout $i \in \llbracket 1, k \rrbracket$, l'inclusion : $\text{Im}(v_i) \subset E_{\lambda_i}(f)$.

- c) Pour tout $j \in \llbracket 1, k \rrbracket$, calculer la somme : $\sum_{i=1}^k L_i(\lambda_j)$. En déduire que les endomorphismes $v_1, v_2, \dots, v_k$ constituent une partition de l'identité de E .

- d) Établir pour tout $i \in \llbracket 1, k \rrbracket$, l'égalité : $\text{Im}(v_i) = E_{\lambda_i}(f)$. Identifier l'endomorphisme v_1 .

3. Soit k endomorphismes $u_1, u_2, \dots, u_k$ de E qui constituent une partition de l'identité de E .

Pour tout $i \in \llbracket 1, k \rrbracket$, on note r_i le rang de u_i .

- a) Établir les relations : $E = \sum_{i=1}^k \text{Im}(u_i)$ et $n \leq \sum_{i=1}^k r_i$.

- b) Montrer que les sous-espaces vectoriels $\text{Im}(u_1), \text{Im}(u_2), \dots, \text{Im}(u_k)$ sont en somme directe si et seulement

$$\text{si on a : } n = \sum_{i=1}^k r_i.$$

- c) Dans cette question, on cherche à montrer l'équivalence des propriétés (1), (2) et (3) suivantes :

$$(1) \quad n = \sum_{i=1}^k r_i.$$

- (2) Les endomorphismes $u_1, u_2, \dots, u_k$ sont des projecteurs.

- (3) Pour tout $(i, j) \in \llbracket 1, k \rrbracket^2$, avec $i \neq j$, on a : $u_i \circ u_j = 0_{\mathcal{L}(E)}$.

(i) En utilisant la trace des matrices de projecteurs, justifier l'implication (2) $\implies$ (1).

(ii) À l'aide de la question 3.b) et en écrivant, pour $x \in E$, les vecteurs $u_1(x), u_2(x), \dots, u_k(x)$ comme des sommes de k vecteurs, établir l'implication (1) $\implies$ (3).

(iii) Conclure en établissant une troisième implication.

Partie II. Représentation matricielle d'un projecteur orthogonal.

4.a) Soit p un endomorphisme de E et P la matrice de p dans la base $\mathcal{B}$.

Montrer que p est un projecteur orthogonal si et seulement si on a : $P^2 = P$ et ${}^tP = P$.

b) Soit f un endomorphisme de E et M la matrice de f dans la base $\mathcal{B}$.

Établir l'existence d'un réel α et d'un projecteur orthogonal p tels que $f = \alpha p$, si et seulement si on a : $\text{tr}(M) M^2 = \text{tr}(M^2) M$ et ${}^tM = M$, où $\text{tr}(M)$ et $\text{tr}(M^2)$ sont les traces respectives de M et M^2 .

5.a) Écrire en *Scilab* une fonction "function t=tr(A)" qui calcule la trace d'une matrice carrée A .

b) La fonction "issym" suivante permet de tester si une matrice carrée A de taille n donnée est symétrique.

```
(1) function b=issym(n,A)
(2)   b=%T; // affectation de la valeur booléenne True à la variable b.
(3)   for i=1:n-1
(4)     for j=i+1:n
(5)       b=b & A(i,j)==A(j,i)
(6)     end;
(7)   end;
(8) endfunction
```

Préciser la signification de la ligne (5) du code et donner un exemple d'utilisation de la fonction "issym" en indiquant les valeurs d'entrée ainsi que la valeur de sortie obtenue.

c) La fonction "orthoproj" suivante, dont une ligne de code est incomplète, permet de tester si, pour une matrice carrée M de taille n donnée, il existe un réel α et un projecteur orthogonal p pour lesquels M est la matrice de l'endomorphisme αp dans une base orthonormale. Cette fonction utilise les deux fonctions précédentes (questions 5.a) et 5.b)) et s'appuie sur la condition nécessaire et suffisante de la question 4.b).

```
(1) function b=orthoproj(n,M)
(2)   A=tr(M)*M^2;
(3)   B=tr(M^2)*M;
(4)   b=issym(n,M);
(5)   if b then
(6)     for i=1:n
(7)       for j=i:n
(8)         b=.....
(9)       end;
(10)    end;
(11)  end;
(12) endfunction
```

Compléter la ligne (8) du code et donner les valeurs de sortie obtenues par application de cette fonction aux deux matrices $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ et $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.

Les définitions et notations suivantes concernent les questions 6 à 9.

Pour tout vecteur $x \in E$, on note X la matrice colonne de ses coordonnées dans la base $\mathcal{B}$.

Soit $\mathcal{F} = (s_1, s_2, \dots, s_k)$ une famille de k vecteurs de E et F le sous-espace vectoriel de E engendré par $\mathcal{F}$.

On note S la matrice de $\mathcal{M}_{n,k}(\mathbf{R})$ dont les colonnes sont, dans cet ordre, $S_1, S_2, \dots, S_k$.

On rappelle que p_F est le projecteur orthogonal d'image F .

6.a) Montrer que les deux matrices S et tSS ont le même rang.

b) Soit $y \in E$. Montrer que $y \in F$ si et seulement si il existe une matrice $Z \in \mathcal{M}_{k,1}(\mathbf{R})$ telle que $Y = SZ$.

c) Soit $y \in E$. Montrer que $y \in F^\perp$ si et seulement si la matrice colonne tSY est nulle.

- d) Soit $x \in E$ et $y = p_F(x)$. Établir l'existence d'une matrice $Z \in \mathcal{M}_{k,1}(\mathbf{R})$ telle que $Y = SZ$ et ${}^tSX = {}^tSSZ$.
e) En déduire l'expression de la matrice de p_F dans la base $\mathcal{B}$ en fonction de S lorsque la famille $\mathcal{F}$ est libre.
7. Soit M une matrice symétrique de $\mathcal{M}_k(\mathbf{R})$. On appelle *inverse de Penrose-Moore de M* toute matrice N de $\mathcal{M}_k(\mathbf{R})$ qui vérifie les quatre propriétés suivantes :

$$MNM = M ; \quad NMN = N ; \quad {}^t(MN) = MN ; \quad {}^t(NM) = NM .$$

- a) Établir l'existence d'une matrice $Q \in \mathcal{M}_k(\mathbf{R})$ et de réels $\rho_1, \rho_2, \dots, \rho_k$ qui vérifient la relation suivante :

$$M = Q \operatorname{Diag}(\rho_1, \rho_2, \dots, \rho_k) {}^tQ .$$

- b) On note h l'application de $\mathbf{R}$ dans $\mathbf{R}$ telle que : $\forall t \in \mathbf{R}, h(t) = \begin{cases} 1/t & \text{si } t \neq 0 \\ 0 & \text{si } t = 0 \end{cases}$. On note $M^{(-)}$ la matrice définie par : $M^{(-)} = Q \operatorname{Diag}(h(\rho_1), h(\rho_2), \dots, h(\rho_k)) {}^tQ$.
Montrer que $M^{(-)}$ est une inverse de Penrose-Moore de M .
- c) Soit N une inverse de Penrose-Moore de M .
(i) Justifier les égalités : $N = M {}^tNN$ et $M^2N = M$.
(ii) Soit U une matrice de $\mathcal{M}_k(\mathbf{R})$. On suppose que M^2U est nulle. Montrer que MU est nulle.
(iii) On pose : $U = N - M^{(-)}$. Justifier que $M^{(-)}$ est l'unique inverse de Penrose-Moore de M .
8. On note $({}^tSS)^{(-)}$ l'unique inverse de Penrose-Moore de la matrice tSS et on pose : $P = S({}^tSS)^{(-)} {}^tS$.
a) Montrer que les matrices P et S ont le même rang.
b) Justifier que P est la matrice de p_F dans la base $\mathcal{B}$ et que son expression généralise la formule trouvée dans la question 6.e) lorsque la famille $\mathcal{F}$ est libre.
9. *Exemple.* On suppose que : $k = 2, s_1 = (\alpha_1, \alpha_2, \dots, \alpha_n), s_2 = (\beta_1, \beta_2, \dots, \beta_n), s_1 \neq 0_E$ et tSS non inversible.
a) Établir l'existence d'un réel θ tel que pour tout $i \in \llbracket 1, n \rrbracket$, on a : $\beta_i = \theta \alpha_i$.
b) Déterminer une matrice carrée Q pour laquelle la matrice ${}^tQ {}^tSSQ$ est diagonale.
c) En déduire l'inverse de Penrose-Moore de la matrice tSS .
d) Soit $x = x_1e_1 + x_2e_2 + \dots + x_ne_n$ un vecteur de E . Calculer $p_F(x)$.

Partie III. Application probabiliste.

Dans cette partie, $E = \mathbf{R}^n$ et on suppose que toutes les variables aléatoires et tous les vecteurs aléatoires considérés sont définis sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$.

Pour tout entier $d \geq 1$, on dit qu'une variable aléatoire C suit la loi du khi-deux de paramètre d , notée $\chi^2(d)$, si la variable aléatoire $\frac{C}{2}$ suit la loi $\gamma\left(\frac{d}{2}\right)$.

On appelle *variable gaussienne* toute variable aléatoire X qui suit une loi normale ou qui est certaine, et on note sa loi $\mathcal{G}(\mu, \sigma^2)$, où μ est l'espérance de X et σ l'écart-type de X .

Autrement dit, pour tout couple $(\mu, \sigma) \in \mathbf{R} \times \mathbf{R}_+$, une variable aléatoire X suit la loi $\mathcal{G}(\mu, \sigma^2)$, soit lorsque $\sigma > 0$ et X suit la loi normale $\mathcal{N}(\mu, \sigma^2)$, soit lorsque $\sigma = 0$ et $\mathbf{P}([X = \mu]) = 1$.

10. Soit $(\mu, \sigma) \in \mathbf{R} \times \mathbf{R}_+^*$ et soit $X_1, X_2, \dots, X_n$ des variables aléatoires mutuellement indépendantes et de même loi normale $\mathcal{N}(\mu, \sigma^2)$. Montrer que la variable aléatoire $\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2$ suit la loi $\chi^2(n)$.

Si $G_1, G_2, \dots, G_n$ sont des variables aléatoires réelles telles que pour tout $a = (a_1, a_2, \dots, a_n) \in \mathbf{R}^n$, la variable aléatoire $\sum_{i=1}^n a_i G_i$ est une *variable gaussienne centrée*, alors on dit que le vecteur aléatoire $(G_1, G_2, \dots, G_n)$ est un *vecteur gaussien* et on note G la matrice colonne de composantes $G_1, G_2, \dots, G_n$.

11. Soit $(G_1, G_2, \dots, G_n)$ un vecteur gaussien, M une matrice de $\mathcal{M}_n(\mathbf{R})$ et $(H_1, H_2, \dots, H_n)$ un vecteur aléatoire tel que la matrice colonne H de composantes $H_1, H_2, \dots, H_n$ vérifie : $H = M G$.

a) Montrer que $(H_1, H_2, \dots, H_n)$ est un vecteur gaussien.

b) Justifier que pour tout $(i, j) \in \llbracket 1, n \rrbracket^2$, la variable aléatoire $G_i G_j$ admet une espérance, notée $\mathbf{E}(G_i G_j)$.

On note alors $\Lambda(G)$ la matrice de $\mathcal{M}_n(\mathbf{R})$ définie par : $\Lambda(G) = (\mathbf{E}(G_i G_j))_{1 \leq i, j \leq n}$ et on admet dans la suite que la loi d'un vecteur gaussien $(G_1, G_2, \dots, G_n)$ est caractérisée par la matrice $\Lambda(G)$.

Autrement dit, si $(G_1, G_2, \dots, G_n)$ et $(R_1, R_2, \dots, R_n)$ sont deux vecteurs gaussiens vérifiant $\Lambda(G) = \Lambda(R)$,

alors ils ont la même loi, c'est-à-dire : $\forall (x_1, x_2, \dots, x_n) \in \mathbf{R}^n$, $\mathbf{P}\left(\bigcap_{i=1}^n [G_i \leq x_i]\right) = \mathbf{P}\left(\bigcap_{i=1}^n [R_i \leq x_i]\right)$.

12. On suppose que $G_1, G_2, \dots, G_n$ sont des variables aléatoires mutuellement indépendantes et de même loi normale $\mathcal{N}(0, 1)$.

a) Montrer que $(G_1, G_2, \dots, G_n)$ est un vecteur gaussien. Déterminer $\Lambda(G)$.

b) Soit Q une matrice orthogonale de $\mathcal{M}_n(\mathbf{R})$ et $(H_1, H_2, \dots, H_n)$ un vecteur aléatoire tel que la matrice colonne H de composantes $H_1, H_2, \dots, H_n$ vérifie : $H = Q G$.

Montrer que les variables aléatoires $H_1, H_2, \dots, H_n$ sont mutuellement indépendantes et de même loi normale $\mathcal{N}(0, 1)$.

13. Soit $(G_1, G_2, \dots, G_n)$ un vecteur gaussien dont les composantes $G_1, G_2, \dots, G_n$ sont mutuellement indépendantes et de variance égale à 1.

Soit $P_1, P_2, \dots, P_k$ des matrices symétriques de $\mathcal{M}_n(\mathbf{R})$ de rangs respectifs $r_1, r_2, \dots, r_k$.

On suppose que $\sum_{i=1}^k P_i = I_n$ et $\sum_{i=1}^k r_i = n$.

a) Justifier que $P_1, P_2, \dots, P_k$ sont des matrices de projecteurs orthogonaux de $\mathbf{R}^n$ dans la base canonique de $\mathbf{R}^n$ dont les images sont deux à deux orthogonales.

b) En déduire l'existence d'une matrice orthogonale Q de $\mathcal{M}_n(\mathbf{R})$ pour laquelle chacune des matrices $Q P_1 Q, Q P_2 Q, \dots, Q P_k Q$ est diagonale.

c) On suppose que $r_1 \neq 0$. Montrer que la variable aléatoire ${}^t G P_1 G$ suit la loi $\chi^2(r_1)$.

d) Montrer que les variables aléatoires ${}^t G P_1 G, {}^t G P_2 G, \dots, {}^t G P_k G$ sont mutuellement indépendantes.

14. Soit q et m deux entiers supérieurs ou égaux à 2 et $(X_{i,j})_{\substack{1 \leq i \leq q \\ 1 \leq j \leq m}}$ une famille de $q \times m$ variables aléatoires mutuellement indépendantes et de même loi normale $\mathcal{N}(0, 1)$.

On pose : $\bar{X} = \frac{1}{q \times m} \sum_{i=1}^q \sum_{j=1}^m X_{i,j}$ et pour tout $j \in \llbracket 1, m \rrbracket$, $Z_j = \frac{1}{q} \sum_{i=1}^q X_{i,j}$.

a) Déterminer les lois respectives des variables aléatoires $\bar{X}$ et $\sum_{i=1}^q \sum_{j=1}^m (X_{i,j} - \bar{X})^2$ et établir l'indépendance de ces deux variables aléatoires.

b) Déterminer les lois respectives des variables aléatoires $\sum_{i=1}^q \sum_{j=1}^m (X_{i,j} - Z_j)^2$ et $q \sum_{j=1}^m (Z_j - \bar{X})^2$ et établir l'indépendance de ces deux variables aléatoires.

