



# BANQUE COMMUNE D'ÉPREUVES

Concepteurs : H.E.C. – E.S.C.P. – E.A.P.

CODE ÉPREUVE :

283

CCIP\_M2\_S

OPTION : SCIENTIFIQUE

## MATHEMATIQUES II

Mardi 10 Mai 2005, 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.

Dans tout le problème,  $n$  et  $r$  désignent des entiers strictement positifs. On note  $\mathcal{M}_{n,r}(\mathbb{R})$ , l'ensemble des matrices rectangulaires à  $n$  lignes et  $r$  colonnes à coefficients réels. Pour  $n = r$ , on pose  $\mathcal{M}_n(\mathbb{R}) = \mathcal{M}_{n,n}(\mathbb{R})$ . Pour tout entier  $n$  de  $\mathbb{N}^*$ , on identifie  $\mathbb{R}^n$  et  $\mathcal{M}_{n,1}(\mathbb{R})$ .

La transposée d'une matrice  $A$  appartenant à  $\mathcal{M}_{n,r}(\mathbb{R})$  est notée  ${}^tA$ . On pourra également la noter  $A^T$ .

On étudie dans ce problème, quelques propriétés du *modèle linéaire*, qui constitue l'instrument de base de l'économétrie.

### Partie I. Trace et matrices aléatoires

Pour toute matrice  $M$  appartenant à  $\mathcal{M}_n(\mathbb{R})$ , on appelle *trace de  $M$* , notée  $\text{tr}(M)$ , la somme de ses coefficients diagonaux ; ainsi, si  $M = (m_{i,j})_{1 \leq i,j \leq n}$ ,  $\text{tr}(M) = \sum_{i=1}^n m_{i,i}$ .

**On rappelle** les trois résultats suivants (que les candidats n'ont pas à démontrer) :

- l'application  $\text{tr}$  qui, à toute matrice  $M$  de  $\mathcal{M}_n(\mathbb{R})$ , associe sa trace, est une application linéaire de  $\mathcal{M}_n(\mathbb{R})$  dans  $\mathbb{R}$  ;
- si  $A$  est une matrice de  $\mathcal{M}_{n,r}(\mathbb{R})$  et  $B$  une matrice de  $\mathcal{M}_{r,n}(\mathbb{R})$ , alors  $\text{tr}(AB) = \text{tr}(BA)$  ;
- si  $M$  et  $N$  sont deux matrices semblables de  $\mathcal{M}_n(\mathbb{R})$ , alors  $\text{tr}(M) = \text{tr}(N)$ .

1. Soit  $M$  une matrice de  $\mathcal{M}_n(\mathbb{R})$  possédant  $q$  valeurs propres ( $1 \leq q \leq n$ ) notées  $\lambda_1, \lambda_2, \dots, \lambda_q$ . Pour tout entier  $i$  de  $\llbracket 1, q \rrbracket$ , on désigne par  $n_i$  la dimension du sous-espace propre associé à la valeur propre  $\lambda_i$ .

a) On suppose que la matrice  $M$  est diagonalisable sur  $\mathbb{R}$ . Montrer que  $\text{tr}(M) = \sum_{i=1}^q n_i \lambda_i$ .

b) On suppose que la matrice  $M = (m_{i,j})_{1 \leq i,j \leq n}$  de  $\mathcal{M}_n(\mathbb{R})$  est symétrique. Montrer les égalités suivantes :

$$\text{tr}({}^tMM) = \text{tr}(M^2) = \sum_{i=1}^q n_i \lambda_i^2 = \sum_{i=1}^n \sum_{j=1}^n m_{i,j}^2$$

2. Pour tout entier  $i$  de  $\llbracket 1, n \rrbracket$  et pour tout entier  $j$  de  $\llbracket 1, r \rrbracket$ , on considère des variables aléatoires réelles  $Z_{i,j}$  définies sur un espace probabilisé  $(\Omega, \mathcal{A}, \mu)$ . On définit la *matrice aléatoire*  $Z$ , à  $n$  lignes et  $r$  colonnes, en associant à tout  $\omega$  de  $\Omega$ , la matrice :

$$Z(\omega) = \begin{pmatrix} Z_{1,1}(\omega) & \dots & Z_{1,r}(\omega) \\ \vdots & \ddots & \vdots \\ Z_{n,1}(\omega) & \dots & Z_{n,r}(\omega) \end{pmatrix} = (Z_{i,j}(\omega))_{\substack{1 \leq i \leq n \\ 1 \leq j \leq r}}$$

On suppose que les  $nr$  variables aléatoires  $Z_{i,j}$  admettent une espérance  $E(Z_{i,j})$ , et on définit l'espérance de la matrice  $Z$ , notée  $E(Z)$ , comme la matrice de  $\mathcal{M}_{n,r}(\mathbb{R})$  dont les éléments sont les espérances  $E(Z_{i,j})$ , soit  $E(Z) = (E(Z_{i,j}))_{\substack{1 \leq i \leq n \\ 1 \leq j \leq r}}$ .

Si  $Z$  et  $W$  sont deux matrices aléatoires à  $n$  lignes et  $r$  colonnes admettant chacune une espérance, et si  $\lambda$  est réel, on remarquera que  $E(\lambda Z + W) = \lambda E(Z) + E(W)$ .

Dans le cas où  $n = r$ , on appelle trace de  $Z$ , notée  $\text{tr}(Z)$ , la variable aléatoire définie par  $\text{tr}(Z) = \sum_{i=1}^n Z_{i,i}$  et si  $n = r = 1$ , la matrice aléatoire  $Z$  coïncide avec la variable aléatoire  $Z$  et on a  $\text{tr}(Z) = Z$ .

Dans le cas où  $r = 1$  et  $n$  est quelconque, si  $T = {}^t(T_1 \dots T_n)$  et  $W = {}^t(W_1 \dots W_n)$  sont deux vecteurs aléatoires de  $\mathbb{R}^n$ , et si  $\lambda$  est un réel quelconque, on définit le vecteur aléatoire  $\lambda T + W$  de  $\mathbb{R}^n$  par

$$\lambda T + W = {}^t(\lambda T_1 + W_1 \dots \lambda T_n + W_n)$$

a) Soit  $Z$  une matrice aléatoire à  $n$  lignes et  $r$  colonnes admettant une espérance  $E(Z)$ . On considère une matrice  $A$  de  $\mathcal{M}_{r,n}(\mathbb{R})$ . Montrer que  $E(AZ) = AE(Z)$ . Soit  $B$  un élément de  $\mathcal{M}_{r,q}(\mathbb{R})$ , avec  $q \in \mathbb{N}^*$ . Montrer que  $E(ZB) = E(Z)B$ .

b) Soit  $Z$  une matrice aléatoire à  $n$  lignes et  $n$  colonnes admettant une espérance  $E(Z)$ . Établir les deux égalités :  $E({}^tZ) = {}^t(E(Z))$  et  $E(\text{tr}(Z)) = \text{tr}(E(Z))$ .

3. Dans cette question,  $Y$  désigne un vecteur aléatoire de  $\mathbb{R}^n$ , noté  $Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$ , admettant une espérance  $E(Y)$

et une matrice de variance-covariance notée  $V(Y)$ .

On rappelle que  $V(Y) = E[(Y - E(Y)) \times {}^t(Y - E(Y))]$ .

**On admet** que la définition et les propriétés de la matrice de variance-covariance  $V(Y)$  d'un vecteur aléatoire discret restent valables pour un vecteur aléatoire dont les composantes sont des variables aléatoires quelconques (discrètes ou à densité).

Ainsi, en supposant que pour tout  $i$  de  $\llbracket 1, n \rrbracket$  et pour tout  $j$  de  $\llbracket 1, n \rrbracket$ , la variable aléatoire  $Y_i Y_j$  possède un moment d'ordre 1 au moins, on définit la covariance de  $Y_i$  et  $Y_j$  par  $\text{cov}(Y_i, Y_j) = E(Y_i Y_j) - E(Y_i)E(Y_j)$ , et si  $Y_i$  et  $Y_j$  sont indépendantes, alors  $\text{cov}(Y_i, Y_j) = 0$ .

a) Montrer que, pour tout vecteur aléatoire  $Y$  de  $\mathbb{R}^n$ ,  $V(Y) = E(Y {}^t Y) - E(Y)E({}^t Y)$ .

b) Soit  $B$  une matrice de  $\mathcal{M}_{r,n}(\mathbb{R})$ . Justifier l'égalité  $V(BY) = BV(Y) {}^t B$ .

c) Soit  $A$  une matrice de  $\mathcal{M}_n(\mathbb{R})$ . On pose  $m = E(Y)$  et  $J = V(Y)$ . Établir les égalités :  $E({}^t Y A Y) = \text{tr}(A \cdot E(Y {}^t Y))$  et  $E({}^t Y A Y) = \text{tr}(A J) + {}^t m A m$ .

## Partie II. Le modèle linéaire

Dans les parties II.A et II.B,  $n$  et  $k$  sont deux entiers donnés qui vérifient  $1 \leq k < n$ . L'espace vectoriel  $\mathbb{R}^n$  est muni de sa structure euclidienne canonique. Toutes les variables aléatoires considérées sont définies sur un espace probabilisé  $(\Omega, \mathcal{A}, \mu)$  et admettent des moments d'ordre au moins 2.

On considère un échantillon de  $n$  individus extrait d'une population donnée. Ces individus sont décrits à l'aide de  $k$  variables statistiques réelles (caractères)  $C_1, C_2, \dots, C_k$ .

Pour tout entier  $j$  de  $\llbracket 1, k \rrbracket$ , chaque caractère  $C_j$  fait l'objet de  $n$  observations notées  $x_{1,j}, \dots, x_{n,j}$ . On définit ainsi une application linéaire  $f$  de  $\mathbb{R}^k$  dans  $\mathbb{R}^n$ , dont la matrice dans les bases canoniques de  $\mathbb{R}^k$  et  $\mathbb{R}^n$  est la matrice  $X = (x_{i,j})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq k}}$  de  $\mathcal{M}_{n,k}(\mathbb{R})$ . On suppose que le rang de  $X$  est égal à  $k$ .

Soit  $U = \begin{pmatrix} U_1 \\ \vdots \\ U_n \end{pmatrix}$  un vecteur aléatoire de  $\mathbb{R}^n$ , dont les composantes  $U_1, \dots, U_n$  sont des variables aléatoires réelles

définies sur  $(\Omega, \mathcal{A}, \mu)$ , mutuellement indépendantes et de même loi. On suppose que  $E(U) = 0_n$  et  $V(U) = \sigma^2 I_n$ , où  $0_n$  désigne le vecteur nul de  $\mathbb{R}^n$ ,  $I_n$  la matrice identité de  $\mathcal{M}_n(\mathbb{R})$  et  $\sigma$  un réel strictement positif inconnu.

Soit  $\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix}$  un vecteur non nul de  $\mathbb{R}^k$  dont les composantes  $\alpha_1, \dots, \alpha_k$  sont inconnues ( $\alpha$  est un paramètre vectoriel)

On considère un vecteur aléatoire non nul,  $Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$  de  $\mathbb{R}^n$  tel que, pour tout  $i$  de  $[[1, n]]$ , la variable aléatoire

$$Y_i \text{ définie sur } (\Omega, \mathcal{A}, \mu) \text{ s'écrit } Y_i = \sum_{j=1}^k x_{i,j} \alpha_j + U_i.$$

Sous forme matricielle, le modèle linéaire s'écrit  $Y = X\alpha + U$ . On s'intéresse dans cette partie II, à l'étude de quelques propriétés de ce modèle, liées à l'estimation des paramètres inconnus  $\alpha_1, \alpha_2, \dots, \alpha_k$  et  $\sigma^2$ .

Pour cela, on désigne par  $y$  et on note  $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ , le vecteur non nul de  $\mathbb{R}^n$  qui représente la réalisation sur l'échantillon considéré du vecteur aléatoire  $Y$ ; ainsi, pour tout  $i$  de  $[[1, n]]$ ,  $y_i$  est la réalisation de la variable aléatoire  $Y_i$ .

Soit  $u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$  le vecteur de  $\mathbb{R}^n$ , dit vecteur d'écart, défini par  $u = y - X\alpha$ .

### A. Quelques résultats algébriques

1. On considère l'endomorphisme  $h$  de  $\mathbb{R}^k$  dont la matrice  $H$ , dans la base canonique de  $\mathbb{R}^k$ , est définie par  $H = {}^t X X$ .

a) Montrer que  $H$  est une matrice symétrique réelle de  $\mathcal{M}_k(\mathbb{R})$ .

b) En étudiant le noyau de  $h$ , montrer que le rang de  $h$  est égal à  $k$ . En déduire que la matrice  $H$  est inversible. On notera  $H^{-1}$  son inverse.

2. Dans cette question, on veut trouver, en fonction de  $y$  et  $X$ , les vecteurs  $\alpha$  de  $\mathbb{R}^k$  qui minimisent  $\|u\|$ .

Montrer que ce problème admet une unique solution  $\hat{\alpha}$  définie par  $\hat{\alpha} = \begin{pmatrix} \hat{\alpha}_1 \\ \vdots \\ \hat{\alpha}_n \end{pmatrix} = H^{-1} {}^t X y$ .

3. Soit  $p$  le projecteur orthogonal de  $\mathbb{R}^n$  sur le sous-espace vectoriel engendré par les colonnes de la matrice  $X$ . On note  $P$  la matrice de  $p$  dans la base canonique de  $\mathbb{R}^n$ .

a) Montrer que  $p(y) = X\hat{\alpha}$ . En déduire que  $P = XH^{-1} {}^t X$ . Vérifier que  $P = P^2 = {}^t P$ .

b) Établir que le rang de  $P$  et la trace de  $P$  sont égaux. Quelle est leur valeur commune ?

c) Montrer que les colonnes de  $X$  constituent une base de vecteurs propres de la matrice  $P$ , associés à la valeur propre 1.

d) Montrer qu'il existe une matrice  $S$  de  $\mathcal{M}_n(\mathbb{R})$ , orthogonale, telle que  $P = S D {}^t S$ , où  $D = (d_{i,j})_{1 \leq i, j \leq n}$  est la matrice diagonale définie par :

$$\begin{cases} d_{i,i} = 1 & \text{si } 1 \leq i \leq k \\ d_{i,j} = 0 & \text{sinon} \end{cases}$$

Préciser les  $k$  premières colonnes de  $S$ .

4. Soit  $\hat{u}$  le vecteur de  $\mathbb{R}^n$  défini par  $\hat{u} = y - X\hat{\alpha}$ .

a) On pose  $Q = I_n - P$ . Montrer que  $\hat{u} = Qu$ . Vérifier que  $Q = Q^2 = {}^t Q$ . Calculer la trace de  $Q$ .

b) Exprimer  ${}^t \hat{u} \hat{u}$  et  ${}^t y Q y$  en fonction de  $Q$  et  $u$ .

5. Par définition, on dit qu'une matrice  $A$  symétrique réelle d'ordre  $n$  est positive, si pour tout vecteur  $z$  de  $\mathbb{R}^n$ , on a  ${}^t z A z \geq 0$ .

a) Montrer que  $A$ , symétrique réelle, est positive si et seulement si ses valeurs propres sont positives ou nulles.

b) Soit  $L$  une matrice appartenant à  $\mathcal{M}_{n,k}(\mathbb{R})$ . Établir que  ${}^t L L$  est symétrique réelle positive.

### B. Estimation des paramètres $\alpha_1, \alpha_2, \dots, \alpha_k$ et $\sigma^2$

1. Soit  $\hat{G}$  le vecteur aléatoire de  $\mathbb{R}^k$  défini par :  $\hat{G} = H^{-1} {}^t X Y$ .

a) Établir que  $E(Y) = X\alpha$ , et que  $V(Y) = \sigma^2 I_n$ . En déduire que  $E(\hat{G}) = \alpha$  ( $\hat{G}$  est un estimateur sans biais de  $\alpha$ , tandis que  $\hat{\alpha}$  est une estimation sans biais de  $\alpha$ ).

b) Montrer que  $V(\hat{G}) = \sigma^2 H^{-1}$ .

2. On veut montrer dans cette question, que dans l'ensemble des estimateurs sans biais du paramètre  $\alpha$ , de la forme  ${}^tBY$ , où  $B$  est une matrice quelconque, non nulle de  $\mathcal{M}_{n,k}(\mathbb{R})$ , l'estimateur  $\widehat{G}$  est *optimal* dans le sens suivant : tout autre estimateur  $G^*$  sans biais du paramètre  $\alpha$ , de la forme  ${}^tBY$  est tel que la matrice  $V(G^*) - V(\widehat{G})$  est positive.

Soit  $B$  une matrice non nulle de  $\mathcal{M}_{n,k}(\mathbb{R})$ . On considère le vecteur aléatoire  $\widehat{C} = {}^tBY$

a) Quelle condition doit satisfaire la matrice  $B$  pour que, pour tout vecteur  $\alpha$  de  $\mathbb{R}^k$ ,  $\widehat{C}$  soit un estimateur sans biais de  $\alpha$  ?

b) En supposant cette condition vérifiée, on pose  ${}^tF = {}^tB - H^{-1} {}^tX$ . Calculer  ${}^tFX$ , et montrer que la matrice  $V(\widehat{C}) - V(\widehat{G})$  est positive.

3. On désigne par  $\widehat{U}$  le vecteur aléatoire de  $\mathbb{R}^n$  défini par  $\widehat{U} = \begin{pmatrix} \widehat{U}_1 \\ \vdots \\ \widehat{U}_n \end{pmatrix} = Y - X\widehat{G}$ .

a) Montrer que  $\widehat{U} = QU$ .

b) Déterminer  $E(\widehat{U})$  et  $V(\widehat{U})$ . Les variables aléatoires  $\widehat{U}_1, \dots, \widehat{U}_n$  sont-elles indépendantes ?

c) Montrer que  ${}^t\widehat{U}\widehat{U} = \sum_{i=1}^n \widehat{U}_i^2 = {}^tUQU = {}^tYQY$ .

d) Calculer  $E({}^t\widehat{U}\widehat{U})$ . En déduire que la variable aléatoire  $s_n$  définie par  $s_n = \frac{{}^t\widehat{U}\widehat{U}}{n-k} = \frac{{}^tYQY}{n-k}$  est un estimateur sans biais de  $\sigma^2$ .

### C. Étude d'une suite d'estimateurs

Dans cette partie,  $k$  est fixé dans  $\mathbb{N}^*$ . On veut montrer que la suite d'estimateurs  $(s_n)_{n \geq k+1}$  de  $\sigma^2$ , est convergente.

On suppose que, pour tout  $i$  de  $\llbracket 1, n \rrbracket$ , la variable aléatoire  $U_i$  possède des moments d'ordre 3 et 4 avec  $E(U_i^3) = 0$  et  $E(U_i^4) = 3\sigma^4$ . On pose  $Q = (q_{i,j})_{1 \leq i,j \leq n}$ .

1. Établir que  ${}^tUQU = \sum_{i=1}^n \sum_{j=1}^n q_{i,j} U_i U_j$ .

2. Montrer que  $E[({}^tUQU)^2] = \sigma^4[(\text{tr}(Q))^2 + 2 \text{tr}(Q^2)]$ . En déduire que  $E[({}^tUQU)^2] = \sigma^4(n-k)(n-k+2)$ .

3. Calculer la variance  $V(s_n)$  de la variable aléatoire  $s_n$ . Conclure.