



PRÉPASIMPLE

prepasimple.fr

HEC — 2005

Corrigé de mathématiques approfondies

Mathématiques approfondies — Prépa ECG

Annales de concours · Corrigé

Corrigé

Conventions utilisées dans tout le corrigé.

Pour a et b dans $\mathbb{R}^n$, le produit scalaire canonique s'écrit

$$\langle a, b \rangle = {}^t a b = \sum_{i=1}^n a_i b_i$$

et la norme associée $\|a\| = \sqrt{{}^t a a}$.

Dans la partie II, on note $X_1, \dots, X_k$ les colonnes de X et $\mathcal{V} = \text{Vect}(X_1, \dots, X_k)$ le sous-espace vectoriel de $\mathbb{R}^n$ engendré par les colonnes de X . Pour $x = {}^t(x_1 \dots x_k) \in \mathbb{R}^k$, on a

$$f(x) = Xx = \sum_{j=1}^k x_j X_j,$$

donc $\mathcal{V} = \text{Im}(f)$.

Comme $\text{rg}(X) = k$, la famille $(X_1, \dots, X_k)$ est libre et $\dim \mathcal{V} = k$. Par le théorème du rang appliqué à $f : \mathbb{R}^k \rightarrow \mathbb{R}^n$, $\dim \text{Ker}(f) = k - \text{rg}(f) = 0$: l'application f est injective, c'est-à-dire que pour $x \in \mathbb{R}^k$, $Xx = 0_n$ implique $x = 0_k$.

Partie I. Trace et matrices aléatoires

- 1) a) On construit une base de vecteurs propres adaptée aux sous-espaces propres, puis on utilise l'invariance de la trace par similitude.

Les réels $\lambda_1, \dots, \lambda_q$ sont les valeurs propres deux à deux distinctes de M et E_{λ_i} désigne le sous-espace propre associé, de dimension n_i . La somme des sous-espaces propres est directe, et comme M est diagonalisable sur $\mathbb{R}$, cette somme est égale à $\mathbb{R}^n$:

$$\mathbb{R}^n = E_{\lambda_1} \oplus E_{\lambda_2} \oplus \cdots \oplus E_{\lambda_q}, \text{ donc}$$

$$\sum_{i=1}^q n_i = n.$$

En concaténant une base de chacun des E_{λ_i} , on obtient une base $(v_1, \dots, v_n)$ de $\mathbb{R}^n$ formée de vecteurs propres de M : les n_1 premiers vecteurs sont associés à λ_1 , les n_2 suivants à λ_2 , et ainsi de suite. Soit R la matrice inversible dont les colonnes sont $v_1, \dots, v_n$. Comme Mv_i est égal à v_i multiplié par la valeur propre associée, on a $MR = RD$, où D est la matrice diagonale dont la diagonale contient λ_1 répété n_1 fois, puis λ_2 répété n_2 fois, ..., puis λ_q répété n_q fois. Ainsi $M = RDR^{-1}$: les matrices M et D sont semblables.

Par le troisième résultat rappelé, $\text{Tr}(M) = \text{Tr}(D)$, et la trace de D est la somme de ses coefficients diagonaux :

$$\text{Tr}(M) = \text{Tr}(D) = \sum_{i=1}^q n_i \lambda_i$$

Conclusion

si M est diagonalisable sur $\mathbb{R}$, alors

$$\text{Tr}(M) = \sum_{i=1}^q n_i \lambda_i.$$

b) On utilise le théorème spectral pour se ramener à a), puis on calcule $\text{Tr}({}^tMM)$ directement à partir des coefficients.

Première égalité. Comme M est symétrique, ${}^tM = M$, donc ${}^tMM = M^2$ et $\text{Tr}({}^tMM) = \text{Tr}(M^2)$.

Deuxième égalité. D'après le théorème spectral (admis au programme), la matrice symétrique réelle M est diagonalisable sur $\mathbb{R}$. Avec les notations de a), $M = RDR^{-1}$, donc :

$$M^2 = RDR^{-1}RDR^{-1} = RD^2R^{-1}$$

La matrice D^2 est diagonale ; sa diagonale contient λ_1^2 répété n_1 fois, ..., λ_q^2 répété n_q fois.
Les matrices M^2 et D^2 étant semblables, elles ont même trace :

$$\text{Tr}(M^2) = \text{Tr}(D^2) = \sum_{i=1}^q n_i \lambda_i^2$$

Troisième égalité. Pour toute matrice M de $\mathcal{M}_n(\mathbb{R})$, le coefficient d'indice (i, i) de tMM vaut

$$\sum_{j=1}^n ({}^tM)_{i,j} m_{j,i} = \sum_{j=1}^n m_{j,i}^2.$$

Donc :

$$\text{Tr}({}^tMM) = \sum_{i=1}^n \sum_{j=1}^n m_{j,i}^2 = \sum_{i=1}^n \sum_{j=1}^n m_{i,j}^2$$

(en échangeant les noms des indices dans une somme finie double).

Conclusion

pour M symétrique,

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

On retiendra que l'égalité

$$\text{Tr}({}^tMM) = \sum_{i,j} m_{i,j}^2$$

est vraie pour toute matrice carrée, et que

$$\text{Tr}(M^2) = \sum_{i,j} m_{i,j}^2$$

pour toute matrice symétrique ; elle servira en II.C.2).

- 2) a) On calcule chaque coefficient et on utilise la linéarité de l'espérance. On établit le résultat un peu plus généralement pour $A \in \mathcal{M}_{s,n}(\mathbb{R})$, avec $s \in \mathbb{N}^*$ quelconque (l'énoncé correspond à $s = r$) : ce cas général sera utilisé ensuite.

Soit $A = (a_{i,l}) \in \mathcal{M}_{s,n}(\mathbb{R})$. La matrice aléatoire AZ appartient à $\mathcal{M}_{s,r}(\mathbb{R})$ et, pour $i \in \llbracket 1, s \rrbracket$ et $j \in \llbracket 1, r \rrbracket$, son coefficient d'indice (i, j) est :

$$(AZ)_{i,j} = \sum_{l=1}^n a_{i,l} Z_{l,j}$$

C'est une combinaison linéaire à coefficients réels de variables aléatoires admettant une espérance ; par linéarité de l'espérance, elle admet une espérance et :

$$E((AZ)_{i,j}) = \sum_{l=1}^n a_{i,l} E(Z_{l,j}) = (AE(Z))_{i,j}$$

Donc AZ admet une espérance et $E(AZ) = AE(Z)$.

De même, pour $B = (b_{l,j}) \in \mathcal{M}_{r,q}(\mathbb{R})$, la matrice ZB appartient à $\mathcal{M}_{n,q}(\mathbb{R})$ et, pour $i \in \llbracket 1, n \rrbracket$ et $j \in \llbracket 1, q \rrbracket$:

$$E((ZB)_{i,j}) = E\left(\sum_{l=1}^r Z_{i,l}b_{l,j}\right) = \sum_{l=1}^r E(Z_{i,l})b_{l,j} = (E(Z)B)_{i,j}$$

Conclusion

$E(AZ) = AE(Z)$ pour toute matrice réelle A à n colonnes, et $E(ZB) = E(Z)B$ pour toute matrice réelle B à r lignes. En combinant les deux, $E(AZB) = AE(Z)B$.

b) On raisonne encore coefficient par coefficient.

Pour $i, j \in \llbracket 1, n \rrbracket$, le coefficient d'indice (i, j) de tZ est $Z_{j,i}$, qui admet une espérance. Donc :

$$E({}^tZ)_{i,j} = E(Z_{j,i}) = (E(Z))_{j,i} = ({}^t(E(Z)))_{i,j}$$

d'où $E({}^tZ) = {}^t(E(Z))$.

La variable aléatoire

$$\text{Tr}(Z) = \sum_{i=1}^n Z_{i,i}$$

est une somme finie de variables admettant une espérance ; par linéarité :

$$E(\text{Tr}(Z)) = \sum_{i=1}^n E(Z_{i,i}) = \text{Tr}(E(Z))$$

Conclusion

$E({}^tZ) = {}^t(E(Z))$ et $E(\text{Tr}(Z)) = \text{Tr}(E(Z))$. La même preuve montre que $E({}^tZ) = {}^t(E(Z))$ pour une matrice aléatoire rectangulaire, en particulier $E({}^tY) = {}^t(E(Y))$ pour un vecteur aléatoire Y .

- 3) Question adaptée au programme actuel : la matrice de variance-covariance d'un vecteur aléatoire n'étant plus au programme, l'énoncé adapté conserve sa définition $V(Y) = E[(Y - E(Y)) {}^t(Y - E(Y))]$ et admet seulement les propriétés usuelles de l'espérance et de la covariance pour des variables aléatoires réelles quelconques (voir le relevé des adaptations).

Remarque préliminaire (existence et coefficients de $V(Y)$). Notons

$m = E(Y) = {}^t(m_1 \dots m_n)$. Pour $i, j \in \llbracket 1, n \rrbracket$, le coefficient d'indice (i, j) de $(Y - m) {}^t(Y - m)$ est :

$$(Y_i - m_i)(Y_j - m_j) = Y_i Y_j - m_j Y_i - m_i Y_j + m_i m_j$$

Par hypothèse $Y_i Y_j$, Y_i et Y_j admettent une espérance, donc cette variable aussi, et par linéarité :

$$\begin{aligned} E[(Y_i - m_i)(Y_j - m_j)] &= E(Y_i Y_j) - m_j E(Y_i) - m_i E(Y_j) + m_i m_j = E(Y_i Y_j) - E(Y_i)E(Y_j) \\ &= \text{Cov}(Y_i, Y_j) \end{aligned}$$

Ainsi $V(Y)$ existe et c'est la matrice $(\text{Cov}(Y_i, Y_j))_{1 \leq i, j \leq n}$; elle est symétrique (car $\text{Cov}(Y_i, Y_j) = \text{Cov}(Y_j, Y_i)$) et ses coefficients diagonaux sont les variances $V(Y_i)$.

- a) On développe le produit et on utilise la linéarité de l'espérance des matrices aléatoires.

Méthode 1 — Calcul matriciel.

En notant $m = E(Y)$, qui est un vecteur constant :

$$(Y - m)^t(Y - m) = Y^tY - Y^tm - m^tY + m^tm$$

Chaque terme admet une espérance : Y^tY a pour coefficients les Y_iY_j , et les autres termes ont des coefficients combinaisons linéaires des Y_i ou constants. Par linéarité, puis par 2) a) avec les matrices constantes ${}^tm \in \mathcal{M}_{1,n}(\mathbb{R})$ (multipliée à droite) et $m \in \mathcal{M}_{n,1}(\mathbb{R})$ (multipliée à gauche) :

$$V(Y) = E(Y^tY) - E(Y)^tm - mE({}^tY) + m^tm$$

Or $E(Y) = m$ et, d'après 2) b), $E({}^tY) = {}^tm$. Donc :

$$V(Y) = E(Y^tY) - m^tm - m^tm + m^tm = E(Y^tY) - m^tm = E(Y^tY) - E(Y)E({}^tY)$$

Méthode 2 — Coefficient par coefficient.

D'après la remarque préliminaire, le coefficient d'indice (i, j) de $V(Y)$ est $\text{Cov}(Y_i, Y_j) = E(Y_iY_j) - E(Y_i)E(Y_j)$. Or $E(Y_iY_j)$ est le coefficient d'indice (i, j) de $E(Y^tY)$, et $E(Y_i)E(Y_j)$ est le coefficient d'indice (i, j) du produit de la colonne $E(Y)$ par la ligne $E({}^tY)$. Les deux matrices ont donc les mêmes coefficients.

La méthode 1 est un calcul global qui imite la formule de Koenig-Huygens ; la méthode 2 ramène l'égalité matricielle à cette même formule, appliquée à chaque covariance.

Conclusion

$$V(Y) = E(Y {}^t Y) - E(Y)E({}^t Y).$$

b) On calcule l'espérance de BY , puis on factorise B dans le produit définissant la variance.

Soit $B = (b_{i,l}) \in \mathcal{M}_{r,n}(\mathbb{R})$. Les composantes

$$(BY)_i = \sum_{l=1}^n b_{i,l} Y_l$$

sont des combinaisons linéaires des Y_l ; leurs produits deux à deux sont des combinaisons linéaires des $Y_l Y_{l'}$, qui admettent une espérance. Donc BY admet une espérance et une matrice de variance-covariance.

Méthode 1 — Calcul matriciel.

D'après 2) a), $E(BY) = Bm$ avec $m = E(Y)$. Donc $BY - E(BY) = B(Y - m)$ et, puisque ${}^t(B(Y - m)) = {}^t(Y - m) {}^t B$:

$$(BY - E(BY)) {}^t (BY - E(BY)) = B (Y - m) {}^t (Y - m) {}^t B$$

En appliquant 2) a) à la matrice aléatoire $Z = (Y - m) {}^t (Y - m)$, à n lignes et n colonnes, avec la matrice $B \in \mathcal{M}_{r,n}(\mathbb{R})$ à gauche et ${}^t B \in \mathcal{M}_{n,r}(\mathbb{R})$ à droite :

$$V(BY) = B E [(Y - m) {}^t (Y - m)] {}^t B = B V(Y) {}^t B$$

Méthode 2 — Bilinearité de la covariance.

Pour $i, j \in \llbracket 1, r \rrbracket$, par bilinéarité de la covariance :

$$\text{Cov}((BY)_i, (BY)_j) = \text{Cov}\left(\sum_{l=1}^n b_{i,l} Y_l, \sum_{l'=1}^n b_{j,l'} Y_{l'}\right) = \sum_{l=1}^n \sum_{l'=1}^n b_{i,l} \text{Cov}(Y_l, Y_{l'}) b_{j,l'}$$

Le membre de droite est le coefficient d'indice (i, j) du produit $BV(Y)^t B$, puisque $(^t B)_{l',j} = b_{j,l'}$. D'après la remarque préliminaire appliquée à BY , le membre de gauche est le coefficient d'indice (i, j) de $V(BY)$.

La méthode 1 utilise uniquement la linéarité de l'espérance matricielle établie en 2) ; la méthode 2 montre que la formule n'est que la bilinéarité de la covariance écrite matriciellement.

Conclusion

$$V(BY) = BV(Y)^t B.$$

c) On écrit le réel $^t Y A Y$ comme une trace pour faire apparaître la matrice $Y^t Y$.

Existence. Avec $A = (a_{i,j})$,

$$^t Y A Y = \sum_{i=1}^n \sum_{j=1}^n a_{i,j} Y_i Y_j$$

est une combinaison linéaire des $Y_i Y_j$, donc admet une espérance. De plus, d'après a), $E(Y^t Y) = V(Y) + E(Y)E(^t Y) = J + m^t m$.

Méthode 1 — Utilisation de la trace.

La matrice tYAY est de taille 1×1 , donc égale à sa trace. On applique $\text{Tr}(CD) = \text{Tr}(DC)$ avec $C = {}^tY \in \mathcal{M}_{1,n}$ et $D = AY \in \mathcal{M}_{n,1}$:

$${}^tYAY = \text{Tr}({}^tY(AY)) = \text{Tr}((AY){}^tY) = \text{Tr}(AY{}^tY)$$

Par 2) b) puis 2) a) appliqué à la matrice aléatoire $Y{}^tY$:

$$E({}^tYAY) = E(\text{Tr}(AY{}^tY)) = \text{Tr}(E(AY{}^tY)) = \text{Tr}(A \cdot E(Y{}^tY))$$

Ensuite, par linéarité de la trace :

$$E({}^tYAY) = \text{Tr}(A(J + m{}^tm)) = \text{Tr}(AJ) + \text{Tr}((Am){}^tm)$$

et, à nouveau par $\text{Tr}(CD) = \text{Tr}(DC)$ avec $C = Am \in \mathcal{M}_{n,1}$ et $D = {}^tm \in \mathcal{M}_{1,n}$,
 $\text{Tr}((Am){}^tm) = \text{Tr}({}^tmAm) = {}^tmAm$ (matrice 1×1).

Méthode 2 — Calcul en coordonnées.

Par linéarité de l'espérance et définition de la covariance :

$$E({}^tYAY) = \sum_{i=1}^n \sum_{j=1}^n a_{i,j} E(Y_i Y_j) = \sum_{i=1}^n \sum_{j=1}^n a_{i,j} (\text{Cov}(Y_i, Y_j) + m_i m_j)$$

D'une part

$$\sum_{i,j} a_{i,j} m_i m_j = {}^t m A m.$$

D'autre part, J étant symétrique, $\text{Cov}(Y_i, Y_j) = J_{j,i}$, donc

$$\sum_i \sum_j a_{i,j} J_{j,i} = \sum_i (AJ)_{i,i} = \text{Tr}(AJ).$$

Enfin

$$\sum_{i,j} a_{i,j} E(Y_i Y_j) = \sum_i (A E(Y {}^t Y))_{i,i} = \text{Tr}(A \cdot E(Y {}^t Y))$$

car $E(Y {}^t Y)$ est symétrique.

La méthode 1 est l'astuce de la trace, très efficace pour les formes quadratiques aléatoires ; la méthode 2 est un calcul direct, plus long mais sans outil particulier.

Conclusion

$$E({}^t Y A Y) = \text{Tr}(A \cdot E(Y {}^t Y)) = \text{Tr}(AJ) + {}^t m A m.$$

Partie II. Le modèle linéaire

A. Quelques résultats algébriques

1) a) On calcule la transposée de H .

$H = {}^t X X$ est le produit d'une matrice de $\mathcal{M}_{k,n}(\mathbb{R})$ par une matrice de $\mathcal{M}_{n,k}(\mathbb{R})$: c'est une matrice réelle de $\mathcal{M}_k(\mathbb{R})$. De plus :

$${}^t H = {}^t ({}^t X X) = {}^t X {}^t ({}^t X) = {}^t X X = H$$

Conclusion

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

- b) On montre que le noyau de h est réduit à $\{0\}$ grâce à une norme, puis on applique le théorème du rang.

Soit $x \in \mathbb{R}^k$ tel que $h(x) = 0$, c'est-à-dire $Hx = 0_k$. Alors :

$$0 = {}^t x H x = {}^t x {}^t X X x = {}^t (X x) (X x) = \|X x\|^2$$

donc $Xx = 0_n$, et par injectivité de f (voir les conventions), $x = 0_k$. Ainsi $\text{Ker}(h) = \{0\}$.

Par le théorème du rang appliqué à l'endomorphisme h de $\mathbb{R}^k$:

$$\text{rg}(h) = k - \dim \text{Ker}(h) = k.$$

Un endomorphisme d'un espace de dimension finie k et de rang k est surjectif, donc bijectif; sa matrice H est inversible.

Conclusion

$\text{rg}(h) = k$ et H est inversible.

- 2) On cherche les $\alpha \in \mathbb{R}^k$ minimisant $\|y - X\alpha\|$. Comme la norme est positive et que $t \mapsto t^2$ est strictement croissante sur $[0, +\infty[$, cela revient à minimiser $\|y - X\alpha\|^2$.

Rectification de l'énoncé : le vecteur $\hat{\alpha}$ appartient à $\mathbb{R}^k$; ses composantes sont $\widehat{\alpha}_1, \dots, \widehat{\alpha}_k$ (le texte original indiquait $\widehat{\alpha}_n$ pour la dernière composante).

Posons $\hat{\alpha} = H^{-1} {}^t X y$, bien défini d'après 1) b). Il vérifie l'équation dite normale :

$${}^t X (y - X \hat{\alpha}) = {}^t X y - H \hat{\alpha} = {}^t X y - {}^t X y = 0_k$$

Méthode 1 — Décomposition orthogonale et théorème de Pythagore.

Soit $\alpha \in \mathbb{R}^k$. On écrit $y - X\alpha = (y - X\hat{\alpha}) + X(\hat{\alpha} - \alpha)$. Les deux termes sont orthogonaux :

$$\langle X(\hat{\alpha} - \alpha), y - X\hat{\alpha} \rangle = {}^t(\hat{\alpha} - \alpha) {}^tX(y - X\hat{\alpha}) = {}^t(\hat{\alpha} - \alpha) 0_k = 0$$

Par le théorème de Pythagore :

$$\|y - X\alpha\|^2 = \|y - X\hat{\alpha}\|^2 + \|X(\hat{\alpha} - \alpha)\|^2 \geq \|y - X\hat{\alpha}\|^2$$

Donc $\hat{\alpha}$ réalise le minimum. De plus, il y a égalité si et seulement si $X(\hat{\alpha} - \alpha) = 0_n$, c'est-à-dire, par injectivité de f , si et seulement si $\alpha = \hat{\alpha}$. Le minimum est donc atteint en $\hat{\alpha}$ et seulement en ce point.

Méthode 2 — Fonction de plusieurs variables.

Soit $g : \mathbb{R}^k \rightarrow \mathbb{R}$, $g(\alpha) = \|y - X\alpha\|^2$. En développant :

$$g(\alpha) = {}^tyy - 2 {}^t\alpha {}^tXy + {}^t\alpha H\alpha = {}^tyy - 2 \sum_{l=1}^k \alpha_l ({}^tXy)_l + \sum_{l=1}^k \sum_{l'=1}^k h_{l,l'} \alpha_l \alpha_{l'}$$

où l'on a utilisé ${}^t\alpha {}^tXy = {}^tyX\alpha$ (réel égal à son transposé). La fonction g est polynomiale en $(\alpha_1, \dots, \alpha_k)$, donc de classe C^2 sur l'ouvert convexe $\mathbb{R}^k$.

Gradient. Pour $l \in \llbracket 1, k \rrbracket$, en utilisant la symétrie de H :

$$\partial_l g(\alpha) = -2({}^t X y)_l + \sum_{l'=1}^k h_{l,l'} \alpha_{l'} + \sum_{l'=1}^k h_{l',l} \alpha_{l'} = 2(H\alpha - {}^t X y)_l$$

donc $\nabla g(\alpha) = 2(H\alpha - {}^t X y)$. Il s'annule si et seulement si $H\alpha = {}^t X y$, c'est-à-dire, H étant inversible, si et seulement si $\alpha = \hat{\alpha}$: g possède un unique point critique, $\hat{\alpha}$.

Hessienne. Pour tous l, l' , $\partial_{l'} \partial_l g(\alpha) = 2h_{l,l'}$, donc la hessienne de g vaut $2H$ en tout point de $\mathbb{R}^k$.

Signe des valeurs propres. La matrice $2H$ est symétrique réelle. Soit λ une valeur propre de H et $x \neq 0_k$ un vecteur propre associé. Alors :

$$\lambda \|x\|^2 = {}^t x(\lambda x) = {}^t x H x = \|Xx\|^2 > 0$$

car $Xx \neq 0_n$ par injectivité de f . Donc $\lambda > 0$, et les valeurs propres de la hessienne $2H$ sont strictement positives en tout point.

Conclusion de la méthode 2. La hessienne de g a toutes ses valeurs propres positives en tout point de l'ouvert convexe $\mathbb{R}^k$, donc tout point critique de g est un minimum global : g atteint son minimum global en $\hat{\alpha}$. Réciproquement, un point où g atteint son minimum sur l'ouvert $\mathbb{R}^k$ est un point critique (condition nécessaire du premier ordre), donc c'est $\hat{\alpha}$: le minimum est atteint en un unique point.

La méthode 1 est géométrique et donne l'écart exact $\|X(\hat{\alpha} - \alpha)\|^2$ au minimum ; la méthode 2 est la méthode systématique d'optimisation (gradient, hessienne, convexité), utile quand aucune décomposition orthogonale n'est visible.

Conclusion

le problème admet une unique solution, $\hat{\alpha} = H^{-1} {}^t X y$. On remarque que le raisonnement n'utilise aucune propriété particulière de y : il est valable pour tout vecteur de $\mathbb{R}^n$ à la place de y .

- 3) a) On identifie $p(y)$ grâce à une caractérisation de la projection orthogonale sur $\mathcal{V}$, puis on en déduit la matrice P .

D'après les conventions, p est le projecteur orthogonal sur $\mathcal{V} = \text{Im}(f) = \{X\beta; \beta \in \mathbb{R}^k\}$.

Méthode 1 — Caractérisation par minimisation de la distance.

Le vecteur $p(y)$ est l'unique vecteur z de $\mathcal{V}$ tel que $\|y - z\|$ soit minimal parmi les vecteurs de $\mathcal{V}$. Or les vecteurs de $\mathcal{V}$ sont exactement les $X\beta$, $\beta \in \mathbb{R}^k$, et d'après 2) la quantité $\|y - X\beta\|$ est minimale pour $\beta = \hat{\alpha}$. Le vecteur $X\hat{\alpha}$ appartient à $\mathcal{V}$ et réalise ce minimum, donc $p(y) = X\hat{\alpha}$.

Méthode 2 — Caractérisation par l'orthogonalité.

Le vecteur $p(y)$ est l'unique vecteur z de $\mathcal{V}$ tel que $y - z$ appartienne à $\mathcal{V}^\perp$. Le vecteur $X\hat{\alpha}$ appartient à $\mathcal{V}$ et, pour tout $j \in \llbracket 1, k \rrbracket$, $\langle X_j, y - X\hat{\alpha} \rangle$ est la j -ième composante de ${}^tX(y - X\hat{\alpha})$, qui est nulle d'après l'équation normale établie en 2). Le vecteur $y - X\hat{\alpha}$ est orthogonal à la famille génératrice $(X_1, \dots, X_k)$ de $\mathcal{V}$, donc à $\mathcal{V}$. Ainsi $p(y) = X\hat{\alpha}$.

La méthode 1 réutilise directement la question 2) ; la méthode 2 est autonome et s'appuie sur l'équation normale.

Conclusion

$$p(y) = X\hat{\alpha} = XH^{-1}{}^tXy.$$

Expression de P . Comme remarqué à la fin de 2), ce raisonnement vaut pour tout vecteur $z \in \mathbb{R}^n$ à la place de y : pour tout $z \in \mathbb{R}^n$, $Pz = p(z) = XH^{-1}{}^tXz$. En appliquant cette égalité aux vecteurs de la base canonique, on obtient que les deux matrices ont les mêmes colonnes, donc $P = XH^{-1}{}^tX$.

Vérifications. Comme H est symétrique et inversible, ${}^t(H^{-1}) = ({}^tH)^{-1} = H^{-1}$. Alors :

$$P^2 = XH^{-1}({}^tXX)H^{-1}{}^tX = XH^{-1}HH^{-1}{}^tX = XH^{-1}{}^tX = P$$

$${}^tP = {}^t({}^tX) {}^t(H^{-1}) {}^tX = XH^{-1}{}^tX = P$$

Conclusion

$$P = XH^{-1}{}^tX \text{ et } P = P^2 = {}^tP.$$

b) On calcule séparément le rang et la trace.

Méthode 1 — Calcul direct.

Rang. Le rang de P est celui de p , c'est-à-dire $\dim \text{Im}(p)$. Le projecteur orthogonal sur $\mathcal{V}$ a pour image $\mathcal{V}$ (tout vecteur de $\mathcal{V}$ est son propre projeté). Donc $\text{rg}(P) = \dim \mathcal{V} = k$.

Trace. On applique $\text{Tr}(CD) = \text{Tr}(DC)$ avec $C = X \in \mathcal{M}_{n,k}(\mathbb{R})$ et $D = H^{-1}{}^tX \in \mathcal{M}_{k,n}(\mathbb{R})$:

$$\text{Tr}(P) = \text{Tr}(X(H^{-1}{}^tX)) = \text{Tr}(H^{-1}{}^tXX) = \text{Tr}(H^{-1}H) = \text{Tr}(I_k) = k$$

Méthode 2 — Par les éléments propres et la question I.1) a).

La matrice P est symétrique réelle, donc diagonalisable sur $\mathbb{R}$ (théorème spectral). Si λ est une valeur propre de P et $x \neq 0$ un vecteur propre associé, alors

$\lambda x = Px = P^2x = \lambda^2x$, donc $\lambda^2 = \lambda$ et $\lambda \in \{0, 1\}$. De plus $\text{Ker}(P - I_n) = \text{Im}(P)$: si

$Px = x$ alors $x \in \text{Im}(P)$, et si $x = Pz$ alors $Px = P^2z = Pz = x$. Comme

$\text{Im}(P) = \mathcal{V} \neq \{0\}$, la valeur 1 est valeur propre, d'espace propre de dimension $\text{rg}(P)$.

Par I.1) a), la valeur propre éventuelle 0 contribuant pour 0 :

$$\text{Tr}(P) = 1 \times \dim \text{Ker}(P - I_n) = \text{rg}(P) = \dim \mathcal{V} = k$$

La méthode 1 utilise la forme explicite de P ; la méthode 2 montre plus généralement que la trace d'un projecteur symétrique est égale à son rang.

Conclusion

$$\text{rg}(P) = \text{Tr}(P) = k.$$

c) On calcule PX .

$$PX = XH^{-1} {}^tXX = XH^{-1}H = X$$

En lisant cette égalité colonne par colonne, $PX_j = X_j$ pour tout $j \in \llbracket 1, k \rrbracket$. Chaque X_j est non nul (la famille des colonnes est libre), donc c'est un vecteur propre de P associé à la valeur propre 1.

La famille $(X_1, \dots, X_k)$ est libre et formée de k vecteurs du sous-espace propre $E_1(P) = \text{Ker}(P - I_n)$. Or, d'après b) (méthode 2) ou directement, $E_1(P) = \text{Im}(P) = \mathcal{V}$ est de dimension k . Une famille libre de k vecteurs dans un espace de dimension k en est une base.

Conclusion

les colonnes de X forment une base du sous-espace propre de P associé à la valeur propre 1 (qui est $\mathcal{V}$). Comme $k < n$, ce n'est pas une base de $\mathbb{R}^n$: les autres vecteurs propres, associés à 0, forment $\mathcal{V}^\perp$.

- d) On cherche une base orthonormée de $\mathbb{R}^n$ formée de vecteurs propres de P , les k premiers associés à 1.

Méthode 1 — Bases orthonormées adaptées à $\mathcal{V}$ et $\mathcal{V}^\perp$.

Le sous-espace $\mathcal{V}$, muni du produit scalaire induit, est un espace euclidien de dimension k : il possède une base orthonormée $(s_1, \dots, s_k)$. Le supplémentaire orthogonal $\mathcal{V}^\perp$ est de dimension $n - k \geq 1$ et possède aussi une base orthonormée $(s_{k+1}, \dots, s_n)$. Chaque s_i ($i \leq k$) est orthogonal à chaque s_j ($j > k$), donc $(s_1, \dots, s_n)$ est une famille orthonormée de n vecteurs de $\mathbb{R}^n$; elle est libre, donc c'est une base orthonormée de $\mathbb{R}^n$.

Soit S la matrice dont les colonnes sont $s_1, \dots, s_n$. C'est la matrice de passage de la base canonique (orthonormée) à la base orthonormée $(s_1, \dots, s_n)$, donc S est orthogonale : ${}^tSS = I_n$, c'est-à-dire $S^{-1} = {}^tS$.

Par définition de la projection orthogonale sur $\mathcal{V}$: $p(s_i) = s_i$ si $i \leq k$ (car $s_i \in \mathcal{V}$) et $p(s_i) = 0$ si $i > k$ (car $s_i \in \mathcal{V}^\perp$). La matrice de p dans la base $(s_1, \dots, s_n)$ est donc D . Par la formule de changement de base :

$$P = SDS^{-1} = SD{}^tS$$

Méthode 2 — Théorème spectral.

P est symétrique réelle : il existe une matrice orthogonale S_0 et une matrice diagonale Δ telles que $P = S_0\Delta{}^tS_0$, soit $PS_0 = S_0\Delta$. Les coefficients diagonaux de Δ sont des valeurs propres de P , donc valent 0 ou 1 (voir b)), et la colonne j de S_0 est un vecteur propre associé au j -ième coefficient diagonal. Comme $\text{Tr}(\Delta) = \text{Tr}(P) = k$ (matrices semblables), exactement k coefficients diagonaux valent 1. Réordonnons les colonnes de S_0 en plaçant d'abord les k colonnes associées à 1 : on obtient une matrice S dont les colonnes $s_1, \dots, s_n$ forment toujours une famille orthonormée, donc S est orthogonale, et $Ps_i = s_i$ pour $i \leq k$, $Ps_i = 0$ pour $i > k$. Ainsi $PS = SD$, d'où $P = SDS^{-1} = SD{}^tS$.

La méthode 1 construit S à partir de la géométrie du problème ; la méthode 2 applique le théorème spectral sans construction.

Les k premières colonnes de S . Réciproquement, si S est orthogonale et $P = SD^tS$, alors $PS = SD$, donc les k premières colonnes $s_1, \dots, s_k$ de S vérifient $Ps_i = s_i$: elles appartiennent à $E_1(P) = \mathcal{V}$. Elles sont orthonormées (colonnes d'une matrice orthogonale), donc libres, et $\dim \mathcal{V} = k$.

Conclusion

il existe S orthogonale telle que $P = SD^tS$, et les k premières colonnes de S forment une base orthonormée du sous-espace $\mathcal{V}$ engendré par les colonnes de X (qui est l'espace propre de P associé à 1). Elles ne sont en général pas égales aux colonnes de X , qui ne sont pas nécessairement orthonormées ; toute base orthonormée de $\mathcal{V}$ convient.

4) a) On utilise $QX = 0$, puis les propriétés de P .

D'après 3) c), $PX = X$, donc $QX = X - PX = 0$. Comme $u = y - X\alpha$:

$$Qu = Qy - QX\alpha = Qy = y - Py = y - p(y) = y - X\hat{\alpha} = \hat{u}$$

en utilisant 3) a). Ensuite, puisque $P^2 = P$ et ${}^tP = P$:

$$Q^2 = I_n - 2P + P^2 = I_n - P = Q \text{ et } {}^tQ = I_n - {}^tP = I_n - P = Q$$

Enfin, par linéarité de la trace et 3) b) :

$$\text{Tr}(Q) = \text{Tr}(I_n) - \text{Tr}(P) = n - k$$

Conclusion

$$\hat{u} = Qu, Q = Q^2 = {}^tQ \text{ et } \text{Tr}(Q) = n - k.$$

b) On utilise $Q = Q^2 = {}^tQ$ et $Qy = Qu$.

D'une part :

$${}^t\hat{u} \hat{u} = {}^t(Qu)(Qu) = {}^t_u {}^tQ Qu = {}^t_u Q^2 u = {}^t_u Qu$$

D'autre part, on a vu en a) que $Qy = Qu$, donc :

$${}^t_y Qy = {}^t_y Q^2 y = {}^t_y {}^tQ Qy = {}^t(Qy)(Qy) = {}^t(Qu)(Qu) = {}^t_u Qu$$

Conclusion

$${}^t\hat{u} \hat{u} = {}^t_y Qy = {}^t_u Qu \text{ (et cette quantité vaut aussi } \|\hat{u}\|^2 \text{)}.$$

5) a) On utilise le théorème spectral. La définition et la preuve ci-dessous valent pour une matrice symétrique réelle d'ordre quelconque (on l'utilisera à l'ordre k).

Soit A symétrique réelle d'ordre n . Par le théorème spectral, il existe une matrice orthogonale S et une matrice diagonale $\Delta = \text{diag}(\mu_1, \dots, \mu_n)$ telles que $A = S\Delta {}^tS$; les réels $\mu_1, \dots, \mu_n$ sont les valeurs propres de A (avec répétitions), et toute valeur propre de A figure parmi eux.

Sens direct. Supposons A positive. Soit μ une valeur propre de A et $z \neq 0$ un vecteur propre associé. Alors :

$$0 \leq {}^t_z Az = {}^t_z (\mu z) = \mu \|z\|^2$$

et $\|z\|^2 > 0$, donc $\mu \geq 0$.

Réciproque. Supposons $\mu_i \geq 0$ pour tout i . Soit $z \in \mathbb{R}^n$ et $w = {}^tS z = {}^t(w_1 \dots w_n)$. Alors :

$${}^t_z A z = {}^t_z S \Delta {}^tS z = {}^t_w \Delta w = \sum_{i=1}^n \mu_i w_i^2 \geq 0$$

Conclusion

une matrice symétrique réelle est positive si et seulement si toutes ses valeurs propres sont positives ou nulles.

b) On vérifie la symétrie puis on reconnaît une norme au carré.

tLL appartient à $\mathcal{M}_k(\mathbb{R})$ et ${}^t({}^tLL) = {}^tL {}^t({}^tL) = {}^tLL$: elle est symétrique réelle. Pour tout $z \in \mathbb{R}^k$:

$${}^t_z {}^tLL z = {}^t(Lz)(Lz) = \|Lz\|^2 \geq 0$$

Conclusion

tLL est symétrique réelle positive (au sens de la définition, à l'ordre k).

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

1) a) On utilise la linéarité de l'espérance (I.2)) et la définition de V .

Pour tout i ,

$$Y_i = \sum_{j=1}^k x_{i,j} \alpha_j + U_i,$$

où

$$\sum_j x_{i,j} \alpha_j$$

est une constante. Par linéarité,

$$E(Y_i) = \sum_j x_{i,j} \alpha_j + E(U_i) = (X\alpha)_i,$$

puisque $E(U) = 0_n$. Donc $E(Y) = X\alpha$.

Ensuite $Y - E(Y) = U = U - E(U)$, donc :

$$V(Y) = E[(Y - E(Y))^t (Y - E(Y))] = E[(U - E(U))^t (U - E(U))] = V(U) = \sigma^2 I_n$$

Enfin, $\widehat{G} = (H^{-1} {}^t X)Y$ avec $H^{-1} {}^t X \in \mathcal{M}_{k,n}(\mathbb{R})$ constante. Par I.2) a) :

$$E(\widehat{G}) = H^{-1} {}^t X E(Y) = H^{-1} {}^t X X \alpha = H^{-1} H \alpha = \alpha$$

$\widehat{G}$ ne dépend que de Y et de la matrice connue X : c'est un estimateur de α , et sa réalisation sur l'échantillon est $H^{-1} {}^t X y = \hat{\alpha}$.

Conclusion

$E(Y) = X\alpha$, $V(Y) = \sigma^2 I_n$ et $E(\widehat{G}) = \alpha$: $\widehat{G}$ est un estimateur sans biais de α et $\hat{\alpha}$ en est l'estimation associée à l'échantillon observé.

b) On applique I.3) b) avec $B = H^{-1} {}^t X \in \mathcal{M}_{k,n}(\mathbb{R})$ (les composantes de Y admettent des moments d'ordre 2).

$$V(\widehat{G}) = (H^{-1} {}^tX) V(Y) {}^t(H^{-1} {}^tX) = \sigma^2 H^{-1} {}^tX X {}^t(H^{-1}) = \sigma^2 H^{-1} H H^{-1} = \sigma^2 H^{-1}$$

en utilisant ${}^t(H^{-1}) = H^{-1}$ (voir A.3 a)).

Conclusion

$$V(\widehat{G}) = \sigma^2 H^{-1}.$$

2) a) On calcule l'espérance de $\widehat{C}$ par I.2) a).

$E(\widehat{C}) = {}^tBE(Y) = {}^tBX\alpha$. L'estimateur $\widehat{C}$ est sans biais pour tout $\alpha \in \mathbb{R}^k$ si et seulement si ${}^tBX\alpha = \alpha$ pour tout $\alpha \in \mathbb{R}^k$.

Si ${}^tBX = I_k$, c'est vrai. Réciproquement, si c'est vrai pour tout α , en prenant pour α le j -ième vecteur e_j de la base canonique de $\mathbb{R}^k$, on obtient que la j -ième colonne de tBX est e_j , pour tout j : ${}^tBX = I_k$.

Conclusion

$\widehat{C}$ est un estimateur sans biais de α pour tout α si et seulement si ${}^tBX = I_k$ (condition équivalente, par transposition, à ${}^tXB = I_k$). Une telle matrice B est automatiquement non nulle.

b) On calcule tFX , puis on développe $V(\widehat{C})$ en écrivant B à l'aide de F .

Calcul de tFX :

$${}^tFX = {}^tBX - H^{-1} {}^tXX = I_k - H^{-1}H = I_k - I_k = 0$$

Par transposition, ${}^tXF = {}^t({}^tFX) = 0$.

Variance de $\widehat{C}$. Par I.3) b) avec la matrice ${}^tB \in \mathcal{M}_{k,n}(\mathbb{R})$ et B.1) a) :

$$V(\widehat{C}) = {}^tB(\sigma^2 I_n) {}^t({}^tB) = \sigma^2 {}^tBB$$

Or ${}^tB = {}^tF + H^{-1} {}^tX$ et, en transposant (avec ${}^t(H^{-1}) = H^{-1}$), $B = F + XH^{-1}$. Donc :

$${}^tBB = ({}^tF + H^{-1} {}^tX)(F + XH^{-1}) = {}^tFF + ({}^tFX)H^{-1} + H^{-1}({}^tXF) + H^{-1}({}^tXX)H^{-1}$$

Les deux termes du milieu sont nuls et le dernier vaut $H^{-1}HH^{-1} = H^{-1}$. Ainsi

${}^tBB = {}^tFF + H^{-1}$ et, avec B.1) b) :

$$V(\widehat{C}) - V(\widehat{G}) = \sigma^2({}^tFF + H^{-1}) - \sigma^2 H^{-1} = \sigma^2 {}^tFF$$

Comme $F \in \mathcal{M}_{n,k}(\mathbb{R})$, la matrice tFF est symétrique réelle positive d'après A.5) b).

Pour tout $z \in \mathbb{R}^k$, ${}^tz(\sigma^2 {}^tFF)z = \sigma^2 {}^tz {}^tFFz \geq 0$ car $\sigma^2 > 0$, donc $\sigma^2 {}^tFF$ est aussi symétrique réelle positive.

Conclusion

${}^tFX = 0$ et $V(\widehat{C}) - V(\widehat{G}) = \sigma^2 {}^tFF$ est positive : $\widehat{G}$ est optimal parmi les estimateurs sans biais de la forme tBY . En particulier, en prenant pour z le j -ième vecteur de la base canonique, $V(\widehat{C}_j) \geq V(\widehat{G}_j)$ pour chaque composante j .

3) a) On reconnaît P dans $X\widehat{G}$.

$X\widehat{G} = XH^{-1} {}^tXY = PY$, donc $\widehat{U} = Y - PY = QY$. Comme $Y = X\alpha + U$ et $QX = 0$ (A.4) a)) :

$$\widehat{U} = QX\alpha + QU = QU$$

Conclusion

$$\widehat{U} = QU.$$

b) On applique I.2) a) et I.3) b) avec la matrice constante Q , puis on étudie les covariances.

Espérance et variance. $E(\widehat{U}) = QE(U) = Q0_n = 0_n$ et, puisque ${}^tQ = Q$ et $Q^2 = Q$:

$$V(\widehat{U}) = QV(U){}^tQ = \sigma^2QQ = \sigma^2Q$$

Indépendance. D'après la remarque préliminaire de I.3), pour $i \neq j$:

$$\text{Cov}(\widehat{U}_i, \widehat{U}_j) = \sigma^2 q_{i,j}$$

Si la matrice Q possède un coefficient non diagonal $q_{i,j} \neq 0$, alors $\text{Cov}(\widehat{U}_i, \widehat{U}_j) \neq 0$, donc $\widehat{U}_i$ et $\widehat{U}_j$ ne sont pas indépendantes (des variables indépendantes ont une covariance nulle), et a fortiori $\widehat{U}_1, \dots, \widehat{U}_n$ ne sont pas mutuellement indépendantes.

Exemple : $n = 2, k = 1, X = {}^t(1 \ 1)$. Alors $H = 2$,

$$P = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

$$Q = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

et $\text{Cov}(\widehat{U}_1, \widehat{U}_2) = -\frac{\sigma^2}{2} \neq 0$: $\widehat{U}_1 = \frac{U_1 - U_2}{2}$ et $\widehat{U}_2 = -\widehat{U}_1$ ne sont pas indépendantes.

Cas particulier où Q est diagonale. Alors $Q^2 = Q$ donne $q_{i,i}^2 = q_{i,i}$, donc $q_{i,i} \in \{0, 1\}$, et $\widehat{U}_i = q_{i,i} U_i$. Chaque $\widehat{U}_i$ est une fonction de la seule variable U_i ; par le lemme des coalitions, $\widehat{U}_1, \dots, \widehat{U}_n$ sont mutuellement indépendantes. Ce cas se produit par exemple si les colonnes de X sont les k premiers vecteurs de la base canonique : alors $\widehat{U} = {}^t(0 \dots 0 U_{k+1} \dots U_n)$.

Conclusion

$E(\widehat{U}) = 0_n$ et $V(\widehat{U}) = \sigma^2 Q$. En général, les variables $\widehat{U}_1, \dots, \widehat{U}_n$ ne sont pas indépendantes : elles le sont si et seulement si Q est diagonale, ce qui est exceptionnel.

c) On utilise $\widehat{U} = QU = QY$ et $Q = Q^2 = {}^tQ$.

Le produit de la ligne ${}^t\widehat{U}$ par la colonne $\widehat{U}$ vaut

$${}^t\widehat{U}\widehat{U} = \sum_{i=1}^n \widehat{U}_i^2.$$

Ensuite :

$${}^t\widehat{U}\widehat{U} = {}^t(QU)(QU) = {}^tU {}^tQQU = {}^tUQ^2U = {}^tUQU$$

et, comme $\widehat{U} = QY$ (vu en a)) :

$${}^tYQY = {}^tYQ^2Y = {}^tY {}^tQQY = {}^t(QY)(QY) = {}^t\widehat{U}\widehat{U}$$

Conclusion

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

d) On applique I.3) c) au vecteur U avec $A = Q$.

Les U_i admettent des moments d'ordre 2, donc chaque $U_i U_j$ admet une espérance (car $|U_i U_j| \leq \frac{1}{2}(U_i^2 + U_j^2)$) : les hypothèses de I.3) sont satisfaites, avec $m = E(U) = 0_n$ et $J = V(U) = \sigma^2 I_n$. D'après c) et I.3) c) :

$$E({}^t \widehat{U} \widehat{U}) = E({}^t U Q U) = \text{Tr}(Q \sigma^2 I_n) + {}^t 0_n Q 0_n = \sigma^2 \text{Tr}(Q) = \sigma^2(n - k)$$

d'après A.4) a). Comme $n - k \geq 1$, $s_n = \frac{{}^t Y Q Y}{n - k}$ est bien défini ; il ne dépend que de Y et de la matrice connue X (par Q), donc c'est un estimateur, et par linéarité :

$$E(s_n) = \frac{\sigma^2(n - k)}{n - k} = \sigma^2$$

Conclusion

$$E({}^t \widehat{U} \widehat{U}) = (n - k)\sigma^2 \text{ et } s_n \text{ est un estimateur sans biais de } \sigma^2.$$

C. Étude d'une suite d'estimateurs

Dans cette partie, pour chaque $n \geq k + 1$, la matrice X (de rang k), les matrices P et Q ainsi que le vecteur $U = {}^t(U_1 \dots U_n)$ dépendent de n ; on garde les notations de la partie B. On utilise $E(U_i) = 0$ et $E(U_i^2) = V(U_i) = \sigma^2$ (coefficients diagonaux de $V(U) = \sigma^2 I_n$), ainsi que $E(U_i^3) = 0$ et $E(U_i^4) = 3\sigma^4$.

1) On développe le produit matriciel.

La i -ième composante de QU est

$$\sum_{j=1}^n q_{i,j} U_j,$$

donc :

$${}^tUQU = \sum_{i=1}^n U_i (QU)_i = \sum_{i=1}^n U_i \sum_{j=1}^n q_{i,j} U_j = \sum_{i=1}^n \sum_{j=1}^n q_{i,j} U_i U_j$$

Conclusion

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

- 2) On développe le carré en une somme quadruple et on calcule chaque espérance grâce à l'indépendance.

D'après 1), en développant le produit de deux sommes finies :

$$({}^tUQU)^2 = \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{m=1}^n q_{i,j} q_{l,m} U_i U_j U_l U_m$$

Existence et calcul de $E(U_i U_j U_l U_m)$. En regroupant les indices égaux, $U_i U_j U_l U_m$ est un produit de puissances (d'exposants au plus 4) de variables U_a d'indices distincts.

Chaque U_a admet des moments d'ordre 1 à 4. Par le lemme des coalitions, ces puissances de variables d'indices distincts sont mutuellement indépendantes, donc leur produit admet une espérance égale au produit des espérances. On distingue selon la répartition des quatre indices :

- si $i = j = l = m$: $E(U_i^4) = 3\sigma^4$;
- si les indices forment deux paires de valeurs distinctes, ce qui correspond à l'un des trois cas disjoints $(i = j, l = m, i \neq l)$, $(i = l, j = m, i \neq j)$, $(i = m, j = l, i \neq j)$: l'espérance vaut $E(U_a^2)E(U_b^2) = \sigma^4$;
- sinon, une des valeurs d'indice apparaît exactement une fois (répartitions $3 + 1$, $2 + 1 + 1$ ou $1 + 1 + 1 + 1$) : en notant a cet indice, U_a est indépendante du produit des autres facteurs, et l'espérance vaut $E(U_a) \times (\dots) = 0$.

Par conséquent :

$$E[({}^tUQU)^2] = 3\sigma^4 \sum_i q_{i,i}^2 + \sigma^4 \left(\sum_{i \neq l} q_{i,i} q_{l,l} + \sum_{i \neq j} q_{i,j} q_{i,j} + \sum_{i \neq j} q_{i,j} q_{j,i} \right)$$

Comme Q est symétrique, $q_{i,j} q_{j,i} = q_{i,j}^2$. En répartissant

$$3\sigma^4 \sum_i q_{i,i}^2$$

en

$$\sigma^4 \sum_i q_{i,i}^2 + 2\sigma^4 \sum_i q_{i,i}^2 :$$

$$E[({}^tUQU)^2] = \sigma^4 \left(\sum_i q_{i,i}^2 + \sum_{i \neq l} q_{i,i} q_{l,l} \right) + 2\sigma^4 \left(\sum_i q_{i,i}^2 + \sum_{i \neq j} q_{i,j}^2 \right)$$

$$E[({}^tUQU)^2] = \sigma^4 \left(\sum_{i=1}^n q_{i,i} \right)^2 + 2\sigma^4 \sum_{i=1}^n \sum_{j=1}^n q_{i,j}^2$$

Le premier terme est $\sigma^4(\text{Tr}(Q))^2$. Comme Q est symétrique, I.1) b) donne

$$\sum_{i,j} q_{i,j}^2 = \text{Tr}(Q^2).$$

D'où :

$$E[({}^tUQU)^2] = \sigma^4 [(\text{Tr}(Q))^2 + 2 \text{Tr}(Q^2)]$$

Application. D'après A.4) a), $Q^2 = Q$ et $\text{Tr}(Q) = n - k$, donc $\text{Tr}(Q^2) = n - k$ et :

$$E[({}^tUQU)^2] = \sigma^4 [(n - k)^2 + 2(n - k)] = \sigma^4(n - k)(n - k + 2)$$

Conclusion

$$E[({}^tUQU)^2] = \sigma^4 [(\text{Tr}(Q))^2 + 2 \text{Tr}(Q^2)] = \sigma^4(n - k)(n - k + 2).$$

3) On applique la formule de Koenig-Huygens, puis l'inégalité de Bienaymé-Tchebychev.

D'après B.3) c), $s_n = \frac{{}^tUQU}{n - k}$. D'après 2), s_n^2 admet une espérance, donc s_n admet une variance, et :

$$E(s_n^2) = \frac{E[({}^tUQU)^2]}{(n - k)^2} = \frac{\sigma^4(n - k + 2)}{n - k}$$

Avec $E(s_n) = \sigma^2$ (B.3) d)) et la formule de Koenig-Huygens :

$$V(s_n) = E(s_n^2) - (E(s_n))^2 = \sigma^4 \frac{(n - k + 2) - (n - k)}{n - k} = \frac{2\sigma^4}{n - k}$$

Conclusion sur la convergence. Soit $\varepsilon > 0$. Comme s_n admet une variance et $E(s_n) = \sigma^2$, l'inégalité de Bienaymé-Tchebychev donne, pour tout $n \geq k + 1$:

$$0 \leq \mu(|s_n - \sigma^2| \geq \varepsilon) \leq \frac{V(s_n)}{\varepsilon^2} = \frac{2\sigma^4}{(n - k)\varepsilon^2}$$

Le majorant tend vers 0 quand $n \rightarrow +\infty$ (k est fixé), donc par encadrement

$\lim_{n \rightarrow +\infty} \mu(|s_n - \sigma^2| \geq \varepsilon) = 0$. C'est exactement la condition suffisante du cours : un estimateur sans biais dont la variance tend vers 0 est convergent.

Conclusion

$V(s_n) = \frac{2\sigma^4}{n-k}$, et la suite $(s_n)_{n \geq k+1}$ converge en probabilité vers σ^2 : c'est une suite d'estimateurs sans biais et convergente de σ^2 .

Relevé des adaptations

- I.3), préambule : la phrase « 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) » est remplacée par l'admission explicite des propriétés de l'espérance et de la covariance (linéarité et croissance de l'espérance, espérance d'un produit de variables indépendantes, lemme des coalitions, bilinéarité et symétrie de la covariance) pour des variables aléatoires réelles quelconques admettant les moments utilisés. Justification : la matrice de variance-covariance d'un vecteur aléatoire ne figure pas au programme, et l'espérance n'y est pas définie dans un cadre général ; la définition de $V(Y)$ donnée par l'énoncé est conservée et seules les propriétés du programme nécessaires à la suite sont admises.
- II.A.2) : rectification d'une coquille, la dernière composante de $\hat{\alpha}$ est $\hat{\alpha}_k$ et non $\hat{\alpha}_n$.
Justification : $\hat{\alpha}$ appartient à $\mathbb{R}^k$.