ANNALES DE L'I. H. P., SECTION B

J.-L. PHILOCHE

À propos du théorème de Gauss-Markov

Annales de l'I. H. P., section B, tome 7, nº 4 (1971), p. 271-281
<http://www.numdam.org/item?id=AIHPB_1971__7_4_271_0>

Ann. Inst. Henri Poincaré, Section B :
Vol. VII, n° 4, 1971, p. 271-281. *Calcul des Probabilités et Statistique.*

A propos du théorème de Gauss-Markov

par

J.-L. PHILOCHE [1]

SUMMARY. — The purpose of this paper is the study of possible extension of the mean square procedure in estimation to the case where the variance-covariance matrix is degenerate.

Introduction. — Cette note montre de quelle façon s'étend la méthode classique des moindres carrés dans le cas où la matrice des covariances en cause est quelconque, éventuellement dégénérée.

Quoique ce problème ait déjà été étudié dans la littérature statistique, il a paru intéressant de revenir sur le sujet : quelques arguments euclidiens naturels permettent en effet d'exposer, et de relier, avec précision, les principaux résultats relatifs à cette question ; ces arguments permettent également de simplifier l'ensemble des démonstrations.

On commence (§ 1) par rassembler les propriétés algébriques en cause, puis (§ 2) on donne les énoncés des théorèmes de Gauss-Markov et de Kruskal ; on commente ensuite ces énoncés (§ 3), et l'on relie le point de vue adopté à la méthodologie fondée sur les ellipsoïdes indicateurs (§ 4) [2] ; enfin (§ 5) on donne les démonstrations de tous les énoncés présentés.

1. Forme algébrique du théorème de Gauss-Markov.

On désigne par $\mathscr{X}$ un espace euclidien, par $(.,.)$ le produit scalaire de $\mathscr{X}$, par $\mathscr{L}(\mathscr{X})$ l'ensemble des opérateurs linéaires sur $\mathscr{X}$. On rappelle que, pour tout $T \in \mathscr{L}(\mathscr{X})$, l'adjoint T^* de T ($T^* \in \mathscr{L}(\mathscr{X})$) est défini par :

$$(1.1) \qquad (Tx, y) = (x, T^*y), \quad \text{pour tous} \quad x, y \in \mathscr{X}.$$

[1] Je tiens à remercier MM. Philippe COURRÈGE et Edmond MALINVAUD avec qui j'ai pu débattre plusieurs fois, avec grand profit, du sujet de cette note.

[2] Cette méthodologie fournit le cadre de la présentation de E. MALINVAUD [*6*].

On désigne par $\mathbb{Q}^+$ le sous-ensemble de $\mathscr{L}(\mathscr{X})$ constitué par les opérateurs symétriques positifs, c'est-à-dire par les éléments $Q \in \mathscr{L}(\mathscr{X})$ satisfaisant :

$$Q = Q^*, \tag{1.2}$$

$$(Qx, x) \geqslant 0, \quad \text{pour tout} \quad x \in \mathscr{X}. \tag{1.3}$$

PROPOSITION 1. — *Étant donné* $Q \in \mathbb{Q}^+$, *on définit sur* Im Q (3) *un produit scalaire, noté* $(.,.)_Q$, *par la relation :*

$$(Qx, Qy)_Q = (Qx, y), \quad \textit{pour tous} \quad x, y \in \mathscr{X}. \tag{1.4}$$

Dans le cas où Q *est inversible,* $(.,.)_Q$ *est déterminé, sur* $\mathscr{X}$ *entier, par :*

$$(x, y)_Q = (Q^{-1}x, y), \quad \textit{pour tous} \quad x, y \in \mathscr{X}. \tag{1.5}$$

(démonstration au § 5).

Étant donné un sous-espace vectoriel L de $\mathscr{X}$ et un élément $Q \in \mathbb{Q}^+$, on définit en premier lieu $\mathscr{B}_L$ par :

$$T \in \mathscr{B}_L \Leftrightarrow T \in \mathscr{L}(\mathscr{X}) \quad \text{et} \quad Tx = x, \quad \text{pour tout} \quad x \in L; \tag{1.6}$$

on introduit ensuite la classe $\mathscr{G}(Q, L)$ des opérateurs $T \in \mathscr{B}_L$ qui satisfont de plus les deux conditions :

$$T(\operatorname{Im} Q) \subset \operatorname{Im} Q, \tag{1.7}$$

$$(x - Tx, y)_Q = 0, \quad \text{pour tous} \quad x \in \operatorname{Im} Q \quad \text{et} \quad y \in L_\cap \operatorname{Im} Q. \tag{1.8}$$

On notera qu'un élément de $\mathscr{G}(Q, L)$, on dira aussi un *opérateur de Gauss-Markov*, n'est autre chose qu'un opérateur de $\mathscr{B}_L$ dont la restriction à à Im Q coïncide avec le projecteur orthogonal, au sens de $(.,.)_Q$, de Im Q sur $L_\cap$ Im Q (on constate d'autre part facilement que $\mathscr{G}(Q, L)$ est toujours non vide).

Cela étant, pour tout $Q \in \mathbb{Q}^+$, on se propose, en particulier, de minimiser la fonction $T \to \operatorname{Tr}[TQT^*]$ (4), quand T parcourt $\mathscr{B}_L$.

PROPOSITION 2 (forme algébrique du théorème de Gauss-Markov). — *On suppose donnés un espace euclidien* $\mathscr{X}$, *un sous-espace vectoriel* L *de* $\mathscr{X}$ *et un élément* $Q \in \mathbb{Q}^+$.

---

(3) Pour $T \in \mathscr{L}(\mathscr{X})$, les notations Im T et Ker T désignent respectivement l'image et le noyau de T.

(4) On rappelle que, pour $A \in \mathscr{L}(\mathscr{X})$, la trace de A est la quantité

$$\operatorname{Tr} A = \sum_{i=1}^{d} (Ae_i, e_i),$$

où $(e_i)_{1 \leqslant i \leqslant d}$ est une base orthonormale de $\mathscr{X}$.

Concernant un opérateur $T_0 \in \mathscr{B}_L$, *les propriétés* (G-M), (*i*) et (*ii*) *suivantes sont équivalentes :*

(G-M) $$T_0 \in \mathscr{G}(Q, L);$$
(*i*) $$TQT^* - T_0QT_0^* \in \mathbb{Q}^+, \quad \textit{pour tout} \quad T \in \mathscr{B}_L;$$
(*ii*) $$\mathrm{Tr}\,[TQT^*] - \mathrm{Tr}\,[T_0QT_0^*] \geqslant 0, \quad \textit{pour tout} \quad T \in \mathscr{B}_L.$$

(démonstration au § 5).

2. Forme statistique du théorème de Gauss-Markov.

On désigne par X l'application identique de $\mathscr{X}$. On suppose donnée une famille $\{ P; P \in \mathscr{P} \}$ de probabilités sur $\mathscr{X}$ (toujours muni de sa tribu borélienne) une famille $\{ m_P ; P \in \mathscr{P} \}$ d'éléments de L et un élément $Q \in \mathbb{Q}^+$ de telle sorte que, pour chaque $P \in \mathscr{P}$, la variable aléatoire X soit du second ordre par rapport à P, avec m_P comme espérance et Q comme opérateur de covariance (5):

(2.1) $$E_P[X] = m_P \in L, \quad \text{pour tout} \quad P \in \mathscr{P}$$
(2.2) $$D_P[X] = Q \quad , \quad \text{pour tout} \quad P \in \mathscr{P};$$

on déduit alors de la proposition 2 (§ 1) le résultat suivant :

PROPOSITION 2 *bis* (forme statistique du théorème de Gauss-Markov).

(1) *Concernant un opérateur* $T_0 \in \mathscr{B}_L$, *les propriétés* (G-M), (*j*) *et* (*jj*) *suivantes sont équivalentes* (6) *:*

(G-M) $$T_0 \in \mathscr{G}(Q, L);$$
(*j*) $$E_P[((T_0X, x) - (m_P, x))^2] = \operatorname*{Inf}_{T \in \mathscr{B}_L} E_P[((TX, x) - (m_P, x))^2], \quad \textit{pour tout} \quad P \in \mathscr{P} \quad \textit{et tout} \quad x \in \mathscr{X};$$
(*jj*) $$E_P[\| T_0X - m_P \|^2] = \operatorname*{Inf}_{T \in \mathscr{B}_L} E_P[\| TX - m_P \|^2], \quad \textit{pour tout} \quad P \in \mathscr{P}.$$

---

(5) L'espérance $E_P[\phi]$ et l'opérateur de covariance $D_P[\phi]$ d'une variable aléatoire ϕ, à valeur dans $\mathscr{X}$ et du second ordre pour P, sont définis par :

$$E_P[\phi] = \int \phi(x) P(dx) \quad \text{et, pour tous} \quad x_1, x_2 \in \mathscr{X},$$

$$(D_P[\phi] \cdot x_1, x_2) = \int (\phi(x) - E_P[\phi], x_1)(\phi(x) - E_P[\phi], x_2) P(dx).$$

(6) TX désigne l'application $x \to Tx$ de $\mathscr{X}$ dans lui-même, et $\| x \|^2 = (x, x)$ le carré de la norme de x dans $\mathscr{X}$.

(2) *Si deux opérateurs* T_0 *et* T_1 *vérifient la propriété* (G-M), *les variables aléatoires* T_0X *et* T_1X sont P-*presque partout égales, pour tout* $P \in \mathscr{P}$.

(Démonstration au § 5).

Désignant par T_L le projecteur orthogonal de $\mathscr{X}$ sur L, il est intéressant de préciser la proposition précédente pour cet opérateur particulier:

PROPOSITION 3 (théorème de Kruskal) (7). — *Pour que* T_L *satisfasse la condition* (G-M), *il faut et il suffit que* Q *laisse* L *invariant :*

$$QL \subset L.$$

(Démonstration au § 5).

REMARQUE 1. — La littérature statistique s'intéressant à la méthode des moindres carrés généralisés n'utilise pas explicitement (à notre connaissance) le produit scalaire $(.,.)_Q$, l'identification précise de la classe $\mathscr{G}(Q, L)$ en pâtit généralement (examiner à ce propos l'article [*5*] donnant le résultat repris dans la proposition 3).

REMARQUE 2. — Toute question de critère mise à part (sur ce point, voir le § 4), on signale également que l'énoncé du théorème de Gauss-Markov disponible chez E. Malinvaud [*6*] est un peu moins précis que celui de la proposition 2 *bis*: on notera en particulier que chez cet auteur la condition (1.7) restreint, *a priori*, la classe des opérateurs compétitifs.

3. **Insertion dans la problématique statistique.**

Considérant que X est une variable aléatoire suivant une loi P inconnue, dont la covariance $D_P[X] = Q$ est donnée et dont l'espérance $E_P[X] = m_P$ reste située dans un sous-espace vectoriel L de $\mathscr{X}$, on se propose d'estimer la valeur du paramètre inconnu m_P comme fonction linéaire d'une observation $x \in \mathscr{X}$ de la variable X.

Cherchant ainsi à déterminer un opérateur $T \in \mathscr{L}(\mathscr{X})$ de telle sorte que la valeur Tx puisse être considérée comme une « bonne » estimation de m_P, on peut tout d'abord (premier critère) préférer T_1 à T_2 si

$$(C_1) \quad E_P[\| T_1X - m_P \|^2] \leqslant E_P[\| T_2X - m_P \|^2], \quad \text{pour tout} \quad P \in \mathscr{P};$$

si l'on désire se ramener à une situation scalaire, et que, plus précisément, on considère que la locution « TX estime m_P » signifie en fait que, pour

(7) Des résultats de nature voisine sont aussi donnés par G. WATSON [*10*] et G. ZYSKIND [*11*].

tout x, (TX, x) estime (m_P, x), on peut aussi (second critère) préférer T_1 à T_2 si

$$(C_2) \quad E_P[((T_1X, x) - (m_P, x))^2] \leqslant E_P[((T_2X, x) - (m_P, x))^2] \quad , \text{ pour tout } P \in \mathscr{P} \text{ et tout } x \in \mathscr{X}.$$

On constate alors que la proposition 2 *bis* permet de sélectionner dans $\mathscr{B}_L$ un estimateur optimal aussi bien pour (C_1) que pour (C_2).

On observera que le fait de réduire $\mathscr{L}(\mathscr{X})$ à $\mathscr{B}_L$ est, en soi, fort naturel : c'est l'idée qu'il convient d'estimer m_P par l'observation elle-même quand celle-ci « tombe » là où précisément se trouve la moyenne inconnue [8]. Dans le cas où l'hypothèse (2.1) est renforcée en

$$(3.1) \qquad L = \{ m_P ; m \in \mathscr{P} \},$$

on peut aussi, pour réduire $\mathscr{L}(\mathscr{X})$ à $\mathscr{B}_L$, tirer argument des équivalences suivantes (faciles à établir) :

$$(3.2) \quad T \in \mathscr{B}_L \Leftrightarrow T \in \mathscr{L}(\mathscr{X}) \text{ et } E_P[TX] = m_P, \text{ pour tout } P \in \mathscr{P},$$

$$(3.3) \quad T \in \mathscr{B}_L \Leftrightarrow T \in \mathscr{L}(\mathscr{X}) \text{ et } \sup_{P \in \mathscr{P}} E_P[\| TX - m_P \|^2] < +\infty ;$$

la première de ces équivalences signifie, en effet, que $\mathscr{B}_L$ est exactement constitué des estimateurs non biaisés de $\mathscr{L}(\mathscr{X})$, la seconde que l'erreur quadratique moyenne d'un opérateur linéaire n'appartenant pas à $\mathscr{B}_L$ est non bornée quand P varie.

Enfin, pour un L fixé, on observera qu'un même opérateur peut très bien être de Gauss-Markov pour une famille assez large de covariance Q ; dans cet ordre d'idées, on notera que la condition $QL \subset L$ du théorème de Kruskal est sensiblement plus générale que la condition usuelle $Q = \sigma^2 I_{\mathscr{X}}$ cautionnant d'ordinaire les bonnes propriétés du projecteur orthogonal T_L.

REMARQUE 1. — La nature du produit scalaire $(., .)_Q$ intervenant dans la définition d'un estimateur de Gauss-Markov est éclairée par l'examen du cas particulier où Q est covariance d'une loi normale, éventuellement singulière ; dans cette situation, $(., .)_Q$ n'est pas autre chose que la forme quadratique apparaissant dans la densité de P par rapport à la mesure de Lebesgue $\lambda_{\sigma(P)}$ du support $\sigma(P)$ de P :

$$P(dx) = K \exp\left[-\frac{1}{2}(x - m_P, x - m_P)_Q \right] \lambda_{\sigma(P)}(dx).$$

[8] Cette idée semble très proche du principe de consistance de Fisher.

REMARQUE 2. — Dans les modèles usuels $\mathscr{X}$ est très généralement rapporté à une base et l'on a affaire à un schéma paramétrique, en ce sens que, pour $\Theta = \mathrm{R}^d$ et $\mathrm{C} \in \mathscr{L}(\Theta, \mathscr{X})$, on a

$$\mathscr{P} = \{ \mathrm{P}_\theta ; \theta \in \Theta \} \quad \text{et} \quad \mathrm{L} = \{ \mathrm{C}\theta ; \theta \in \Theta \} ;$$

sur cet aspect des choses on peut consulter de nombreux auteurs, en particulier W. Kruskal [*4*], E. Malinvaud [*6*], R. C. Rao [*7*], H. Scheffé [*8*] et G. A. F. Seber [*9*] ([9]).

4. Utilisation des ellipsoïdes indicateurs.

Pour tout $\Gamma \in \mathbb{Q}^+$, on définit l'ellipsoïde indicateur $\varepsilon(\Gamma)$ de Γ par :

$$x \in \varepsilon(\Gamma) \Leftrightarrow x \in \operatorname{Im} \Gamma \quad \text{et} \quad (x, x)_\Gamma \leqslant 1, \tag{4.1}$$

(dans le cas où Γ est inversible, (1.5) montre que

$$x \in \varepsilon(\Gamma) \Leftrightarrow x \in \mathscr{X} \quad \text{et} \quad (\Gamma^{-1}x, x) \leqslant 1).$$

PROPOSITION 4. — *Étant donné* $\Gamma \in \mathbb{Q}^+$, $\varepsilon(\Gamma)$ *est identique au polaire* ([10]) *de l'ensemble* $\{ x ; x \in \mathscr{X}$ *et* $(\Gamma x, x) \leqslant 1 \}$.

(Démonstration au § 5).

PROPOSITION 5. — *Étant donnés* Γ_1 *et* $\Gamma_2 \in \mathbb{Q}^+$, *les deux propriétés suivantes sont équivalentes :*

(*I*) $\qquad \Gamma_2 - \Gamma_1 \in \mathbb{Q}^+ ;$

(*II*) $\qquad \varepsilon(\Gamma_2) \supset \varepsilon(\Gamma_1).$

(Démonstration au § 5).

COMMENTAIRES. — Définissant l'ellipsoïde indicateur $\varepsilon_\mathrm{P}[\mathrm{Y}]$ d'un vecteur aléatoire Y, du second ordre pour P, comme l'ellipsoïde indicateur de son opérateur de covariance ([11]), il est intéressant de comparer les divers estimateurs de m_P en terme d'inclusion de leurs ellipsoïdes indicateurs, ainsi que le fait E. Malinvaud [*6*].

([9]) Le travail [*1*] s'efforce, pour sa part, à faire le point sur les liens entre la régression linéaire et la méthode des moindres carrés.

([10]) Pour $\mathrm{B} \subset \mathscr{X}$, le polaire B^0 de B est défini par

$$x \in \mathrm{B}^0 \Leftrightarrow x \in \mathscr{X} \quad \text{et} \quad |(x, y)| \leqslant 1 \quad \text{pour tout} \quad y \in \mathrm{B}.$$

([11]) L'idée d'utiliser un ellipsoïde indicateur en calcul des probabilités est due à G. DARMOIS [*2*].

La proposition 4 montre déjà que la définition donnée dans [*6*] équivaut à la définition (à notre sens plus simple) donnée en (4.1).

D'autre part, compte tenu de ce que la condition (j) de la proposition 2 *bis* n'est (voir le § 5) qu'une transcription probabiliste de la condition (i) de la proposition (2), il résulte clairement de la proposition 5 que le théorème de Gauss-Markov (proposition 2 *bis*) peut être donné en remplaçant la propriété (j) par la propriété (j') équivalente :

(j') $\varepsilon_P[T_0X] \subset \varepsilon_P[TX]$, pour tout $T \in \mathscr{B}_L$ et tout $P \in \mathscr{P}$.

La démarche du § 2 est ainsi bien située par rapport à celle fondée sur la notion d'ellipsoïde indicateur : on n'en dira pas plus sur une notion dont l'intérêt est essentiellement d'ordre géométrique, renvoyant à [*6*] pour tous autres compléments.

5. Démonstrations des différentes propositions.

Préliminaire. — Désignant par $V^\perp$ l'orthogonal de $V \subset \mathscr{X}$, on rappelle que tout $Q \in \mathbb{Q}^+$ satisfait les deux relations :

$$\text{Im } Q = (\text{Ker } Q)^\perp \tag{5.1}$$

et

$$\text{Ker } Q = \{ x ; x \in \mathscr{X} \text{ et } (Qx, x) = 0 \} \tag{5.2}$$ (12).

Démonstration de la proposition 1.

Soit $Q \in \mathbb{Q}^+$; utilisant la symétrie de Q, on vérifie aisément tout d'abord que la quantité (Qx, y) ne dépend que de Qx et Qy ; la bilinéarité et la positivité de $(.,.)_Q$ sont évidentes ; le fait que $(x, x)_Q$ ne s'annule, sur Im Q, que dans le cas où x est nul est conséquence directe de (5.2). Enfin, quand Q est inversible, (1.5) se déduit clairement de (1.4).

Dans ce qui suit, il sera utile de remarquer que (1.4) entraîne :

$$(Qx, y)_Q = (x, y), \quad \text{pour tous } x \in \mathscr{X},\ y \in \text{Im } Q. \tag{5.3}$$

Avant d'établir la proposition 2, on introduira une notation : pour tout $x \in \mathscr{X}$, on définit $x \otimes x$ ($x \otimes x \in \mathscr{L}(\mathscr{X})$) par

$$(x \otimes x)y = (x, y)x, \quad \text{pour tout } y \in \mathscr{X}. \tag{5.4}$$

(12) Pour les résultats élémentaires d'algèbre admis ici ou là, on peut, par exemple, consulter HALMOS [*3*].

LEMME. — *Soit* $Q \in \mathbb{Q}^+$, *si* $(x_i)_{1 \leqslant i \leqslant q}$ *est une base de* Im Q, *orthonormale pour* $(.,.)_Q$, *on a*

$$Q = \sum_{i=1}^{q} x_i \otimes x_i.$$

Démonstration. — Pour $x, y \in \mathscr{X}$, on constate en développant Qx et Qy par rapport à la base donnée de Im Q que

$$(Qx, Qy)_Q = \sum_{i=1}^{q} (Qx, x_i)_Q (Qy, x_i)_Q ;$$

exploitant (1.4) et (5.3), on en déduit que, pour tous $x, y \in \mathscr{X}$,

$$(Qx, y) = \sum_{i=1}^{q} (x_i, x)(y_i, y) = \left(\left[\sum_{i=1}^{q} x_i \otimes x_i\right] x, y\right);$$

ce qui établit bien le résultat annoncé.

Démonstration de la proposition 2.

De façon préliminaire, notant $r = \dim (L_\cap \operatorname{Im} Q)$, on fera choix (à l'aide du procédé d'orthogonalisation de Schmidt) d'une base $(x_i)_{1 \leqslant i \leqslant q}$ de Im Q, orthonormale pour $(.,.)_Q$ et telle que $(x_i)_{1 \leqslant i \leqslant r}$ soit une base de $L_\cap$ Im Q. Pour cette base, on aura, d'après le lemme :

$$Q = \sum_{i=1}^{q} x_i \otimes x_i$$

et

$$(5.5)\quad TQT^* = \sum_{i=1}^{r} x_i \otimes x_i + \sum_{i=r+1}^{q} Tx_i \otimes Tx_i, \quad \text{pour tout} \quad T \in \mathscr{B}_L ;$$

(On a obtenu (5.5) en utilisant le résultat facile :

$$T \in \mathscr{L}(\mathscr{X}) \quad \text{et} \quad x \in \mathscr{X} \Rightarrow T(x \otimes x)T^* = Tx \otimes Tx).$$

La relation (5.5) entraîne évidemment :

$$(5.6)\quad (TQT^*x, x) = \sum_{i=1}^{r} (x_i, x)^2 + \sum_{i=r+1}^{q} (Tx_i, x)^2$$

$$, \quad \text{pour tout} \quad T \in \mathscr{B}_L \quad \text{et tout} \quad x \in \mathscr{X},$$

et aussi

$$(5.7)\quad \mathrm{Tr}[\mathrm{TQT}^*] = \sum_{i=1}^{r} \| x_i \|^2 + \sum_{i=r+1}^{q} \| \mathrm{T}x_i \|^2, \quad \text{pour tout} \quad \mathrm{T} \in \mathscr{B}_{\mathrm{L}}.$$

On va maintenant établir la proposition conformément au schéma :

$$(\text{G-M}) \Rightarrow (i) \Rightarrow (ii) \Rightarrow (\text{G-M}).$$

a) (G-M) $\Rightarrow$ (*i*). Étant donné T_0 qui satisfait (G-M), compte tenu du choix de la base $(x_i)_{1 \leqslant i \leqslant q}$, il est clair que $\mathrm{T}_0 x_i = 0$, pour $r + 1 \leqslant i \leqslant q$; d'où résulte, avec (5.6), que

$$(5.8)\quad (\mathrm{T}_0\mathrm{QT}_0^* x, x) \leqslant (\mathrm{TQT}^* x, x), \quad \text{pour tout} \quad \mathrm{T} \in \mathscr{B}_{\mathrm{L}} \quad \text{et tout} \quad x \in \mathscr{X},$$

ce qui n'est qu'une autre écriture de (*i*).

b) (*i*) $\Rightarrow$ (*ii*) s'obtient immédiatement à partir de (5.8) en faisant parcourir à x une base orthonormale de $\mathscr{X}$ et en sommant.

c) (*ii*) $\Rightarrow$ (G-M). Soit T_0 satisfaisant (*ii*), choisissant un élément T_1 qui satisfait (G-M), donc aussi (*ii*), on obtient en exploitant (5.7) :

$$\sum_{i=1}^{r} \| x_i \|^2 + \sum_{i=r+1}^{q} \| \mathrm{T}_0 x_i \|^2 = \mathrm{Tr}[\mathrm{T}_0\mathrm{QT}_0^*] = \mathrm{Tr}[\mathrm{T}_1\mathrm{QT}_1^*] = \sum_{i=1}^{r} \| x_i \|^2 ;$$

donc $\mathrm{T}_0 x_i = 0$, pour $r + 1 \leqslant i \leqslant q$: T_0, coïncidant avec T_1 sur Im Q, satisfait par conséquent (G-M), ce qui achève la démonstration.

Démonstration de la proposition 2-bis.

Remarquant que, pour tout $\mathrm{T} \in \mathscr{B}_{\mathrm{L}}$, la variable aléatoire TX est du second ordre par rapport à P, pour tout $\mathrm{P} \in \mathscr{P}$, et que

$$\mathrm{E_P}[\mathrm{TX}] = \mathrm{T}m_{\mathrm{P}} = m_{\mathrm{P}} \quad \text{et} \quad \mathrm{D_P}[\mathrm{TX}] = \mathrm{TQT}^*, \quad \text{pour tout} \quad \mathrm{P} \in \mathscr{P},$$

on note que la partie (1) de la proposition est une simple traduction probabiliste de la proposition 2.

Le point (2) s'obtient à son tour en observant que toute probabilité $\mathrm{P} \in \mathscr{P}$ est portée par le sous-espace affine $m_{\mathrm{P}} + \mathrm{Im}\, \mathrm{D_P}[\mathrm{X}]$, puis en notant que ces divers sous-espaces sont nécessairement inclus dans L + Im Q, sous espace vectoriel de $\mathscr{X}$ sur lequel coïncident automatiquement tous les opérateurs de Gauss-Markov.

Démonstration de la proposition 3.

Pour tout sous-espace V de Im Q, on note $V^{\#}$ le sous-espace de Im Q orthogonal à V au sens de $(.,\ .)_Q$. Il est clair que T_L ne satisfait la condition (G-M) que si et seulement si $(L_\cap \operatorname{Im} Q)^{\#} \subset \operatorname{Ker} T_L$, soit encore, tenant compte des relations $\operatorname{Ker} T_L = L^\perp$ et $(L_\cap \operatorname{Im} Q)^{\#} \subset \operatorname{Im} Q$, si et seulement si

$$(5.9) \qquad (L_\cap \operatorname{Im} Q)^{\#} \subset L^\perp_\cap \operatorname{Im} Q.$$

Or il résulte de (5.3) que $L^\perp_\cap \operatorname{Im} Q = (QL)^{\#}$; il ne reste donc plus qu'à transformer (5.9) en passant aux orthogonaux pour $(.,\ .)_Q$, dans Im Q, pour obtenir l'équivalence annoncée [13].

Démonstration des propositions 4 et 5.

Notant $B_\Gamma = \{ x ; x \in \mathscr{X} \text{ et } (\Gamma x, x) \leqslant 1 \}$, on établira en premier lieu l'identité $\varepsilon(\Gamma) = B^0_\Gamma$ (proposition 4).

Tout $x \in \varepsilon(\Gamma)$ est du type $x = \Gamma z$ avec $(x, x)_\Gamma = (\Gamma z, \Gamma z)_\Gamma \leqslant 1$, or, utilisant l'inégalité de Cauchy-Schwartz,

$$(x, y)^2 = (\Gamma z, y)^2 = (\Gamma z, \Gamma y)^2_\Gamma \leqslant (\Gamma z, \Gamma z)_\Gamma (\Gamma y, \Gamma y)_\Gamma = (x, x)_\Gamma (\Gamma y, y) ;$$

on en déduit que, pour tout $x \in \varepsilon(\Gamma)$ et tout $y \in B_\Gamma$, $(x, y)^2 \leqslant 1$, d'où l'inclusion $\varepsilon(\Gamma) \subset B^0_\Gamma$.

Inversement, utilisant (5.1), un raisonnement facile d'homothétie donne déjà $B^0_\Gamma \subset \operatorname{Im} \Gamma$; l'inclusion $B^0_\Gamma \subset \varepsilon(\Gamma)$ s'obtient donc en constatant que

$$x \in \operatorname{Im} \Gamma \quad \text{et} \quad (x, x)_\Gamma > 1 \;\Rightarrow\; x \notin B^0_\Gamma :$$

en effet, pour $x = \Gamma z$ tel que $(x, x)_\Gamma = (\Gamma z, z) = c > 1$, on note que

$$y = c^{1/2} z \in B_\Gamma,$$

tandis que $|(x, y)| = c^{-1/2}(\Gamma z, z) = c^{1/2} > 1$.

La proposition 4 étant établie, la proposition 5 n'en est qu'un simple corollaire : un raisonnement d'homothétie donne en effet aussitôt l'équivalence entre la condition (l) et la condition $B_{Q_1} \supset B_{Q_2}$, on conclut par passage au polaire et avec la proposition 4.

---

[13] Le fait que cette démonstration soit sensiblement plus simple que celle donnée dans [5] est à rapprocher de la remarque 1 du paragraphe 2.

RÉFÉRENCES BIBLIOGRAPHIQUES

[1] Ph. Courrège, J.-L. Philoche et P. Priouret, *Régression linéaire et estimation par la méthode des moindres carrés (à paraître).*

[2] G. Darmois, Sur les limites de la dispersion de certaines estimations. *Rev. instit. intern. stat.*, 13e année, 1943, p. 9-15.

[3] P. Halmos, *Finite dimensionnal vector spaces*, 2e édit., Van Nostrand, 1958.

[4] W. Kruskal, The coordinate-free approach to Gauss-Markov estimation, and its application to missing and extra observations. *Proc. of the IVth Berk. Sym. on math. statist. and prob.*, vol. **1**, 1961.

[5] W. Kruskal, When are Gauss-Markov and least squares estimators identical? A coordinate-free approach. *Ann. of Math. Stat.*, vol. **39**, 1968, p. 70-75.

[6] E. Malinvaud, *Méthodes statistiques de l'économétrie.* 2e édit., Dunod, 1969.

[7] C. R. Rao, *Linear statistical inference and its applications.* Wiley, 1965.

[8] H. Scheffé, *The analysis of variance.* 2e édit., Wiley, 1961.

[9] G. A. F. Seber, *The linear hypotheses : a general theory.* Griffin, 1966.

[10] G. S. Watson, Linear least squares regression. *Ann. of Math. Statist.*, vol. **38**, 1967, p. 1679-1699.

[11] G. Zyskind, On canonical, non-negative covariance matrices and best and simple least squares linear estimator in linear models. *Ann. of Math. Statist.*, vol. **38**, 1967, p. 1092-1109.

Manuscrit reçu le 25 mars 1971.

Jean-Louis Philoche
5, rue Poliveau 75-Paris 5e.