

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

PH. COURRÈGE

J.-L. PHILOCHE

P. PRIOURET

Régression linéaire et estimation par la méthode des moindres carrés

Annales de l'I. H. P., section B, tome 7, n° 4 (1971), p. 253-270

http://www.numdam.org/item?id=AIHPB_1971__7_4_253_0

© Gauthier-Villars, 1971, tous droits réservés.

L'accès aux archives de la revue « Annales de l'I. H. P., section B » (<http://www.elsevier.com/locate/anihpb>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

Régression linéaire et estimation par la méthode des moindres carrés

par

Ph. COURRÈGE, J.-L. PHILOCHE et P. PRIOURET

SUMMARY. — The purpose of this paper is to present in a unified way the general features of linear regression, mean square estimation and Bayesian estimation. The emphasis is on the relations between these methods. Hilbertian formalism is systematically used.

On se propose ici, en un exposé autonome, de présenter dans le cadre commun du modèle statistique linéaire ⁽¹⁾ ces trois méthodes que constituent la régression linéaire, la méthode des moindres carrés et l'estimation bayésienne. On insiste spécialement sur les rapports existant entre ces approches, rapports qui sont généralement peu explicités dans la littérature statistique.

On commence (§ 1) par des énoncés qui condensent les principales propriétés mathématiques en cause. Puis (§ 2), on commente ces propriétés du point de vue statistique en les situant par rapport aux trois méthodes ci-dessus. On donne enfin (§ 3) une démonstration complète des énoncés du § 1 en utilisant systématiquement le formalisme hilbertien qui est naturel en la matière.

⁽¹⁾ On se limite ici au cas de dimension finie. On reprendra le cas de dimension infinie dans un prochain travail avec Alain BENSOUSSAN qui a déjà traité du sujet dans son mémoire « Identification et filtrage », cahier n° 1 de l'IRIA, février 1969.

§ 1. Énoncé des résultats mathématiques en cause.

1.1. **Mise en place de la situation.** — On désigne par \mathcal{X} et \mathcal{Y} des espaces de Hilbert réel de dimension finie, et par X et Y les projections de l'espace produit $\mathcal{X} \times \mathcal{Y}$ sur \mathcal{X} et \mathcal{Y} respectivement. On suppose donnée une probabilité de transition $x \rightarrow P_x$ de \mathcal{X} dans \mathcal{Y} (toujours supposés munis de leurs tribus boréliennes), un élément m de \mathcal{Y} , une application linéaire C de \mathcal{X} dans \mathcal{Y} et un opérateur linéaire symétrique positif Σ sur \mathcal{Y} de telle sorte que, pour chaque $x \in \mathcal{X}$, P_x soit du second ordre sur \mathcal{Y} avec $m + Cx$ comme moyenne et Σ comme opérateur de covariance :

$$(1.1) \quad \int_{\mathcal{Y}} y P_x(dy) = m + Cx, \quad \text{pour tout } x \in \mathcal{X}.$$

$$(1.2) \quad \int_{\mathcal{Y}} (y_1, y - m - Cx)(y_2, y - m - Cx) P_x(dy) = (\Sigma y_1, y_2),$$

pour tous $x \in \mathcal{X}$, $y_1 \in \mathcal{Y}$ et $y_2 \in \mathcal{Y}$ ⁽²⁾ ⁽³⁾.

On désigne de plus par \mathcal{P} l'ensemble des mesures de probabilité du second ordre sur \mathcal{X} et, pour chaque $\pi \in \mathcal{P}$, par P_π la mesure de probabilité sur $\mathcal{X} \times \mathcal{Y}$ canoniquement associée à π et à la transition $(P_x)_{x \in \mathcal{X}}$ en posant, pour chaque fonction numérique borélienne bornée ϕ sur $\mathcal{X} \times \mathcal{Y}$,

$$(1.3) \quad \int_{\mathcal{X} \times \mathcal{Y}} \phi(x, y) P_\pi(dx \times dy) = \int_{\mathcal{X}} \pi(dx) \int_{\mathcal{Y}} P_x(dy) \phi(x, y);$$

On notera $E_\pi[\phi]$ la quantité ainsi définie.

Désignant, pour chaque $\pi \in \mathcal{P}$, par ξ_π la moyenne de π et par Γ_π son opérateur de covariance, on remarque que P_π est du second ordre sur $\mathcal{X} \times \mathcal{Y}$ ⁽⁴⁾ et admet pour moyenne et opérateur de covariance les quantités μ_π et Δ_π données par,

$$(1.4) \quad \mu_\pi = (\xi_\pi, m + C\xi_\pi)$$

$$(1.5) \quad \Delta_\pi = \begin{pmatrix} \Gamma_\pi & \Gamma_\pi C^* \\ C\Gamma_\pi & C\Gamma_\pi C^* + \Sigma \end{pmatrix} \quad (5).$$

(ce dernier point sera établi au n° 3.1).

⁽²⁾ Les produits scalaires sur \mathcal{X} et \mathcal{Y} sont notés de la même manière : (x_1, x_2) , (y_1, y_2) .

⁽³⁾ En ce qui concerne l'interprétation statistique de ces notions, voir le paragraphe 2.

⁽⁴⁾ Muni de la structure hilbertienne associée au produit scalaire

$$((x_1, y_1), (x_2, y_2)) = (x_1, x_2) + (y_1, y_2).$$

⁽⁵⁾ C^* désigne l'opérateur linéaire adjoint de C défini par $(Cx, y) = (x, C^*y)$.

1.2. **Cas de la régression linéaire.** — On désignera par \mathcal{R} l'ensemble des applications affines R de \mathcal{Y} dans \mathcal{X} (c'est-à-dire des applications de la forme $Ry = \xi + Sy$ où $\xi \in \mathcal{X}$ et $S \in \mathcal{L}(\mathcal{Y}, \mathcal{X})$). Les éléments de \mathcal{R} seront simplement appelés *estimateurs de X*. Cela étant, on cherche à minimiser $E_\pi[\|X - RY\|^2]$ ⁽⁶⁾, avec π donnée :

THÉORÈME 1. — *On suppose donnée une mesure de probabilité π appartenant à \mathcal{P} . Alors,*

(1) *Il existe une application affine $R_1 \in \mathcal{R}$ ayant les propriétés équivalentes suivantes ⁽⁶⁾ :*

- (i) $E_\pi[\|X - R_1 Y\|^2] = \inf_{R \in \mathcal{R}} E_\pi[\|X - RY\|^2]$;
 - (ii) $E_\pi[(x, X) - (x, R_1 Y)]^2 = \inf_{R \in \mathcal{R}} E_\pi[(x, X) - (x, RY)]^2$, pour tout $x \in \mathcal{X}$;
 - (iii) $E_\pi[R_1 Y] = \xi_\pi (= E_\pi[X])$, et
- (1.6) $E_\pi[(y, Y)(x, X - R_1 Y)] = 0$, pour tous $x \in \mathcal{X}$ et $y \in \mathcal{Y}$ ⁽⁷⁾ ;
- (iv) $R_1 y = \xi_\pi + S_1(y - m - C\xi_\pi)$ ($y \in \mathcal{Y}$),

où $S_1 \in \mathcal{L}(\mathcal{Y}, \mathcal{X})$ est tel que

(1.7) $\Gamma_\pi C^* = S_1(C\Gamma_\pi C^* + \Sigma)$ ⁽⁸⁾.

(2) *De plus, si Σ et Γ_π sont inversibles, et si on pose, pour $x \in \mathcal{X}$ et $y \in \mathcal{Y}$,*

(1.8) $J_\pi(x, y) = (\Gamma_\pi^{-1}(x - \xi_\pi), x - \xi_\pi) + (\Sigma^{-1}(y - m - Cx), y - m - Cx)$,

les propriétés (i) à (iv) ci-dessus sont équivalentes aux propriétés (vi) et (vii) ci-dessous :

(vi) $J_\pi(R_1 y, y) = \inf_{x \in \mathcal{X}} J_\pi(x, y)$, pour tout $y \in \mathcal{Y}$;

(vii) *Pour tout $y \in \mathcal{Y}$,*

(1.9) $R_1 y = \xi_\pi + (\Gamma_\pi^{-1} + C^* \Sigma^{-1} C)^{-1} C^* \Sigma^{-1} (y - m - C\xi_\pi)$

(démonstration au n° 3.2).

⁽⁶⁾ RY désigne l'application $(x, y) \rightarrow Ry$ de $\mathcal{X} \times \mathcal{Y}$ dans \mathcal{X} et $\|x\|$ la norme de x dans \mathcal{X} .

⁽⁷⁾ Autrement dit, les variables aléatoires Y et $X - R_1 Y$ sont non corrélées.

⁽⁸⁾ En particulier, cette équation a toujours une solution S_1 au moins ; et, cette solution est unique dès que $C\Gamma_\pi C^* + \Sigma$ est inversible ; ce qui est le cas si Σ est inversible. Par ailleurs les propriétés (i) - (iv) déterminent de façon unique la classe de la fonction $R_1 Y$ dans l'espace $L^2_{\mathcal{X}}(\mathcal{X} \times \mathcal{Y}, P_\pi)$ (voir le n° 3.2a).

1.3. Cas de la méthode des moindres carrés. Théorème de Gauss-Markov. — On cherche encore à minimiser $E_{\pi}[\|X - RY\|^2]$ mais, contrairement à la situation du théorème 1, π n'est plus donnée et on souhaite un résultat valable pour tout $\pi \in \mathcal{P}$. Désignant par $\mathcal{R}^{\#}$ l'ensemble des applications affines R de \mathcal{Y} dans \mathcal{X} (de la forme $Ry = \xi + Sy$) pour lesquelles $SCx = x$ pour tout $x \in \mathcal{X}$ ⁽⁹⁾, on a,

THÉORÈME 2. — *On suppose que Σ est inversible et que C est injective. Alors,*

(1) *Concernant une application affine $R \in \mathcal{R}$, les propriétés suivantes sont équivalentes*

- (h) $\sup_{\pi \in \mathcal{P}} E_{\pi}[\|X - RY\|^2] < +\infty$
 (hh) $\sup_{\pi \in \mathcal{P}} E_{\pi}[(x, X) - (x, RY)]^2 < +\infty$, pour tout $x \in \mathcal{X}$
 (hhh) $R \in \mathcal{R}^{\#}$.

De plus, pour tout $R \in \mathcal{R}^{\#}$ (de la forme $Ry = \xi + Sy$), on a,

$$(1.10) \quad E_{\pi}[\|X - RY\|^2] = \|\xi + Sm\|^2 + \text{Tr}(S\Sigma S^*) \quad \text{pour tout } \pi \in \mathcal{P};$$

et, pour tout $x \in \mathcal{X}$,

$$(1.11) \quad E_{\pi}[(x, X) - (x, RY)]^2 = (x, \xi + Sm)^2 + (S\Sigma S^*x, x), \quad \text{pour tout } \pi \in \mathcal{P}.$$

(2) *Il existe une application affine $R_2 \in \mathcal{R}^{\#}$ et une seule ayant les propriétés équivalentes suivantes :*

- (j) $E_{\pi}[\|X - R_2Y\|^2] = \inf_{R \in \mathcal{R}^{\#}} E_{\pi}[\|X - RY\|^2]$, pour tout $\pi \in \mathcal{P}$;
 (jj) Pour tout $x \in \mathcal{X}$,
 $E_{\pi}[(x, X) - (x, R_2Y)]^2 = \inf_{R \in \mathcal{R}^{\#}} E_{\pi}[(x, X) - (x, RY)]^2$, pour tout $\pi \in \mathcal{P}$;
 (jjj) $J(R_2y, y) = \inf_{x \in \mathcal{X}} J(x, y)$,

où, pour $x \in \mathcal{X}$ et $y \in \mathcal{Y}$, on a posé,

$$(1.12) \quad J(x, y) = (\Sigma^{-1}(y - m - Cx), y - m - Cx).$$

En outre l'application affine R_2 ainsi caractérisée est définie par

$$(jv) \quad R_2y = (C^*\Sigma^{-1}C)^{-1}C^*\Sigma^{-1}(y - m) \quad (y \in \mathcal{Y}) \quad (10).$$

(démonstration au n° 3.4).

⁽⁹⁾ $\mathcal{R}^{\#}$ est non vide seulement si C est injectif, ce qui est postulé ci-dessous.

⁽¹⁰⁾ L'injectivité de C entraîne que $C^*\Sigma^{-1}C$ est inversible.

D'autre part, désignant, pour chaque $x \in \mathcal{X}$, par $\mathcal{R}_x^\#$ l'ensemble des fonctions numériques affines r sur y telles que

$$E_\theta[rY] = (x, \theta) \quad \text{pour tout } \theta \in \mathcal{X}$$

[en notant ici E_θ au lieu de E_{ε_θ} (ε_θ mesure de Dirac de $\theta \in \mathcal{X}$)], on a,

COROLLAIRE (Théorème de Gauss-Markov). — *L'application R_2 définie par (jv) ⁽¹¹⁾ est l'unique élément de $\mathcal{R}^\#$ ayant l'une ou l'autre des propriétés (d'ailleurs équivalentes) (k) ou (kk) ci-dessous :*

(k) $E_\theta[\|\theta - R_2 Y\|^2] = \inf_{R \in \mathcal{R}^\#} E_\theta[\|\theta - RY\|^2]$, pour tout $\theta \in \mathcal{X}$;

(kk) Pour tout $x \in \mathcal{X}$,

(1.13) $E_\theta[\|(x, \theta) - (x, R_2 Y)\|^2] = \inf_{r \in \mathcal{R}_x^\#} E_\theta[\|(x, \theta) - rY\|^2]$, pour tout $\theta \in \mathcal{X}$.

De plus, on a

(kkk) $E_\theta[R_2 Y] = \theta$, pour tout $\theta \in \mathcal{X}$.

(démonstration au n° 3.4).

1.4. Comparaison des estimateurs R_π et $R^\#$. — Supposant ici que Σ est inversible et que C est injective, on désignera, pour chaque $\pi \in \mathcal{P}$, par R_π l'estimateur introduit par le théorème 1,

(1.14) $R_\pi y = \xi_\pi + \Gamma_\pi C^*(C\Gamma_\pi C^* + \Sigma)^{-1}(y - m - C\xi_\pi) \quad (y \in \mathcal{Y})$,

et par $R^\#$ l'estimateur introduit par le théorème 2,

(1.15) $R^\# y = (C^*\Sigma^{-1}C)^{-1}C^*\Sigma^{-1}(y - m) \quad (y \in \mathcal{Y})$.

Les estimateurs R_π et $R^\#$ ainsi définis seront appelés ici respectivement la *régression linéaire* ⁽¹²⁾ de X sur Y (associée à la répartition *a priori* π de X) et l'*estimateur de Gauss-Markov*.

Ces deux estimateurs peuvent être situés mutuellement comme suit :

THÉORÈME 3. — *On suppose que Σ est inversible et que C est injective, alors,*

(l) Pour tout $\pi \in \mathcal{P}$,

(1.16) $E_\pi[\|X - R_\pi Y\|^2] < E_\pi[\|X - R^\# Y\|^2]$,

⁽¹¹⁾ Sous les hypothèses du théorème 2 ci-dessus. On peut étendre le théorème de Gauss-Markov au cas où ces hypothèses ne sont pas satisfaites, voir à ce sujet le travail de J.-L. PHILOCHE ci-dessous, p. 271.

⁽¹²⁾ Le terme affine conviendrait mieux que le terme linéaire, mais ce dernier est consacré par l'usage des statisticiens.

et,

$$(1.17) \quad E_{\pi}[(x, X) - (x - R_{\pi} Y)]^2 \leq E_{\pi}[(x, X) - (x, R^{\#} Y)]^2, \text{ pour tout } x \in \mathcal{X},$$

avec au moins pour un $x \in \mathcal{X}$ l'inégalité stricte; en particulier $R_{\pi} \neq R^{\#}$.

(II) Pour tout $\pi \in \mathcal{P}$,

$$(1.18) \quad \text{Sup}_{\pi' \in \mathcal{P}} E_{\pi'}[\|X - R_{\pi'} Y\|^2] = E_{\pi}[\|X - R^{\#} Y\|^2], \text{ et,}$$

$$(1.19) \quad \text{Sup}_{\pi' \in \mathcal{P}} E_{\pi'}[(x, X) - (x, R_{\pi'} Y)]^2 = E_{\pi}[(x, X) - (x, R^{\#} Y)]^2 \text{ pour tout } x \in \mathcal{X}.$$

(III) Soit (π_n) une suite d'éléments de \mathcal{P} dont les moyennes sont les mêmes et dont les opérateurs de covariance Γ_{π_n} sont inversibles. Dans ces conditions, pour que

$$(1.20) \quad \lim_{n \rightarrow \infty} R_{\pi_n} y = R^{\#} y \text{ pour tout } y \in \mathcal{Y},$$

il faut et suffit que $\lim_{n \rightarrow \infty} \Gamma_{\pi_n}^{-1} = 0$ ⁽¹³⁾; et on a alors,

$$(1.21) \quad \lim_{n \rightarrow \infty} E_{\pi_n}[\|X - R_{\pi_n} Y\|^2] = \text{Sup}_{\pi' \in \mathcal{P}} E_{\pi'}[\|X - R_{\pi'} Y\|^2], \text{ et,}$$

$$(1.22) \quad \lim_{n \rightarrow \infty} E_{\pi_n}[(x, X) - (x, R_{\pi_n} Y)]^2 = \text{Sup}_{\pi' \in \mathcal{P}} E_{\pi'}[(x, X) - (x, R_{\pi'} Y)]^2, \text{ pour tout } x \in \mathcal{X}.$$

(démonstration au n° 3.5).

§ 2. Interprétation statistique des énoncés du paragraphe 1.

2.1. Description de la situation : interprétation du modèle statistique linéaire $(\mathcal{Y}, (P_x)_{x \in \mathcal{X}})$. — On envisage un système physique \mathcal{S} dont chaque état est repéré par le couple $(x, y) \in \mathcal{X} \times \mathcal{Y}$ des valeurs que prennent les deux variables X et Y ⁽¹⁴⁾, et on suppose, d'abord que seule la variable Y est susceptible d'être mesurée (« X est une variable cachée du système ») et ensuite que, lorsque X prend la valeur $x \in \mathcal{X}$, une mesure de Y est régie

⁽¹³⁾ On est dans ce cas si l'on prend pour π_n une probabilité (par exemple gaussienne) telle que $\xi_{\pi_n} = 0$ et $\Gamma_{\pi_n} = nI_x$ pour tout n : autrement dit, R_{π_n} tend vers $R^{\#}$ lorsque π_n s'étale uniformément dans \mathcal{X} .

⁽¹⁴⁾ Le substantif « variable » est à entendre ici au sens de « caractéristique du système \mathcal{S} », et a priori pas toujours au sens mathématiquement fixé du terme « variable aléatoire » employé en calcul des probabilités.

par la probabilité P_x sur \mathcal{Y} ⁽¹⁵⁾; la transition $(P_x)_{x \in \mathcal{X}}$ de \mathcal{X} dans \mathcal{Y} étant supposée connue (au n° 1.1, c'est une donnée) ⁽¹⁶⁾.

Cela étant, on cherche à estimer la valeur de la variable cachée X en fonction affine d'une mesure $y \in \mathcal{Y}$ obtenue pour la variable Y . Plus précisément, on cherche à déterminer une application affine R de \mathcal{Y} dans \mathcal{X} de telle sorte que la quantité $Ry \in \mathcal{X}$ soit une estimation de la valeur de la variable X dans une situation où une mesure de Y a fourni comme résultat la valeur $y \in \mathcal{Y}$. Afin de choisir l'estimateur R de X de façon optimale, il est nécessaire de préciser d'abord le rôle joué par la variable X et ensuite ce que l'on sait (ou que l'on peut « raisonnablement » postuler) sur elle *a priori* ⁽¹⁷⁾. On distingue alors traditionnellement deux cas selon que X apparaît comme une variable aléatoire de répartition plus ou moins connue (cas de la régression linéaire) ou comme un paramètre fixé mais inconnu (cas de la méthode des moindres carrés).

2.2. Cas de la régression linéaire. — On admet dans ce cas que la variable X est une variable aléatoire au même titre que Y ⁽¹⁸⁾, de telle sorte que l'on peut considérer le couple (X, Y) comme une variable aléatoire à valeurs dans $\mathcal{X} \times \mathcal{Y}$. La répartition P_π de cette variable aléatoire est alors entièrement déterminée par la donnée de la répartition π de X et de la répartition conditionnelle de Y dans X , laquelle est précisée par la transition $(P_x)_{x \in \mathcal{X}}$ [on note que les caractéristiques du second ordre de P_π ne dépendent que de celles de π et de (P_x) ainsi que l'indiquent les relations (1.4) et (1.5) du n° 1.1]. Cela étant, deux cas se présentent selon que l'on considère π comme connue ou inconnue :

a) Lorsqu'on suppose connue la répartition π de X , on peut déterminer l'estimateur optimal au moyen des critères (i) ou (ii) (théorème 1, n° 1.2); et il est clair alors que l'estimateur R_π obtenu (« la régression linéaire de X sur Y ») dépend de π (par l'intermédiaire des caractéristiques du second ordre ξ_π et Γ_π de π ; formule (1.14), n° 1.4).

⁽¹⁵⁾ Autrement dit, le processus de la mesure de Y est considéré comme un tirage au sort régi par la probabilité P_x . On se placera ici en aval de cette formulation, sans chercher, en amont, à préciser la conception de la probabilité qu'elle sous-entend (potentielle ou fréquentielle, etc...).

⁽¹⁶⁾ Voir aussi le n° 2.7 à ce sujet.

⁽¹⁷⁾ C'est-à-dire comme une connaissance sur les mécanismes du système déjà acquise avant l'expérience en cause (mesure de Y). En ce sens, le modèle $(\mathcal{Y}, (P_x)_{x \in \mathcal{X}})$ introduit ci-dessus est aussi un *a priori*.

⁽¹⁸⁾ Autrement dit, dans ce cas, aussi bien X que Y sont conditionnées par les valeurs de paramètres (ou caractéristiques) du système dont on ne peut s'assurer de la *stabilité* dans les conditions de l'expérience qui permet de mesurer Y .

b) Lorsqu'on ne suppose pas que la répartition π de X est connue, on est amené à chercher un estimateur R ayant une propriété d'optimalité *uniforme* lorsque π décrit \mathcal{P} . La propriété (1) du théorème 2 (n° 1.3) introduit alors la classe $\mathcal{R}^\#$ [un peu plus large que celle des estimateurs sans biais, le terme de translation ξ n'étant pas nécessairement égal à -5 sm ; début du n° 1.3], et, dans cette classe, on peut déterminer l'estimateur optimal au moyen des critères (j) ou (jj) (propriété (2) du théorème 2, n° 1.3). On note que l'estimateur $R^\#$ ainsi obtenu (« l'estimateur de Gauss-Markov ») est sans biais [propriété (kkk) du théorème de Gauss-Markov (corollaire du théorème 2)].

Ainsi, *via* le théorème 2, on voit clairement comment se situent mutuellement et se distinguent la régression linéaire R_π et l'estimateur de Gauss-Markov $R^\#$ dans le cadre commun du modèle statistique linéaire $(\mathcal{Y}, (P_x)_{x \in \mathcal{X}})$ introduit au n° 2.1. Ce cadre est aussi celui de la méthode des moindres carrés que l'on va examiner maintenant.

2.3. Cas de la méthode des moindres carrés. — On admet dans ce cas que la variable X constitue un paramètre déterminé ⁽¹⁹⁾ du système \mathcal{S} considéré, mais de valeur *a priori* inconnue. La méthode des moindres carrés est alors basée sur le théorème de Gauss-Markov (corollaire du théorème 2, n° 1.3) : cherchant un estimateur ayant une propriété d'optimalité uniforme lorsque le paramètre inconnu θ décrit \mathcal{X} , on peut déterminer l'estimateur optimal $R^\#$ au moyen des critères (k) ou (kk) ; le critère traditionnel (kk) signifiant que, pour tout $x \in \mathcal{X}$, l'estimateur numérique affine $y \rightarrow (x, R^\# y)$ donne à la variable aléatoire numérique $(x, R^\# Y)$ une variance minimum parmi la classe des estimateurs numériques affines sous biais (propriété (kkk)) de la forme linéaire (x, \cdot) sur \mathcal{X} . On note que l'estimateur $R^\#$ ne dépend que des caractéristiques du second ordre m , C et Σ du modèle linéaire $(\mathcal{Y}, (P_x)_{x \in \mathcal{X}})$ et même, fait important en pratique, est invariant par les homothéties $\Sigma \rightarrow \sigma^2 \Sigma$ ($\sigma > 0$) effectuées sur Σ .

2.4. Méthode bayésienne. — Elle consiste ici à partir de la situation du n° 2.3 où la variable X apparaît comme un paramètre déterminé du système considéré et à introduire en plus une information *a priori* concernant X qui est représentée par une probabilité $\pi \in \mathcal{P}$. Associant à π la probabilité P_π sur $\mathcal{X} \times \mathcal{Y}$, on peut déterminer un estimateur affine optimal R_π comme au n° 2.2a au moyen des critères (i) ou (ii) (théorème 1).

⁽¹⁹⁾ C'est-à-dire une caractéristique du système dont la valeur est entièrement déterminée par les conditions de l'expérience qui permet de mesurer Y .

Ainsi, du point de vue mathématique, rien ne distingue ce cas de celui de la régression linéaire examiné au n° 2.2a; et on pourrait faire une remarque analogue en ce qui concerne le n° 2.2b et le théorème 2.

2.5. Le rapport entre les estimateurs R_π et $R^\#$ est précisé par le théorème 3 (n° 1.4) : pour chaque π , R_π est strictement meilleur que $R^\#$ [pour les critères adoptés, propriétés (I) et (II)], et R_π peut s'approcher arbitrairement près de $R^\#$ lorsque la mesure π s'étale assez régulièrement dans \mathcal{X} [propriété (III) ⁽²⁰⁾]. Cette dernière propriété est particulièrement explicite lorsque R_π et $R^\#$ sont déterminés en minimisant les formes quadratiques J_π et J respectivement (propriété (vi) du théorème 1 et (jjj) du théorème 2).

2.6. **Cas gaussien.** — Lorsque la loi P_π est gaussienne, l'estimateur R_π est optimal (par exemple pour le critère consistant à minimiser $E_\pi[||X - RY||^2]$) non seulement dans la classe \mathcal{R} des applications affines de \mathcal{Y} dans \mathcal{X} , mais encore dans la classe des applications mesurables V de \mathcal{Y} dans \mathcal{X} telles que $E_\pi[||V||^2] < +\infty$. De plus, R_π coïncide alors avec l'estimateur bayésien de X basé sur Y , c'est-à-dire que $R_\pi Y$ est alors l'espérance conditionnelle de X dans Y . Enfin, en ce qui concerne la méthode des moindres carrés, lorsque chacune des lois P_x est gaussienne, l'estimateur de Gauss-Markov $R^\#$ coïncide avec l'estimateur fourni par la méthode du maximum de vraisemblance.

2.7. **Retour sur la régression linéaire.** — La situation générale de régression linéaire de X sur Y correspond à la donnée d'une mesure de probabilité du second ordre P sur $\mathcal{X} \times \mathcal{Y}$ de moyenne $\mu = (\xi, \eta)$ et d'opérateur de covariance $\Delta = \begin{pmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{pmatrix}$. La situation envisagée ici [où μ et Δ sont donnés respectivement par les relations (1.4) et (1.5) (n° 1.1)] est, en fait, équivalente à la situation générale dès que Δ_{11} est inversible : il suffit pour le voir de poser $\Gamma_\pi = \Delta_{11}$, $C = \Delta_{21}\Delta_{11}^{-1}$ et $\Sigma = \Delta_{22} - \Delta_{21}\Delta_{11}^{-1}\Delta_{12}$, l'opérateur Σ ainsi défini étant positif car il s'écrit aussi $\Lambda\Delta\Lambda^*$, où Λ désigne l'application linéaire $(x, y) \rightarrow y - Cx$ de $\mathcal{X} \times \mathcal{Y}$ sur \mathcal{Y} . La présentation adoptée ici correspond simplement à la mise en évidence dans les données de la régression linéaire de Y sur X , laquelle est constituée (quelle que soit π) par l'application affine $x \rightarrow m + Cx$ de \mathcal{X} dans \mathcal{Y} .

⁽²⁰⁾ On peut donner une expression de cette propriété en termes d'optimalité bayésienne généralisée de $R^\#$ par rapport à la mesure de Lebesgue sur \mathcal{X} . Mais l'utilité statistique d'une telle généralisation est discutable.

§ 3. Démonstrations des énoncés du paragraphe 1.

3.1. **Étude préliminaire de P_π .** — On note $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, $z = (x, y)$ l'élément courant de \mathcal{Z} et $Z = (X, Y)$ l'application identique de \mathcal{Z} . On établit tout d'abord le résultat indiqué à la fin du n° 1.1 :

LEMME. — *Les hypothèses étant celles du n° 1.1, P_π est une probabilité du second ordre sur \mathcal{Z} admettant μ_π et Δ_π pour moyenne et opérateur de covariance, autrement dit :*

a) $E_\pi Z = \mu_\pi$.

b) $E_\pi[(\zeta, Z - \mu_\pi)(\zeta', Z - \mu_\pi)] = (\Delta_\pi \zeta, \zeta')$ pour tous $\zeta, \zeta' \in \mathcal{Z}$.

En effet, le point a) résulte de ce que, pour tout $\zeta = (\xi, \eta) \in \mathcal{Z}$, la fonction $(\zeta, Z) = (\xi, X) + (\eta, Y)$ est intégrable sur \mathcal{Z} , et,

$$\begin{aligned} E_\pi(\zeta, Z) &= \int_{\mathcal{X}} \pi(dx) \int_{\mathcal{Y}} P_x(dy)[(\xi, x) + (\eta, y)] \\ &= \int_{\mathcal{X}} \pi(dx)[(\xi, x) + (\eta, m + Cx)] \\ &= (\xi, \xi_\pi) + (\eta, m + C\xi_\pi) = (\zeta, \mu_\pi). \end{aligned}$$

Pour établir le point b), il suffit, puisque Δ_π est symétrique, de le vérifier quand $\zeta' = \zeta$. Or, pour $\zeta = (\xi, \eta) \in \mathcal{Z}$, on a,

$$(\zeta, Z - \mu_\pi)^2 = (\xi, X - \xi_\pi)^2 + (\eta, Y - m - C\xi_\pi)^2 + 2(\xi, X - \xi_\pi)(\eta, Y - m - C\xi_\pi).$$

L'intégrabilité de $(\zeta, Z - \mu_\pi)^2$ résultant aussitôt de celle des deux premiers termes du membre de droite, on les considérera en premier lieu.

Tout d'abord

$$\begin{aligned} E_\pi[(\xi, X - \xi_\pi)^2] &= \int_{\mathcal{X}} \pi(dx) \int_{\mathcal{Y}} P_x(dy)(\xi, x - \xi_\pi)^2 \\ &= \int_{\mathcal{X}} \pi(dx)(\xi, x - \xi_\pi)^2 = (\Gamma_\pi \xi, \xi); \end{aligned}$$

puis,

$$E_\pi[(\eta, Y - m - C\xi_\pi)^2] = \int_{\mathcal{X}} \pi(dx) \int_{\mathcal{Y}} P_x(dy)(\eta, y - m - C\xi_\pi)^2;$$

mais, faisant intervenir, pour chaque $x \in \mathcal{X}$, la moyenne $m + Cx$ de P_x , on voit que,

$$\int_{\mathcal{X}} P_x(dy)(\eta, y - m - C\xi_\pi)^2 = (\Sigma\eta, \eta) + (\eta, C(x - \xi_\pi))^2;$$

d'où,

$$E_{\pi}[(\eta, Y - m - C\xi_{\pi})^2] = (\Sigma\eta, \eta) + \int_{\mathcal{X}} \pi(dx)(C^*\eta, x - \xi_{\pi})^2 = ((\Sigma + C\Gamma_{\pi}C^*)\eta, \eta).$$

Un calcul analogue donne enfin :

$$\begin{aligned} E_{\pi}[(\zeta, X - \xi_{\pi})(\eta, Y - m - C\xi_{\pi})] &= \int_{\mathcal{X}} \pi(dx)(\zeta, x - \xi_{\pi}) \int_{\mathcal{Y}} P_x(dy)(\eta, y - m - C\xi_{\pi}) \\ &= \int_{\mathcal{X}} \pi(dx)(\zeta, x - \xi_{\pi})(\eta, C(x - \xi_{\pi})) = \int_{\mathcal{X}} \pi(dx)(\zeta, x - \xi_{\pi})(C^*\eta, x - \xi_{\pi}) \\ &= (C\Gamma_{\pi}\zeta, \eta); \end{aligned}$$

et il ne reste plus qu'à regrouper tous les termes pour constater que

$$E_{\pi}[(\zeta, Z - \mu_{\pi})^2] = (\Delta_{\pi}\zeta, \zeta).$$

CQFD

De plus, les deux relations suivantes interviendront dans la suite : pour tous $A, B \in \mathcal{L}(\mathcal{Z}, \mathcal{Z})$ et tous $z_1, z_2 \in \mathcal{Z}$, les fonctions (z_1, AZ) et (z_2, BZ) sont de carré intégrable sur \mathcal{Z} , et,

$$(3.1) \quad E_{\pi}[(z_1, A(Z - \mu_{\pi}))(z_2, B(Z - \mu_{\pi}))] = (B\Delta_{\pi}A^*z_1, z_2);$$

il en résulte, en faisant parcourir à $z_1 = z_2$ une base orthonormale de \mathcal{Z} et en sommant que

$$(3.2) \quad E_{\pi}(A(Z - \mu_{\pi}), B(Z - \mu_{\pi})) = \text{Tr } B\Delta_{\pi}A^*.$$

3.2. Démonstration du théorème 1. — Pour établir la propriété (1) on commencera par montrer l'existence de $R_1 \in \mathcal{R}$ satisfaisant (i), puis on montrera les équivalences annoncées selon le schéma :

$$(i) \Rightarrow (iv) \Rightarrow (iii) \Rightarrow (ii) \Rightarrow (i).$$

a) Existence de $R_1 \in \mathcal{R}$ satisfaisant (i) : on considère l'espace de Hilbert $L^2_{\mathcal{X}}(\mathcal{Z}, P_{\pi})$ avec son produit scalaire canonique,

$$(f(Z), g(Z))_2 = E_{\pi}(f(Z), g(Z)).$$

L'ensemble $H = \{RY, R \in \mathcal{R}\}$ est un sous-espace de dimension finie, donc un sous-espace fermé de $L^2_{\mathcal{X}}(\mathcal{Z}, P_{\pi})$, et la projection R_1Y de X sur H vérifie

$$E_{\pi}[\|X - R_1Y\|^2] \leq E_{\pi}[\|X - RY\|^2], \text{ pour tout } R \in \mathcal{R};$$

d'où l'existence de R_1 satisfaisant (i).

b) (i) \Rightarrow (iv): Soit $R_1 \in \mathcal{R}$ satisfaisant (i); $R_1 Y$ est la projection de X sur H dans $L_{\mathbb{R}}^2(\mathcal{Z}, P_{\pi})$; donc $R_1 Y$ est caractérisé dans H par la relation,

$$(X - R_1 Y, RY)_2 = 0, \text{ pour tout } R \in \mathcal{R};$$

mais cette condition entraîne en particulier que,

$$E_{\pi}(X - R_1 Y, \xi) = 0, \text{ pour tout } \xi \in \mathcal{X},$$

et

$$E_{\pi}(X - R_1 Y, SY) = 0, \text{ pour tout } S \in \mathcal{L}(\mathcal{Y}, \mathcal{X}).$$

La première de ces relations entraîne que $E_{\pi}[R_1 Y] = \xi_{\pi} = E_{\pi}[X]$, ce qui permet d'écrire R_1 sous la forme,

$$(3.3) \quad R_1 y = E_{\pi}[X] + S_1(y - E_{\pi}[Y]) \quad (y \in \mathcal{Y}) \text{ avec } S_1 \in \mathcal{L}(\mathcal{Y}, \mathcal{X});$$

et la deuxième donne alors,

$$(3.4) \quad E_{\pi}(X - E_{\pi}[X] - S_1(Y - E_{\pi}[Y]), S(Y - E_{\pi}[Y])) = 0 \text{ pour tout } S \in \mathcal{L}(\mathcal{Y}, \mathcal{X}).$$

Or, pour un choix convenable de A et B , (3.2) (n° 3.1) entraîne que,

$$E_{\pi}(X - E_{\pi}[X] - S_1(Y - E_{\pi}[Y]), S(Y - E_{\pi}[Y])) = \text{Tr} (S[CF_{\pi} - (CF_{\pi}C^* + \Sigma)S_1^*]);$$

de sorte que (3.4) s'écrit aussi,

$$\text{Tr} (S[CF_{\pi} - (CF_{\pi}C^* + \Sigma)S_1^*]) = 0, \text{ pour tout } S \in \mathcal{L}(\mathcal{Y}, \mathcal{X});$$

d'où $CF_{\pi} - (CF_{\pi}C^* + \Sigma)S_1^* = 0$, comme on le voit en faisant parcourir à S l'ensemble des applications de rang 1 de $\mathcal{L}(\mathcal{Y}, \mathcal{X})$; et par passage à l'adjoint, on obtient la relation (1.7), puis la propriété (iv) d'après (3.3) et (1.4).

c) (iv) \Rightarrow (iii): de la première relation de (iv) on tire $E_{\pi}[X - R_1 Y] = 0$; ce qui donne,

$$E_{\pi}[(y, Y)(x, X - R_1 Y)] = E_{\pi}[(y, Y - E_{\pi}Y)(x, X - E_{\pi}X - S_1(Y - E_{\pi}Y))];$$

et, pour A et B bien choisis, (3.1) donne cette fois-ci

$$E_{\pi}[(y, Y - E_{\pi}Y)(x, X - E_{\pi}X - S_1(Y - E_{\pi}Y))] = [(CF_{\pi} - (CF_{\pi}C^* + \Sigma)S_1^*)x, y],$$

d'où (iii) en vertu de (1.7).

d) (iii) \Rightarrow (ii): Pour $Ry = \xi + Sy$ et utilisant $E_{\pi}[X - R_1 Y] = 0$, on note que, pour tout $x \in \mathcal{X}$,

$$E_{\pi}[(x, RY)(x, X - R_1 Y)] = E_{\pi}[(S^*x, Y)(x, X - R_1 Y)] = 0;$$

on en déduit que, pour tout $x \in \mathcal{X}$, dans l'espace $L_{\mathbb{R}}^2(\mathcal{Z}, P_{\pi})$, $(x, R_1 Y)$ est

la projection de (x, X) sur le sous-espace $\{(x, RY); R \in \mathcal{R}\}$, la propriété (ii) en résulte.

e) Enfin (ii) \Rightarrow (i) est immédiat, car

$$(3.5) \quad E_{\pi}[\|X - RY\|^2] = \sum_{i=1}^d E_{\pi}[(e_i, X - RY)^2],$$

où $(e_i)_{1 \leq i \leq d}$ désigne une base orthonormale de \mathcal{X} .

Pour établir la propriété (2), supposant ici que Σ et Γ_{π} sont inversibles, on va montrer successivement que

$$(iv) \Leftrightarrow (vii) \quad \text{et que} \quad (vi) \Leftrightarrow (vii).$$

f) (iv) \Leftrightarrow (vii): il suffit de remarquer que, sous l'hypothèse que Σ et Γ_{π} sont inversibles, on a,

$$(3.6) \quad \Gamma_{\pi} C^* (C \Gamma_{\pi} C^* + \Sigma)^{-1} = (\Gamma_{\pi}^{-1} + C^* \Sigma^{-1} C)^{-1} C^* \Sigma^{-1}$$

En effet, la relation (3.6) équivaut à

$$(\Gamma_{\pi}^{-1} + C^* \Sigma^{-1} C) \Gamma_{\pi} C^* = C^* \Sigma^{-1} (C \Gamma_{\pi} C^* + \Sigma);$$

et cette dernière relation est visiblement une identité.

g) (vi) \Leftrightarrow (vii): On introduit en premier lieu un produit scalaire $(\cdot, \cdot)_{\pi}$ sur $\mathcal{X} \times \mathcal{Y}$ en posant, pour $z = (x, y)$ et $z' = (x', y') \in \mathcal{X} \times \mathcal{Y}$:

$$(z, z')_{\pi} = (\Gamma_{\pi}^{-1} x, x') + (\Sigma^{-1}((x - y), (x' - y'));$$

on constate alors que, pour $x \in \mathcal{X}$ et $y \in \mathcal{Y}$,

$$J_{\pi}(x, y) = (\mu_{\pi} - (x, y), \mu_{\pi} - (x, y))_{\pi};$$

pour tout y , $J_{\pi}(x, y)$ fournit donc le carré de la distance (prise au sens de $(\cdot, \cdot)_{\pi}$) entre μ_{π} et le point courant de la variété affine $V_y = \{(x, y); x \in \mathcal{X}\}$. Pour chaque $y \in \mathcal{Y}$, un seul $x_1 \in \mathcal{X}$ réalise donc le minimum de la fonction $x \rightarrow J_{\pi}(x, y)$, à savoir l'élément x_1 tel que (x_1, y) soit la projection de μ_{π} sur V_y :

$$(\mu_{\pi} - (x_1, y), (x, 0))_{\pi} = 0, \quad \text{pour tout } x \in \mathcal{X}.$$

Après explicitation, on constate que cette dernière relation équivaut à

$$\Gamma_{\pi}^{-1}(x_1 - \xi_{\pi}) + C^* \Sigma^{-1}(y - m - Cx_1) = 0,$$

équation dont le point $x_1 = R_1 y$ donné par (1.9) est la seule solution. Ceci achève la démonstration du théorème 1.

3.3. **Démonstration du théorème 2.** — On établira la propriété (1) en montrant simultanément les équivalences

$$(h) \Leftrightarrow (hhh) \quad \text{et} \quad (hh) \Leftrightarrow (hhh).$$

Pour $R \in \mathcal{R}$, du type $Ry = \xi + Sy$, on remarque, en introduisant

$$E_\pi[RY] = \xi + S[E_\pi Y],$$

que

$$(3.7) \quad E_\pi[\|X - RY\|^2] = \|E_\pi[X - RY]\|^2 + E_\pi[\|X - E_\pi X - S(Y - E_\pi Y)\|^2]$$

et que, pour tout $x \in \mathcal{X}$

$$(3.8) \quad E_\pi[((x, X) - (x, RY))^2] \\ = (x, E_\pi[X - RY])^2 + E_\pi[(x, X - E_\pi X - S(Y - E_\pi Y))^2].$$

Cela étant, pour un bon choix de $A = B$, les relations (3.2) et (3.1) donnent

$$E_\pi[\|X - E_\pi X - S(Y - E_\pi Y)\|^2] = \text{Tr}(S\Sigma S^*) + \text{Tr}((I_x - SC)\Gamma_\pi(I_x - C^*S^*))$$

et

$$E_\pi[(x, X - E_\pi X - S(Y - E_\pi Y))^2] = (S\Sigma S^*x, x) + ((I_x - SC)\Gamma_\pi(I_x - C^*S^*)x, x);$$

si bien, qu'explicitant les relations (3.7) et (3.8), on obtient

$$(3.9) \quad E_\pi[\|X - RY\|^2] = \|(I_x - SC)\xi_\pi - \xi - Sm\|^2 + \text{Tr}(S\Sigma S^*) \\ + \text{Tr}((I_x - SC)\Gamma_\pi(I_x - C^*S^*)),$$

et, pour tout $x \in \mathcal{X}$,

$$(3.10) \quad E_\pi[((x, X) - (x, RY))^2] = (x, (I_x - SC)\xi_\pi - \xi - Sm)^2 \\ + (S\Sigma S^*x, x) + ((I_x - SC)\Gamma_\pi(I_x - C^*S^*)x, x).$$

On déduit aussitôt des relations (3.9) et (3.10) que les propriétés (h) et (hh) équivalent chacune à (hhh) [en effet si $SC \neq I_x$ et quand Γ_π parcourt l'ensemble des opérateurs symétriques positifs sur \mathcal{X} , le terme

$$\text{Tr}((I_x - SC)\Gamma_\pi(I_x - C^*S^*))$$

(respectivement le terme $((I_x - SC)\Gamma_\pi(I_x - C^*S^*)x, x)$ pour un $x \in \mathcal{X}$ au moins) n'est pas borné].

La propriété (2) sera obtenue en établissant les équivalences annoncées conformément au schéma

$$(j) \Rightarrow (jv) \Rightarrow (jj) \Rightarrow (j) \quad \text{et} \quad (jv) \Leftrightarrow (jj).$$

a) $(j) \Rightarrow (jv)$: On constate en premier lieu que $H^\# = \{RY; R \in \mathcal{R}^\#\}$ est un sous-espace affine de dimension finie de $L_{\mathcal{X}}^2(\mathcal{Z}, P_\pi)$, et que le sous-

espace vectoriel parallèle à $H^\#$ est exactement constitué des $\xi + SY$ pour lesquels $\xi \in \mathcal{X}$ et $S \in D_C$, en définissant D_C par,

$$S \in D_C \Leftrightarrow S \in \mathcal{L}(\mathcal{Y}, \mathcal{X}) \text{ et } SC = 0.$$

On remarque alors que, pour R_2 vérifiant (j), R_2Y est la projection de X sur $H^\#$ et satisfait donc

$$(X - R_2Y, \xi + SY)_2 = 0 \text{ pour tout } \xi \in \mathcal{X}, \text{ tout } S \in D_C;$$

d'où, utilisant le fait que $R_2 \in \mathcal{R}^\#$, c'est-à-dire que $R_2y = \xi_2 + S_2y$ avec

$$(3.11) \quad S_2C = I_{\mathcal{X}},$$

on obtient, en procédant comme dans la démonstration du théorème 1,

$$(3.12) \quad E_\pi[X - R_2Y] = -\xi_2 - S_2m = 0,$$

et

$$(3.13) \quad \text{Tr}(S\Sigma S_2^*) = 0, \text{ pour tout } S \in D_C.$$

Vérifiant ensuite que $y \otimes x \in D_C$ dès que $C^*y = 0$ ⁽²¹⁾, puis explicitant $\text{Tr}(S\Sigma S_2^*)$ quand $S = y \otimes x$, on déduit de (3.13) que

$$(S_2\Sigma y, x) = 0, \text{ pour tout } x \in \mathcal{X} \text{ et tout } y \in \mathcal{Y} \text{ tel que } C^*y = 0;$$

il en résulte que $S_2\Sigma y = 0$ pour tout y tel que $C^*y = 0$; donc il existe $\Lambda \in \mathcal{L}(\mathcal{Y}, \mathcal{X})$ vérifiant $S_2\Sigma = \Lambda C^*$; donc aussi $S_2 = \Lambda C^*\Sigma^{-1}$ et, en portant dans (3.11), $\Lambda C^*\Sigma^{-1}C = I_{\mathcal{X}}$; et finalement, $S_2 = (C^*\Sigma^{-1}C)^{-1}C^*\Sigma^{-1}$, ce qui, joint à la relation (3.12), achève d'établir (jv).

b) (jv) \Rightarrow (jj): Définissant $R_2y = \xi_2 + S_2y$ par (jv), on commence par noter que,

$$(3.14) \quad S\Sigma S_2^* = 0 \text{ pour tout } S \in D_C;$$

exploitant ensuite la nullité de $E_\pi[X - R_2Y]$, on remarque, en utilisant successivement (3.1), le fait que $S_2C = I_{\mathcal{X}}$ et (3.14), que pour tous $\xi \in \mathcal{X}$, $S \in D_C$ et $x \in \mathcal{X}$, $E_\pi[(x, X - R_2Y)(x, \xi + SY)] = 0$; ce qui montre précisément, pour tout $\pi \in \mathcal{P}$ et tout $x \in \mathcal{X}$, que, dans $L_{\mathbb{R}}^2(\mathcal{L}, P_\pi)$, (x, R_2Y) est la projection de (x, X) sur le sous-espace affine $\{(x, RY); R \in \mathcal{R}^\#\}$, d'où la condition (jj).

⁽²¹⁾ Pour $y \in \mathcal{Y}$ et $x \in \mathcal{X}$, $y \otimes x$ est l'élément de $\mathcal{L}(\mathcal{Y}, \mathcal{X})$ défini par

$$(y \otimes x)y' = (y, y')x \text{ pour tout } y' \in \mathcal{Y}.$$

c) $(jj) \Rightarrow (j)$ est immédiat d'après (3.5) (n° 3.2e).

d) Enfin $(jv) \Leftrightarrow (jjj)$: Introduisant sur \mathcal{Y} le produit scalaire

$$(y, y')_{\Sigma} = (\Sigma^{-1}y, y'),$$

on note que $J(x, y) = (y - (m + Cx), y - (m + Cx))_{\Sigma}$ fournit le carré de la distance de y au point courant de la variété affine $\{m + Cx; x \in \mathcal{X}\} \subset \mathcal{Y}$; procédant alors comme au n° 3.2g, on constate, étant donné $y \in \mathcal{Y}$, qu'un élément $x_2 \in \mathcal{X}$ minimise la fonction $x \rightarrow J(x, y)$ si et seulement si

$$(y - (m + Cx_2), Cx_2)_{\Sigma} = 0, \text{ pour tout } x \in \mathcal{X};$$

or, cette dernière condition équivaut à $x_2 = R_2y$ où R_2y est défini par (jv) ; d'où l'équivalence de (jjj) et (jv) et le théorème 2.

3.4. Démonstration du théorème de Gauss-Markov. — L'équivalence $(k) \Leftrightarrow (jv)$ s'obtient en notant que chacune des propriétés (k) et (jv) est équivalente à (j) (d'après le théorème 2).

L'implication $(jv) \Rightarrow (kkk)$ n'est aussi qu'une simple vérification; on en déduit en particulier que, R_2 étant définie par (jv) ,

$$(3.15) \quad E_{\theta}(x, R_2Y) = (x, \theta) \text{ pour tout } x \in \mathcal{X} \text{ et tout } \theta \in \mathcal{X}.$$

On établira l'équivalence de (kk) et de (jv) à l'aide du lemme suivant :

LEMME. — Pour tout $x \in \mathcal{X}$, tout $r \in \mathcal{R}_x^{\#}$, il existe $R \in \mathcal{R}^{\#}$ tel que :

$$ry = (x, Ry), \text{ pour tout } y \in \mathcal{Y}.$$

En effet, on note en premier lieu que, pour $r \in \mathcal{R}_x^{\#}$, il existe $\eta \in \mathcal{Y}$ satisfaisant $C^*\eta = x$ et tel que $ry = (\eta, y - m)$ pour tout $y \in \mathcal{Y}$. Désignant alors par M un supplémentaire de $\text{Ker } C^*$ dans \mathcal{Y} tel que $\eta \in M$, on constate que C^* est un isomorphisme de M sur \mathcal{X} . Introduisant alors l'unique élément S^* de $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ tel que $S^*C^*y = y$ pour tout $y \in M$, on note que $S^*x = \eta$ et que $C^*S^* = I_{\mathcal{X}}$; d'où le résultat en définissant $R \in \mathcal{R}^{\#}$ par,

$$Ry = S(y - m) \text{ pour tout } y \in \mathcal{Y}.$$

On déduit du lemme que, pour tout $x \in \mathcal{X}$ et tout $\theta \in \mathcal{X}$,

$$\inf_{r \in \mathcal{R}_x^{\#}} E_{\theta}[(x, \theta) - rY]^2 \geq \inf_{R \in \mathcal{R}^{\#}} E_{\theta}[(x, \theta) - (x, RY)]^2,$$

de sorte que, grâce à la relation (3.15) et l'implication $(jjj) \Rightarrow (jj)$,

$$\inf_{r \in \mathcal{R}_x^{\#}} E_{\theta}[(x, \theta) - rY]^2 = \inf_{R \in \mathcal{R}^{\#}} E_{\theta}[(x, \theta) - (x, RY)]^2;$$

d'où le fait que (kk) équivaut à (jj) donc aussi à (jv) , et le théorème de Gauss-Markov.

Remarque. — La substance algébrique du théorème de Gauss-Markov peut être résumée ainsi: Σ étant inversible et C injective,

$$S_2 = (C^*\Sigma^{-1}C)^{-1}C^*\Sigma^{-1}$$

est l'unique application linéaire S de \mathcal{Y} dans \mathcal{X} minimisant $\text{Tr}(SES^*)$ ou, de façon équivalente (SES^*x, x) pour tout $x \in \mathcal{X}$. En outre, pour chaque $y \in \mathcal{Y}$, S_2y est l'unique élément x de \mathcal{X} minimisant la quantité

$$(\Sigma^{-1}(y - Cx), y - Cx).$$

3.5. Démonstration du théorème 3. — *a)* Il résulte tout d'abord des propriétés (i) et (j), (ii) et (jj) (théorèmes 1 et 2) de R_π et $R^\#$ respectivement que

$$(3.16) \quad E_\pi[||X - R_\pi Y||^2] \leq E_\pi[||X - R^\# Y||^2],$$

et

$$(3.17) \quad E_\pi[((x, X) - (x, R_\pi Y))^2] \leq E_\pi[((x, X) - (x, R^\# Y))^2] \text{ pour tout } x \in \mathcal{X},$$

l'égalité dans (3.16) (resp. dans (3.21) pour tout $x \in \mathcal{X}$) ne pouvant avoir lieu que si $R_\pi = R^\#$ d'après l'équivalence de (i) et de (iv) (resp. de (ii) et de (iv)). Or, on va voir que $R_\pi \neq R^\#$ pour tout $\pi \in \mathcal{P}$; il en résultera la propriété (l) du théorème 3. Pour montrer que $R_\pi \neq R^\#$, il suffit d'utiliser le fait que $R^\#$ est invariant par les homothéties effectuées sur Σ ce qui n'est pas le cas de R_π , ainsi qu'on le vérifie aussitôt.

b) Concernant la propriété (lll), l'équivalence de $\lim_{n \rightarrow \infty} \Gamma_{\pi_n}^{-1} = 0$ et de la relation (1.20) s'obtient aisément lorsque l'on prend R_{π_n} sous la forme (1.9) (propriété (vii)). Pour établir les relations (1.21) et (1.22) sous l'hypothèse $\lim_{n \rightarrow \infty} \Gamma_{\pi_n}^{-1} = 0$, on constate, en portant l'expression (1.9) de R_{π_n} dans (3.9) et (3.10), (n° 3.3) que,

$$(3.18) \quad E_\pi[||X - R_{\pi_n} Y||^2] = \text{Tr}(S_n \Sigma S_n^*) + \text{Tr}((I_{\mathcal{X}} - (I_{\mathcal{X}} + \Lambda_n)^{-1})\Gamma_{\pi_n}(I_{\mathcal{X}} - (I_{\mathcal{X}} + \Lambda_n^*)^{-1}))$$

et

$$(3.19) \quad E_\pi[((x, X) - (x, R_{\pi_n} Y))^2] = (S_n \Sigma S_n^* x, x) + ((I_{\mathcal{X}} - (I_{\mathcal{X}} + \Lambda_n)^{-1})\Gamma_{\pi_n}(I_{\mathcal{X}} - (I_{\mathcal{X}} + \Lambda_n^*)^{-1})x, x),$$

où l'on a posé

$$S_n = (\Gamma_{\pi_n}^{-1} + C^*\Sigma^{-1}C)^{-1}C^*\Sigma^{-1} \quad \text{et} \quad \Lambda_n = (C^*\Sigma^{-1}C)^{-1}\Gamma_{\pi_n}^{-1}.$$

On observe alors que, sous la condition $\lim_{n \rightarrow \infty} \Gamma_{\pi_n}^{-1} = 0$,

$$\lim_{n \rightarrow \infty} S_n \Sigma S_n^* = C^*\Sigma^{-1}C,$$

et [en utilisant l'identité $(I_{\mathcal{X}} + \Lambda_n)^{-1} = I_{\mathcal{X}} - (I_{\mathcal{X}} + \Lambda_n)^{-1}\Lambda_n$,

$$\lim_{n \rightarrow \infty} (I_{\mathcal{X}} - (I_{\mathcal{X}} + \Lambda_n)^{-1})\Gamma_{\pi_n}(I_{\mathcal{X}} - (I_{\mathcal{X}} + \Lambda_n^*)^{-1}) = 0;$$

de sorte que

$$\lim_{n \rightarrow \infty} E_{\pi_n}[\|X - R_n Y\|^2] = \text{Tr}(C^* \Sigma^{-1} C),$$

et

$$\lim_{n \rightarrow \infty} E_{\pi_n}[(x, X) - (x, R_{\pi_n} Y)]^2 = (C^* \Sigma^{-1} C x, x), \text{ pour tout } x \in \mathcal{X};$$

et il ne reste plus qu'à utiliser les relations (1.10) et (1.11) avec $R = R^\#$ pour obtenir

$$(3.20) \quad \lim_{n \rightarrow \infty} E_{\pi_n}[\|X - R_n Y\|^2] = E_{\pi}[\|X - R^\# Y\|^2], \text{ pour tout } \pi \in \mathcal{P},$$

et

$$(3.21) \quad \lim_{n \rightarrow \infty} E_{\pi_n}[(x, X) - (x, R_{\pi_n} Y)]^2 = E_{\pi}[(x, X) - (x, R_{\#} Y)]^2,$$

pour tout $\pi \in \mathcal{P}$, tout $x \in \mathcal{X}$.

Rapprochant enfin (3.20) de (1.16) et (3.21) de (1.17), on obtient les relations (1.18) et (1.19) (propriété *(II)*) et les relations (1.21) et (1.22) (propriété *(III)*), ce qui achève la démonstration du théorème 3.

Manuscrit reçu le 25 mars 1971.

Philippe COURRÈGE,
85, boulevard de Port-Royal 75-Paris 13^e.

Jean-Louis PHILOCHE,
5, rue Poliveau 75-Paris 5^e.

Pierre PRIOURET,
26, rue du Commandant-Mouchotte 75-Paris 14^e.