

# 1. Modèles linéaires. Estimateurs.

Objekttyp: **Chapter**

Zeitschrift: **L'Enseignement Mathématique**

Band (Jahr): **6 (1960)**

Heft 1: **L'ENSEIGNEMENT MATHÉMATIQUE**

PDF erstellt am: **14.09.2024**

## **Nutzungsbedingungen**

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern. Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden. Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

## **Haftungsausschluss**

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

0, 34. Rappelons enfin que, si l'on nomme « gaussienne » toute aléatoire normale de moyenne nulle et de variance égale à 1, on a les énoncés suivants (cfr. [III], chap. 18):

- a) la somme des carrés de  $p$  aléatoires gaussiennes indépendantes est une aléatoire  $\chi^2$  à  $p$  degrés de liberté (en abrégé,  $\chi_p^2$ );
- b) si  $\mathbf{x}$  est une aléatoire gaussienne et  $\mathbf{u}$  une aléatoire  $\chi_p^2$  indépendante de  $\mathbf{x}$ , le quotient  $\mathbf{x}/\sqrt{(\mathbf{u}/p)}$  est une aléatoire de Student à  $p$  degrés de liberté (en abrégé,  $t_p$ );
- c) si  $\mathbf{u}$  et  $\mathbf{v}$  sont des aléatoires  $\chi^2$ , respectivement à  $m$  et  $n$  degrés de liberté, indépendantes, le quotient  $(\mathbf{u}/m):(\mathbf{v}/n)$  est une aléatoire  $\mathbf{F}$  de Snedecor à  $(m, n)$  degrés de liberté (en abrégé,  $\mathbf{F}_{m,n}$ ); il est souvent plus commode d'utiliser alors le fait que  $\mathbf{v}/(\mathbf{u} + \mathbf{v})$  est une aléatoire  $\beta_{p,q}$  avec  $p = n/2$ ,  $q = m/2$ , et de se référer aux tables de la distribution  $\beta$  (en général, en effet,  $n \geq m$ , donc  $p \geq q$ , comme dans les tables de Pearson [IV]); on notera que les grandes valeurs de  $\mathbf{F}$  correspondent aux petites valeurs de  $\beta$ .

## 1. MODÈLES LINÉAIRES. ESTIMATEURS.

### 1, 1. Définitions.

1, 11. Considérons une expérience aléatoire dont le résultat est un  $n$ -uple ordonné de nombres réels, toutes les valeurs a priori possibles, de  $-\infty$  à  $+\infty$ , étant en effet à prendre en considération. Structurons l'ensemble des observations possibles en un espace vectoriel euclidien sur le corps des réels en postulant que, si  $\alpha = (a_1, \dots, a_n)$  et  $\beta = (b_1, \dots, b_n)$  sont deux observations, on a

pour la somme:

$$\alpha + \beta = (a_1 + b_1, \dots, a_n + b_n), \quad (1)$$

pour le produit par un scalaire:

$$p \alpha = (p a_1, \dots, p a_n), \quad (2)$$

pour le produit scalaire:

$$\langle \alpha, \beta \rangle = a_1 b_1 + \dots + a_n b_n. \quad (3)$$

L'espace des observations, ainsi structuré, sera désigné par  $\mathbf{V}$ , et son dual par  $\mathbf{V}^*$ ;  $\mathbf{V}$  et  $\mathbf{V}^*$  sont évidemment de dimension  $n$ .

Les  $n$  observations  $(1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$  forment évidemment une base de  $\mathbf{V}$ ; elle sera désignée par  $\mathfrak{B}$ , et sa duale par  $\mathfrak{B}^*$ . La base  $\mathfrak{B}$  n'a, bien entendu, aucun privilège de droit; son lien particulièrement étroit avec la forme même des observations lui confère cependant un privilège de fait, qui se traduit notamment en ceci, que c'est par rapport à  $\mathfrak{B}$  que les relations (1), (2), (3) explicitent la définition des opérations fondamentales de  $\mathbf{V}$ . Sauf mention expresse du contraire, les représentations de vecteurs de  $\mathbf{V}$  ou  $\mathbf{V}^*$  seront toujours censées être faites par rapport à  $\mathfrak{B}$  ou  $\mathfrak{B}^*$ ; comme la relation (3) implique que  $\mathfrak{B}$  est orthonormale, les conventions du § 0, 2 seront appliquées.

1, 12. A côté de l'espace des observations, nous considérerons un autre espace vectoriel,  $\mathbf{B}$ , de dimension  $p \leq n$ , dit « espace des paramètres », et son dual,  $\mathbf{B}^*$ , « espace des combinaisons paramétriques ». Il n'existe pas, en général, de base « naturelle » qui soit à  $\mathbf{B}$  ce que  $\mathfrak{B}$  est à  $\mathbf{V}$ ; aussi les représentations des vecteurs de  $\mathbf{B}$  seront-elles notées par un symbole rappelant la base utilisée (la dualité des bases de  $\mathbf{B}$  et  $\mathbf{B}^*$  allant de soi).

1, 13. Le caractère aléatoire des observations a pour conséquence l'existence d'une catégorie d'épreuve  $\mathbf{C}$ , munie d'une mesure probabiliste  $\mathbf{Pr}$ ; en tant qu'éléments aléatoires, les observations constituent un vecteur aléatoire de  $\mathbf{V}$  (application mesurable de  $\mathbf{C}$  dans  $\mathbf{V}$ ). La mesure  $\mathbf{Pr}$  est décrite d'une façon explicite par référence à la base  $\mathfrak{B}$ , par le postulat que les composantes par rapport à  $\mathfrak{B}$  du vecteur aléatoire  $\mathfrak{x}$  représentant les observations sont  $n$  variables aléatoires normales, indépendantes, de même variance  $\sigma^2$ , dont les moyennes sont  $n$  éléments bien déterminés de  $\mathbf{B}^*$ :

$$\mathbf{E} e_i^* \mathfrak{x} \equiv e_i^* \int_{\mathbf{C}} x(\lambda) \mathbf{Pr}(d\lambda) = e_i^* \mathfrak{A} \mathfrak{b} \quad (x \in \mathbf{V}, \mathfrak{b} \in \mathbf{B}),$$

l'application  $\mathfrak{A}$  étant de rang  $r (\leq p)$ .

On voit donc que l'opérateur « valeur moyenne dans  $\mathbf{C}$  » induit une application linéaire de  $\mathbf{V}^*$  dans  $\mathbf{B}^*$ ; si on la désigne par  $\mathfrak{E}$ , on a, pour tout  $l^* \in \mathbf{V}^*$ ,

$$\mathbf{E} (l^* \mathfrak{z}) = l^* \mathbf{E} \mathfrak{z} = l^* \mathfrak{z} b, \quad (4)$$

d'où

$$\mathfrak{C}: l^* \rightarrow l^* \mathfrak{z} b \quad . \quad 8)$$

$\mathfrak{C}$  est évidemment de rang  $r$ .

1, 14. Le noyau de  $\mathfrak{C}$  (sous-espace  $V_0$  de  $V^*$  formé des vecteurs  $l^*$  tels que  $\mathfrak{C} l^*$  soit identiquement nul <sup>9)</sup>) est appelé « espace des erreurs » <sup>10)</sup>; il est de dimension  $n - \text{rg } \mathfrak{C} = n - r$ .

Le complément orthogonal de  $V_0$  (sous-espace  $V_+$  de  $V^*$  formé des vecteurs  $m^*$  pour lesquels  $m^* l = 0$  pour tout  $l^* \in V_0$ ) est appelé « espace des estimatrices ». Une estimatrice est donc, par définition, une fonctionnelle linéaire des observations, orthogonale à toute fonctionnelle dont la moyenne est identiquement <sup>9)</sup> nulle. On remarquera que,  $V^*$  étant somme directe de  $V_0$  et  $V_+$ , à toute fonctionnelle linéaire des observations dont la moyenne n'est pas identiquement <sup>9)</sup> nulle correspond une et une seule fonctionnelle de  $V_+$  ayant identiquement <sup>9)</sup> même moyenne qu'elle.

Enfin, l'image de  $\mathfrak{C}$  (sous-espace  $B_+$  de  $B^*$  formé des combinaisons paramétriques  $b^*$  pour lesquelles il existe un vecteur  $l^* \in V^*$  tel que  $\mathbf{E} l^* \mathfrak{z} = b^* b$ ) s'appelle « espace des combinaisons (paramétriques) estimables ». Nous noterons  $B_0$  un complément quelconque de  $B_+$ .

On sait que la restriction de  $\mathfrak{C}$  à  $V_+$  est un isomorphisme de  $V_+$  sur  $B_+$ ; on peut donc énoncer que

*toute combinaison estimable est la moyenne d'une et une seule estimatrice, et réciproquement.*

### 1, 2. Estimateurs privilégiés.

1, 21. Si l'on a, pour  $l^* \in V^*$ ,  $\mathbf{E} l^* \mathfrak{z} = f^* b$ ,  $f^* b$  est évidemment une combinaison estimable, et  $l^* \mathfrak{z}$  en est un estimateur fidèle (au sens de la théorie statistique de l'estimation); si  $m^* \in V_0$ ,  $(l^* + m^*) \mathfrak{z}$  est aussi un estimateur fidèle de  $f^* b$ ; pour distinguer, parmi tous ces estimateurs fidèles de  $f^* b$ , l'unique estimatrice, celle-ci est dite « estimateur privilégié de  $f^* b$  », et désignée par  $l_f^* \mathfrak{z}$  (donc, par définition,  $\mathbf{E} l_f^* \mathfrak{z} = f^* b$  et  $l_f^* \in V_+$ ).

1, 22. Théorème. Parmi tous les estimateurs fidèles de la combinaison estimable  $f^* b$ , l'estimateur privilégié a la variance minimum.

Soit en effet  $m^*$  tel que  $E m^* \varepsilon = f^* b$ . On a, en posant  $\tilde{\varepsilon} = \varepsilon - E \varepsilon$ ,

$$\begin{aligned} \text{var } m^* \varepsilon &= E [m^* \varepsilon - E m^* \varepsilon] [m^* \varepsilon - E m^* \varepsilon] \\ &= E (m^* \tilde{\varepsilon}) (\tilde{\varepsilon}^* m)^* = m^* (E \tilde{\varepsilon} \tilde{\varepsilon}^*) m. \end{aligned} \quad (5)$$

La définition des propriétés distributionnelles de  $\varepsilon$  faisant intervenir la base  $\mathfrak{B}$ , introduisons cette base pour un calcul explicite:

$$\begin{aligned} E \tilde{\varepsilon} \tilde{\varepsilon}^* &= E \tilde{\varepsilon}_P (\tilde{\varepsilon}_P)^T = E \begin{vmatrix} \tilde{\varepsilon}_1^2 & \tilde{\varepsilon}_1 \tilde{\varepsilon}_2 \dots & \tilde{\varepsilon}_1 \tilde{\varepsilon}_n \\ \vdots & & \\ \tilde{\varepsilon}_n \tilde{\varepsilon}_1 \dots & & \tilde{\varepsilon}_n^2 \end{vmatrix} \\ &= \mathfrak{J}_n \sigma^2 \end{aligned}$$

d'où

$$\text{var } m^* \varepsilon = (m^* m) \sigma^2.$$

Soit alors

$$m^* = l_f^* + m_0^*,$$

de sorte que

$$l_f^* m_0 = m_0^* l_f = 0;$$

on a

$$\text{var } m^* \varepsilon = (l_f^* l_f) \sigma^2 + (m_0^* m_0) \sigma^2 \geq (l_f^* l_f) \sigma^2 = \text{var } l_f^* \varepsilon,$$

l'inégalité étant d'ailleurs stricte si  $m_0^* \neq 0$ .

### 1, 3. Exécution des calculs.

1, 31. Pour l'exécution effective des calculs, il importe d'introduire une base dans chacun des espaces considérés; dans ce paragraphe,  $V$  et  $V^*$  sont rapportés aux bases  $\mathfrak{B}$  et  $\mathfrak{B}^*$ ,  $B$  et  $B^*$  sont rapportés à des bases déterminées  $\mathfrak{H}$  et  $\mathfrak{H}^*$ .

1, 32. Pour que  $l^* \in V_0$ , il est nécessaire et suffisant que  $l^* \mathfrak{A} = 0$ ; donc

*l'espace des erreurs est engendré par ceux des vecteurs de  $V^*$  qui sont orthogonaux aux colonnes de la matrice  $\mathfrak{A}$  (plus explicitement:  $\mathfrak{A}_{P,H}$ ).*

Il en résulte immédiatement que l'espace des estimateurs est engendré par les lignes de la matrice  $\mathcal{A}^T$ , donc

toute estimateur est de la forme  $l^* \mathcal{A}^T \mathfrak{z}$ .

Comme  $\mathbf{E} l^* \mathcal{A}^T \mathfrak{z} = l^* \mathcal{A}^T \mathbf{E} \mathfrak{z} = l^* \mathcal{A}^T \mathcal{A} \mathfrak{b}_H$ , cet énoncé, à son tour, entraîne celui-ci:

toute combinaison estimable est de la forme  $l^* \mathcal{A}^T \mathcal{A} \mathfrak{b}_H$  et réciproquement.

Ainsi la correspondance biunivoque entre estimateurs et combinaisons estimables est clairement mise en évidence:

- a) si  $f^T \mathfrak{b}_H$  est estimable, il existe nécessairement un vecteur  $m_f^T \in \mathbf{B}^*$  tel que  $f^T = m_f^T \mathcal{A}^T \mathcal{A}$ ;
- b) l'estimateur privilégié de  $f^T \mathfrak{b}_H$  est alors  $m_f^T \mathcal{A}^T \mathfrak{z}$  (de sorte que  $l_f^* = m_f^T \mathcal{A}^T$ );
- c) la variance de cet estimateur vaut  $[m_f^T (\mathcal{A}^T \mathcal{A}) m_f] \sigma^2$ .

1, 33. Supposons que  $r = p$ .  $\mathcal{A}$ ,  $\mathcal{A}^T$ , et  $\mathcal{A}^T \mathcal{A}$  sont alors des matrices de rang  $p$ , et le système de dimension  $p$  en l'inconnue  $\hat{\mathfrak{b}}_H$

$$\mathcal{A}^T \mathcal{A} \hat{\mathfrak{b}}_H = \mathcal{A}^T \mathfrak{z} \quad (6)$$

détermine entièrement cette inconnue. Celle-ci jouit de la précieuse propriété que voici:

*l'estimateur privilégié de  $f^T \mathfrak{b}_H$  n'est autre que  $f^T \hat{\mathfrak{b}}_H$ .*

En effet, soit  $f^T = m_f^T \mathcal{A}^T \mathcal{A}$ ; on a

$$l_f^* \mathfrak{z} = m_f^T \mathcal{A}^T \mathfrak{z} = (m_f^T \mathcal{A}^T \mathcal{A}) \hat{\mathfrak{b}}_H = f^T \hat{\mathfrak{b}}_H, \quad \text{q.e.d.}$$

1, 34. Si  $r < p$ , le système (6) ne détermine pas univoquement l'inconnue  $\hat{\mathfrak{b}}_H$ . Pourtant, il reste vrai que, quelle que soit la détermination choisie pour  $\hat{\mathfrak{b}}_H$ , l'estimateur privilégié de la combinaison estimable  $f^T \mathfrak{b}_H$  est  $f^T \hat{\mathfrak{b}}_H$ ; en d'autres termes, les premiers membres de (6) mettent en évidence  $r$  combinaisons estimables particulières qui constituent une base de  $\mathbf{B}_+$ .

Le système (6) est dit « système normal »; il faut évidemment se garder d'y vouloir introduire l'inverse de  $\mathcal{A}^T \mathcal{A}$  lorsque  $r < p$  (cfr. § 3).

1, 4. *Moindres carrés.*

A partir de la relation

$$\mathbf{E} \mathbf{x} = \mathfrak{A} \mathfrak{b}_H ,$$

le théorème de Gauss-Markov conduit à introduire le vecteur-estimateur  $\mathfrak{b}^*$  défini en fonction de l'observation  $\mathbf{x}$  par la condition que

$$S^2(\mathfrak{b}) \equiv (\mathfrak{x} - \mathfrak{A} \mathfrak{b})^T (\mathfrak{x} - \mathfrak{A} \mathfrak{b})$$

soit minimum pour  $\mathfrak{b} = \mathfrak{b}^*$ . Or, comme

$$\frac{d}{d b_{i,H}} \mathfrak{b}_H = \mathbf{e}_{i,H} , \quad \frac{d}{d b_{i,H}} \mathfrak{b}_H^T = \mathbf{e}_{i,H}^T ,$$

et

$$S^2(\mathfrak{b}) = \mathfrak{x}^* \mathfrak{x} - \mathfrak{b}_H^T \mathfrak{A}^T \mathfrak{x} - \mathfrak{x}^* \mathfrak{A} \mathfrak{b}_H + \mathfrak{b}_H^T \mathfrak{A}^T \mathfrak{A} \mathfrak{b}_H ,$$

on a

$$\frac{d S^2(\mathfrak{b})}{d b_{i,H}} = 2 (\mathfrak{A}^T \mathfrak{A} \mathfrak{b}_H - \mathfrak{A}^T \mathfrak{x}) .$$

Les conditions  $\frac{d S^2(\mathfrak{b}^*)}{d b_{i,H}} = 0$  conduisent donc au système

$$\mathfrak{A}^T \mathfrak{A} \mathfrak{b}_H^* = \mathfrak{A}^T \mathbf{x} ,$$

identique au système normal. Il en résulte que, si  $r = p$ , les estimateurs de moindres carrés ne sont autres que les estimateurs privilégiés. Si  $r < p$ , les deux méthodes conduisent aux mêmes combinaisons estimables fondamentales.

## 2. DISTRIBUTIONS ET ÉPREUVES D'HYPOTHÈSES.

2, 1. *Sommes des carrés.*

2, 11. Soit  $\mathbf{U}^*$  un sous-espace vectoriel de  $\mathbf{V}^*$ ; on appelle « somme de carrés due à  $\mathbf{U}^*$  », et on note  $\mathbf{SCU}^*$  <sup>11)</sup>, le carré scalaire de la projection orthogonale de  $\mathbf{x}$  sur le dual  $\mathbf{U}$  de  $\mathbf{U}^*$ . La dimension de  $\mathbf{U}^*$  est, par définition, le « nombre de degrés de liberté » de  $\mathbf{SCU}^*$ .