1. Modèles linéaires. Estimateurs.

Objekttyp: **Chapter**

Zeitschrift: **L'Enseignement Mathématique**

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

Heft 1: **L'ENSEIGNEMENT MATHÉMATIQUE**

PDF erstellt am: **20.09.2024**

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 χ^2 à p degrés de liberté (en abrégé, χ_p^2);

b) si $\mathbf{x}$ est une aléatoire gaussienne et $\mathbf{u}$ une aléatoire χ_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, $\mathbf{t}_p$);

c) si $\mathbf{u}$ et $\mathbf{v}$ sont des aléatoires χ^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 β (en général, en effet, $n \geqslant m$, donc $p \geqslant q$, comme dans les tables de Pearson [IV]); on notera que les grandes valeurs de $\mathbf{F}$ correspondent aux petites valeurs de β.

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 $\alpha = (a_1, \ldots, a_n)$ et $\beta = (b_1, \ldots, b_n)$ sont deux observations, on a

pour la somme:

$$\alpha + \beta = (a_1 + b_1, \ldots, a_n + b_n) \ , \tag{1}$$

pour le produit par un scalaire:

$$p\,\alpha = (p\,a_1, \ldots, p\,a_n) \ , \tag{2}$$

pour le produit scalaire:

$$< \alpha, \beta > = a_1 b_1 + \ldots + a_n b_n \ . \tag{3}$$

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

Les n observations (1, 0, ..., 0), ..., (0, ..., 0, 1) forment évidemment une base de V; elle sera désignée par $\mathfrak{P}$, et sa duale par $\mathfrak{P}^\star$. La base $\mathfrak{P}$ 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 $\mathfrak{P}$ que les relations (1), (2), (3) explicitent la définition des opérations fondamentales de V. Sauf mention expresse du contraire, les représentations de vecteurs de V ou $\mathsf{V}^\star$ seront toujours censées être faites par rapport à $\mathfrak{P}$ ou $\mathfrak{P}^\star$; comme la relation (3) implique que $\mathfrak{P}$ est orthonormale, les conventions du § 0, 2 seront appliquées.

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

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

$$\mathbf{E}\, \mathfrak{e}_i^\star\, \mathfrak{x} = \mathfrak{e}_i^\star \int_{\mathsf{C}} \mathfrak{x}(\lambda)\, \mathrm{Pr}\,(d\lambda) = \mathfrak{e}_i^\star\, \mathfrak{A}\, \mathfrak{b} \quad (\mathfrak{x} \in \mathsf{V},\ \mathfrak{b} \in \mathsf{B}) \ ,$$

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

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

$$\mathbf{E}\,(\mathfrak{l}^\star\,\mathfrak{x}) = \mathfrak{l}^\star\,\mathbf{E}\,\mathfrak{x} = \mathfrak{l}^\star\,\mathfrak{A}\,\mathfrak{b}\ , \tag{4}$$

d'où

$$\mathfrak{E}:\ \mathfrak{l}^\star \to \mathfrak{l}^\star\,\mathfrak{A}\,\mathfrak{b}\ .$$ [8]

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

1, 14. Le noyau de $\mathfrak{E}$ (sous-espace V_0 de $\mathsf{V}^\star$ formé des vecteurs $\mathfrak{l}^\star$ tels que $\mathfrak{E}\,\mathfrak{l}^\star$ soit identiquement nul [9]) est appelé «espace des erreurs » [10]); il est de dimension $n - \operatorname{rg}\mathfrak{E} = n - r$.

Le complément orthogonal de V_0 (sous-espace V_+ de $\mathsf{V}^\star$ formé des vecteurs $\mathfrak{m}^\star$ pour lesquels $\mathfrak{m}^\star\,\mathfrak{l} = 0$ pour tout $\mathfrak{l}^\star \in \mathsf{V}_0$) 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 [9]) nulle. On remarquera que, $\mathsf{V}^\star$ étant somme directe de V_0 et V_+, à toute fonctionnelle linéaire des observations dont la moyenne n'est pas identiquement [9]) nulle correspond une et une seule fonctionnelle de V_+ ayant identiquement [9]) même moyenne qu'elle.

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

On sait que la restriction de $\mathfrak{E}$ 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 $\mathfrak{l}^\star \in \mathsf{V}^\star$, $\mathbf{E}\,\mathfrak{l}^\star\,\mathfrak{x} = \mathfrak{f}^\star\,\mathfrak{b}$, $\mathfrak{f}^\star\,\mathfrak{b}$ est évidemment une combinaison estimable, et $\mathfrak{l}^\star\,\mathfrak{x}$ en est un estimateur fidèle (au sens de la théorie statistique de l'estimation); si $\mathfrak{m}^\star \in \mathsf{V}_0$, $(\mathfrak{l}^\star + \mathfrak{m}^\star)\,\mathfrak{x}$ est aussi un estimateur fidèle de $\mathfrak{f}^\star\,\mathfrak{b}$; pour distinguer, parmi tous ces estimateurs fidèles de $\mathfrak{f}^\star\,\mathfrak{b}$, l'unique estimatrice, celle-ci est dite « estimateur privilégié de $\mathfrak{f}^\star\,\mathfrak{b}$ », et désignée par $\mathfrak{l}_f^\star\,\mathfrak{x}$ (donc, par définition, $\mathbf{E}\,\mathfrak{l}_f^\star\,\mathfrak{x} = \mathfrak{f}^\star\,\mathfrak{b}$ et $\mathfrak{l}_f^\star \in \mathsf{V}_+$).

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

Soit en effet $\mathfrak{m}^\star$ tel que $\mathbf{E}\,\mathfrak{m}^\star\,\mathfrak{x} = \mathfrak{f}^\star\,\mathfrak{b}$. On a, en posant $\tilde{\mathfrak{x}} = \mathfrak{x} - \mathbf{E}\,\mathfrak{x}$,

$$\begin{aligned} \operatorname{var} \mathfrak{m}^\star\,\mathfrak{x} &= \mathbf{E}\,[\mathfrak{m}^\star\,\mathfrak{x} - \mathbf{E}\,\mathfrak{m}^\star\,\mathfrak{x}]\,[\mathfrak{m}^\star\,\mathfrak{x} - \mathbf{E}\,\mathfrak{m}^\star\,\mathfrak{x}] \\ &= \mathbf{E}\,(\mathfrak{m}^\star\,\tilde{\mathfrak{x}})\,(\tilde{\mathfrak{x}}^\star\,\mathfrak{m})^\star = \mathfrak{m}^\star\,(\mathbf{E}\,\tilde{\mathfrak{x}}\,\tilde{\mathfrak{x}}^\star)\,\mathfrak{m}\,. \qquad (5) \end{aligned}$$

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

$$\mathbf{E}\,\tilde{\mathfrak{x}}\,\tilde{\mathfrak{x}}^\star = \mathbf{E}\,\tilde{\mathfrak{x}}_P\,(\tilde{\mathfrak{x}}_P)^{\mathrm{T}} = \mathbf{E} \left\| \begin{matrix} \tilde{\mathbf{x}}_1^2 & \tilde{\mathbf{x}}_1\,\tilde{\mathbf{x}}_2 \ldots & \tilde{\mathbf{x}}_1\,\tilde{\mathbf{x}}_n \\ \vdots & & \\ \tilde{\mathbf{x}}_n\,\tilde{\mathbf{x}}_1 \ldots & & \tilde{\mathbf{x}}_n^2 \end{matrix} \right\|$$

$$= \mathfrak{I}_n\,\sigma^2$$

d'où

$$\operatorname{var} \mathfrak{m}^\star\,\mathfrak{x} = (\mathfrak{m}^\star\,\mathfrak{m})\,\sigma^2\,.$$

Soit alors

$$\mathfrak{m}^\star = \mathfrak{l}_f^\star + \mathfrak{m}_0^\star\,,$$

de sorte que

$$\mathfrak{l}_f^\star\,\mathfrak{m}_0 = \mathfrak{m}_0^\star\,\mathfrak{l}_f = 0\,;$$

on a

$$\operatorname{var} \mathfrak{m}^\star\,\mathfrak{x} = (\mathfrak{l}_f^\star\,\mathfrak{l}_f)\,\sigma^2 + (\mathfrak{m}_0^\star\,\mathfrak{m}_0)\,\sigma^2 \geqslant (\mathfrak{l}_f^\star\,\mathfrak{l}_f)\,\sigma^2 = \operatorname{var} \mathfrak{l}_f^\star\,\mathfrak{x},$$

l'inégalité étant d'ailleurs stricte si $\mathfrak{m}_0^\star \neq 0$.

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 $\mathsf{V}^\star$ sont rapportés aux bases $\mathfrak{P}$ et $\mathfrak{P}^\star$, B et $\mathsf{B}^\star$ sont rapportés à des bases déterminées $\mathfrak{H}$ et $\mathfrak{H}^\star$.

1, 32. Pour que $\mathfrak{l}^\star \in \mathsf{V}_0$, il est nécessaire et suffisant que $\mathfrak{l}^\star\,\mathfrak{A} = 0$; donc

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

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

toute estimatrice est de la forme $\mathfrak{l}^\star\, \mathfrak{A}^T\, \mathfrak{x}$.

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

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

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

a) si $\mathfrak{f}^T\, \mathfrak{b}_H$ est estimable, il existe nécessairement un vecteur $\mathfrak{m}_f^T \in \mathbf{B}^\star$ tel que $\mathfrak{f}^T = \mathfrak{m}_f^T\, \mathfrak{A}^T\, \mathfrak{A}$;

b) l'estimateur privilégié de $\mathfrak{f}^T\, \mathfrak{b}_H$ est alors $\mathfrak{m}_f^T\, \mathfrak{A}^T\, \mathfrak{x}$ (de sorte que $\mathfrak{l}_f^\star = \mathfrak{m}_f^T\, \mathfrak{A}^T$);

c) la variance de cet estimateur vaut $[\mathfrak{m}^T\, (\mathfrak{A}^T\, \mathfrak{A})\, \mathfrak{m}]\, \sigma^2$.

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

$$\mathfrak{A}^T\, \mathfrak{A}\, \hat{\mathfrak{b}}_H = \mathfrak{A}^T\, \mathfrak{x} \tag{6}$$

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

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

En effet, soit $\mathfrak{f}^T = \mathfrak{m}_f^T\, \mathfrak{A}^T\, \mathfrak{A}$; on a

$$\mathfrak{l}_f^\star\, \mathfrak{x} = \mathfrak{m}_f^T\, \mathfrak{A}^T\, \mathfrak{x} = (\mathfrak{m}_f^T\, \mathfrak{A}^T\, \mathfrak{A})\, \hat{\mathfrak{b}}_H = \mathfrak{f}^T\, \hat{\mathfrak{b}}_H\,, \quad \text{q.e.d.}$$

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 $\mathfrak{f}^T\, \mathfrak{b}_H$ est $\mathfrak{f}^T\, \hat{\mathfrak{b}}_H$; 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 $\mathfrak{A}^T\, \mathfrak{A}$ lorsque $r < p$ (cfr. § 3).

1, 4. *Moindres carrés.*

A partir de la relation

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

le théorème de Gauss-Markov conduit à introduire le vecteur-estimateur $\overset{*}{\boldsymbol{\mathfrak{b}}}$ défini en fonction de l'observation $\boldsymbol{\mathfrak{x}}$ 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 $\mathfrak{b} = \overset{*}{\mathfrak{b}}$. Or, comme

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

et

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

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

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

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 $\mathsf{U}^\star$ un sous-espace vectoriel de $\mathsf{V}^\star$; on appelle « somme de carrés due à $\mathsf{U}^\star$ », et on note $\mathbf{SC}\,\mathsf{U}^\star$ [11]), le carré scalaire de la projection orthogonale de $\boldsymbol{\mathfrak{x}}$ sur le dual U de $\mathsf{U}^\star$. La dimension de $\mathsf{U}^\star$ est, par définition, le « nombre de degrés de liberté » de $\mathbf{SC}\,\mathsf{U}^\star$.