Section 3 :
Algèbre des matrices de projection et notation matricielle du modèle

STAT-F-406

Master en sciences mathématiques, Master en statistique

ACTU-F4001

Master en sciences actuarielles

Davy Paindaveine

Université libre de Bruxelles

2024–2025

Contenu du chapitre

Matrices de projection

Soient $v_1, \ldots, v_k \in \mathbb{R}^n$ linéairements indépendants, où $k \leq n$ (de sorte que $V \stackrel{\text{def}}{=} (v_1 \ \ldots \ v_k)$ soit une matrice $n \times k$ de rang maximal).

Théorème 1

La projection orthogonale sur l'espace vectoriel $\mathcal{M}(V)$ engendré par $v_1, \ldots, v_k \in \mathbb{R}^n$ a pour matrice $P_V = V(V'V)^{-1}V'$.

Preuve: (i) Soit $x \in \mathcal{M}(V)$, de sorte que $x = V\alpha$ pour un certain $\alpha \in \mathbb{R}^k$. Alors $V(V'V)^{-1}V'x = V(V'V)^{-1}V'V\alpha = V\alpha = x$.

(ii) Soit $x \in (\mathcal{M}(V))^\perp$, c'est-à-dire $x'y = 0$ pour tout $y \in \mathcal{M}(V)$. Alors $v_j'x = 0$ pour tout $j = 1, \ldots, k$, c'est-à-dire $V'x = 0$. Donc $V(V'V)^{-1}V'x = 0$. $\square$

Matrices de projection

Notons que $P_V = V(V'V)^{-1}V'$ est symétrique et idempotente.

Donc $P_V = O\Lambda O'$, où O est orthogonale et $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, et

$$O\Lambda O' = P_V = P_V^2 = (O\Lambda O')(O\Lambda O') = O\Lambda^2 O',$$

ce qui implique que $\lambda_j = 1$ ou 0 pour tout j.

Le nombre de λ_j égaux à 1 vaut alors

$$\begin{aligned} \sum_{j=1}^n \lambda_j &= \mathrm{tr}[\Lambda] = \mathrm{tr}[O'O\Lambda] = \mathrm{tr}[O\Lambda O'] = \mathrm{tr}[P_V] \\ &= \mathrm{tr}[V(V'V)^{-1}V'] = \mathrm{tr}[(V'V)^{-1}V'V] = \mathrm{tr}[I_k] = k. \end{aligned}$$

Notons qu'en permutant les colonnes de O, on peut toujours faire en sorte que $\Lambda = \mathrm{diag}(1, \ldots, 1, 0, \ldots, 0)$.

Notation matricielle du modèle linéaire

L'équation du modèle linéaire

$$Y_i = X_i'\beta + \varepsilon_i, \quad i = 1, \ldots, n,$$

se réécrit sous forme matricielle

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\varepsilon},$$

où

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} X_1' \\ \vdots \\ X_n' \end{pmatrix} \quad \text{et} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Notons que

$$\hat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i Y_i\right) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

Valeurs prédites

Donc les "prédictions"

$$\hat{Y}_i = X_i'\hat{\beta}, \quad i = 1, \ldots, n,$$

s'écrivent

$$\hat{\mathbf{Y}} = \begin{pmatrix} \hat{Y}_1 \\ \vdots \\ \hat{Y}_n \end{pmatrix} = \begin{pmatrix} X_1'\hat{\beta} \\ \vdots \\ X_n'\hat{\beta} \end{pmatrix} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \stackrel{\text{def}}{=} \mathbf{QY},$$

où $\mathbf{Q}$ est la matrice de la projection orthogonale sur l'espace vectoriel engendré par les k colonnes de $\mathbf{X}$.

Résidus

Donc les "erreurs de prédiction" (ou *résidus*)

$$e_i \stackrel{\text{def}}{=} Y_i - \hat{Y}_i = Y_i - X_i'\hat{\beta}, \quad i = 1, \ldots, n,$$

(voir le slide suivant) s'écrivent donc

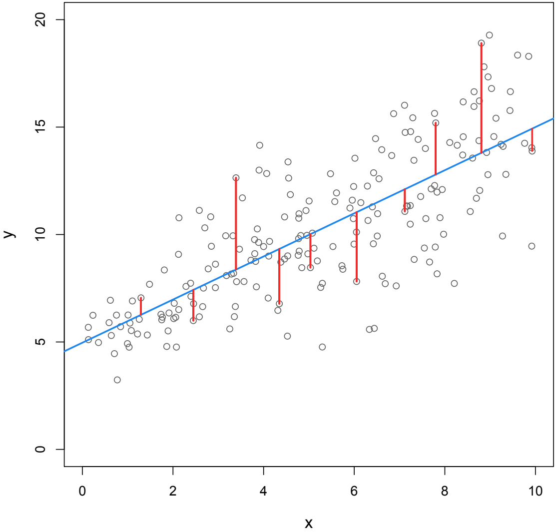
$$\mathbf{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} Y_1 - \hat{Y}_1 \\ \vdots \\ Y_n - \hat{Y}_n \end{pmatrix} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{Q})\mathbf{Y} \stackrel{\text{def}}{=} \mathbf{P}\mathbf{Y},$$

où $\mathbf{P}$ est la matrice de la projection orthogonale sur le complément orthogonal à celui engendré par les colonnes de $\mathbf{X}$.

Si il y a un "intercept" dans le modèle ($X_1 = 1$ p.s.), alors on a toujours

$$\bar{e} = \frac{1}{n}\sum_{i=1}^{n} e_i = \frac{1}{n}\mathbf{e}'\mathbf{1} = \frac{1}{n}(\mathbf{P}\mathbf{Y})'\mathbf{1} = \frac{1}{n}\mathbf{Y}'\mathbf{P}\mathbf{1} = 0.$$

Résidus



Coefficients R^2

Pour $\mathbf{v} = (v_1, \ldots, v_n)'$, notons $\bar{\mathbf{v}} = (\bar{v}, \ldots, \bar{v})' \in \mathbb{R}^n$, où $\bar{v} = \frac{1}{n}\sum_{i=1}^n v_i$.
Si il y a un intercept dans le modèle, alors

$$\|\mathbf{Y} - \bar{\mathbf{Y}}\|^2 = \|\mathbf{Q}(\mathbf{Y} - \bar{\mathbf{Y}}) + \mathbf{P}(\mathbf{Y} - \bar{\mathbf{Y}})\|^2 = \|\mathbf{Q}(\mathbf{Y} - \bar{\mathbf{Y}})\|^2 + \|\mathbf{P}(\mathbf{Y} - \bar{\mathbf{Y}})\|^2$$

$$= \|\mathbf{QY} - \overline{\mathbf{QY}}\|^2 + \|\mathbf{PY}\|^2 = \left\|\hat{\mathbf{Y}} - \bar{\hat{\mathbf{Y}}}\right\|^2 + \|\mathbf{e}\|^2$$

(puisque $\bar{\mathbf{v}} = \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{v}$), de sorte qu'une mesure de fit est le coefficient

$$R^2 = 1 - \frac{\|\mathbf{e}\|^2}{\|\mathbf{Y} - \bar{\mathbf{Y}}\|^2}$$

ou sa version ajustée

$$R_a^2 = 1 - \frac{\|\mathbf{e}\|^2/(n-k)}{\|\mathbf{Y} - \bar{\mathbf{Y}}\|^2/(n-1)},$$

qui est plus adaptée pour le choix de modèles (pourquoi?)

Coefficients R^2