# PC 2 - Modèle linéaire

04/09/2026

On va parler de mod. linéaire, ou de régression.

$Y \simeq f(x)$. On va se baser sur un ensemble d'obs. $(x_i, y_i)_{1 \le i \le n}$. On va mettre un modèle stat sur ce pb

$f(x)$ est 1 C. linéaire des variables dans $x$.

$x \in \mathbb{R}^p$, $y \in \mathbb{R}$

$$f_\beta(x) \simeq \beta_1 x^1 + \beta_2 x^2 + \dots + \beta_p x^p + \beta_0$$

Faire 1 reg linéaire, c'est trouver le "meilleur" $\beta$ .. celui pour lequel $y_i \simeq f_\beta(x_i)$, $\forall i$

En optim : cela se traduirait par :

$$\Phi_n(\beta) = \sum_{i=1}^{n} \left[ y_i - f_\beta(x_i) \right]^2$$

$$\hat{\beta}_n = \underset{\beta}{\operatorname{argmin}}\, \Phi_n(\beta)$$

---

On peut aussi mettre un modèle stat :

$$Y = \underbrace{\langle \beta, x \rangle}_{\beta_0 + \beta_1 x^{(1)} + \dots + \beta_p x^{(p)} \text{ si } x \to [1, x]} + \sigma \varepsilon$$

$\varepsilon \sim \mathcal{N}(0,1)$
$\sigma$ : variance
$\beta$ : effets
$x = (1, x^{(1)}, \dots, x^{(p)})$

Les param. du modèle sont $(\beta, \sigma^2)$

Pour estimer dans 1 modèle stat, on peut alors utiliser

- méthode des moments
- max de vraisemblance.

$$X = \begin{bmatrix} 1 & x_1^{(1)} & \cdots & x_1^{(p)} \\ \vdots & \vdots & & \\ 1 & x_n^{(1)} & \cdots & x_n^{(p)} \end{bmatrix}$$

X est la matrice de design.

Nous on va étudier le cas $p=2$. Mais tout peut se généraliser. On va essayer de rester le + général possible...

$$\mathbf{Y} := \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \in \mathbb{R}^n , \qquad \mathbf{X} := \begin{bmatrix} \mathbf{1} & \underline{\mathbf{x}} \end{bmatrix} \in \mathbb{R}^{n \times 2} , \qquad \beta := \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \in \mathbb{R}^2 .$$

1. Montrer que la matrice $\mathbf{X^T X}$ est inversible.
*Remarque : Puisque $u^T \mathbf{X^T X} u \geq 0$ pour tout $u \in \mathbb{R}^2$, on en déduit que la matrice est définie positive.*

$X^T X$ est 1 matrice de Gramm.

C'est 1 mat positive : soit $u \in \mathbb{R}^{p+1}$ (ici $p+1=2$, $p=1$)

$^T u\, X^T X\, u = \langle Xu, Xu \rangle = \|Xu\|^2 \geq 0$ Donc $X^T X \geq 0$

Etudions l'inversibilité : $X^T X$ est de taille $(p+1)\times(p+1)$, ici $2\times 2$.

$$\det(X^T X) = \begin{vmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{vmatrix} = n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2$$

Cauchy-Schwarz

$$\left(\sum x_i\right)^2 = \langle 1, x \rangle^2 \leq \|1\|^2 . \|x\|^2 = n \sum_{i=1}^n x_i^2$$

On sait qu'il y a égalité lorsque 1 et x sont colinéaires.

Comme $\exists\, i \neq j : x_i \neq x_j$, alors x et 1 ne sont pas colinéaires.

$\det(X^T X) > 0 \neq 0$ et $X^T X$ est inversible.

Remarque : on peut généraliser à $p$ quelconque.
Dans le cas où $n > p$, $X^T X$ est inversible ssi
$[X_1, \dots, X_n]$ est de rang $p$.

2. Déterminer les estimateurs du maximum de vraisemblance $(\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2)$ de $(\beta_1, \beta_2, \sigma^2)$.

Il faut écrire la vraisemblance... $L_n(\theta)$ $\theta = (\beta, \sigma^2)$

$\theta \longmapsto L_n(\theta)$ est la fonction de vraisemblance

$\theta \longmapsto \ell_n(\theta) = \log L_n(\theta)$ ——— log. ———.

$$L_n(\theta) = \underbrace{\prod_{i=1}^{n}}_{\text{indépendance des } n \text{ obs.}} \frac{e^{-(y_i - \langle \beta, x_i \rangle)^2 / 2\sigma^2}}{\sqrt{2\pi}\,\sigma}$$

$$= (2\pi)^{-n/2} \sigma^{-n} e^{-\sum_{i=1}^{n} \frac{(y_i - \langle \beta, x_i \rangle)^2}{2\sigma^2}}$$

$$\ell_n(\theta) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \underbrace{\sum_{i=1}^{n} \frac{(y_i - \langle \beta, x_i \rangle)^2}{2\sigma^2}}_{\frac{\|Y - X\beta\|^2}{2\sigma^2}}$$

$$\boxed{\ell_n(\theta) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{\|Y - X\beta\|^2}{2\sigma^2}}$$

(On pourrait faire de plus général...
$Y = X\beta + \varepsilon$ $\varepsilon \sim N(0, \Sigma^2)$
Mais cela induit des calculs un petit peu + techniques.)

Optimisation en $\sigma^2$ : $\sigma^2 \longmapsto \ell_n(\theta)$ s'écrit comme

$$\sigma^2 \longmapsto -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\|Y - X\beta\|^2 \quad \text{à maximiser.}$$

↓

le log "réparer" :

$$\frac{\partial(\dots)}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{\|Y - X\beta\|^2}{2\sigma^4}.$$

$$\boxed{\hat{\sigma}^2 = \frac{\|Y - X\beta\|^2}{n}}$$

$\hat{\sigma}^2$ : variance empirique des résidus.

Optimisation en $\beta$ : On doit max $\ell_n$ donc on doit minimiser $\|Y - X\beta\|^2 := J(\beta)$

$J$ est 1 forme quadratique en $\beta$, donc convexe (ici évident que ce n'est pas concave ...) a priori.

On va donc étudier les (le) points critiques.

$\nabla J(\beta)$ : on essaye d'écrire les vecteurs des $\frac{\partial J}{\partial \beta_j}$ svp !

$$J(\beta) = \|Y\|^2 - 2\langle Y, X\beta\rangle + \langle X\beta, X\beta\rangle$$
$$= \|Y\|^2 - 2\langle X^T Y, \beta\rangle + \langle X^T X\beta, \beta\rangle$$

$$\nabla J(\beta) = -2X^T Y + 2X^T X\beta.$$

Trouver le point critique, c'est résoudre : $\nabla J(\beta) = 0$

c.à.d : $X^T Y = X^T X\beta$.

$$\boxed{\hat{\beta} = (X^T X)^{-1} X^T Y}$$

Remarque : $D^2 J(\beta) = 2X^T X$. Matrice d'info de Fisher.

L'est. du m.v. est donné par :

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad , \quad \hat{\sigma}^2 = \frac{\| Y - X\hat{\beta} \|^2}{n}$$

Merci d'expliciter $\hat{\beta}$ dans le contexte $p = 1$ !

Pour nous : $(X^T X)^{-1} = \frac{1}{n \sum_i x_i^2 - (\sum_i x_i)^2} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix}$

$$X^T Y = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

$$(X^T X)^{-1} X^T Y = \frac{1}{n \sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} \sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i \\ -\sum x_i \sum y_i + n \sum x_i y_i \end{pmatrix}$$

Remarque : on se ramène souvent au cas où les $x_i$ sont centrés

$$\sum x_i = 0$$

$$\hat{\beta} = \frac{1}{n \sum_{i=1} x_i^2} \begin{pmatrix} \sum x_i^2 \sum y_i \\ n \sum x_i y_i \end{pmatrix}$$

$$= \begin{pmatrix} \sum y_i / n \\ \frac{\sum x_i y_i}{\sum x_i^2} \end{pmatrix}$$

3. Déterminer la loi de $(\hat{\beta}_1, \hat{\beta}_2)$ sous $p_\theta \cdot \mathrm{dLeb}^{\otimes n}$.

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$Y \sim N(X \beta^*, \sigma^{*2} I_n)$$

donc $Y \sim X\beta^* + \sigma^* N(0, I_n)$

$\hat{\beta}$ est 1 combi linéaire de $Y$ :

$$\hat{\beta} = (X^T X)^{-1} X^T \left( X\beta^* + \sigma^* \xi \right) \text{ où } \xi \sim N(0, I_d)$$

$$= \cancel{(X^T X)^{-1}} \cancel{X^T X} \beta^* + \sigma^* (X^T X)^{-1} X^T \xi$$

$$= \beta^* + \sigma^* (X^T X)^{-1} X^T \xi .$$

$\mathbb{E}\,\hat{\beta} = \beta^*$ ! Good !

$$\mathrm{Cov}(\hat{\beta}) = \mathrm{Cov}\left( \sigma^* \underbrace{(X^T X)^{-1} X^T}_{A} \xi \right)$$

$$= \sigma^{*2} \cancel{(X^T X)^{-1}} \cancel{X^T X} (X^T X)^{-1} = \sigma^{*2} (X^T X)^{-1}$$

car :

$$\mathrm{Cov}(A\xi) = \mathbb{E}\left[ A\xi (A\xi)^T \right]$$

$$= \mathbb{E}\, A \xi \xi^T A^T$$

$$= A \underbrace{\mathbb{E}[\xi \xi^T]}_{= I_n} A^T$$

$$= A A^T$$

Donc :

$$\boxed{\hat{\beta} \sim N\left( \beta^*, \sigma^* (X^T X)^{-1} \right)}$$

4. Montrer que, sous $p_\theta \cdot \mathrm{dLeb}^{\otimes n}$, $\hat{\beta}_1$ et $\hat{\beta}_2$ sont indépendants si et seulement si $n^{-1} \sum_{i=1}^n x_i = 0$.

$\hat{\beta}$ est Gaussien, donc $\hat{\beta}_1 \perp\!\!\!\perp \hat{\beta}_2$ ssi cov. nulle ssi $\sum x_i = 0$ car $(X^TX)^{-1} = \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix}$

5. Déterminer la loi de $\hat{\sigma}^2$ sous $p_\theta \cdot \mathrm{dLeb}^{\otimes n}$.

$$\hat{\sigma}^2 = \frac{\|Y - X\hat{\beta}\|^2}{n} \qquad \hat{\beta} = (X^TX)^{-1}X^TY$$

$$Y = X\beta^* + \sigma^*\xi.$$

$$Y - X\hat{\beta} = \left(X\beta^* + \sigma^*\xi - X(X^TX)^{-1}X^T(X\beta^* + \sigma^*\xi)\right)$$

$$= X\beta^* + \sigma^*\xi - X(X^TX)^{-1}X^TX\beta^* - \sigma^* X(X^TX)^{-1}X^T\xi$$

$$= \sigma^*\left(\mathrm{Id} - X(X^TX)^{-1}X^T\right)\xi.$$

$\mathrm{Id} - P$ avec $P = X(X^TX)^{-1}X^T$

$P^2 = P$

$P^T = P$

$P$ est 1 matrice de proj $\perp$.

$Y - X\hat{\beta}$ est un vecteur Gaussien $= \sigma^* \cdot N(0, \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 0 \end{pmatrix})$ avec $\begin{pmatrix} 1 & & \\ & \ddots & \\ & & 0 \end{pmatrix}$ : réduction de la matrice $\mathrm{Id} - P$.

$$\frac{1}{n}\|Y - X\hat{\beta}\|^2 = \frac{\sigma^{*2}}{n} \cdot \chi^2\left(\mathrm{rank}(I - P)\right)$$

$$= \frac{\sigma^{*2}}{n}\chi^2(n - p)$$

À savoir refaire en mode pilote automatique !

**Exercice 2 (Estimation d'un coefficient de mélange ($\star$)).** Soit un $n$-échantillon $(X_1, \dots, X_n)$ du modèle $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \{f_\theta \cdot \mathrm{Leb}, \theta \in ]0,1[\})$ où

$$f_\theta(x) = \frac{\theta}{a}\, \mathbb{1}_{\{[0,a]\}}(x) + \frac{(1-\theta)}{b}\, \mathbb{1}_{\{[0,b]\}}(x).$$

On suppose que $a$ et $b$ sont connus et que $0 < a < b$. La fonction de répartition associée à la densité $f_\theta$ vaut

$$x \mapsto \begin{cases} 0 & x \leq 0 \\ \left(\frac{\theta}{a} + \frac{(1-\theta)}{b}\right) x & 0 \leq x \leq a \\ \theta + \frac{(1-\theta)}{b} x & a \leq x \leq b \\ 1 & x \geq 1 \end{cases}$$

1. Exprimer la fonction de vraisemblance à l'aide de $N_a := \sum_{i=1}^n \mathbb{1}_{X_i \in [0,a]}$ le nombre d'observations à valeur dans $[0, a]$.

Les $X_i$ sont tous $\perp\!\!\!\perp$.

$$L_n(\theta) = \prod_{i=1}^n f_\theta(X_i)$$

$$= \prod_{i=1}^n \left( \frac{\theta}{a} \mathbb{1}_{0 \leq X_i \leq a} + \frac{1-\theta}{b} \mathbb{1}_{0 \leq X_i \leq b} \right)$$

$$= \prod_{i=1}^n \left[ \mathbb{1}_{0 \leq X_i \leq a} \left( \frac{\theta}{a} + \frac{1-\theta}{b} \right) + \mathbb{1}_{X_i > a} \left( \frac{1-\theta}{b} \right) \right]$$

$$= \left( \frac{\theta}{a} + \frac{1-\theta}{b} \right)^{N_a} \left( \frac{1-\theta}{b} \right)^{n - N_a}$$

2. En déduire l'estimateur du maximum de vraisemblance de $\theta$, noté $\hat{\theta}_n^{(1)}$.

On définit $\hat{\theta}_n = \underset{\theta}{\operatorname{argmax}}\, L_n(\theta)$ :

$$l_n(\theta) = N_a \log\left( \frac{\theta}{a} + \frac{1-\theta}{b} \right) + (n - N_a) \log\left( \frac{1-\theta}{b} \right)$$

$$\frac{\partial}{\partial \theta} l_n(\theta) = N_a \frac{\frac{1}{a} - \frac{1}{b}}{\frac{\theta}{a} + \frac{1-\theta}{b}} + \frac{(n - N_a)}{1-\theta}$$

$$N_a \left(\tfrac{1}{a} - \tfrac{1}{b}\right)(1-\theta) - (n - N_a)\left(\frac{\theta}{a} + \frac{1-\theta}{b}\right) = 0$$

$\partial_\theta \ell_n(\theta) = 0 \iff \hat{\theta}_n = \dots$