

Section 2 : Estimateurs des moindres carrés

STAT-F-406

Master en sciences mathématiques, Master en statistique

ACTU-F4001

Master en sciences actuarielles

Davy Paindaveine

Université libre de Bruxelles

2024–2025

Estimateur de β par analogie

Soient $(Y_1, X_1), \dots, (Y_n, X_n)$ des observations i.i.d. engendrées depuis le modèle linéaire faible

$$Y = \beta' X + \varepsilon,$$

où $E[XX'] > 0$ (nous supposons toujours $E[XX'] > 0$ dans la suite).

Alors l'estimateur naturel de $\beta = (E[XX'])^{-1}E[XY]$ est

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

Par la LFGN, $\hat{\beta} \xrightarrow{\text{p.s.}} \beta$ si $n \rightarrow \infty$.

Attention aux notations: X_j, X_i, X_{ij} .

Cas particulier du modèle de position

Le modèle de position est obtenu avec $k = 1$ et $X = 1$ p.s.:

$$Y = \beta + \varepsilon.$$

Dans ce modèle, l'estimateur de

$$\beta = (E[XX'])^{-1}E[XY] = E[Y]$$

ci-dessus est

$$\hat{\beta} = \left(\frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i Y_i \right) = \frac{1}{n} \sum_{i=1}^n Y_i \stackrel{\text{not}}{=} \bar{Y}.$$

Estimateur des moindres carrés ordinaires

Pour chaque valeur de β , on peut considérer les résidus

$$e_i(\beta) = Y_i - \beta' X_i, \quad i = 1, \dots, n.$$

Intuitivement, β est une bonne valeur du paramètre si les résidus sont petits (voir le slide suivant). Il est classique de minimiser

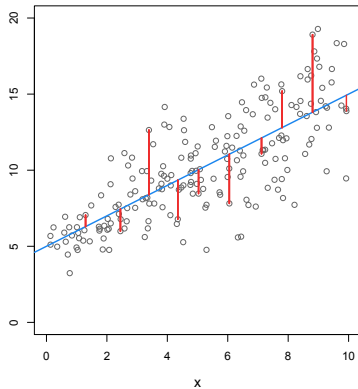
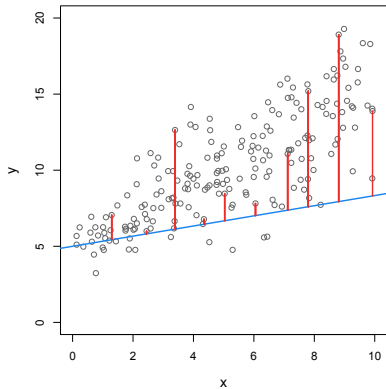
$$\sum_{i=1}^n |e_i(\beta)| \quad \text{ou} \quad \sum_{i=1}^n (e_i(\beta))^2.$$

En fait,

$$\hat{\beta}_{\text{MCO}} \stackrel{\text{def}}{=} \arg \min_{\beta \in \mathbb{R}^k} \sum_{i=1}^n (e_i(\beta))^2 = \hat{\beta},$$

d'où la terminologie **estimateur des moindres carrés (ordinaires)** pour $\hat{\beta}$.

Estimateur des moindres carrés ordinaires



Estimateur des moindres carrés ordinaires

Puisque

$$\begin{aligned}\nabla_{\beta} \sum_{i=1}^n (e_i(\beta))^2 &= \nabla_{\beta} \sum_{i=1}^n (Y_i - \beta' X_i)^2 \\ &= \sum_{i=1}^n 2(Y_i - \beta' X_i)(-X_i) \\ &= -2\left\{ \left(\sum_{i=1}^n X_i Y_i \right) - \left(\sum_{i=1}^n X_i \beta' X_i \right) \right\} \\ &= -2\left\{ \left(\sum_{i=1}^n X_i Y_i \right) - \left(\sum_{i=1}^n X_i X_i' \right) \beta \right\} \\ &= -2n \left\{ \left(\frac{1}{n} \sum_{i=1}^n X_i Y_i \right) - \left(\frac{1}{n} \sum_{i=1}^n X_i X_i' \right) \beta \right\},\end{aligned}$$

la seule valeur de β qui annule le gradient de la fonction objectif

$$\beta \mapsto \sum_{i=1}^n (e_i(\beta))^2$$

est $\hat{\beta}$. En étudiant la matrice hessienne, on voit que cette fonction objectif est convexe sur \mathbb{R}^k , ce qui implique que $\hat{\beta}$ est un minimum global.

Estimateur des moindres carrés ordinaires

L'estimateur

$$\hat{\beta}_{\text{LAD}} \stackrel{\text{def}}{=} \arg \min_{\beta \in \mathbb{R}^k} \sum_{i=1}^n |e_i(\beta)|$$

n'admet pas d'expression explicite, mais il a aussi de bonnes propriétés, notamment en termes de **robustesse**.

Non-biais

Nous allons étudier les propriétés de l'estimateur $\hat{\beta}$.

Théorème 1

Dans un modèle semi-fort, $\hat{\beta}$ est *sans biais* pour β .

Preuve: En notant $E[\cdot|\mathcal{X}]$ l'espérance conditionnelle sachant X_1, \dots, X_n et $Q = \frac{1}{n} \sum_{i=1}^n X_i X_i'$, on a

$$\begin{aligned} E[\hat{\beta}|\mathcal{X}] &= E[Q^{-1}(\frac{1}{n} \sum_{i=1}^n X_i Y_i)|\mathcal{X}] = Q^{-1} \frac{1}{n} \sum_{i=1}^n E[X_i Y_i|\mathcal{X}] \\ &= Q^{-1} \frac{1}{n} \sum_{i=1}^n X_i E[Y_i|\mathcal{X}] = Q^{-1} \frac{1}{n} \sum_{i=1}^n X_i E[X_i' \beta + \varepsilon_i|\mathcal{X}] \\ &= Q^{-1} \frac{1}{n} \sum_{i=1}^n X_i (E[X_i' \beta|\mathcal{X}] + E[\varepsilon_i|\mathcal{X}]) = Q^{-1} \frac{1}{n} \sum_{i=1}^n X_i (X_i' \beta + 0) = \beta \end{aligned}$$

pour tout β . Donc $E[\hat{\beta}] = \beta$ pour tout β . □

Variance

Si $\hat{\beta}$ vise donc bien en moyenne, il pourrait avoir une grande variance.

Théorème 2

Dans un modèle semi-fort, $\text{Var}[\hat{\beta}|\mathcal{X}] = \frac{1}{n^2}Q^{-1} \sum_{i=1}^n \sigma^2(X_i)X_iX_i'Q^{-1}$.

Preuve: En utilisant des notations similaires à la preuve précédente,

$$\begin{aligned}\text{Var}[\hat{\beta}|\mathcal{X}] &= \text{Var}[Q^{-1}(\frac{1}{n} \sum_{i=1}^n X_iY_i)|\mathcal{X}] = \frac{1}{n^2}Q^{-1}\text{Var}[\sum_{i=1}^n X_iY_i|\mathcal{X}]Q^{-1} \\ &= \frac{1}{n^2}Q^{-1} \sum_{i=1}^n \text{Var}[X_iY_i|\mathcal{X}]Q^{-1} = \frac{1}{n^2}Q^{-1} \sum_{i=1}^n X_i \text{Var}[Y_i|\mathcal{X}]X_i'Q^{-1} \\ &= \frac{1}{n^2}Q^{-1} \sum_{i=1}^n X_i \text{Var}[X_i'\beta + \varepsilon_i|\mathcal{X}]X_i'Q^{-1} = \frac{1}{n^2}Q^{-1} \sum_{i=1}^n \sigma^2(X_i)X_iX_i'Q^{-1},\end{aligned}$$

puisque, par définition, $\sigma^2(X_i) = \text{Var}[\varepsilon_i|X_i]$. □

Variance

Théorème 2

Dans un modèle semi-fort, $\text{Var}[\hat{\beta}|\mathcal{X}] = \frac{1}{n^2}Q^{-1} \sum_{i=1}^n \sigma^2(X_i)X_iX_i'Q^{-1}$.

Quelques remarques:

- ▶ Dans un modèle semi-fort homoscédastique,

$$\begin{aligned}\text{Var}[\hat{\beta}|\mathcal{X}] &= \frac{1}{n^2}Q^{-1} \sum_{i=1}^n \sigma^2 X_iX_i'Q^{-1} \\ &= \frac{\sigma^2}{n}Q^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_iX_i'\right)Q^{-1} \\ &= \frac{\sigma^2}{n}Q^{-1}.\end{aligned}$$

- ▶ On a $\text{Var}[\hat{\beta}] = \text{E}[\text{Var}[\hat{\beta}|\mathcal{X}]] + \text{Var}[\text{E}[\hat{\beta}|\mathcal{X}]] = \text{E}[\text{Var}[\hat{\beta}|\mathcal{X}]]$.
- ▶ Si $n \rightarrow \infty$, alors $n \text{Var}[\hat{\beta}|\mathcal{X}] \rightarrow (\text{E}[XX'])^{-1}\text{E}[\sigma^2(X)XX'](\text{E}[XX'])^{-1}$, de sorte que la variance conditionnelle tend vers zéro.

Loi exacte

La variance qui tend vers zéro confirme la convergence de $\hat{\beta}$.
Mais pour faire de l'inférence fondée sur $\hat{\beta}$, on a besoin de sa loi.

Théorème 3

Dans le modèle fort avec normalité, $\hat{\beta}|\mathcal{X} \sim \mathcal{N}_k(\beta, \frac{\sigma^2}{n} Q^{-1})$.

Preuve: Dans le modèle fort avec normalité, $\varepsilon_i|X_i \sim \mathcal{N}(0, \sigma^2)$. Puisque les observations sont i.i.d., on a donc $\varepsilon_i|\mathcal{X} \sim \mathcal{N}(0, \sigma^2)$, ce qui livre

$$(X_i'\beta + \varepsilon_i)|\mathcal{X} \sim \mathcal{N}(X_i'\beta, \sigma^2), \text{ et donc } X_i Y_i|\mathcal{X} \sim \mathcal{N}(X_i X_i'\beta, \sigma^2 X_i X_i').$$

Par additivité des normales indépendantes, on a donc

$$\sum_{i=1}^n X_i Y_i|\mathcal{X} \sim \mathcal{N}(\sum_{i=1}^n X_i X_i'\beta, \sigma^2 \sum_{i=1}^n X_i X_i'),$$

ce qui fournit le résultat (pourquoi?) □

Loi asymptotique

Si on n'est pas dans le modèle restrictif dans lequel on a pu déterminer la loi exacte de $\hat{\beta}$, on peut obtenir sa loi asymptotique.

Théorème 4

Même dans le modèle faible, mais pour autant que $E[\sigma^2(X)\|X\|^2] < \infty$, on a que $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\mathcal{D}} \mathcal{N}_k(0, (E[XX'])^{-1}E[\sigma^2(X)XX'](E[XX'])^{-1})$.

Preuve: Puisque

$$\hat{\beta} = Q^{-1} \frac{1}{n} \sum_{i=1}^n X_i(X_i'\beta + \varepsilon_i) = \beta + Q^{-1} \frac{1}{n} \sum_{i=1}^n X_i \varepsilon_i,$$

le lemme de Slutsky et le TCL fournissent

$$\sqrt{n}(\hat{\beta} - \beta) = Q^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i X_i \xrightarrow{\mathcal{D}} (E[XX'])^{-1} \mathcal{N}_k(0, E[(\varepsilon X)(\varepsilon X)']),$$

ce qui établit le résultat (puisque $E[\varepsilon^2 XX'] = E[\sigma^2(X)XX']$). □

Efficacité (1)

Peut-on faire plus précis que $\hat{\beta}$? En tout cas, pas asymptotiquement dans le modèle fort avec normalité, car...

Théorème 5

Dans le modèle fort avec normalité, $\hat{\beta}$ est l'estimateur du maximum de vraisemblance de β .

Preuve: Dans le modèle considéré, la vraisemblance

$$\begin{aligned}L_{\beta, \sigma^2} &= \prod_{i=1}^n f_{\beta, \sigma^2}^{(Y_i, X_i)}(Y_i, X_i) = \prod_{i=1}^n f_{\beta, \sigma^2}^{Y_i|X_i}(Y_i, X_i) f_{\beta, \sigma^2}^{X_i}(X_i) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(Y_i - \beta'X_i)^2\right) \prod_{i=1}^n f^{X_i}(X_i) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta'X_i)^2\right) \prod_{i=1}^n f^{X_i}(X_i)\end{aligned}$$

est maximale en β lorsque $\sum_{i=1}^n (Y_i - \beta'X_i)^2$ est minimale en β . □

Efficacité (2)

Pour n fixé et pour le modèle semi-fort homoscédastique, on a le

Théorème 6 (Gauss–Markov)

Dans le modèle semi-fort homoscédastique, $\hat{\beta}$ est à **variance conditionnellement minimale** dans la classe \mathcal{C} des estimateurs **linéaires** et **conditionnellement sans biais**.

Soit $\tilde{\beta}$ un estimateur de β . Ici,

$\tilde{\beta}$ est **linéaire** $\stackrel{\text{def}}{\Leftrightarrow} \tilde{\beta} = \sum_{i=1}^n \tilde{W}_i Y_i$ pour des v.a. $\tilde{W}_i = \tilde{W}_i(X_1, \dots, X_n)$

$\tilde{\beta}$ est **conditionnellement sans biais** $\stackrel{\text{def}}{\Leftrightarrow} E[\tilde{\beta} | \mathcal{X}] = \beta$ pour tout $\beta \in \mathbb{R}^k$

Puisque $\hat{\beta} = Q^{-1}(\frac{1}{n} \sum_{i=1}^n X_i Y_i) = \sum_{i=1}^n (Q^{-1} \frac{1}{n} X_i) Y_i = \sum_{i=1}^n \hat{W}_i Y_i$, $\hat{\beta}$ satisfait bien ces deux propriétés. Gauss–Markov affirme qu’il est le **BLUE** (Best Linear Unbiased Estimator).

Efficacité (2)

Preuve: Soit $\tilde{\beta} \in \mathcal{C}$. Notons d'abord que, pour tout $\beta \in \mathbb{R}^k$,

$$\beta = \mathbb{E}[\tilde{\beta}|\mathcal{X}] = \mathbb{E}[\sum_{i=1}^n \tilde{W}_i Y_i | \mathcal{X}] = \sum_{i=1}^n \tilde{W}_i \mathbb{E}[Y_i | \mathcal{X}] = \sum_{i=1}^n \tilde{W}_i X_i' \beta,$$

de sorte que $\sum_{i=1}^n \tilde{W}_i X_i' = I_k$ p.s. (I_k est la matrice identité $k \times k$).

Donc

$$\begin{aligned} \text{Var}[\tilde{\beta}|\mathcal{X}] &= \text{Var}[\sum_{i=1}^n \tilde{W}_i Y_i | \mathcal{X}] = \sum_{i=1}^n \text{Var}[\tilde{W}_i Y_i | \mathcal{X}] \\ &= \sum_{i=1}^n \tilde{W}_i \text{Var}[Y_i | \mathcal{X}] \tilde{W}_i' = \sum_{i=1}^n \tilde{W}_i \text{Var}[X_i' \beta + \varepsilon_i | \mathcal{X}] \tilde{W}_i' \\ &= \sum_{i=1}^n \tilde{W}_i \text{Var}[\varepsilon_i | X_i] \tilde{W}_i' = \sigma^2 \sum_{i=1}^n \tilde{W}_i \tilde{W}_i'. \end{aligned}$$

En particulier, $\text{Var}[\hat{\beta}|\mathcal{X}] = \sigma^2 \sum_{i=1}^n \hat{W}_i \hat{W}_i'$.

Efficacité (2)

On a donc

$$\begin{aligned}\text{Var}[\tilde{\beta}|\mathcal{X}] &= \sigma^2 \sum_{i=1}^n \tilde{W}_i \tilde{W}_i' \\ &= \sigma^2 \sum_{i=1}^n \{\hat{W}_i + (\tilde{W}_i - \hat{W}_i)\} \{\hat{W}_i + (\tilde{W}_i - \hat{W}_i)\}' \\ &= \text{Var}[\hat{\beta}|\mathcal{X}] + T + T' + \sigma^2 \sum_{i=1}^n (\tilde{W}_i - \hat{W}_i)(\tilde{W}_i - \hat{W}_i)' \\ &= \text{Var}[\hat{\beta}|\mathcal{X}] + \sigma^2 \sum_{i=1}^n (\tilde{W}_i - \hat{W}_i)(\tilde{W}_i - \hat{W}_i)' \\ &\geq \text{Var}[\hat{\beta}|\mathcal{X}]\end{aligned}$$

(au sens de l'ordre associé aux matrices semi-définies positives), où on a utilisé le fait que

$$\begin{aligned}T &\stackrel{\text{def}}{=} \sigma^2 \sum_{i=1}^n (\tilde{W}_i - \hat{W}_i) \hat{W}_i' = \sigma^2 \sum_{i=1}^n (\tilde{W}_i - Q^{-1} \frac{1}{n} X_i) (Q^{-1} \frac{1}{n} X_i)' \\ &= \frac{\sigma^2}{n} \left(\sum_{i=1}^n \tilde{W}_i X_i' - Q^{-1} \frac{1}{n} \sum_{i=1}^n X_i X_i' \right) Q^{-1} = 0. \quad \square\end{aligned}$$

Estimateur des moindres carrés pondérés

Dans le modèle semi-fort hétéroscédastique, $\hat{\beta}$ n'est pas efficace.
Comment construire un estimateur efficace?

Le modèle

$$\frac{Y}{\sigma(X)} = \left(\frac{X}{\sigma(X)} \right)' \beta + \frac{\varepsilon}{\sigma(X)},$$

que nous noterons

$$\tilde{Y} = \tilde{X}' \beta + \tilde{\varepsilon}$$

est semi-fort homoscedastique. En effet,

$$\mathbb{E}[\tilde{\varepsilon} | \tilde{X}] = \mathbb{E} \left[\mathbb{E} \left[\frac{\varepsilon}{\sigma(X)} | X \right] | \tilde{X} \right] = \mathbb{E} \left[\frac{1}{\sigma(X)} \mathbb{E}[\varepsilon | X] | \tilde{X} \right] = 0$$

$$\text{Var}[\tilde{\varepsilon} | \tilde{X}] = \mathbb{E}[\tilde{\varepsilon}^2 | \tilde{X}] = \mathbb{E} \left[\mathbb{E} \left[\frac{\varepsilon^2}{\sigma^2(X)} | X \right] | \tilde{X} \right] = \mathbb{E} \left[\frac{1}{\sigma^2(X)} \mathbb{E}[\varepsilon^2 | X] | \tilde{X} \right] = 1.$$

Estimateur des moindres carrés pondérés

Donc l'estimateur des moindres carrés pondérés

$$\begin{aligned}\hat{\beta}_{\text{MCP}} &= \left(\frac{1}{n} \sum_{i=1}^n \tilde{X}_i \tilde{X}_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n \tilde{X}_i \tilde{Y}_i \right) \\ &= \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{\sigma^2(X_i)} X_i X_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{\sigma^2(X_i)} X_i Y_i \right)\end{aligned}$$

est efficace pour β .

Malheureusement, il ne s'agit pas d'un estimateur...

Comment dès lors implémenter cet estimateur?

Estimateur des moindres carrés pondérés

Ceci requiert typiquement de choisir un modèle pour $x \mapsto \sigma^2(x)$.

Par exemple, $\sigma^2(x) = a + x'Bx$, avec $a > 0$ et B semi-définie positive.

Procédure:

- (1) Calculer les résidus $e_i = Y_i - X_i'\hat{\beta}$, $i = 1, \dots, n$, fondés sur $\hat{\beta}$
- (2) Estimer a et B par moindres carrés ordinaires dans le modèle linéaire

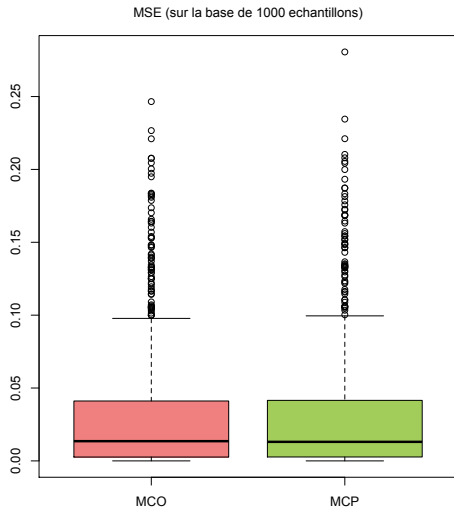
$$e_i^2 = a + X_i'BX_i + \eta_i$$

(convergence car $\varepsilon^2 = a + X'BX + \eta$ est un modèle linéaire semi-fort)

- (3) Calculer $\hat{\beta}_{\text{MCP}}$, fondé sur $\hat{\sigma}^2(x) = \hat{a} + x'\hat{B}x$.

On peut montrer que l'estimateur $\hat{\beta}_{\text{MCP}}$ qui en résulte est (asymptotiquement) efficace malgré les estimations non efficaces aux étapes (1)–(2).

Estimateur des moindres carrés pondérés



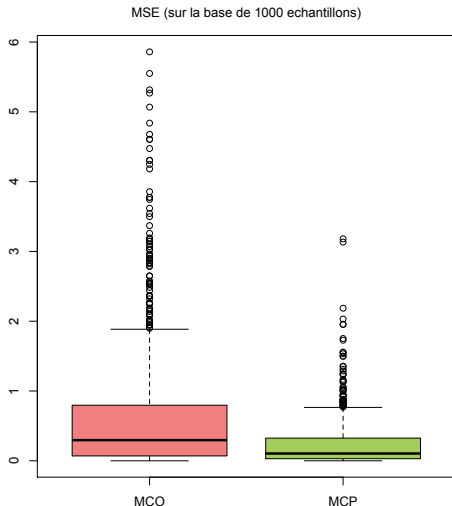
$$Y = 5 + 2X + \varepsilon$$

$$X \sim \text{Unif}(1, 10)$$

$$\varepsilon|X \sim \mathcal{N}(0, 1)$$

$$n = 200$$

Estimateur des moindres carrés pondérés



$$Y = 5 + 2X + \varepsilon$$
$$X \sim \text{Unif}(1, 10)$$
$$\varepsilon|X \sim \mathcal{N}(0, 1+X^2)$$
$$n = 200$$