

TD n°2

stoehr@ceremade.dauphine.fr

Exercice 1. On considère le modèle linéaire homoscédastique sans constante, *i.e.*, pour $i \in \llbracket 1, n \rrbracket$,

$$Y_i = \beta x_i + \varepsilon_i,$$

où $x_1, \dots, x_n$ sont fixés, les erreurs $\varepsilon_1, \dots, \varepsilon_n$ sont indépendantes, centrées et de variance commune σ^2 . On suppose que

$$\sum_{i=1}^n x_i^2 > 0, \quad \sum_{i=1}^n x_i \neq 0.$$

On définit

$$\hat{\beta}_n = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}, \quad \tilde{\beta}_n = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n x_i}.$$

1. Interpréter $\hat{\beta}_n$ et $\tilde{\beta}_n$ comme des Z-estimateurs en précisant leurs fonctions d'estimation. Lequel de ces estimateurs coïncide avec l'estimateur des moindres carrés ?
2. Comparer les propriétés imposées aux résidus par ces deux méthodes.
3. Montrer qu'ils sont sans biais pour β .
4. Comparer les variances de ces estimateurs. Dans quel cas sont-elles égales ?
5. Interpréter le résultat à partir des poids accordés aux observations.

Exercice 2. On considère la régression linéaire simple, *i.e.*, $i \in \llbracket 1, n \rrbracket$,

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

où $x_1, \dots, x_n$ sont des réels fixés qui ne sont pas tous égaux, et les erreurs $\varepsilon_1, \dots, \varepsilon_n$ sont indépendantes, centrées et de variance commune σ^2 . On souhaite déterminer la droite qui ajuste au mieux les observations au sens des moindres carrés, c'est-à-dire résoudre

$$(\hat{\beta}_{0,n}, \hat{\beta}_{1,n}) \in \arg \min_{(a,b) \in \mathbb{R}^2} \sum_{i=1}^n (Y_i - a - bx_i)^2.$$

On note

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i, \quad S_{xx} = \sum_{i=1}^n (x_i - \bar{X})^2, \quad S_{xy} = \sum_{i=1}^n (x_i - \bar{X})(Y_i - \bar{Y}).$$

1. Écrire la matrice de design $\mathbf{X}$ de ce modèle. Montrer qu'elle est de rang 2 et en déduire que l'estimateur des moindres carrés existe et est unique.
2. Donner les deux équations satisfaites par $\hat{\beta}_{0,n}$ et $\hat{\beta}_{1,n}$ pour le problème de minimisation. Dédurre de l'une d'elle une expression de $\hat{\beta}_{0,n}$ en fonction de $\hat{\beta}_{1,n}$, $\bar{X}$ et $\bar{Y}$.

3. Établir les identités

$$S_{xx} = \sum_{i=1}^n x_i^2 - n\bar{X}^2, \quad S_{xy} = \sum_{i=1}^n x_i Y_i - n\bar{X}\bar{Y}.$$

En déduire les expressions de $\hat{\beta}_{0,n}$ et $\hat{\beta}_{1,n}$.

4. Montrer que la droite ajustée passe par le barycentre $(\bar{X}, \bar{Y})$ du nuage de points et que la somme des résidus est nulle.

Exercice 3. On considère le modèle linéaire homoscédastique

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathbb{E}_\theta[\boldsymbol{\varepsilon}] = \mathbf{0}_n, \quad \text{Cov}_\theta[\boldsymbol{\varepsilon}] = \sigma^2 \mathbf{I}_n,$$

où $\boldsymbol{\beta} \in \mathbb{R}^p$ et $\mathbf{X} \in \mathbb{R}^{n \times p}$ est une matrice déterministe de rang p , avec $n > p$. On note $\hat{\boldsymbol{\beta}}_n$ l'estimateur des moindres carrés et

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top, \quad \mathbf{M} = \mathbf{I}_n - \mathbf{H}.$$

On rappelle que $\mathbf{H}$ et $\mathbf{M}$ sont les matrices des projections orthogonales sur $\text{Im}(\mathbf{X})$ et $\text{Im}(\mathbf{X})^\perp$, respectivement.

1. Montrer que

$$\text{SCR} = \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_n\|^2 = \boldsymbol{\varepsilon}^\top \mathbf{M} \boldsymbol{\varepsilon}.$$

2. En utilisant les propriétés de la trace, montrer que

$$\mathbb{E}_\theta[\text{SCR}] = \sigma^2 \text{tr}(\mathbf{M}).$$

3. Montrer que la trace d'une matrice de projection orthogonale est égale à son rang. En déduire la valeur de $\text{tr}(\mathbf{M})$.
4. Construire un estimateur sans biais de σ^2 à partir de SCR. Calculer le biais de SCR/n et expliquer pourquoi on divise par $n-p$ plutôt que par n . Ce résultat nécessite-t-il que les erreurs soient gaussiennes ?

Exercice 4. Soit $p \geq 1$ fixé. On considère une suite de modèles de régression à plan déterministe, *i.e.*, pour tout $n \geq p$, on considère le modèle

$$\mathbf{Y}_n = \mathbf{X}_n \boldsymbol{\beta} + \boldsymbol{\varepsilon}_n, \quad \mathbb{E}_\theta[\boldsymbol{\varepsilon}] = \mathbf{0}_n, \quad \text{Cov}_\theta[\boldsymbol{\varepsilon}] = \sigma^2 \mathbf{I}_n,$$

où $\boldsymbol{\beta} \in \mathbb{R}^p$, $\mathbf{X}_n \in \mathbb{R}^{n \times p}$ est une matrice déterministe dont la i -ème ligne est notée $\mathbf{x}_{n,i}^\top$, et $\sigma^2 \in \mathbb{R}_+^*$. On suppose que $\mathbf{X}_n$ est de rang p pour tout $n \geq p$ et que

$$\frac{1}{n} \mathbf{X}_n^\top \mathbf{X}_n \xrightarrow{n \rightarrow \infty} \mathbf{Q},$$

où $\mathbf{Q}$ est une matrice définie positive. On note $\hat{\boldsymbol{\beta}}_n$ l'estimateur des moindres carrés.

Partie I – Consistance.

1. Rappeler l'expression de $\hat{\beta}_n$. Montrer que

$$\hat{\beta}_n - \beta = (\mathbf{X}_n^\top \mathbf{X}_n)^{-1} \sum_{i=1}^n \mathbf{x}_{n,i} \varepsilon_i.$$

En déduire que $\hat{\beta}_n$ est sans biais et déterminer sa matrice de covariance.

2. Montrer que

$$\mathbb{E} \left[\left\| \hat{\beta}_n - \beta \right\|^2 \right] = \frac{\sigma^2}{n} \operatorname{tr} \left\{ \left(\frac{1}{n} \mathbf{X}_n^\top \mathbf{X}_n \right)^{-1} \right\} \xrightarrow{n \rightarrow \infty} 0.$$

En déduire la consistance de $\hat{\beta}_n$.

Partie II — Normalité asymptotique. On suppose désormais que les variables $(\varepsilon_i)_{i \geq 1}$ sont *i.i.d.*, en plus d'être centrées et de variance σ^2 , et que le plan satisfait également

$$\max_{1 \leq i \leq n} \frac{\|\mathbf{x}_{n,i}\|}{\sqrt{n}} \xrightarrow{n \rightarrow \infty} 0.$$

3. Montrer que

$$\sqrt{n}(\hat{\beta}_n - \beta) = \left(\frac{1}{n} \mathbf{X}_n^\top \mathbf{X}_n \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_{n,i} \varepsilon_i.$$

Pourquoi ne peut-on pas appliquer directement le théorème central limite usuel pour des variables *i.i.d.* à la somme figurant dans cette expression ?

4. Soit $\mathbf{u} \in \mathbb{R}^p \setminus \{\mathbf{0}_p\}$. On pose

$$Z_{n,i} = \frac{\mathbf{u}^\top \mathbf{x}_{n,i} \varepsilon_i}{\sqrt{n}}.$$

Vérifier que $Z_{n,1}, \dots, Z_{n,n}$ sont indépendantes, centrées et de carré intégrable. Montrer que

$$\sum_{i=1}^n \operatorname{Var}[Z_{n,i}] \xrightarrow{n \rightarrow \infty} \sigma^2 \mathbf{u}^\top \mathbf{Q} \mathbf{u} > 0.$$

5. Soit $\delta > 0$. Montrer que

$$\sum_{i=1}^n \mathbb{E} \left[Z_{n,i}^2 \mathbb{1}_{\{|Z_{n,i}| > \delta\}} \right] \leq \left\{ \frac{1}{n} \sum_{i=1}^n (\mathbf{u}^\top \mathbf{x}_{n,i})^2 \right\} \mathbb{E} \left[\varepsilon_1^2 \mathbb{1}_{\left\{ |\varepsilon_1| > \frac{\delta \sqrt{n}}{\|\mathbf{u}\| \max_{1 \leq j \leq n} \|\mathbf{x}_{n,j}\|} \right\}} \right].$$

6. En déduire que la condition de Lindeberg est satisfaite.
7. Appliquer le théorème de Lindeberg–Feller puis le critère de Cramér–Wold pour établir

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_{n,i} \varepsilon_i \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_p(\mathbf{0}_p, \sigma^2 \mathbf{Q}).$$

En conclure que

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_p(\mathbf{0}_p, \sigma^2 \mathbf{Q}^{-1}).$$

Partie III — Pourquoi contrôler le poids des observations ?

8. On considère le cas $p = 1$ et le plan

$$x_{n,1} = \sqrt{n}, \quad x_{n,i} = 0, \quad i \in \llbracket 2, n \rrbracket.$$

Vérifier que la première hypothèse sur le plan est satisfaite, mais pas la seconde. Calculer $\hat{\beta}_n - \beta$ et $\sqrt{n}(\hat{\beta}_n - \beta)$. Si les erreurs prennent les valeurs -1 et 1 avec probabilité $1/2$, l'estimateur est-il consistant ? Est-il asymptotiquement normal à la vitesse $\sqrt{n}$? Interpréter.

Annexe

Théorème de Lindeberg–Feller. Pour chaque n , soient $Z_{n,1}, \dots, Z_{n,n}$ des variables aléatoires réelles indépendantes, centrées et de carré intégrable. Supposons que

$$\sum_{i=1}^n \text{Var}[Z_{n,i}] \xrightarrow{n \rightarrow \infty} s^2 \in]0, +\infty[$$

et que, pour tout $\delta > 0$,

$$\sum_{i=1}^n \mathbb{E} \left[Z_{n,i}^2 \mathbb{1}_{\{|Z_{n,i}| > \delta\}} \right] \xrightarrow{n \rightarrow \infty} 0.$$

Alors,

$$\sum_{i=1}^n Z_{n,i} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, s^2).$$

Théorème de Cramér–Wold. Soient $(\mathbf{Z}_n)_{n \geq 1}$ une suite de vecteurs aléatoires de $\mathbb{R}^p$ et $\mathbf{Z}$ un vecteur aléatoire de $\mathbb{R}^p$. Alors,

$$\mathbf{Z}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathbf{Z} \quad \Longleftrightarrow \quad \forall \mathbf{u} \in \mathbb{R}^p, \quad \mathbf{u}^\top \mathbf{Z}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathbf{u}^\top \mathbf{Z}.$$