Régression linéaire
Convergence des estimateurs

Alain Camanes

1 Moindres Carrés & Estimation

1.1 Le modèle

On considère le modèle de régression linéaire :

$$\forall\ i \in [\![1, n]\!],\ y_i = a_1 x_i + a_0 + \varepsilon_i, \tag{1}$$

où $x_1, \dots, x_n$ sont des variables à valeurs réelles et $(\varepsilon_i)_{i\in\{1,\dots,n\}}$ est une famille de variables aléatoires indépendantes, identiquement distribuées, de carrés intégrables, centrées (i.e. de moyenne nulle) et de variance σ^2.

Remarque 1.1 - Les quantités a_0, a_1, $x_1, \ldots, x_n$ sont des réels. Les quantités $\varepsilon_1, \ldots, \varepsilon_n, y_1, \ldots, y_n$ sont des variables aléatoires. L'objectif est d'estimer les valeurs des paramètres a_0, a_1 et σ^2 connaissant les valeurs $(x_i)_{1 \leqslant i \leqslant n}$ et $(y_i)_{1 \leqslant i \leqslant n}$.

On introduit les notations suivantes :

* les moments empiriques : pour tout $k \in \mathbb{N}$, on note

$$\overline{x^k} = \frac{1}{n} \sum_{i=1}^{n} x_i^k.$$

* les (co)variances empiriques :

$$s_{x,x} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 = \overline{x^2} - \overline{x}^2,$$
$$s_{y,y} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y})^2 = \overline{y^2} - \overline{y}^2,$$
$$s_{x,y} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) = \overline{xy} - \overline{x} \cdot \overline{y}.$$

Remarque 1.2 - On suppose que, pour tout k entier inférieur ou égal à 4, les quantités $\overline{x^k}$ et $s_{x,x}$ restent bornées lorsque n tend vers $+\infty$. Pour prouver la consistance des estimateurs de a_0, a_1 et σ^2, on supposera que ε_1 possède un moment d'ordre 4 et on notera $\mu_4 = \mathbf{E}\left[\varepsilon_1^4\right]$.

La méthode des moindres carrés (cf. Section C) permet d'estimer les paramètres a_1 et a_0 par les quantités :

$$\widehat{a_1} = \frac{s_{x,y}}{s_{x,x}}, \tag{2}$$
$$\widehat{a_0} = \overline{y} - \widehat{a_1}\overline{x}. \tag{3}$$

On introduit alors les variables prédites :

$$\widehat{y_i} = \widehat{a_1} x_i + \widehat{a_0} \tag{4}$$

et l'erreur résiduelle commise :

$$\varepsilon_{i,R} = y_i - \widehat{y_i}, \tag{5}$$
$$s^2 = \frac{1}{n-2} \sum_{i=1}^{n} \left(y_i - \widehat{y_i}\right)^2. \tag{6}$$

Remarque 1.3 - Les quantités $\widehat{a_1}$, $\widehat{a_0}$ et s^2 sont des variables aléatoires dépendant du nombre n de variables explicatives. Nous montrons par la suite que, lorsque n est grand, ces quantités sont proches des valeurs des paramètres a_0, a_1 et σ^2. De plus, on montre que les erreurs commises suivent asymptotiquement des lois normales. Le théorème suivant précise ces notions.

Théorème 1.1 - Convergence des estimateurs

Les variables aléatoires $\widehat{a_1}$, $\widehat{a_0}$ et s^2 sont des estimateurs sans biais et fortement consistants de a_1, a_0 et σ^2 respectivement.

De plus, il existe des suites de réels $s_{1,n}$, $s_{2,n}$ et $s_{3,n}$ telles que les suites de variables aléatoires $\left(\frac{\widehat{a_0}-a_0}{s_{1,n}}\right)$, $\left(\frac{\widehat{a_1}-a_1}{s_{2,n}}\right)$ et $\left(\frac{s^2-\sigma^2}{s_{3,n}}\right)$ convergent en loi vers des lois normales centrées réduites.

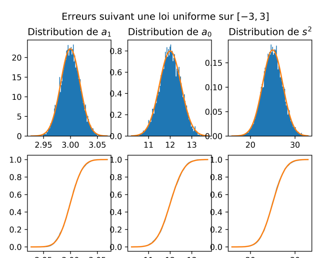

FIGURE 1 – Histogrammes et fonctions de répartitions empiriques pour 10 000 simulations avec des valeurs $(x_1, \ldots, x_{100})$ fixées.

Remarque 1.4 - On distingue (voir la Définition A.1) les estimateurs :

- $*$ consistants : qui convergent en probabilité,
- $*$ fortement consistants : qui convergent presque sûrement.

La convergence presque sûre impliquant la convergence en probabilité, si un estimateur est fortement consistant, alors il est consistant. Nous faisons toutefois le choix par la suite de prouver d'abord la consistance car elle fait appel à des outils plus simples, essentiellement l'inégalité de Bienaymé-Tchebychev (voir le Théorème A.1).

Les propriétés de forte consistance reposent sur une généralisation de la loi forte des grands nombres, via le critère de Kolmogorov, où les variables aléatoires ne sont pas toutes de même loi (voir le

Théorème B.2).

Les convergences en lois reposent sur une généralisation du théorème central limite, via la condition de Lyapunov, où les variables aléatoires ne sont pas toutes de même loi (voir le Théorème B.3).

1.2 Interprétation en termes de projection

Posons

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \widetilde{Y} = Y - \mathbf{E}[Y], X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \text{ et } H = X\left(X^T X\right)^{-1} X^T.$$

Proposition 1.1 - Projection orthogonale

En utilisant les notations précédentes, H est la matrice d'un projecteur orthogonal,

$$Y - \widehat{Y} = (I_n - H)\widetilde{Y} \text{ et } (n-2)s^2 = \left\|Y - \widehat{Y}\right\|^2.$$

Preuve. Voir Annexe C □

1.3 Calculs préliminaires

Proposition 1.2 - Résultats en moyenne

Soit $i \in \{1, \ldots, n\}$. Comme la variable aléatoire ε_i est centrée,

$$\mathbf{E}[y_i] = a_1 x_i + a_0 \text{ et } \mathbf{E}[\overline{y}] = a_1 \overline{x} + a_0.$$

Comme la variable aléatoire ε_i est de variance σ^2 et que les (ε_i) sont indépendantes,

$$\mathbf{V}(y_i) = \sigma^2, \mathbf{V}(\overline{y}) = \frac{\sigma^2}{n} \text{ et } \mathrm{Cov}(y_i, \varepsilon_j) = \mathrm{Cov}(y_i, y_j) = \sigma^2 \delta_{i,j},$$

où $\delta_{i,j}$, le symbole de Kronecker, vaut 1 si $i = j$ et 0 sinon.

Preuve. Pour les espérances, on utilise la linéarité de l'espérance et l'hypothèse $\mathbf{E}[\varepsilon_i] = 0$. Pour la variance, comme $a_1 x_i + a_0$ est déterministe,

$$\mathbf{V}(y_i) = \mathbf{V}(a_1 x_i + a_0 + \varepsilon_i) = \mathbf{V}(\varepsilon_i).$$

L'indépendance permet alors d'obtenir $\mathbf{V}(\overline{y}) = \frac{1}{n^2} \sum_{i=1}^{n} \mathbf{V}(y_i)$.

D'après la bilinéarité de la covariance,

$$\mathrm{Cov}\,(y_i, \varepsilon_j) = \mathrm{Cov}\,(a_1 x_i + a_0 + \varepsilon_i, \varepsilon_j) = \mathrm{Cov}\,(\varepsilon_i, \varepsilon_j) = \sigma^2 \delta_{i,j}.$$

□

1.4 Calculs de moments

Proposition 1.3 - Sommes de v.a. indépendantes

- $\widehat{a_1}$ est une somme de variables aléatoires indépendantes :

$$\widehat{a_1} = a_1 + \frac{1}{n s_{x,x}} \sum_{i=1}^{n} (x_i - \overline{x}) \varepsilon_i. \tag{7}$$

- $\widehat{a_0}$ est une somme de variables aléatoires indépendantes :

$$\widehat{a_0} = a_0 + \frac{1}{n s_{x,x}} \sum_{i=1}^{n} \left[\overline{x^2} - x_i \overline{x} \right] \varepsilon_i. \tag{8}$$

- $\widehat{a_1}$ et $\overline{y}$ sont non corrélées : $\mathrm{Cov}\,(\overline{y}, \widehat{a_1}) = 0$.
- L'erreur quadratique s'exprime ainsi :

$$s^2 = \frac{1}{n-2} \left(\sum_{i=1}^{n} (\varepsilon_i - \overline{\varepsilon})^2 - n s_{x,x} (a_1 - \widehat{a_1})^2 \right). \tag{9}$$

Preuve.

- D'après la définition de $\widehat{a_1}$ rappelée dans la relation (2),

$$\begin{aligned} n s_{x,x} \left(\widehat{a_1} - a_1 \right) &= \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) - n s_{x,x} a_1 \\ &= \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y} - a_1 x_i + a_1 \overline{x}) \\ &= \sum_{i=1}^{n} (x_i - \overline{x})(\varepsilon_i + a_0 - \overline{\varepsilon} - a_0) \\ &= \sum_{i=1}^{n} (x_i - \overline{x})(\varepsilon_i - \overline{\varepsilon}) = \sum_{i=1}^{n} (x_i - \overline{x}) \varepsilon_i. \end{aligned}$$

- D'après la définition de $\widehat{a_0}$ rappelée dans la relation (2) puis la relation (7) précédente,

$$\begin{aligned}\widehat{a_0}-a_0&=\overline{y}-\widehat{a_1}\overline{x}-(\overline{y}-a_1\overline{x}-\overline{\varepsilon})\\&=(a_1-\widehat{a_1})\overline{x}+\overline{\varepsilon}\\&=-\frac{\overline{x}}{ns_{x,x}}\sum_{i=1}^n(x_i-\overline{x})\varepsilon_i+\frac{1}{n}\sum_{i=1}^n\varepsilon_i.\end{aligned}$$

Ainsi,

$$\begin{aligned}ns_{x,x}\left(\widehat{a_0}-a_0\right)&=\sum_{i=1}^n\left(s_{x,x}-\overline{x}(x_i-\overline{x})\right)\varepsilon_i=\sum_{i=1}^n\left(\overline{x^2}-\overline{x}^2-x_i\overline{x}+\overline{x}^2\right)\varepsilon_i\\&=\sum_{i=1}^n\left(\overline{x^2}-x_i\overline{x}\right)\varepsilon_i=\sum_{i=1}^n\left(\overline{x^2}-x_i\overline{x}\right)\varepsilon_i.\end{aligned}$$

∗ En utilisant la bilinéarité de la covariance et la relation (7),

$$\begin{aligned}n^2s_{x,x}\mathrm{Cov}\left(\overline{y},\widehat{a_1}\right)&=\sum_{i=1}^n\sum_{j=1}^n\underbrace{\mathrm{Cov}\left(y_i,y_j\right)}_{\sigma^2\delta_{i,j}}(x_j-\overline{x})\\&=\sum_{i=1}^n\sigma^2(x_i-\overline{x})=0.\end{aligned}$$

∗ En reprenant les définitions,

$$y_i-\widehat{y_i}=(a_1-\widehat{a_1})x_i+(a_0-\widehat{a_0})+\varepsilon_i.$$

Or, $a_0=\overline{y}-a_1\overline{x}-\overline{\varepsilon}$ et $\widehat{a_0}=\overline{y}-\widehat{a_1}\overline{x}$. Ainsi,

$$\varepsilon_{i,R}=y_i-\widehat{y_i}=(a_1-\widehat{a_1})(x_i-\overline{x})+(\varepsilon_i-\overline{\varepsilon}).\tag{10}$$

En injectant dans la définition de s^2,

$$\begin{aligned}(n-2)s^2&=\sum_{i=1}^n\varepsilon_{i,R}^2\\&=\sum_{i=1}^n\left[(a_1-\widehat{a_1})^2(x_i-\overline{x})^2+(\varepsilon_i-\overline{\varepsilon})^2+\cdots\right.\\&\qquad\left.\cdots+2(a_1-\widehat{a_1})(x_i-\overline{x})(\varepsilon_i-\overline{\varepsilon})\right].\end{aligned}$$

Or, d'après le point précédent,

$$\sum_{i=1}^n(x_i-\overline{x})(\varepsilon_i-\overline{\varepsilon})=ns_{x,x}(\widehat{a_1}-a_1).$$

Ainsi,

$$(n-2)s^2=\sum_{i=1}^n(\varepsilon_i-\overline{\varepsilon})^2-ns_{x,x}(a_1-\widehat{a_1})^2.$$

□

Proposition 1.4 - Moments de $\widehat{a_1}$

La variable aléatoire $\widehat{a_1}$ possède un moment d'ordre 2 :

$$\mathbf{E}\left[\widehat{a_1}\right] = a_1 \text{ et } \mathbf{V}\left(\widehat{a_1}\right) = \frac{\sigma^2}{n s_{x,x}}.$$

Preuve. En utilisant la relation (7) de décomposition de $\widehat{a_1}$ comme somme de v.a. i.i.d.,

$$\mathbf{E}\left[\widehat{a_1}\right] = a_1 + \frac{1}{n s_{x,x}} \sum_{i=1}^{n} (x_i - \overline{x}) \mathbf{E}\left[\varepsilon_i\right] = a_1.$$

Toujours en utilisant la relation (7),

$$\mathbf{V}\left(\widehat{a_1}\right) = \frac{1}{n^2 s_{x,x}^2} \sum_{i=1}^{n} (x_i - \overline{x})^2 \mathbf{V}\left(\varepsilon_i\right) = \frac{\sigma^2}{n s_{x,x}}.$$

□

Proposition 1.5 - Moments de $\widehat{a_0}$

La variable aléatoire $\widehat{a_0}$ possède un moment d'ordre 2 :

$$\mathbf{E}\left[\widehat{a_0}\right] = a_0 \text{ et } \mathbf{V}\left(\widehat{a_0}\right) = \frac{\sigma^2 \overline{x^2}}{n s_{x,x}}.$$

Preuve. En utilisant le calcul de $\mathbf{E}\left[\widehat{a_1}\right]$,

$$\mathbf{E}\left[\widehat{a_0}\right] = \mathbf{E}\left[\overline{y}\right] - \mathbf{E}\left[\widehat{a_1}\right] \overline{x} = a_1 \overline{x} + a_0 - a_1 \overline{x} = a_0.$$

D'après la Proposition 1.3, comme $\widehat{a_1}$ et $\overline{y}$ ne sont pas corrélées, puis en utilisant le calcul de la question précédente,

$$\begin{aligned}
\mathbf{V}\left(\widehat{a_0}\right) &= \mathbf{V}\left(\overline{y} - \widehat{a_1}\overline{x}\right) \\
&= \mathbf{V}\left(\overline{y}\right) + \mathbf{V}\left(\widehat{a_1}\overline{x}\right) - 2\mathrm{Cov}\left(\overline{y}, \widehat{a_1}\overline{x}\right) \\
&= \frac{\sigma^2}{n} + \frac{\overline{x}^2 \sigma^2}{n s_{x,x}} \\
&= \frac{\sigma^2}{n \sum_{i=1}^{n} (x_i - \overline{x})^2} \left[\sum_{i=1}^{n} (x_i - \overline{x})^2 + n \overline{x}^2 \right] \\
&= \frac{\sigma^2}{n \sum_{i=1}^{n} (x_i - \overline{x})^2} \sum_{i=1}^{n} x_i^2 = \frac{\sigma^2 \overline{x^2}}{n s_{x,x}}.
\end{aligned}$$

□

2 Le cas gaussien

On suppose ici que les variables aléatoires $(\varepsilon_i)_{i\in\mathbb{N}}$ suivent une même loi gaussienne $\mathscr{N}(0,\sigma^2)$. La stabilité des lois normales et le théorème de Cochran permettent ici d'effectuer des calculs de lois explicites.

Proposition 2.1 - Lois des estimateurs (cas gaussien)

Si $(\varepsilon_1,\ldots,\varepsilon_n)$ suivent des lois gaussiennes, alors $\widehat{a_1}$ et $\widehat{a_0}$ suivent des lois gaussiennes. De plus, $\frac{(n-2)s^2}{\sigma^2}$ suit une loi du $\chi^2(n-2)$. En particulier, $\mathbf{E}\left[s^2\right]=\sigma^2$ et $\mathbf{V}\left(s^2\right)=\frac{2\sigma^4}{n-2}$.

Preuve. Comme $\widehat{a_1}$ est une somme de v.a. gaussiennes indépendantes (voir la relation (7)), alors $\widehat{a_1}$ suit une loi gaussienne. Les calculs de l'espérance et de la variance de $\widehat{a_1}$ (voir Proposition 1.4) permettent d'obtenir :

$$\widehat{a_1} \hookrightarrow \mathscr{N}\left(a_1, \frac{\sigma^2}{ns_{x,x}}\right).$$

Comme $\widehat{a_0}$ est une somme de v.a. gaussiennes indépendantes (voir la relation (8)), alors $\widehat{a_0}$ suit une loi gaussienne. Les calculs de l'espérance et de la variance de $\widehat{a_0}$ (voir Proposition 1.5) permettent d'obtenir :

$$\widehat{a_0} \hookrightarrow \mathscr{N}\left(a_0, \frac{\sigma^2\overline{x^2}}{ns_{x,x}}\right).$$

L'interprétation en termes de projection orthogonale (cf. Section 1.2) permet d'écrire

$$(n-2)s^2 = \left\|Y-\widehat{Y}\right\|^2 = \|(I_n-H)Y\|^2.$$

Comme (I_n-H) est un projecteur orthogonal sur $\mathrm{Im}(I_n-H)$ et que

$$\dim \mathrm{Im}(I_n-H) = \mathrm{Rg}(I_n-H) = \mathrm{Tr}(I_n-H) = n-2,$$

d'après le théorème de Cochran (voir Théorème B.4),

$$\frac{(n-2)s^2}{\sigma^2} \hookrightarrow \chi^2(n-2).$$

En utilisant les propriétés de la loi du χ^2, on obtient $\mathbf{E}\left[\frac{(n-2)s^2}{\sigma^2}\right]=n-2$, soit

$$\mathbf{E}\left[s^2\right]=\sigma^2$$

et $\mathbf{V}\left(\frac{(n-2)s^2}{\sigma^2}\right)=2(n-2)$, soit $\mathbf{V}\left(s^2\right)=\frac{2\sigma^4}{n-2}$. □

Remarque 2.1 - Dans le cas gaussien, $\mu_4 = \mathbf{E}\left[\varepsilon_1^4\right] = 3\sigma^4$. Le calcul de la variance de s^2 sera généralisé (Voir la relation (11)) à des variables aléatoires de loi quelconque possédant un moment d'ordre 4.

Remarque 2.2 - Dans ce cadre, les normalités de $\widehat{a_1}$ et $\widehat{a_0}$ sont donc évidentes. Comme la loi du χ^2 converge asymptotiquement vers une loi normale, la normalité asymptotique de s^2 en découle.

3 Étude de $\widehat{a_0}$ et $\widehat{a_1}$

3.1 Calculs de biais

Proposition 3.1 - Biais de $\widehat{a_1}$

L'estimateur $\widehat{a_1}$ est un estimateur sans biais de a_1.

Preuve. Conséquence immédiate des calculs de moments de la Proposition 1.4. □

Proposition 3.2 - Biais de $\widehat{a_0}$

L'estimateur $\widehat{a_0}$ est un estimateur sans biais de a_0.

Preuve. Conséquence immédiate des calculs de moments de la Proposition 1.5. □

3.2 Consistances de $\widehat{a_1}$

Proposition 3.3 - Consistance de $\widehat{a_1}$

L'estimateur $\widehat{a_1}$ est consistant.

Preuve. D'après la Proposition 1.4, $\mathbf{V}\left(\widehat{a_1}\right) = \frac{\sigma^2}{n s_{x,x}}$. Comme $s_{x,x}$ est bornée, on obtient bien $\lim\limits_{n\to+\infty} \mathbf{V}\left(\widehat{a_1}\right) = 0$ et la convergence en probabilité de $\widehat{a_1}$ vers son espérance a_1. □

Proposition 3.4 - Forte consistance de $\widehat{a_1}$

L'estimateur $\widehat{a_1}$ est fortement consistant.

Preuve. D'après l'écriture de $\widehat{a_1}$ comme somme de v.a. i.i.d. de la relation (7),

$$\widehat{a_1} - a_1 = \frac{1}{ns_{x,x}} \sum_{i=1}^{n} (x_i - \overline{x})\varepsilon_i.$$

Posons $Z_i = \frac{(x_i - \overline{x})\varepsilon_i}{s_{x,x}}$. Alors, $\mathbf{E}[Z_i] = \frac{x_i - \overline{x}}{s_{x,x}} \mathbf{E}[\varepsilon_i] = 0$ et $\mathbf{V}(Z_i) = \frac{(x_i - \overline{x})^2}{s_{x,x}^2} \sigma^2$. Ainsi,

$$\sum_{i=1}^{n} \frac{\mathbf{V}(Z_i)}{i^2} = \frac{\sigma^2}{s_{x,x}^2} \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{i^2}$$

Posons $V_i = \sum_{j=1}^{i} (x_j - \overline{x})^2$. En utilisant une transformation d'Abel,

$$\begin{aligned}
\sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{i^2} &= \sum_{i=1}^{n} \frac{V_i - V_{i-1}}{i^2} \\
&= \sum_{i=1}^{n} \frac{V_i}{i^2} - \sum_{i=0}^{n-1} \frac{V_i}{(i+1)^2} \\
&= \sum_{i=1}^{n} \left(\frac{1}{i^2} - \frac{1}{(i+1)^2} \right) V_i + \frac{1}{n^2} V_n.
\end{aligned}$$

Comme $\frac{V_n}{n} = s_{x,x}$ est le terme général d'une suite bornée, alors $\lim\limits_{n \to +\infty} \frac{V_n}{n^2} = 0$. De plus,

$$\left(\frac{1}{i^2} - \frac{1}{(i+1)^2} \right) V_i = \frac{2i+1}{i^2(i+1)^2} V_i \sim \frac{2V_i}{i^3}.$$

Comme (V_i/i) est supposée bornée, alors cette quantité est dominée par $1/i^2$ et, d'après les critères de comparaison aux séries de Riemann, il s'agit du terme général d'une série convergente. Finalement, $\sum \frac{(x_i - \overline{x})^2}{i^2}$ converge.

D'après la loi forte des grands nombres de Kolmogorov (voir Théorème B.2), $\widehat{a_1}$ converge presque sûrement vers a_1. □

3.3 Consistances de $\widehat{a_0}$

Proposition 3.5 - Consistance de $\widehat{a_0}$

L'estimateur $\widehat{a_0}$ est faiblement consistant.

Preuve. D'après la Proposition 1.5, $\mathbf{V}(\widehat{a_0}) = \frac{\sigma^2 \overline{x^2}}{ns_{x,x}}$ Ainsi, comme $\overline{x^2}$ et $s_{x,x}$ sont bornées, on obtient bien $\lim\limits_{n \to +\infty} \mathbf{V}(\widehat{a_0}) = 0$ et la convergence en probabilité de $\widehat{a_0}$ vers son espérance a_0. □

Proposition 3.6 - Forte consistance de $\widehat{a_0}$

L'estimateur $\widehat{a_0}$ est fortement consistant.

Preuve. D'après la définition,

$$\widehat{a_0} = \overline{y} - \widehat{a_1}\overline{x} = a_1\overline{x} + a_0 + \overline{\varepsilon} - \widehat{a_1}\overline{x} = (a_1 - \widehat{a_1})\overline{x} + \overline{\varepsilon} + a_0.$$

D'après la loi forte des grands nombres, $\overline{\varepsilon} \to 0$ presque sûrement. D'après la Proposition 3.4, $\widehat{a_1} \to a_1$ presque sûrement. Comme $\overline{x}$ est bornée, en additionnant les convergences presque sûres, $\widehat{a_0} \to a_0$ presque sûrement. □

3.4 Normalités asymptotiques

Proposition 3.7 - Normalité asymptotique de $\widehat{a_1}$

Il y a convergence en loi :

$$\sqrt{\frac{ns_{x,x}}{\sigma^2}}(\widehat{a_1} - a_1) \longrightarrow \mathscr{N}(0,1).$$

Preuve. On reprend l'équation (7) d'écriture de $\widehat{a_1}$ comme somme de v.a. i.i.d. :

$$\widehat{a_1} - a_1 = \frac{1}{ns_{x,x}} \sum_{i=1}^{n} (x_i - \overline{x})\varepsilon_i.$$

On pose $Z_i = (x_i - \overline{x})\varepsilon_i$. Alors, en utilisant l'indépendance,

$$\sigma(n)^2 = \mathbf{V}\left(\sum_{i=1}^{n} Z_i\right) = n\sigma^2 s_{x,x}.$$

Ainsi, en rappelant que $\mu_4 = \mathbf{E}\left[\varepsilon_i^4\right]$,

$$\begin{aligned}\frac{1}{\sigma(n)^4}\sum_{i=1}^{n}\mathbf{E}\left[Z_i^4\right] &= \frac{1}{n^2\sigma^4 s_{x,x}^2}\sum_{i=1}^{n}(x_i - \overline{x})^4\mu_4 \\ &= \frac{\mu_4}{n\sigma^4 s_{x,x}^2} \times \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4,\end{aligned}$$

Comme $\frac{1}{n}\sum\limits_{i=1}^{n}(x_i - \overline{x})^4$ s'exprime à l'aide de $\overline{x}$, $\overline{x^2}$, $\overline{x^3}$ et $\overline{x^4}$, cette quantité est bornée. Ainsi,

$$\lim_{n\to+\infty} \frac{1}{\sigma(n)^4}\sum_{i=1}^{n}\mathbf{E}\left[Z_i^4\right] = 0.$$

La condition de Lyapunov du Théorème B.3 est satisfaite, soit la convergence en loi :

$$\frac{1}{\sqrt{n\sigma^2 s_{x,x}}} \sum_{i=1}^{n} (x_i - \overline{x})\varepsilon_i \longrightarrow \mathcal{N}(0,1).$$

□

Proposition 3.8 - Normalité asymptotique de $\widehat{a_0}$

Il y a convergence en loi :

$$\sqrt{\frac{n s_{x,x}}{\sigma^2 \overline{x^2}}} (\widehat{a_0} - a_0) \longrightarrow \mathcal{N}(0,1).$$

Preuve. En utilisant la relation (8),

$$\widehat{a_0} - a_0 = \frac{1}{n s_{x,x}} \sum_{i=1}^{n} \left(\overline{x^2} - x_i \overline{x} \right) \varepsilon_i.$$

En posant $Z_i = (\overline{x^2} - x_i\overline{x})\varepsilon_i$, alors $\mathbf{E}\left[Z_i\right] = 0$ et

$$\begin{aligned} \sigma(n)^2 = \sum_{i=1}^{n} \mathbf{V}\left(Z_i\right) &= \sigma^2 \sum_{i=1}^{n} (\overline{x^2} - x_i \overline{x})^2 \\ &= \sigma^2 \left[n(\overline{x^2})^2 - 2n\overline{x}^2\overline{x^2} + n\overline{x}^2\overline{x^2} \right] \\ &= n\sigma^2\overline{x^2}\left(\overline{x^2} - \overline{x}^2 \right) = n\sigma^2 s_{x,x}\overline{x^2}. \end{aligned}$$

Ainsi, en utilisant un raisonnement similaire à celui de la démonstration précédente,

$$\frac{1}{\sigma(n)^4} \sum_{i=1}^{n} \mathbf{E}\left[Z_i^4\right] = \frac{\mu_4}{n^2\sigma^4 s_{x,x}^2 (\overline{x^2})^2} \sum_{i=1}^{n} \left(\overline{x^2} - x_i \overline{x} \right)^4 \to 0.$$

La condition de Lyapunov du Théorème B.3 est satisfaite, donc

$$\frac{1}{\sqrt{n\sigma^2 s_{x,x}\overline{x^2}}} \sum_{i=1}^{n} (\overline{x^2} - x_i\overline{x})\varepsilon_i \longrightarrow \mathcal{N}(0,1),$$

soit la convergence en loi attendue. □

4 Étude de s^2

4.1 Calcul de biais (I)

Proposition 4.1 - Biais de s^2

L'estimateur s^2 est un estimateur sans biais de σ^2.

Preuve. Comme $\mathbf{E}[\widehat{a_1}] = a_1$ et que les variables $(\varepsilon_i)_{i\in\mathbb{N}}$ sont centrées,

$$\begin{aligned}
\mathbf{E}\left[(a_1 - \widehat{a_1})^2\right] &= \mathbf{V}(\widehat{a_1}) = \frac{\sigma^2}{ns_{x,x}},\\
\mathbf{E}\left[(\varepsilon_i - \overline{\varepsilon})^2\right] &= \mathbf{V}(\varepsilon_i - \overline{\varepsilon}) = \mathbf{V}\left(\frac{n-1}{n}\varepsilon_i + \frac{1}{n}\sum_{j\neq i}\varepsilon_j\right)\\
&= \frac{(n-1)^2}{n^2}\sigma^2 + \frac{n-1}{n^2}\sigma^2 = \frac{n-1}{n}\sigma^2.
\end{aligned}$$

Finalement, d'après la relation (9) de décomposition de s^2,

$$\begin{aligned}
\mathbf{E}\left[(n-2)s^2\right] &= \mathbf{E}\left[\sum_{i=1}^{n}(\varepsilon_i - \overline{\varepsilon})^2 - ns_{x,x}(a_1 - \widehat{a_1})^2\right]\\
&= (n-1)\sigma^2 - \sigma^2 = (n-2)\sigma^2.
\end{aligned}$$

Ainsi, $\mathbf{E}\left[s^2\right] = \sigma^2$ et s^2 est un estimateur sans biais de σ^2. □

4.2 Calcul de biais (II)

On propose une autre stratégie (voir Zhanxiong 2019) pour calculer le biais de s^2 qui repose sur l'interprétation en termes de projection.

Preuve. En remarquant qu'une norme est un réel, elle est égale à sa trace et on obtient :

$$\begin{aligned}
\mathbf{E}\left[(n-2)s^2\right] &= \mathbf{E}\left[\left\|Y - \widehat{Y}\right\|^2\right] = \mathbf{E}\left[\left\|(I_n - H)\widetilde{Y}\right\|^2\right]\\
&= \mathbf{E}\left[\operatorname{Tr}\left(\widetilde{Y}^T(I_n - H)\widetilde{Y}\right)\right] = \mathbf{E}\left[\operatorname{Tr}\left((I_n - H)\widetilde{Y}\widetilde{Y}^T\right)\right]\\
&= \operatorname{Tr}\left((I_n - H)\mathbf{E}\left[\widetilde{Y}\widetilde{Y}^T\right]\right) = \operatorname{Tr}\left((I_n - H)\sigma^2 I\right)\\
&= \sigma^2 \operatorname{Tr}(I_n - H) = (n-2)\sigma^2.
\end{aligned}$$

□

4.3 Consistance

Proposition 4.2 - Consistance de s^2

L'estimateur s^2 est faiblement consistant et $\mathbf{V}\left(s^2\right) \sim_{n\to+\infty} \frac{\mu_4-\sigma^4}{n}$.

Preuve. Notons $P = (p_{i,j})_{1\leqslant i,j\leqslant n} = I_n - H$ et rappelons que $\widetilde{Y}\widetilde{Y}^T = (\varepsilon_i\varepsilon_j)_{1\leqslant i,\, j\leqslant n}$. Alors,

$$\begin{aligned}
\mathbf{E}\left[(n-2)^2 s^4\right] &= \mathbf{E}\left[\left(\widetilde{Y}^T(I_n-H)\widetilde{Y}\right)^2\right] \\
&= \mathbf{E}\left[\operatorname{Tr}\left(\widetilde{Y}^T(I_n-H)\widetilde{Y}\widetilde{Y}^T(I_n-H)\widetilde{Y}\right)\right] \\
&= \mathbf{E}\left[\operatorname{Tr}\left((I_n-H)\widetilde{Y}\widetilde{Y}^T(I_n-H)\widetilde{Y}\widetilde{Y}^T\right)\right] \\
&= \mathbf{E}\left[\sum_{i=1}^n\sum_{k=1}^n\left[(I_n-H)\widetilde{Y}\widetilde{Y}^T\right]_{i,k}\left[(I_n-H)\widetilde{Y}\widetilde{Y}^T\right]_{k,i}\right] \\
&= \mathbf{E}\left[\sum_{i=1}^n\sum_{k=1}^n\sum_{\ell=1}^n\sum_{m=1}^n p_{i,\ell}\varepsilon_\ell\varepsilon_k p_{k,m}\varepsilon_m\varepsilon_i\right].
\end{aligned}$$

Remarquons que, si un indice est distinct de tous les autres, par indépendance et centrage, alors le terme est nul. Les termes restants sont donc les suivants :

* $i = k = \ell = m$,
* $(i = k) \neq (\ell = m)$,
* $(i = \ell) \neq (k = m)$,
* $(i = m) \neq (k = \ell)$.

Comme $P^T = P$, alors $p_{i,\ell} = p_{\ell,i}$ et

$$\begin{aligned}
\mathbf{E}\left[(n-2)^2 s^4\right] &= \sum_{i=1}^n p_{i,i}^2 \mathbf{E}\left[\varepsilon_i^4\right] + \sum_{i=1}^n \sum_{\ell \neq i} p_{i,\ell}^2 \mathbf{E}\left[\varepsilon_i^2\right] \mathbf{E}\left[\varepsilon_\ell^2\right] + \cdots \\
&\qquad \cdots + \sum_{i=1}^n \sum_{k \neq i} p_{i,i} p_{k,k} \mathbf{E}\left[\varepsilon_i^2\right] \mathbf{E}\left[\varepsilon_k^2\right] + \cdots \\
&\qquad \cdots + \sum_{i=1}^n \sum_{\ell \neq i} p_{i,\ell} p_{\ell,i} \mathbf{E}\left[\varepsilon_i^2\right] \mathbf{E}\left[\varepsilon_\ell^2\right] \\
&= \mu_4 \sum_{i=1}^n p_{i,i}^2 + \sigma^4 \sum_{i=1}^n \sum_{\ell \neq i} \left(p_{i,\ell}^2 + p_{i,i} p_{\ell,\ell} + p_{i,\ell} p_{\ell,i}\right) \\
&= \mu_4 \sum_{i=1}^n p_{i,i}^2 + \sigma^4 \sum_{i=1}^n \left(\sum_{\ell \neq i} \left(2 p_{i,\ell} p_{\ell,i} + p_{i,i} p_{\ell,\ell}\right) \right) \\
&= \left(\mu_4 - 3\sigma^4\right) \sum_{i=1}^n p_{i,i}^2 + \sigma^4 \sum_{i=1}^n \sum_{\ell=1}^n \left(2 p_{i,\ell} p_{\ell,i} + p_{i,i} p_{\ell,\ell}\right) \\
&= \left(\mu_4 - 3\sigma^4\right) \sum_{i=1}^n p_{i,i}^2 + \sigma^4 \left(2 \operatorname{Tr}(P^2) + \operatorname{Tr}(P)^2\right).
\end{aligned}$$

Comme P est un projecteur, $\operatorname{Tr}(P^2) = \operatorname{Tr}(P) = \operatorname{Tr}(I_n - H) = n - 2$ et

$$\mathbf{E}\left[(n-2)^2 s^4\right] = \left(\mu_4 - 3\sigma^4\right) \sum_{i=1}^n p_{i,i}^2 + n(n-2)\sigma^4.$$

Ainsi,

$$\begin{aligned}
\mathbf{V}\left(s^2\right) &= \frac{\mu_4 - 3\sigma^4}{(n-2)^2} \sum_{i=1}^n p_{i,i}^2 + \frac{n}{n-2}\sigma^4 - \sigma^4 \\
&= \frac{\mu_4 - 3\sigma^4}{(n-2)^2} \sum_{i=1}^n p_{i,i}^2 + \frac{2}{n-2}\sigma^4
\end{aligned}$$

Rappelons enfin que $P = I - X(X^TX)^{-1}X^T$. Ainsi,

$$\begin{aligned}
X^TX &= n\begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{x^2} \end{pmatrix} \\
\left(X^TX\right)^{-1} &= \frac{1}{n(\overline{x^2} - \overline{x}^2)}\begin{pmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix} \\
\left(X^TX\right)^{-1}X^T &= \frac{1}{ns_{x,x}}\begin{pmatrix} \overline{x^2} - x_1\overline{x} & \cdots & \overline{x^2} - x_n\overline{x} \\ -\overline{x} + x_1 & \cdots & -\overline{x} + x_n \end{pmatrix} \\
p_{i,i} &= 1 - \frac{1}{ns_{x,x}}\left(\overline{x^2} - x_i\overline{x} - x_i\overline{x} + x_i^2\right) \\
p_{i,i}^2 &= 1 - 2\frac{\overline{x^2} - 2x_i\overline{x} + x_i^2}{ns_{x,x}} + \frac{(\overline{x^2} - 2x_i\overline{x} + x_i^2)^2}{n^2s_{x,x}^2} \\
\sum_{i=1}^n p_{i,i}^2 &= n - 4 + \frac{\sum_{i=1}^n(\overline{x^2} - 2x_i\overline{x} + x_i^2)^2}{n^2s_{x,x}^2} \\
&= n - 4 + \frac{3(\overline{x^2})^2 - 4\overline{x^3}\cdot\overline{x} + \overline{x^4}}{ns_{x,x}^2}.
\end{aligned}$$

Ainsi,

$$\begin{aligned}
\mathbf{V}\left(s^2\right) &= \frac{\mu_4 - 3\sigma^4}{n-2} + \frac{\mu_4 - 3\sigma^4}{(n-2)^2}\left(\frac{3(\overline{x^2})^2 - 4\overline{x^3}\cdot\overline{x} + \overline{x^4}}{ns_{x,x}^2} - 2\right) + \frac{2}{n-2}\sigma^4 \quad (11) \\
&= \frac{\mu_4 - \sigma^4}{n-2} + \frac{\mu_4 - 3\sigma^4}{(n-2)^2}\left(\frac{3(\overline{x^2})^2 - 4\overline{x^3}\cdot\overline{x} + \overline{x^4}}{ns_{x,x}^2} - 2\right). \quad (12)
\end{aligned}$$

Comme les quantités $\overline{x}$, $\overline{x^2}$, $\overline{x^3}$ et $\overline{x^4}$ sont bornées, alors $\mathbf{V}\left(s^2\right) \to 0$ et l'estimateur s^2 converge en probabilité vers σ^2. □

> **Remarque 4.1 - D**ans le cas où les variables aléatoires sont gaussiennes, alors $\mu_4 = 3\sigma^4$. On retrouve bien la variance obtenue dans le cadre de la loi $\chi^2(n-2)$ (voir Section 2).

Proposition 4.3 - Forte consistance de s^2

L'estimateur s^2 est fortement consistant.

Preuve. Nous allons pour cela utiliser la relation (9) :

$$s^2 = \frac{1}{n-2}\left(\sum_{i=1}^n(\varepsilon_i - \overline{\varepsilon})^2 - ns_{x,x}(a_1 - \widehat{a}_1)^2\right).$$

D'après la consistance de l'estimateur $\widehat{a_1}$,

$$\lim_{n\to+\infty}(a_1-\widehat{a_1})=0 \text{ p.s.}$$

De plus,

$$\frac{1}{n-2}\sum_{i=1}^{n}(\varepsilon_i-\overline{\varepsilon})^2=\frac{1}{n-2}\sum_{i=1}^{n}\varepsilon_i^2-\frac{1}{n(n-2)}\left(\sum_{i=1}^{n}\varepsilon_i\right)^2.$$

D'après la loi forte des grands nombres,

$$\lim_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i=0 \text{ p.s. et } \lim_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2=\sigma^2 \text{ p.s.}$$

Comme $s_{x,x}$ est bornée, en additionnant les convergences presque sûres, $\lim\limits_{n\to+\infty} s^2=\sigma^2$ p.s. □

Proposition 4.4 - Normalité asymptotique de s^2

Il y a convergence en loi :

$$\sqrt{\frac{n}{\mu_4-\sigma^4}}(s^2-\sigma^2)\longrightarrow\mathscr{N}(0,1).$$

Preuve. Rappelons que d'après la relation (9) de décomposition de s^2,

$$s^2=\frac{1}{n-2}\left(\sum_{i=1}^{n}(\varepsilon_i-\overline{\varepsilon}^2)-ns_{x,x}(a_1-\widehat{a_1})^2\right).$$

Ainsi,

$$(n-2)s^2=\sum_{i=1}^{n}\varepsilon_i^2-n\overline{\varepsilon}^2-ns_{x,x}(a_1-\widehat{a_1})^2$$

$$\sqrt{n}(s^2-\sigma^2)=\frac{\sqrt{n}}{n-2}\sum_{i=1}^{n}(\varepsilon_i^2-\sigma^2)-\frac{n\sqrt{n}}{n-2}\overline{\varepsilon}^2-\frac{n\sqrt{n}s_{x,x}}{n-2}(a_1-\widehat{a_1})^2+\frac{2\sqrt{n}}{n-2}\sigma^2.$$

Étudions la convergence de chacun de ces termes.

* Comme σ^2 est constante, $\frac{2\sqrt{n}}{n-2}\sigma^2\longrightarrow 0$ presque sûrement, donc en probabilité.
* D'après le théorème de la limite centrale,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i\longrightarrow\mathscr{N}(0,1).$$

Ainsi, d'après le théorème de Slutsky B.1,

$$\frac{n\sqrt{n}}{n-2}\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\right)^2 = \frac{\sqrt{n}}{n-2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i\right)^2$$
$$\longrightarrow 0 \text{ en probabilité}$$

* D'après la normalité asymptotique de $\widehat{a_1}$ (voir Proposition 3.7),

$$\frac{\sqrt{n s_{x,x}}}{\sigma}(\widehat{a_1} - a_1) \longrightarrow \mathcal{N}(0,1).$$

Ainsi, d'après le théorème de Slutsky B.1 et le caractère borné de $s_{x,x}$,

$$\frac{n\sqrt{n}}{n-2} s_{x,x}(a_1 - \widehat{a_1})^2 = \frac{\sqrt{n}}{n-2} s_{x,x}\left(\sqrt{n}(a_1 - \widehat{a_1})\right)$$
$$\longrightarrow 0 \text{ en probabilité}$$

* En posant $\widetilde{\varepsilon_i} = \varepsilon_i^2 - \sigma^2$, alors $\sum\limits_{i=1}^{n}\widetilde{\varepsilon_i}$ est une somme de variables aléatoires indépendantes, centrées et de variance

$$\mathbf{V}\left(\widetilde{\varepsilon_i}\right) = \mathbf{E}\left[(\varepsilon_i^2 - \sigma^2)^2\right] = \mathbf{E}\left[\varepsilon_i^4\right] - 2(\sigma^2)^2 + \sigma^4 = \mu_4 - \sigma^4.$$

Ainsi, d'après le théorème central limite,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{\varepsilon_i} \longrightarrow \mathcal{N}(0, \mu_4 - \sigma^4).$$

Finalement, le premier terme converge en loi vers une loi normale et les autres convergent en probabilité vers 0 donc, d'après le théorème de Slutsky B.1,

$$\sqrt{n}(s^2 - \sigma^2) \longrightarrow \mathcal{N}(0, \mu_4 - \sigma^4).$$

□

A Estimateurs

Définition A.1 - Estimateurs

Soit θ_n un estimateur de θ.

* θ_n est un estimateur *sans biais* si $\mathbf{E}\left[\theta_n\right] = \theta$.
* θ_n est un estimateur *consistant* si (θ_n) converge en probabilités vers θ, i.e.

$$\forall\ \varepsilon > 0,\ \lim_{n\to+\infty} \mathbf{P}\left(|\theta_n - \theta| \geqslant \varepsilon\right) = 0.$$

* θ_n est un estimateur *fortement consistant* si (θ_n) converge presque

sûrement vers θ, i.e.

$$\mathbf{P}\left(\lim_{n\to+\infty}\theta_n=\theta\right)=1.$$

Proposition A.1 - Consistance & Variance

Si θ_n est sans biais, possède un moment d'ordre 2 et $\mathbf{V}(\theta_n)\longrightarrow 0$, alors θ est un estimateur consistant.

Preuve. En effet, d'après l'inégalité de Bienaymé-Tchebychev,

$$0\leqslant\mathbf{P}\left(|\theta_n-\theta|\geqslant\varepsilon\right)\leqslant\frac{\mathbf{V}(\theta_n)}{\varepsilon^2}.$$

On conclut à l'aide du théorème d'encadrement. □

Proposition A.2 - Normalité asymptotique

Si $\frac{1}{s_n}(\theta_n-\theta)$ converge en loi vers une loi normale, alors on peut construire un intervalle de confiance pour la valeur de θ.

Preuve. Notons Φ la fonction de répartition de la loi normale centrée réduite. Comme $\left(\frac{\theta-\theta_n}{s_n}\right)$ converge en loi vers la loi normale centrée réduite, alors pour tout $\alpha<\beta$,

$$\lim_{n\to+\infty}\mathbf{P}\left(\frac{\theta-\theta_n}{s_n}\in[\alpha,\beta]\right)=\Phi(\beta)-\Phi(\alpha)$$
$$\lim_{n\to+\infty}\mathbf{P}\left(\theta\in[\theta_n+\alpha s_n,\theta_n+\beta s_n]\right)=\Phi(\beta)-\Phi(\alpha).$$

□

B Théorèmes probabilistes

Théorème B.1 - Théorème de Slutsky (cf. Athreya et Lahiri 2006 p. 290)

Soit a un réel et $(X_n)_{n\in\mathbb{N}}$, $(Y_n)_{n\in\mathbb{N}}$ deux suites de variables aléatoires telles que $X_n\longrightarrow X$ en loi et $Y_n\longrightarrow a$ en probabilité. Alors,

(*i*) $X_n+Y_n\longrightarrow X+a$ en loi,

(*ii*) $X_nY_n\longrightarrow aX$ en loi,

Théorème B.2 - Critère de Kolmogorov (cf. Athreya et Lahiri 2006 p. 259)

Soit $(X_i)_{i\in\mathbb{N}}$ une suite de variables aléatoires indépendantes telles que $\mathbf{E}[X_n] = 0$ et $\sum \frac{\mathbf{V}(X_n)}{n^2}$ converge. Alors, $\frac{1}{n}(X_1 + \cdots + X_n)$ converge presque sûrement vers 0.

Théorème B.3 - Condition de Lyapunov (cf. Athreya et Lahiri 2006 p. 348)

Pour tout $n \in \mathbb{N}$, on suppose que $(X_{n,j})_{1\leqslant j\leqslant r_n}$ est une suite de variables aléatoires indépendantes. On suppose que les $X_{n,j}$ sont centrées, possèdent un moment d'ordre 2 et on note $\sigma_{n,j}^2 = \mathbf{V}(X_{n,j})$. On pose $s_n^2 = \sum\limits_{j=1}^{r_n} \sigma_{n,j}^2$.
S'il existe $\delta > 0$ tel que

$$\lim_{n\to+\infty} \frac{1}{s_n^{2+\delta}} \sum_{j=1}^{r_n} \mathbf{E}\left[|X_{n,j}|^{2+\delta}\right] = 0,$$

alors, il y a la convergence en loi

$$\frac{1}{s_n} \sum_{j=1}^{r_n} X_{n,j} \longrightarrow \mathscr{N}(0,1).$$

Théorème B.4 - Théorème de Cochran

Soit $X = (X_1, \ldots, X_n)$ un vecteur gaussien de loi $\mathscr{N}(\mu, \sigma^2 I_n)$ et F un sous-espace vectoriel de $\mathbb{R}^n$ de dimension d. On note p_F (resp. $p_{F^\perp}$) la projection orthogonale sur F (resp. $F^\perp$).

* Les variables aléatoires $p_F(X)$ et $p_{F^\perp}(X)$ sont indépendantes et suivent des lois normales.
* Les variables aléatoires $\frac{\|p_F(X-\mu)\|^2}{\sigma^2}$ et $\frac{\left\|p_{F^\perp}(X-\mu)\right\|^2}{\sigma^2}$ sont indépendantes et de lois respectives $\chi^2(d)$ et $\chi^2(n-d)$.

C Moindres carrés

Soient $n \in \mathbb{N}^\star$ et $((x_i, y_i))_{1\leqslant i\leqslant n} \in (\mathbb{R}^2)^n$. On suppose que les réels $x_1, \ldots, x_n$ ne sont pas tous égaux à une même valeur.

On rappelle que

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \widetilde{Y} = Y - \mathbf{E}\left[Y\right], X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

$$H = X\left(X^T X\right)^{-1} X^T.$$

On cherche les réels λ et μ qui minimisent la fonction

$$f : (\lambda, \mu) \mapsto \sum_{i=1}^{n} (\lambda x_i + \mu - y_i)^2.$$

En notant $h = \begin{pmatrix} \mu \\ \lambda \end{pmatrix}$, on remarque que

$$f(h) = \|Xh - Y\|^2 = h^T X^T X h - h^T X^T Y - Y^T X h + Y^T Y.$$

Ainsi,

$$\nabla f(h) = 2X^T X h - 2X^T Y.$$

On remarque que la hessienne de f est

$$X^T X = n \begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{x^2} \end{pmatrix}.$$

L'inégalité de Cauchy-Schwarz assure que $X^T X$ est inversible et de déterminant strictement positif. La fonction f est ainsi convexe.

Ainsi, l'unique point critique de f est le point satisfaisant $\nabla f(\widehat{h}) = 0$ soit

$$\begin{aligned} \widehat{h} = \begin{pmatrix} \widehat{a_0} \\ \widehat{a_1} \end{pmatrix} &= \left(X^T X\right)^{-1} X^T Y \\ &= \frac{1}{n(\overline{x^2} - (\overline{x})^2)} \begin{pmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix} \begin{pmatrix} n\overline{y} \\ n\overline{xy} \end{pmatrix} \\ &= \frac{1}{\overline{x^2} - (\overline{x})^2} \begin{pmatrix} \overline{x^2} \cdot \overline{y} - \overline{x} \cdot \overline{xy} \\ \overline{xy} - \overline{x} \cdot \overline{y} \end{pmatrix} = \begin{pmatrix} \overline{y} - \frac{s_{x,y}}{s_{x,x}} \overline{x} \\ \frac{s_{x,y}}{s_{x,x}} \end{pmatrix} \end{aligned}$$

Comme f est convexe, elle atteint un minimum en $\widehat{h}$ et

$$\widehat{Y} = X\widehat{h} = X\left(X^T X\right)^{-1} X^T Y = HY.$$

Proposition C.1 - Projecteur orthogonal

La matrice H est la matrice d'un projecteur orthogonal et $\mathrm{Tr}(H) = 2$.

Preuve. On remarque que $H^T = H$ et $H^2 = H$. De plus,

$$\begin{aligned}\mathrm{Tr}(H) &= \mathrm{Tr}\left(X(X^TX)^{-1}X^T\right) \\ &= \mathrm{Tr}\left((X^TX)^{-1}X^TX\right) = \mathrm{Tr}(I_2) = 2.\end{aligned}$$

□

Alors,

$$\begin{aligned}\widehat{Y} &= HY \\ (Y - \widehat{Y}) &= (I_n - H)Y.\end{aligned}$$

Remarquons enfin que

$$\begin{aligned}&\mathbf{E}\left[y_i\right] = a_1x_i + a_0, \\ &\mathbf{E}\left[\widehat{y_i}\right] = \mathbf{E}\left[\widehat{a_1}\right]x_i + \mathbf{E}\left[\widehat{a_0}\right] = a_1x_i + a_0,\end{aligned}$$

soit $\mathbf{E}\left[Y\right] = \mathbf{E}\left[\widehat{Y}\right]$. En remarquant que $\widetilde{Y} = \begin{pmatrix}\varepsilon_1 \\ \vdots \\ \varepsilon_n\end{pmatrix}$, alors $\widetilde{Y}\widetilde{Y}^T = (\varepsilon_i\varepsilon_j)_{1\leqslant i,j\leqslant n}$ et

$$\begin{aligned}Y - \widehat{Y} &= Y - \mathbf{E}\left[Y\right] - (\widehat{Y} - \mathbf{E}\left[\widehat{Y}\right]) \\ &= \widetilde{Y} - H(Y - \mathbf{E}\left[Y\right]) \\ &= (I_n - H)\widetilde{Y}.\end{aligned}$$

Références

Athreya, K.B. et S.N. Lahiri (2006). *Measure Theory and Probability Theory*. Springer Texts in Statistics. Springer. isbn : 9780387329031.

Zhanxiong (2019). *StackExchange - Mathematics - Proof $E[\hat{\sigma^2}] = E\left(\frac{1}{n-2}\sum(y_i - y_i^2)\right) = \sigma^2$: Linear Regression*. url : https://math.stackexchange.com/questions/3100557/proof-e-hat-sigma-2-e-left-frac1n-2-sigmay-i-haty-i2-right (visité le 19/02/2024).