Série 2 STT5100

## Exercice 1

Observations : $(x_i, Y_i)_{i=1,\dots,n}$

Modèle : $Y_i = \alpha + x_i\beta + \varepsilon_i$ $\begin{cases} E[\varepsilon_i] = 0 \\ Var[\varepsilon_i] = \sigma^2 \\ Cov(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j \end{cases}$

Question : Meilleur estimateur linéaire non biaisé de $\alpha$

Étapes : Soit $\hat{\alpha}$ un tel estimateur

① Estimateur linéaire $\rightarrow$ $\hat{\alpha} = \sum_{i=1}^{n} a_i Y_i$

② Estimateur non biaisé en $\alpha$

$$\mathbb{E}[\hat{\alpha}] = \alpha$$

③ Meilleur estimateur $\rightarrow$ estimateur de variance minimale

(De plus petite variance parmi tous les estimateurs linéaires en $\alpha$).

Subtilités : Pour tout vecteur $w \in \mathbb{R}^n$,

$$\sum_{i=1}^{n} (w_i - \bar{w}) = 0.$$

En effet,

$$\sum (w_i - \bar{w}) = \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} \bar{w} = n\bar{w} - n\bar{w} = 0$$

Solution :

(2) $\mathbb{E}[\hat{\alpha}] = \alpha \rightarrow \mathbb{E}[\sum a_i Y_i] = \alpha$

$\rightarrow \sum_i a_i(\alpha + x_i\beta) = \alpha$ car $\mathbb{E}[\sum a_i Y_i] = \sum_i a_i \mathbb{E}[Y_i]$.

$\rightarrow \alpha \sum_i a_i + \beta \sum_i a_i x_i = \alpha$

Cette égalité n'est vérifiée que si

$$\begin{cases} \sum_i a_i = 1 \\ \sum a_i x_i = 0 \end{cases}$$

(3) $Var[\hat{\alpha}] = Var[\sum_i a_i Y_i] = \sum_i a_i^2 Var[Y_i]$

$Var[\hat{\alpha}] = \sigma^2 \sum_i a_i^2$

$\hat{\alpha}$ est de variance minimale et donc est solution du problème de minimisation

$$\min Var[\hat{\sigma}]$$
s.c.
$$\sum a_i = 1,$$
$$\sum a_i x_i = 0$$

$$\Updownarrow$$

$$\hat{\alpha} = \operatorname{argmin} \sum_i a_i^2$$
s.c. $\sum a_i = 1$,
$\sum a_i x_i = 0$

La résolution de ce problème d'optimisation par la méthode du Lagrangien (KKT) est fastidieuse. Nous proposons ci-dessous une autre méthode de résolution.

$\sum_{i=1}^{n} a_i = 1 \Rightarrow \exists\, k, \{b_i, i=1,\dots,n\}$ constantes tels que

$$a_i = \frac{1}{n} + k(b_i - \bar{b})$$

(On vérifie aisément que
$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} \left\{\frac{1}{n} + k(b_i - \bar{b})\right\} = 1)$$

* Cherchons de tels $k$ et $\{b_i, i=1,\dots,n\}$

$$\sum_i a_i x_i = 0 \rightarrow \sum_{i=1}^{n} \left\{\frac{1}{n} + k(b_i - \bar{b})\right\} x_i = 0$$

$$\rightarrow \frac{1}{n}\sum_i x_i + k\sum_i (b_i - \bar{b}) x_i = 0$$

$$\rightarrow \frac{1}{n}\sum_i x_i + k\sum_i (b_i - \bar{b})(x_i - \bar{x}) = 0$$

$$\left(\sum_{i=1}^{n}(b_i - \bar{b})(x_i - \bar{x}) = \sum_{i=1}^{n}(b_i - \bar{b}) x_i\right)$$

$$\rightarrow \bar{x} + k\sum_i (b_i - \bar{b})(x_i - \bar{x}) = 0$$

d'où

$$k = -\frac{\bar{x}}{\sum_{i=1}^{n}(b_i - \bar{b})(x_i - \bar{x})}$$

$$a_i = \frac{1}{n} + k(b_i - \bar{b}) \rightarrow a_i = \frac{1}{n} - \frac{\bar{x}(b_i - \bar{b})}{\sum_i (b_i - \bar{b})(x_i - \bar{x})}$$

d'où

$$\sum_{i=1}^{n} a_i^2 = \sum_i \left\{ \frac{1}{n} - \frac{\bar{x}(b_i - \bar{b})}{\sum (b_i - \bar{b})(x_i - \bar{x})} \right\}^2$$

$$= \sum_i \left\{ \frac{1}{n^2} + \frac{\bar{x}^2 (b_i - \bar{b})^2}{\left[\sum (b_i - \bar{b})(x_i - \bar{x})\right]^2} - \frac{2}{n} \frac{\bar{x}(b_i - \bar{b})}{\sum_i (b_i - \bar{b})(x_i - \bar{x})} \right\}$$

$$= \sum_i \frac{1}{n^2} + \frac{\bar{x}^2 (b_i - \bar{b})^2}{\left[\sum (b_i - \bar{b})(x_i - \bar{x})\right]^2} - \frac{2}{n} \frac{\bar{x} \overbrace{\sum (b_i - \bar{b})}^{0}}{\sum_i (b_i - \bar{b})(x_i - \bar{x})}$$

$$= \frac{1}{n} + \bar{x}^2 \frac{\sum_{i=1}^{n} (b_i - \bar{b})^2}{\left[\sum_{i=1}^{n} (b_i - \bar{b})(x_i - \bar{x})\right]^2}$$

Minimiser $\sum a_i^2$ revient donc à minimiser

$$\bar{x}^2 \frac{\sum_{i=1}^{n} (b_i - \bar{b})^2}{\left[\sum_{i=1}^{n} (b_i - \bar{b})(x_i - \bar{x})\right]^2}$$

De l'inégalité de Cauchy-Swarz, on sait que

$$\sum_i \left[(b_i - \bar{b})(x_i - \bar{x})\right]^2 \leq \sum_i (b_i - \bar{b})^2 \sum_i (x_i - \bar{x})^2$$

Scanned by CamScanner

$$\frac{\sum_i (b_i - \bar{b})^2}{\left[\sum_i (b_i - \bar{b})(x_i - \bar{x})\right]^2} \geq \frac{\sum_i (b_i - \bar{b})^2}{\sum_i (b_i - \bar{b})^2 \sum_i (x_i - \bar{x})^2}$$

Il en suit que

$$\sum_i a_i^2 \geq \frac{\bar{x}^2}{\sum_i (x_i - \bar{x})^2} := A$$

Le minimum A est atteint à $b_i = x_i$

En effet, remplacer $b_i$ par $x_i$ dans $\sum a_i^2$.
~~et utiliser~~ On a alors

$$\frac{\sum (b_i - \bar{b})^2}{\left(\sum (b_i - \bar{b})(x_i - \bar{x})\right)^2} = \frac{\sum_i (x_i - \bar{x})^2}{\left(\sum_i (x_i - \bar{x})(x_i - \bar{x})\right)^2}$$

$$= \frac{\sum_i (x_i - \bar{x})^2}{\left(\sum_i (x_i - \bar{x})^2\right)^2}$$

$$= \frac{1}{\sum_i (x_i - \bar{x})^2}$$

On a donc

$$\begin{cases} b_i = x_i \\ b_a = - \dfrac{\bar{x}}{\sum_i (b_i - \bar{b})(x_i - \bar{x})} \end{cases}$$

Scanned by CamScanner

Remplaçons $b_i = x_i$ et obtenons la valeur minimale dans $a_i$, c'est-à-dire

$$a_i = \frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

et d'où

$$\hat{\alpha} = \sum_i a_i Y_i$$

$$= \sum_i \left\{ \frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \right\} Y_i$$

$$= \frac{1}{n} \sum_i Y_i - \sum_i \frac{(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2} Y_i$$

$$= \bar{Y} - \hat{\beta}\bar{x}, \text{ car } \left( \hat{\beta} = \frac{\sum_i (x_i - \bar{x}) Y_i}{\sum_i (x_i - \bar{x})^2} = \frac{\sum_i (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_i (x_i - \bar{x})^2} \right.$$

Conclusion :

$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}$ est l'estimateur des moindres carrés.

# Exercice 2

Observations : $(x_i, Y_i)_{i=1,\dots,n}$

Modèle : $Y_i = \alpha + x_i \beta + \varepsilon_i$

$x_1, \dots, x_n$ fixés, constants

$\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, les $\varepsilon_i$ i.i.d.

Modèle reparamétrisé :

$$Y_i^* = \alpha^* + (x_i - \bar{x})\beta^* + \varepsilon_i$$

On veut montrer que

(i) $\hat{\beta} = \hat{\beta}^*$

(ii) $\hat{\alpha} \neq \hat{\alpha}^*$, $\hat{\alpha} = \bar{Y}$, distribution de $\hat{\alpha}^*$

(iii) $\hat{\alpha}^*$ et $\hat{\beta}^*$ sont non corrélés et de ce fait indépendants sous la normalité.

Solution :

(i) Posons $z_i = x_i - \bar{x}$ pour tout $i = 1, \dots, n$

Alors $\bar{z} = \frac{1}{n}\sum_i z_i = \frac{1}{n}\sum_i (x_i - \bar{x}) = 0$

et on a :

$$\hat{\beta} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{\sum_i z_i (y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$$

$$= \frac{\sum_i (z_i - \bar{z})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \hat{\beta}^* \quad (\text{car } \bar{z} = 0)$$

Scanned by CamScanner

(ii) $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = \bar{y} - \hat{\beta}^*\bar{x}$ (car $\hat{\beta} = \hat{\beta}^*$)

$\hat{\alpha}^* = \bar{y} - \hat{\beta}^*\bar{z} = \bar{y}$ (car $\bar{z} = 0$)

On a clairement

$\hat{\alpha} \neq \hat{\alpha}^*$ puisque $\hat{\beta}^* \neq 0$

* On vient de montrer que

$\hat{\alpha}^* = \bar{y}$ et $\hat{\alpha} \neq \hat{\alpha}^*$

* Distribution de $\hat{\alpha}^*$

$\hat{\alpha} = \bar{y} - \hat{\beta}^*\bar{x} = \hat{\alpha}^* - \hat{\beta}^*\bar{x}$ car $\hat{\alpha}^* = \bar{y}$

d'où

$$\hat{\alpha}^* = \hat{\alpha} + \hat{\beta}^*\bar{x}.$$

$$E[\hat{\alpha}^*] = E[\hat{\alpha} + \hat{\beta}^*\bar{x}]$$

$$= \alpha + \beta\bar{x}$$

$$Var[\hat{\alpha}^*] = Var[\bar{y}] = \frac{1}{n^2}\sum_i \cancel{\text{Var}}\, Var(y_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Donc

$$\hat{\alpha}^* \sim \mathcal{N}(\alpha + \beta\bar{x}, \sigma^2/n)$$

(iii) ~~$\hat{\alpha}^*$ est linéaire en $\hat{\alpha}$~~

$Cov(\hat{\alpha}^*, \hat{\beta}^*) = Cov(\hat{\alpha}^*, \hat{\beta})$ car $\hat{\beta}^* = \hat{\beta}$

et $\hat{\alpha}^*$ est linéaire en $\hat{\beta}$

D'où

$$\mathrm{Cov}(\hat{\alpha}^*, \hat{\beta}^*) = \mathrm{Cov}(\hat{\alpha} + \hat{\beta}\bar{x}, \hat{\beta})$$

$$= \mathrm{Cov}(\hat{\alpha}, \hat{\beta}) + \bar{x}\,\mathrm{Cov}(\hat{\beta}, \hat{\beta})$$

$$= \mathrm{Cov}(\hat{\alpha}, \hat{\beta})$$

$$= -\sigma^2 \sum_i \frac{1}{n}\left(\frac{z_i}{\sum_i z_i^2}\right)$$

$$= 0 \quad \text{car } \bar{z} = 0$$

Question 3 :

$Y_i = x_i^2 \theta + \varepsilon_i$ , $x_1, \dots, x_n$ fixés, constants

$\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ , i.i.d. pour tout $i$

(i) Estimateur des moindres carré de $\theta$, $\hat{\theta}$.

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n} \varepsilon_i^2$$

Rappel : Comment résoudre le problème de minimisation

$$\min_x f(x) \quad ?$$

$\rightarrow$ Résoudre en $x$, $f'(x) = 0$.

La solution de cette équation, notons la $\hat{x}$, pour être un minimum doit vérifier :

$\rightarrow$ $f''(\hat{x}) \geq 0$.

Posons

$$f(\theta) = \sum_i \varepsilon_i^2 = \sum_i \left(Y_i - x_i^2 \theta\right)^2.$$

$$f'(\theta) = -2 \sum_i \left(Y_i - x_i^2 \theta\right) x_i^2$$

$$f'(\theta) = 0 \Rightarrow \sum_i \left(Y_i - x_i^2 \theta\right) x_i^2 = 0$$

$$\Rightarrow \sum_i x_i^2 Y_i - \theta \sum_i x_i^4 = 0$$

d'où $\hat{\theta}_{LS} = \sum_{i=1}^{n} x_i^2 Y_i \Big/ \sum_{i=1}^{n} x_i^4$

L.S. = Least Square

(ii) Estimateur des moindres carrés $\hat{\theta}_{MLE}$ de $\theta$.

Vraisemblance:

$$L(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ \sum_i \frac{(y_i - \theta x_i^2)^2}{2\sigma^2} \right\}$$

Log-vraisemblance

$$\log(L(\theta)) = \ell(\theta) = -\log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_i (y_i - \theta x_i^2)^2$$

$$\ell'(\theta) = -\frac{2}{2\sigma^2} \sum_i (y_i - \theta x_i^2) x_i^2$$

$$\ell'(\theta) = 0 \Rightarrow \sum_i (y_i - \theta x_i^2) x_i^2 = 0$$

$$\Rightarrow \boxed{\hat{\theta}_{MLE} = \sum_{i=1}^{n} x_i^2 y_i \Big/ \sum_{i=1}^{n} x_i^4}$$

(iii) Meilleur estimateur sans biais de $\theta$

$$\ell''(\theta) = -\frac{1}{\sigma^2} \sum_i x_i^4$$

La borne de Cramer-Rao est :

$$BCR = -E\left[\frac{d^2}{d\theta^2}\ell(\theta)\right]$$

$$= +\frac{1}{\sigma^2}\sum_{i=1}^{n} x_i^4$$

Proposition :

$$\left(Var[\hat{\theta}] = BCR\right) \Rightarrow \left(\hat{\theta} \text{ est le meilleur estimateur de } \theta\right).$$

- On vérifie aisément que
  $$E[\hat{\theta}] = \theta \qquad \left(\hat{\theta} = \hat{\theta}_{LS} = \hat{\theta}_{MLE}\right)$$

- $$Var[\hat{\theta}] = Var\left[\sum_i y_i x_i^2 \Big/ \sum_i x_i^4\right].$$

$$= \sum_i x_i^4 \, Var[y_i] \Big/ \left(\sum_i x_i^4\right)^2.$$

$$= \sigma^2 \sum_i x_i^4 \Big/ \left(\sum_i x_i^4\right)^2.$$

$$= \sigma^2 \Big/ \sum_i x_i^4$$

$$= BCR$$

Donc $\hat{\theta}$ est le meilleur estimateur sans biais de $\theta$.