ULB

Section 4 : Estimation de la variance

STAT-F-406

Master en sciences mathématiques, Master en statistique

ACTU-F4001

Master en sciences actuarielles

Davy Paindaveine

Université libre de Bruxelles

2023–2024

Motivation

On a vu que, dans le modèle semi-fort homoscédastique,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\mathcal{D}} \mathcal{N}_k\big(0, \sigma^2(\mathrm{E}[XX'])^{-1}\big).$$

Donc pour tout $j = 1, \ldots, k$,

$$\frac{\sqrt{n}(\hat{\beta}_j - \beta_j)}{\sigma\sqrt{((\mathrm{E}[XX'])^{-1})_{jj}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

de sorte que, pour tout $\alpha \in]0,1[$,

$$P\left[\hat{\beta}_j - \frac{z_{\alpha/2}}{\sqrt{n}}\sigma\sqrt{((\mathrm{E}[XX'])^{-1})_{jj}} \leq \beta_j \leq \hat{\beta}_j + \frac{z_{\alpha/2}}{\sqrt{n}}\sigma\sqrt{((\mathrm{E}[XX'])^{-1})_{jj}}\right] \to 1-\alpha$$

quand $n \to \infty$. Nous avons presque un intervalle de confiance pour β_j au niveau de confiance asymptotique $1 - \alpha$.

Motivation

Bien entendu, $Q = \frac{1}{n}\sum_{i=1}^{n} X_i X_i' \overset{\text{p.s.}}{\to} \mathrm{E}[XX']$ (LFGN). Si $\hat{\sigma}^2$ est un estimateur (au moins faiblement) convergent de σ^2, alors on aura

$$\frac{\sqrt{n}(\hat{\beta}_j - \beta_j)}{\hat{\sigma}\sqrt{(Q^{-1})_{jj}}} = \frac{\sigma\sqrt{((\mathrm{E}[XX'])^{-1})_{jj}}}{\hat{\sigma}\sqrt{(Q^{-1})_{jj}}} \times \frac{\sqrt{n}(\hat{\beta}_j - \beta_j)}{\sigma\sqrt{((\mathrm{E}[XX'])^{-1})_{jj}}} \overset{\mathcal{D}}{\to} \mathcal{N}(0,1),$$

de sorte que, pour tout $\alpha \in]0,1[$,

$$P\left[\hat{\beta}_j - \frac{z_{\alpha/2}}{\sqrt{n}}\hat{\sigma}\sqrt{(Q)^{-1})_{jj}} \leq \beta_j \leq \hat{\beta}_j + \frac{z_{\alpha/2}}{\sqrt{n}}\hat{\sigma}\sqrt{(Q)^{-1})_{jj}}\right] \to 1-\alpha$$

quand $n \to \infty$. Ceci fournit un vrai intervalle de confiance pour β_j au niveau de confiance asymptotique $1-\alpha$.

Estimation de la variance

Comment estimer σ^2 (dans le modèle homoscédastique)?

Puisque $\sigma^2 = \mathrm{Var}[\varepsilon|X] = \mathrm{E}[\varepsilon^2|X] - (\mathrm{E}[\varepsilon|X])^2 = \mathrm{E}[\varepsilon^2|X]$, on a

$$\sigma^2 = \mathrm{E}[\varepsilon^2],$$

de sorte que

$$\sigma^2 \approx \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \beta' X_i)^2 \approx \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\beta}' X_i)^2 \stackrel{\text{def}}{=} \hat{\sigma}^2.$$

Notons que la LFGN ne permet pas d'affirmer que $\hat{\sigma}^2 \stackrel{\text{p.s.}}{\to} \sigma^2$ (pourquoi?)

Remarque: en notation matricielle, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 = \frac{1}{n}\mathbf{e}'\mathbf{e} = \frac{1}{n}\|\mathbf{e}\|^2$.

Biais

Le résultat suivant montre que $\hat{\sigma}^2$ est un estimateur biaisé, mais asymptotiquement non biaisé, de σ^2.

Théorème 1

Dans le modèle semi-fort homoscédastique, $\mathrm{E}[\hat{\sigma}^2] = \frac{n-k}{n}\sigma^2$.

Preuve: Notons $\mathrm{E}[\cdot|\mathcal{X}]$ l'espérance conditionnelle sachant $X_1, \ldots, X_n$. Puisque $\mathbf{e} = \mathbf{PY} = \mathbf{P}(\mathbf{X}\beta + \boldsymbol{\varepsilon}) = \mathbf{P}\boldsymbol{\varepsilon}$, on a

$$\mathrm{E}[n\hat{\sigma}^2|\mathcal{X}] = \mathrm{E}[\mathbf{e}'\mathbf{e}|\mathcal{X}] = \mathrm{E}[(\mathbf{P}\boldsymbol{\varepsilon})'(\mathbf{P}\boldsymbol{\varepsilon})|\mathcal{X}] = \mathrm{E}[\boldsymbol{\varepsilon}'\mathbf{P}'\mathbf{P}\boldsymbol{\varepsilon}|\mathcal{X}]$$

$$= \mathrm{E}[\boldsymbol{\varepsilon}'\mathbf{P}\boldsymbol{\varepsilon}|\mathcal{X}] = \mathrm{E}[\mathrm{tr}[\boldsymbol{\varepsilon}'\mathbf{P}\boldsymbol{\varepsilon}]|\mathcal{X}] = \mathrm{E}[\mathrm{tr}[\mathbf{P}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}']|\mathcal{X}] = \mathrm{tr}[\mathrm{E}[\mathbf{P}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathcal{X}]]$$

$$= \mathrm{tr}[\mathbf{P}\mathrm{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathcal{X}]] = \sigma^2\mathrm{tr}[\mathbf{P}] = \sigma^2\mathrm{tr}[\mathbf{I} - \mathbf{Q}] = \sigma^2(n-k),$$

ce qui, en prenant l'espérance, établit le résultat. □

Biais

Bien entendu, un estimateur non biaisé de σ^2 est alors

$$\tilde{\sigma}^2 = \frac{n}{n-k}\hat{\sigma}^2 = \frac{1}{n-k}\sum_{i=1}^{n}(Y_i - \hat{\beta}' X_i)^2.$$

Dans le modèle de position ($k = 1$ et $X_1 = 1$ p.s.),

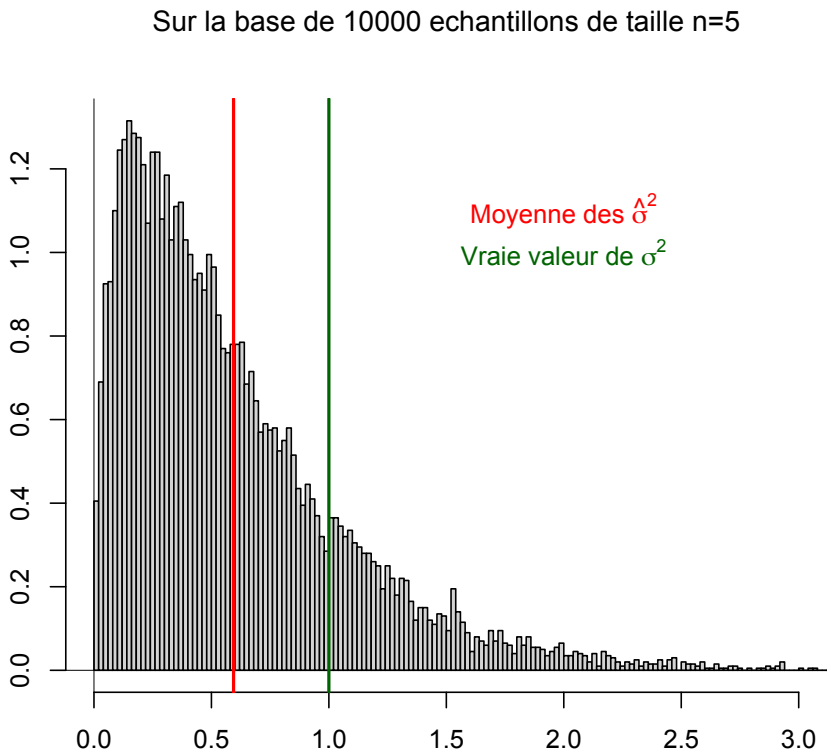
$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\beta})^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = s^2$$

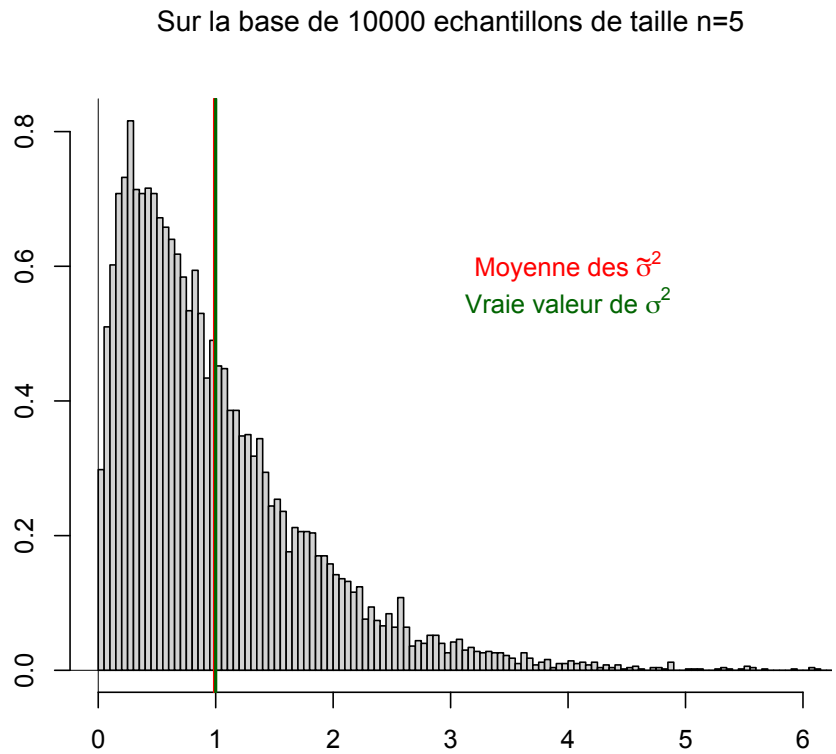
est la variance empirique et

$$\tilde{\sigma}^2 = \frac{n}{n-1}\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = S^2$$

est son habituelle version non biaisée.

Biais


Sur la base de 10000 echantillons de taille n=5
Moyenne des $\hat{\sigma}^2$
Vraie valeur de σ^2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0


Sur la base de 10000 echantillons de taille n=5
Moyenne des $\tilde{\sigma}^2$
Vraie valeur de σ^2
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6

Loi exacte

Théorème 2

Dans le modèle fort avec normalité (*avec* $n > k$), $\frac{n\hat{\sigma}^2}{\sigma^2}|\mathcal{X} \sim \chi^2_{n-k}$.

Preuve: En procédant comme dans la preuve précédente, on obtient

$$\frac{n\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2}\boldsymbol{\varepsilon}'\mathbf{P}\boldsymbol{\varepsilon} = \frac{1}{\sigma^2}\boldsymbol{\varepsilon}'\mathbf{O}\boldsymbol{\Lambda}\mathbf{O}'\boldsymbol{\varepsilon} = \left(\frac{1}{\sigma}\mathbf{O}'\boldsymbol{\varepsilon}\right)'\boldsymbol{\Lambda}\left(\frac{1}{\sigma}\mathbf{O}'\boldsymbol{\varepsilon}\right) \stackrel{\text{def}}{=} \boldsymbol{\eta}'\boldsymbol{\Lambda}\,\boldsymbol{\eta},$$

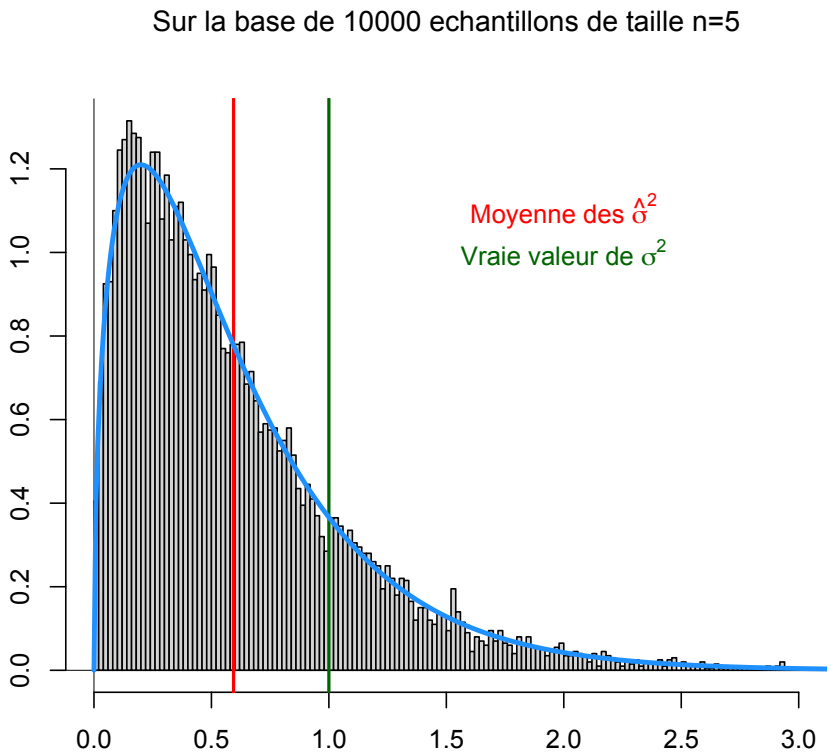
où $\mathbf{O}$ est $n \times n$ orthogonale et $\boldsymbol{\Lambda} = \text{diag}(1, \ldots, 1, 0, \ldots, 0)$ contient $n - k$ entrées non nulles. Puisque

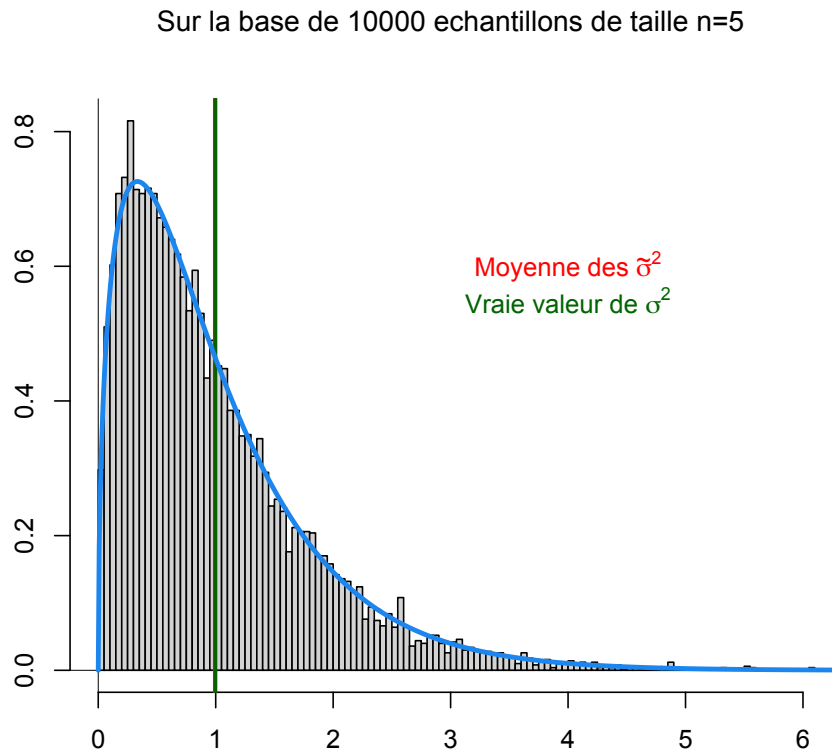
$$\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)'|\mathcal{X} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$$

(pourquoi?), on a

$$\frac{n\hat{\sigma}^2}{\sigma^2} = \boldsymbol{\eta}'\boldsymbol{\Lambda}\boldsymbol{\eta} = \eta_1^2 + \ldots + \eta_{n-k}^2|\mathcal{X} \sim \chi^2_{n-k}. \qquad \square$$

Loi exacte


Sur la base de 10000 echantillons de taille n=5
Moyenne des $\hat{\sigma}^2$
Vraie valeur de σ^2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0


Sur la base de 10000 echantillons de taille n=5
Moyenne des $\tilde{\sigma}^2$
Vraie valeur de σ^2
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6

Lemme de Fisher

Dans le modèle de position ($k = 1$ et $X_1 = 1$ p.s.) avec normalité, le lemme de Fisher dit que

(i) $\bar{Y} \sim \mathcal{N}(\beta, \frac{\sigma^2}{n})$

(ii) $\frac{ns^2}{\sigma^2} \sim \chi^2_{n-1}$

(iii) $\bar{Y} \perp\!\!\!\perp s^2$.

Dans le modèle fort avec normalité, on a

(i) $\hat{\beta}|\mathcal{X} \sim \mathcal{N}(\beta, \frac{\sigma^2}{n} Q^{-1})$

(ii) $\frac{n\hat{\sigma}^2}{\sigma^2}|\mathcal{X} \sim \chi^2_{n-k}$

(iii) ?

Un "lemme de Fisher" complet pour le modèle linéaire permettrait de construire des intervalles de confiance et des tests exacts...

Indépendance conditionnelle de $\hat{\beta}$ et $\hat{\sigma}^2$

Théorème 3

Dans le modèle fort avec normalité, $\hat{\beta} \perp\!\!\!\perp \hat{\sigma}^2$ conditionnellement à $\mathcal{X}$.

Preuve: Notons d'abord que, puisque $\mathbf{PX} = 0$ (donc aussi $\mathbf{X}'\mathbf{P} = 0$),

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \boldsymbol{\varepsilon}) = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$$

$$= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{P} + \mathbf{Q})\boldsymbol{\varepsilon} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}\boldsymbol{\varepsilon} \stackrel{\text{def}}{=} g_{\mathcal{X}}(\mathbf{Q}\boldsymbol{\varepsilon}),$$

tandis que $\hat{\sigma}^2 = \frac{1}{n}(\mathbf{P}\boldsymbol{\varepsilon})'(\mathbf{P}\boldsymbol{\varepsilon}) \stackrel{\text{def}}{=} h_{\mathcal{X}}(\mathbf{P}\boldsymbol{\varepsilon})$. Le résultat découle donc du fait que, puisque

$$\begin{pmatrix} \mathbf{P}\boldsymbol{\varepsilon} \\ \mathbf{Q}\boldsymbol{\varepsilon} \end{pmatrix} | \mathcal{X} \sim \mathcal{N}_{2n}\left(\mathbf{0}, \sigma^2 \begin{pmatrix} \mathbf{PP}' & \mathbf{PQ}' \\ \mathbf{QP}' & \mathbf{QQ}' \end{pmatrix} = \sigma^2 \begin{pmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{pmatrix}\right),$$

on a que $\mathbf{P}\boldsymbol{\varepsilon} \perp\!\!\!\perp \mathbf{Q}\boldsymbol{\varepsilon}$ conditionnellement à $\mathcal{X}$. □

Loi asymptotique

Nous n'avons pas montré que $\hat{\sigma}^2$ est un estimateur convergent dans le modèle semi-fort homoscédastique[1], ce qui motive le résultat suivant.

Théorème 4

Dans le modèle semi-fort homoscédastique, $\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathrm{Var}[\varepsilon^2])$, *pour autant que* $\mathrm{E}[\varepsilon^4] < \infty$.

Preuve: Puisque

$$\hat{\sigma}^2 - \sigma^2 = \frac{1}{n}\sum_{i=1}^n (e_i^2 - \sigma^2) = \frac{1}{n}\sum_{i=1}^n (\varepsilon_i^2 - \sigma^2) + \frac{1}{n}\sum_{i=1}^n (e_i^2 - \varepsilon_i^2),$$

on a

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (\varepsilon_i^2 - \sigma^2) + \frac{1}{\sqrt{n}}\sum_{i=1}^n (e_i^2 - \varepsilon_i^2) \stackrel{\mathrm{def}}{=} T_1 + T_2,$$

où le TCL livre $T_1 \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathrm{Var}[\varepsilon^2])$.

[1] Dans le fort avec normalité, la convergence découle de la loi exacte (pourquoi?)

Loi asymptotique

Comme $e_i = Y_i - \hat{\beta}' X_i$ et $\varepsilon_i = Y_i - \beta' X_i$, on a que

$$
\begin{aligned}
T_2 &= \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n (e_i^2 - \varepsilon_i^2) = \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n (e_i - \varepsilon_i)(e_i + \varepsilon_i) \\
&= \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n (\beta' X_i - \hat{\beta}' X_i)(2Y_i - \hat{\beta}' X_i - \beta' X_i) \\
&= \tfrac{1}{\sqrt{n}} (\beta - \hat{\beta})' \textstyle\sum_{i=1}^n X_i (2Y_i - X_i'(\hat{\beta} + \beta)) \\
&= \tfrac{1}{\sqrt{n}} (\beta - \hat{\beta})' \big\{ 2 \textstyle\sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i X_i' (\hat{\beta} + \beta) \big\} \\
&= \sqrt{n} (\beta - \hat{\beta})' \big\{ 2Q\hat{\beta} - Q(\hat{\beta} + \beta) \big\} \\
&= \sqrt{n} (\beta - \hat{\beta})' \big\{ Q\hat{\beta} - Q\beta \big\} \\
&= -\sqrt{n} (\hat{\beta} - \beta)' Q (\hat{\beta} - \beta)
\end{aligned}
$$

tend vers zéro en loi (pourquoi?), ce qui établit le résultat (pourquoi?) □