Propriétés des Estimateurs MLE et MCO

Simon-Pierre Boucher

14 janvier 2025

Table des matières

1 Propriétés des Estimateurs MLE et MCO

1.1 Introduction

Les estimateurs par Maximum de Vraisemblance (MLE) et par Moindres Carrés Ordinaires (MCO) sont deux méthodes couramment utilisées pour estimer les paramètres d'un modèle linéaire. Sous certaines conditions, ces deux estimateurs coïncident et présentent des propriétés intéressantes.

1.2 Estimateur BLUE

Définition 1.1 (BLUE)**. BLUE** signifie *Best Linear Unbiased Estimator* (Meilleur Estimateur Linéaire Sans Biais).

Un estimateur **BLUE** possède les propriétés suivantes :

— **Sans biais** : L'espérance de l'estimateur est égale à la vraie valeur du paramètre.

— **Variance minimale** : L'estimateur a la variance la plus petite possible parmi les estimateurs linéaires sans biais.

Proposition 1.1. *L'estimateur des moindres carrés ordinaires (MCO) est **BLUE**.*

Proposition 1.2. *L'estimateur $\hat{\sigma}^2$ obtenu par Maximum de Vraisemblance est biaisé vers le bas.*

Proposition 1.3. *Il existe une alternative sans biais pour estimer σ^2 donnée par :*

$$\hat{S}^2 = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{T - K}$$

1.3 Propriétés BLUE

Définition 1.2 (Sans biais)**.** Un estimateur $\hat{\theta}$ est **sans biais** si :

$$E(\hat{\theta}) = \theta$$

Définition 1.3 (Efficacité)**.** Un estimateur est **efficace** s'il atteint la borne de Cramér-Rao, c'est-à-dire que sa variance est égale à l'inverse de la matrice d'information de Fisher.

1.4 Borne de Cramér-Rao

Théorème 1.1 (Borne de Cramér-Rao)**.** *Pour tout estimateur régulier et sans biais, sa variance est bornée par l'inverse de la matrice d'information :*

$$Var(\hat{\theta}) \geq [I(\theta)]^{-1}$$

Démonstration. La borne de Cramér-Rao est obtenue en utilisant les propriétés des estimateurs sans biais et les inégalités de Fisher. La matrice d'information de Fisher $I(\theta)$ est définie comme :

$$I(\theta) = -E\left[\frac{\partial^2 \log L(\theta)}{\partial \theta^2}\right] = E\left[\left(\frac{\partial \log L(\theta)}{\partial \theta}\right)^2\right]$$

Ainsi, l'inverse de la matrice d'information donne une borne inférieure sur la variance des estimateurs sans biais. □

1.5 Matrice d'Information et Hessienne

Définition 1.4 (Matrice d'Information)**.** La matrice d'information est la matrice hessienne des dérivées secondes de la fonction de log-vraisemblance.

Pour une fonction $f(x_1, x_2)$, la hessienne est représentée par $H_{i,j}(f)$:

$$H_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

Dans le cas d'un modèle linéaire avec deux paramètres (β et σ^2), il y aura quatre dérivées secondes possibles :

- $i = 1,\ j = 1$: $\frac{\partial^2 f}{\partial \beta^2}$
- $i = 1,\ j = 2$: $\frac{\partial^2 f}{\partial \beta \partial \sigma^2}$
- $i = 2,\ j = 1$: $\frac{\partial^2 f}{\partial \sigma^2 \partial \beta}$
- $i = 2,\ j = 2$: $\frac{\partial^2 f}{\partial (\sigma^2)^2}$

1.6 Matrice d'Information dans le Modèle Linéaire Estimé par MLE

Dans le cas du modèle linéaire estimé par MLE, la matrice d'information est une matrice 2×2 avec les dérivées secondes calculées comme suit :

$$I(\beta, \sigma^2) = -E\begin{bmatrix} \frac{\partial^2 L}{\partial \beta \partial \beta'} & \frac{\partial^2 L}{\partial \beta \partial \sigma^2} \\ \frac{\partial^2 L}{\partial \sigma^2 \partial \beta'} & \frac{\partial^2 L}{\partial (\sigma^2)^2} \end{bmatrix}$$

Les dérivées secondes sont résolues comme suit :

1.6.1 En haut à gauche

$$\frac{\partial^2 L}{\partial\beta\partial\beta'} = -\frac{1}{\sigma^2}(X'X)$$

1.6.2 En haut à droite

$$\frac{\partial^2 L}{\partial\beta\partial\sigma^2} = -\frac{1}{\sigma^4}(X'u)$$

Sachant que $u = Y - X\hat{\beta}$.

1.6.3 En bas à gauche

$$\frac{\partial^2 L}{\partial\sigma^2\partial\beta'} = -\frac{1}{\sigma^4}(u'X)$$

1.6.4 En bas à droite

$$\frac{\partial^2 L}{\partial(\sigma^2)^2} = \frac{T}{2\sigma^4}$$

1.7 Inverse de la Matrice d'Information

La matrice d'information et son inverse sont données par :

$$I(\beta, \sigma^2) = \begin{bmatrix} \frac{1}{\sigma^2}(X'X) & 0 \\ 0 & \frac{T}{2\sigma^4} \end{bmatrix}$$
$$I^{-1}(\beta, \sigma^2) = \begin{bmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{T} \end{bmatrix}$$

1.8 Propriétés de l'Estimateur $\hat{\beta}$

1.8.1 Espérance de $\hat{\beta}$

Proposition 1.4. *L'estimateur $\hat{\beta}$ est sans biais.*

Démonstration. L'estimateur $\hat{\beta}$ est donné par :

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Sachant que $Y = X\beta + u$, on peut écrire :

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'[X\beta + u] \\
&= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u \\
&= \beta + (X'X)^{-1}X'u
\end{aligned}$$

En appliquant l'espérance :

$$\begin{aligned}
E(\hat{\beta}) &= E\left(\beta + (X'X)^{-1}X'u\right) \\
&= \beta + (X'X)^{-1}X'E(u) \\
&= \beta \quad \text{(puisque } E(u) = 0\text{)}
\end{aligned}$$

□

1.8.2 Variance de $\hat{\beta}$

Proposition 1.5. *La variance de $\hat{\beta}$ est donnée par :*

$$Var(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

Démonstration. La variance de $\hat{\beta}$ est :

$$\begin{aligned}
\text{Var}(\hat{\beta}) &= E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right] \\
&= E\left[((X'X)^{-1}X'u)((X'X)^{-1}X'u)'\right] \\
&= (X'X)^{-1}X'E(uu')X(X'X)^{-1}
\end{aligned}$$

Sachant que $E(uu') = \sigma^2 I_T$:

$$\begin{aligned}
\text{Var}(\hat{\beta}) &= (X'X)^{-1}X'(\sigma^2 I_T)X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}
\end{aligned}$$

□

Conclusion : La variance de $\hat{\beta}$ atteint la borne de Cramér-Rao, c'est-à-dire que $\hat{\beta}$ est un estimateur efficace.

2 Propriétés de $\hat{\sigma}^2$

2.1 Distribution et Espérance

Proposition 2.1. *La statistique suivante suit une loi du chi carré :*

$$\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{\sigma^2} \sim \chi^2(T - K)$$

Démonstration. Les résidus $\hat{u} = Y - X\hat{\beta}$ sont normalement distribués sous les hypothèses du modèle linéaire. La somme des carrés des résidus suit alors une loi du chi carré avec $T - K$ degrés de liberté, où T est le nombre d'observations et K le nombre de paramètres estimés. □

2.2 Espérance et Variance de $\hat{\sigma}^2$

Proposition 2.2. *L'espérance de $\hat{\sigma}^2$ est biaisée et donnée par :*

$$E(\hat{\sigma}^2) = \frac{T - K}{T}\sigma^2$$

Démonstration. L'estimateur $\hat{\sigma}^2$ est défini par :

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{T}$$

En multipliant le numérateur et le dénominateur par σ^2, on obtient :

$$\begin{aligned} E(\hat{\sigma}^2) &= E\left(\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{T}\right) \\ &= \frac{\sigma^2}{T} E\left(\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{\sigma^2}\right) \\ &= \frac{\sigma^2}{T}(T - K) \\ &= \frac{T - K}{T}\sigma^2 \end{aligned}$$

□

Proposition 2.3. *La variance de $\hat{\sigma}^2$ est donnée par :*

$$Var(\hat{\sigma}^2) = \frac{2\sigma^4}{T}$$

Démonstration. La variance de $\hat{\sigma}^2$ est :

$$\begin{aligned}
\mathrm{Var}(\hat{\sigma}^2) &= \mathrm{Var}\left(\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{T}\right) \\
&= \left(\frac{\sigma^2}{T}\right)^2 \mathrm{Var}\left(\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{\sigma^2}\right) \\
&= \left(\frac{\sigma^2}{T}\right)^2 \cdot 2(T-K) \quad (\text{car } \mathrm{Var}(\chi^2_{T-K}) = 2(T-K)) \\
&= \frac{2\sigma^4(T-K)}{T^2} \\
&= \frac{2\sigma^4}{T}\left(1 - \frac{K}{T}\right)
\end{aligned}$$

□

2.3 Estimateur de la Variance Sans Biais

Proposition 2.4. *L'estimateur sans biais de la variance est donné par :*

$$\hat{S}^2 = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{T - K}$$

et satisfait $E(\hat{S}^2) = \sigma^2$.

Démonstration. L'estimateur $\hat{S}^2$ est défini par :

$$\hat{S}^2 = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{T - K}$$

En multipliant le numérateur et le dénominateur par σ^2, on obtient :

$$\begin{aligned}
E(\hat{S}^2) &= E\left[\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{T - K}\right] \\
&= \frac{\sigma^2}{T-K} E\left[\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{\sigma^2}\right] \\
&= \frac{\sigma^2}{T-K}(T-K) \\
&= \sigma^2
\end{aligned}$$

□

Proposition 2.5. *La variance de $\hat{S}^2$ est donnée par :*

$$Var(\hat{S}^2) = \frac{2\sigma^4}{T-K}$$

Démonstration. La variance de $\hat{S}^2$ est :

$$\begin{aligned}
\text{Var}(\hat{S}^2) &= \text{Var}\left[\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{T-K}\right] \\
&= \left(\frac{\sigma^2}{T-K}\right)^2 \text{Var}\left[\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{\sigma^2}\right] \\
&= \left(\frac{\sigma^2}{T-K}\right)^2 \cdot 2(T-K) \quad (\text{car } \text{Var}(\chi^2_{T-K}) = 2(T-K)) \\
&= \frac{2\sigma^4(T-K)}{(T-K)^2} \\
&= \frac{2\sigma^4}{T-K}
\end{aligned}$$

□

Remarque : La variance de $\hat{S}^2$ est supérieure à celle donnée par la borne de Cramér-Rao :

$$\frac{2\sigma^4}{T-K} > \frac{2\sigma^4}{T}$$

ce qui indique que $\hat{S}^2$ n'atteint pas la borne de Cramér-Rao.

3 Résumé des Propriétés

- **Estimateur** $\hat{\beta}$:
  - Sans biais : $E(\hat{\beta}) = \beta$
  - Variance : $\text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$
  - Atteint la borne de Cramér-Rao
- **Estimateur** $\hat{\sigma}^2$:
  - Biaisé : $E(\hat{\sigma}^2) = \sigma^2\left(1 - \frac{K}{T}\right)$
  - Variance : $\text{Var}(\hat{\sigma}^2) = \frac{2\sigma^4}{T}\left(1 - \frac{K}{T}\right)$
- **Estimateur** $\hat{S}^2$ (sans biais) :
  - Sans biais : $E(\hat{S}^2) = \sigma^2$
  - Variance : $\text{Var}(\hat{S}^2) = \frac{2\sigma^4}{T-K}$