

# Preuve des Estimateurs du Maximum de Vraisemblance (MLE) pour un Modèle Linéaire

Simon-Pierre Boucher

Janvier 2025

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Modèle de Régression Linéaire</b>	<b>2</b>
<b>3</b>	<b>Fonction de Vraisemblance</b>	<b>2</b>
<b>4</b>	<b>Log-Vraisemblance</b>	<b>3</b>
<b>5</b>	<b>Maximisation de la Log-Vraisemblance</b>	<b>3</b>
5.1	Dérivée de la Log-Vraisemblance par Rapport à $\beta$ . . . . .	3
5.2	Dérivée de la Log-Vraisemblance par Rapport à $\sigma^2$ . . . . .	4
<b>6</b>	<b>Propriétés des Estimateurs MLE</b>	<b>4</b>
<b>7</b>	<b>Conclusion</b>	<b>5</b>

# 1 Introduction

La méthode du Maximum de Vraisemblance (MLE) est une approche statistique puissante permettant d'estimer les paramètres d'un modèle en maximisant la probabilité (ou la vraisemblance) des données observées sous ce modèle. Dans cette section, nous appliquerons cette méthode à un modèle de régression linéaire pour estimer les paramètres  $\beta$  et  $\sigma^2$ .

L'objectif de cette démonstration est de maximiser la log-vraisemblance pour obtenir les estimateurs des paramètres  $\beta$  et  $\sigma^2$ , qui sont ceux qui maximisent la probabilité des données observées.

## 2 Modèle de Régression Linéaire

Le modèle de régression linéaire peut être écrit de manière matricielle comme suit :

$$Y = X\beta + u$$

où :

- $Y$  est un vecteur  $n \times 1$  des valeurs observées de la variable dépendante,
- $X$  est une matrice  $n \times (k + 1)$  des variables explicatives (y compris une colonne de 1 pour le terme constant),
- $\beta$  est un vecteur  $(k + 1) \times 1$  des coefficients de régression à estimer  $(\beta_0, \beta_1, \dots, \beta_k)$ ,
- $u$  est un vecteur  $n \times 1$  des erreurs, supposées suivre une distribution normale  $u \sim N(0, \sigma^2 I)$ , où  $\sigma^2$  est la variance des erreurs et  $I$  est la matrice identité.

Le modèle peut être reformulé comme suit :

$$Y = X\beta + u$$

Cela signifie que la variable dépendante  $Y$  est expliquée par une combinaison linéaire des variables explicatives  $X$ , plus un terme d'erreur  $u$ .

## 3 Fonction de Vraisemblance

Sous l'hypothèse que les erreurs suivent une distribution normale  $u_t \sim N(0, \sigma^2)$ , la densité pour chaque observation  $Y_t$  conditionnée sur  $\beta$  et  $\sigma^2$  est donnée par :

$$f(Y_t | \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(Y_t - X_t^\top \beta)^2}{2\sigma^2} \right)$$



Cela représente la probabilité de l'observation  $Y_t$  donnée un ensemble de paramètres  $\beta$  et  $\sigma^2$ .

En supposant que les erreurs sont indépendantes et identiquement distribuées, la fonction de vraisemblance jointe des observations  $Y_1, Y_2, \dots, Y_n$  est simplement le produit des densités individuelles :

$$f(Y_1, Y_2, \dots, Y_n | \beta, \sigma^2) = \prod_{t=1}^n f(Y_t | \beta, \sigma^2)$$

La fonction de vraisemblance devient alors :

$$L(\beta, \sigma^2 | Y) = \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left( -\frac{(Y - X\beta)^\top (Y - X\beta)}{2\sigma^2} \right)$$

Cela représente la probabilité d'obtenir l'échantillon  $Y$  sous les paramètres  $\beta$  et  $\sigma^2$ .

## 4 Log-Vraisemblance

La log-vraisemblance est souvent utilisée en raison de sa simplicité mathématique. En prenant le logarithme de la fonction de vraisemblance, nous obtenons :

$$\log L(\beta, \sigma^2 | Y) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (Y - X\beta)^\top (Y - X\beta)$$

La log-vraisemblance est maintenant plus facile à manipuler et à maximiser. Nous allons maximiser cette fonction pour obtenir les estimateurs des paramètres  $\beta$  et  $\sigma^2$ .

## 5 Maximisation de la Log-Vraisemblance

L'objectif est de maximiser la log-vraisemblance par rapport à  $\beta$  et  $\sigma^2$ . Pour ce faire, nous devons dériver la log-vraisemblance par rapport à ces paramètres et évaluer les dérivées à zéro.

### 5.1 Dérivée de la Log-Vraisemblance par Rapport à $\beta$

Prenons la dérivée de la log-vraisemblance par rapport à  $\beta$  :

$$\frac{\partial}{\partial \beta} \log L(\beta, \sigma^2 | Y) = \frac{1}{\sigma^2} X^\top (Y - X\beta)$$



Nous égalons cette dérivée à zéro pour obtenir les conditions de premier ordre :

$$X^\top(Y - X\beta) = 0$$

Cela peut être réécrit comme :

$$X^\top X\beta = X^\top Y$$

La solution de cette équation est l'estimateur de  $\beta$  par Maximum de Vraisemblance :

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

C'est l'estimateur des paramètres du modèle de régression.

## 5.2 Dérivée de la Log-Vraisemblance par Rapport à $\sigma^2$

Prenons maintenant la dérivée de la log-vraisemblance par rapport à  $\sigma^2$  :

$$\frac{\partial}{\partial \sigma^2} \log L(\beta, \sigma^2 | Y) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} (Y - X\beta)^\top (Y - X\beta)$$

En égalant cette dérivée à zéro, nous obtenons :

$$\sigma^2 = \frac{(Y - X\hat{\beta})^\top (Y - X\hat{\beta})}{n}$$

Cela donne l'estimateur de la variance  $\sigma^2$  par Maximum de Vraisemblance :

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})^\top (Y - X\hat{\beta})}{n}$$

## 6 Propriétés des Estimateurs MLE

Les estimateurs obtenus par la méthode du Maximum de Vraisemblance possèdent plusieurs propriétés importantes :

- $\hat{\beta}$  est un estimateur sans biais de  $\beta$  car  $E[\hat{\beta}] = \beta$ ,
- $\hat{\beta}$  est l'estimateur linéaire sans biais et efficace (BLUE), ce qui signifie qu'il a la plus petite variance parmi tous les estimateurs sans biais,
- $\hat{\sigma}^2$  est biaisé, mais le biais devient négligeable pour des échantillons de grande taille. Un estimateur sans biais alternatif pour la variance est  $S^2 = \frac{(Y - X\hat{\beta})^\top (Y - X\hat{\beta})}{n-k}$ ,
- La variance de  $\hat{\beta}$  est donnée par  $\text{Var}(\hat{\beta}) = \sigma^2 (X^\top X)^{-1}$ , et atteint la borne de Cramér-Rao, ce qui signifie que cet estimateur est le plus efficace.



## 7 Conclusion

Nous avons démontré comment utiliser la méthode du Maximum de Vraisemblance pour estimer les paramètres d'un modèle de régression linéaire. Les estimateurs obtenus par MLE pour  $\beta$  et  $\sigma^2$  sont les mêmes que ceux des Moindres Carrés Ordinaires (MCO) pour  $\beta$ , mais l'estimateur de la variance  $\sigma^2$  par MLE est légèrement biaisé. Cependant, ce biais est souvent très faible pour des échantillons de grande taille.

La méthode MLE est très puissante car elle permet d'estimer les paramètres du modèle tout en maximisant la probabilité des données observées sous ce modèle. De plus, les estimateurs obtenus par MLE sont efficaces et sans biais, ce qui les rend très utiles dans les applications pratiques.

