Statistique Mathématique

Chapitre 5: Estimation

Mohamed Essaied Hamrita

Mastère de Recherche: Finance & Actuariat
IHEC Sousse

Décembre 2022

1. Vocabulaires
2. Propriétés d'un estimateur
3. Méthodes d'estimation

Soit X une variable aléatoire de densité $f(x,\theta)$ où θ est un paramètre inconnu. Soit $(x_1, x_2, \ldots, x_n)$ un n$-$échantillon *i.i.d* issu de la loi de X. Comment estimer le paramètre θ?

Soit X une variable aléatoire de densité $f(x,\theta)$ où θ est un paramètre inconnu. Soit $(x_1, x_2, \ldots, x_n)$ un n−échantillon *i.i.d* issu de la loi de X. Comment estimer le paramètre θ?

Définition 1

Un estimateur ponctuel est toute fonction $T(x_1, x_2, \ldots, x_n)$ d'un échantillon. Toute statistique est estimateur ponctuel.

Soit X une variable aléatoire de densité $f(x,\theta)$ où θ est un paramètre inconnu. Soit $(x_1, x_2, \ldots, x_n)$ un n−échantillon *i.i.d* issu de la loi de X. Comment estimer le paramètre θ?

Définition 1

Un estimateur ponctuel est toute fonction $T(x_1, x_2, \ldots, x_n)$ d'un échantillon. Toute statistique est estimateur ponctuel.

Exemple 1

Soit $(X_1, X_2, \ldots, X_n)$ un n−échantillon i.i.d issu de la loi normale de paramètres m et σ^2.

- $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ *est un estimateur ponctuel du paramètre m.*

Soit X une variable aléatoire de densité $f(x,\theta)$ où θ est un paramètre inconnu. Soit $(x_1, x_2, \ldots, x_n)$ un n−échantillon *i.i.d* issu de la loi de X. Comment estimer le paramètre θ?

Définition 1

Un estimateur ponctuel est toute fonction $T(x_1, x_2, \ldots, x_n)$ d'un échantillon. Toute statistique est estimateur ponctuel.

Exemple 1

Soit $(X_1, X_2, \ldots, X_n)$ un n−échantillon i.i.d issu de la loi normale de paramètres m et σ^2.

- $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ *est un estimateur ponctuel du paramètre m.*

- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ *est un estimateur ponctuel du paramètre σ^2.*

1. Vocabulaires

2. **Propriétés d'un estimateur**
   - Estimateur sans biais
   - Estimateur convergent
   - Estimateur efficace

3. Méthodes d'estimation

Un estimateur $\widehat{\theta}$ est **une variable aléatoire**.

Un estimateur $\widehat{\theta}$ est **une variable aléatoire**.

conséquences:

- $\widehat{\theta}$ possède une distribution de probabilité **empirique**.

Un estimateur $\widehat{\theta}$ est **une variable aléatoire**.

conséquences:

- $\widehat{\theta}$ possède une distribution de probabilité **empirique**.
- La distribution empirique de $\widehat{\theta}$ est caractérisée par les moments tels que $\mathbb{E}(X)$, $\mathbb{V}(X)$ et $\mathbb{E}(X^k)$.

Un estimateur $\widehat{\theta}$ est **une variable aléatoire**.

conséquences:

- $\widehat{\theta}$ possède une distribution de probabilité **empirique**.
- La distribution empirique de $\widehat{\theta}$ est caractérisée par les moments tels que $\mathbb{E}(X)$, $\mathbb{V}(X)$ et $\mathbb{E}(X^k)$.

Un estimateur $\widehat{\theta}$ est **une variable aléatoire**.

conséquences:

- $\widehat{\theta}$ possède une distribution de probabilité **empirique**.
- La distribution empirique de $\widehat{\theta}$ est caractérisée par les moments tels que $\mathbb{E}(X)$, $\mathbb{V}(X)$ et $\mathbb{E}(X^k)$.

Définition 2

Les propriétés d'un estimateur $\widehat{\theta}$ correspondent aux propriétés de sa densité empirique définie pour un échantillon de taille $n \in \mathbb{N}$.

Un estimateur $\widehat{\theta}$ est **une variable aléatoire**.

conséquences:

- $\widehat{\theta}$ possède une distribution de probabilité **empirique**.
- La distribution empirique de $\widehat{\theta}$ est caractérisée par les moments tels que $\mathbb{E}(X)$, $\mathbb{V}(X)$ et $\mathbb{E}(X^k)$.

Définition 2

Les propriétés d'un estimateur $\widehat{\theta}$ correspondent aux propriétés de sa densité empirique définie pour un échantillon de taille $n \in \mathbb{N}$.

Exemple 2

Soit $(X_1, X_2, \ldots, X_n)$ un $n-$échantillon i.i.d issu de la loi normale de paramètres m et σ^2. $\widehat{m} = \overline{X} \sim \mathbb{N}(m, \frac{\sigma^2}{n})$.

1. Vocabulaires
2. Propriétés d'un estimateur
   - Estimateur sans biais
   - Estimateur convergent
   - Estimateur efficace
3. Méthodes d'estimation

Définition 3

Un estimateur $\widehat{\theta}$ est dit ***sans biais*** *si $\mathbb{E}(\widehat{\theta}) = \theta$.*

Définition 3

Un estimateur $\widehat{\theta}$ est dit ***sans biais*** *si $\mathbb{E}(\widehat{\theta}) = \theta$.*

Exercice 1

Soit le modèle linéaire $Y = X\beta + \varepsilon$, où $\varepsilon \sim N(0, \sigma^2 I_n)$. Montrer que $\widehat{\beta} = (X'X)^{-1}X'Y$ est un estimateur sans bias de β.

Définition 3

Un estimateur $\widehat{\theta}$ est dit ***sans biais*** *si* $\mathbb{E}(\widehat{\theta}) = \theta$.

Exercice 1

Soit le modèle linéaire $Y = X\beta + \varepsilon$, où $\varepsilon \sim N(0, \sigma^2 I_n)$. Montrer que $\widehat{\beta} = (X'X)^{-1}X'Y$ est un estimateur sans bias de β.

Définition 4

Un estimateur $\widehat{\theta}$ est dit ***asymptotiquement sans biais*** *si* $\lim_{n\to\infty} \mathbb{E}(\widehat{\theta}) = \theta$.

Exercice 2

Soit $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ un estimateur de σ^2 d'un échantillon i.i.d issu d'une loi normale de paramètres m et σ^2. Montrer que S^2 est asymptotiquement sans bias.

1. Vocabulaires
2. Propriétés d'un estimateur
   - Estimateur sans biais
   - Estimateur convergent
   - Estimateur efficace
3. Méthodes d'estimation

Définition 5

Un estimateur $\widehat{\theta}$ est dit ***convergent*** *s'il est sans biais et* $\lim\limits_{n\to\infty} \mathbb{V}(\widehat{\theta}) = 0.$

Définition 5

Un estimateur $\widehat{\theta}$ est dit ***convergent*** *s'il est sans biais et $\lim_{n\to\infty} \mathbb{V}(\widehat{\theta}) = 0$.*

Exercice 3

Montrer que $\widehat{m} = \overline{X}$ est un estimateur convergent de m

Définition 5

Un estimateur $\widehat{\theta}$ est dit ***convergent*** *s'il est sans biais et $\lim\limits_{n\to\infty} \mathbb{V}(\widehat{\theta}) = 0$.*

Exercice 3

Montrer que $\widehat{m} = \overline{X}$ est un estimateur convergent de m

$$\begin{aligned}
\mathbb{V}(\overline{X}) &= \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{V}(X_i) \\
&= \frac{\sigma^2}{n} \xrightarrow[n\longrightarrow\infty]{} 0
\end{aligned}$$

1. Vocabulaires

2. Propriétés d'un estimateur
   - Estimateur sans biais
   - Estimateur convergent
   - Estimateur efficace

3. Méthodes d'estimation

Définition 6

Soit $(X_1, X_2, \ldots, X_n)$ *un* $n-$*échantillon i.i.d de densité* $f(X_i, \theta)$ *et soit* $\widehat{\theta}$ *un estimateur sans bias de* θ. $\widehat{\theta}$ *est dit* ***efficace*** *si* $\mathbb{V}(\widehat{\theta}) = I^{-1}(\theta)$ *où* $I(\theta)$ *est appelée* ***quantité d'information de Fisher***.

$I(\theta) = \mathbb{E}\left((\partial L/\partial\theta)^2\right) = -\mathbb{E}\left(\partial^2 L/\partial\theta^2\right)$ *avec L est la fonction log-vraisemblance (sera définie dans la section suivante).*

Définition 6

*Soit $(X_1, X_2, \dots, X_n)$ un $n-$échantillon i.i.d de densité $f(X_i, \theta)$ et soit $\widehat{\theta}$ un estimateur sans bias de θ. $\widehat{\theta}$ est dit **efficace** si $\mathbb{V}(\widehat{\theta}) = I^{-1}(\theta)$ où $I(\theta)$ est appelée **quantité d'information de Fisher**.*
$I(\theta) = \mathbb{E}\left((\partial L/\partial\theta)^2\right) = -\mathbb{E}\left(\partial^2 L/\partial\theta^2\right)$ avec L est la fonction log-vraisemblance (sera définie dans la section suivante).

Définition 7

*Soient $\widehat{\theta}_1$ et $\widehat{\theta}_2$ deux estimateurs d'un paramètre θ. On dit que $\widehat{\theta}_1$ est **plus efficace** que $\widehat{\theta}_2$ si $\mathbb{V}(\widehat{\theta}_1) \leq \mathbb{V}(\widehat{\theta}_2)$.*

Définition 8 (La vraisemblance)

Soit $(X_1, X_2, \ldots, X_n)$ *un* $n-$*échantillon i.i.d de densité* $f(X_i, \theta)$. *On appelle la fonction de vraisemblance la fonction* $L(x_1, x_2, \ldots, x_n, \theta)$ *définie par:*

$$L(x_1, x_2, \ldots, x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

Définition 8 (La vraisemblance)

Soit $(X_1, X_2, \ldots, X_n)$ un $n-$échantillon i.i.d de densité $f(X_i, \theta)$. On appelle la fonction de vraisemblance la fonction $L(x_1, x_2, \ldots, x_n, \theta)$ définie par:

$$L(x_1, x_2, \ldots, x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

En pratique, on utilise plutôt la fonction de log-vraisemblance qui est définie par:

$$\ell(x_1, x_2, \ldots, x_n, \theta) = \ln L = \sum_{i=1}^{n} \ln\left(f(x_i, \theta)\right)$$

Exercice 4

Déterminer le log-vraisemblance d'un $n-$*échantillon* $i.i.d$ *de loi de poisson de paramètre* θ.

Exercice 4

Déterminer le log-vraisemblance d'un $n-$échantillon $i.i.d$ de loi de poisson de paramètre θ.

$$
\begin{aligned}
\ell(x_1, x_2, \ldots, x_n, \theta) &= \sum_{i=1}^{n} \ln\left(f(x_i, \theta)\right) = \sum_{i=1}^{n} \ln\left(\frac{e^{-\theta}\theta^{x_i}}{x_i!}\right) \\
&= \sum_{i=1}^{n} \left[\ln(e^{-\theta}\theta^{x_i}) - \ln(x_i!)\right] \\
&= \sum_{i=1}^{n} \left[-\theta + x_i \ln\theta - \sum_{i=1}^{n} \ln(x_i)\right] \\
&= -n\theta + \ln\theta \sum_{i=1}^{n} x_i - n \sum_{i=1}^{n} \ln(x_i)
\end{aligned}
$$

La fonction log-vraisemblance avec R

```
lPois=function(theta,x){
n=length(x)
ss=sum(x)
ss1=sum(log(factorial(x)))
-n*theta+ss*log(theta)-n*ss1
}
```

Définition 9

L'estimateur du maximum de vraisemblance est celui qui maximise la fonction de vraisemblance:

$$\widehat{\theta}_{MV} = arg \max_{\theta} \ell(x_1, x_2, \ldots, x_n, \theta)$$

Définition 9

L'estimateur du maximum de vraisemblance est celui qui maximise la fonction de vraisemblance:

$$\widehat{\theta}_{MV} = arg \max_{\theta} \ell(x_1, x_2, \ldots, x_n, \theta)$$

Il s'agit d'un problème d'optimisation sans contraintes.
Condition du premier ordre : La dérivé (les dérivées partielles) première(s) s'annule(nt), i.e; $\dfrac{\partial \ell}{\partial \theta} = 0$. Le vecteur colonne $s(\theta, x) = \frac{\partial \ell}{\partial \theta}$ sera appelé la fonction score.

Condition du second ordre: La dérivée seconde est négative ($\dfrac{\partial^2 \ell}{\partial \theta \partial \theta'}$ est définie négative)

Remarque:

$$H = \frac{\partial^2 \ell}{\partial\theta\partial\theta'} = \begin{pmatrix} \frac{\partial^2 \ell}{\partial\theta_1^2} & \frac{\partial^2 \ell}{\partial\theta_1\partial\theta_2} & \cdots & \frac{\partial^2 \ell}{\partial\theta_1\partial\theta_k} \\ \frac{\partial^2 \ell}{\partial\theta_2\partial\theta_1} & \frac{\partial^2 \ell}{\partial\theta_2^2} & \cdots & \frac{\partial^2 \ell}{\partial\theta_2\partial\theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \ell}{\partial\theta_k\partial\theta_1} & \cdots & \cdots & \frac{\partial^2 \ell}{\partial\theta_k^2} \end{pmatrix}$$

La matrice H est appelée matrice hessienne et elle est égale à l'opposé de l'inverse de la matrice des variances-covariance de θ,
$(H(\theta) = -(V(\theta))^{-1} = -I(\theta))$.
Elle est **symétrique** et est définie négative.
Une matrice symétrique est définie négative si ses valeurs propres sont négatives.

Exercice 5

Soit $(x_1, x_2, \ldots, x_{10})$ *un échantillon de taille* $n = 10$ *issu de la loi de poisson de paramètre* θ. *Déterminer un estimateur de* θ *par la méthode de vraisemblance.*

Exercice 5

Soit $(x_1, x_2, \ldots, x_{10})$ *un échantillon de taille* $n = 10$ *issu de la loi de poisson de paramètre* θ. *Déterminer un estimateur de* θ *par la méthode de vraisemblance.*

On a $\ell(x_1, x_2, \ldots, x_n, \theta) = -n\theta + \ln\theta \sum_{i=1}^{n} x_i - n \sum_{i=1}^{n} \ln(x_i)$.

$$\frac{\partial \ell}{\partial \theta} = 0 \Longrightarrow -n + \frac{\sum_{i=1}^{n} x_i}{\theta} = 0.$$

Soit $\widehat{\theta} = \overline{X} = 0.8$.

Vérifions la condition suffisante:

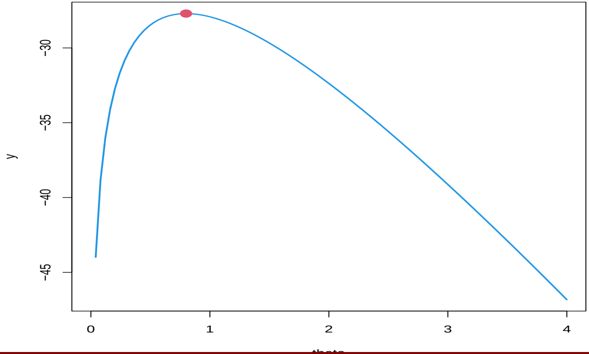
$$\frac{\partial^2 \ell}{\partial \theta^2} = -\frac{\sum_{i=1}^{n} x_i}{\theta^2} < 0.$$

```r
x=c(1, 1, 1, 1, 0, 0, 0, 0, 1, 3)

estim=optimise(lPois, c(0,4), maximum = TRUE, x=x)

estim[[1]]

[1] 0.8000203
```

```
x=c(1, 1, 1, 1, 0, 0, 0, 0, 1, 3)

estim=optimise(lPois, c(0,4), maximum = TRUE, x=x)

estim[[1]]

[1] 0.8000203
```

Représentons le graphique de la fonction de vraisemblance en fonction de θ.

```
theta=seq(0,4,len=100)
y=lPois(theta,x)
plot(theta,y,type="l", col=4,lwd=2)
points(0.8,lPois(0.8,x),col=2, pch=21,lwd=6)
```



Exercice 6

Soit $(x_1, \ldots, x_{10})$ *un échantillon i.i.d de taille* $n = 10$ *issu de la loi normale de paramètre m et sigma*^{2}*. Déterminer les estimateurs de m et de* σ^2 *par la méthode du maximum de vraisemblance et calculer la matrice hessienne.*$x_i =$
$(4.634, 1.227, 2.337, 5.241, 3.597, 4.559, 5.912, 1.711, -0.106, -0.195)$.

Exercice 6

Soit $(x_1, \ldots, x_{10})$ *un échantillon i.i.d de taille* $n = 10$ *issu de la loi normale de paramètre m et sigma*^{2}*. Déterminer les estimateurs de m et de* σ^2 *par la méthode du maximum de vraisemblance et calculer la matrice hessienne.*$x_i =$
$(4.634, 1.227, 2.337, 5.241, 3.597, 4.559, 5.912, 1.711, -0.106, -0.195)$.

On peut vérifier que la vraisemblance de la loi normale est donnée par:

$$\ell(x_1, x_2, \ldots, x_n, \theta) = -\frac{n}{2}\left(\ln(2\pi) + \ln(\sigma^2)\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2$$

Exercice 6

Soit $(x_1, \ldots, x_{10})$ *un échantillon i.i.d de taille* $n = 10$ *issu de la loi normale de paramètre m et sigma*^{2}*. Déterminer les estimateurs de m et de* σ^2 *par la méthode du maximum de vraisemblance et calculer la matrice hessienne.*$x_i =$
$(4.634, 1.227, 2.337, 5.241, 3.597, 4.559, 5.912, 1.711, -0.106, -0.195)$.

On peut vérifier que la vraisemblance de la loi normale est donnée par:

$$\ell(x_1, x_2, \ldots, x_n, \theta) = -\frac{n}{2}\left(\ln(2\pi) + \ln(\sigma^2)\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2$$

$$\frac{\partial \ell}{\partial m} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - m) = 0 \Longleftrightarrow \widehat{m} = \overline{X} = 2.8917 \qquad (1)$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(X_i - m)^2 = 0 \Longleftrightarrow \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = 4.3941$$

$$\frac{\partial^2 \ell}{\partial m^2} = -\frac{n}{\sigma^2}$$

$$\frac{\partial^2 \ell}{\partial \beta^2} = \frac{n}{2\beta^2} - \frac{2}{\beta^3}\sum_{i=1}^{n}(X_i - m)^2, \quad (\beta = \sigma^2)$$

$$\frac{\partial^2 \ell}{\partial m \partial \beta} = -\frac{1}{\beta^2}\sum_{i=1}^{n}(X_i - m)$$

Le hessienne observée est:

$$H = \begin{pmatrix} -\dfrac{n}{\sigma^2} & -\dfrac{1}{\sigma^4}\sum_{i=1}^{n}(X_i - m) \\ -\dfrac{1}{\sigma^4}\sum_{i=1}^{n}(X_i - m) & -\dfrac{n}{2\sigma^4} + \dfrac{1}{\sigma^6}\sum_{i=1}^{n}(X_i - m)^2 \end{pmatrix}$$

La hessienne espérée est:

$$\mathcal{H} = \begin{pmatrix} -\dfrac{n}{\sigma^2} & 0 \\ 0 & -\dfrac{n}{2\sigma^4} \end{pmatrix} = \begin{pmatrix} -2.275 & 0 \\ 0 & -0.2589 \end{pmatrix}$$

En effet, $\mathbb{E}\left(-\frac{n}{\sigma^2}\right) = -\frac{n}{\sigma^2}$;
$\mathbb{E}\left(-\frac{1}{\sigma^4}\sum_{i=1}^n (X_i - m)\right) = -\frac{1}{\sigma^4}\sum_{i=1}^n \mathbb{E}(X_i - m) = 0$;
$\mathbb{E}\left(\frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^n (X_i - m)^2\right) = \frac{n}{2\sigma^4} - \frac{2}{2\sigma^6}\sum_{i=1}^n \mathbb{E}(X_i - m)^2 = \frac{n}{2\sigma^4} - \frac{2n\sigma^2}{2\sigma^6}$.

La hessienne espérée est:

$$\mathcal{H} = \begin{pmatrix} -\dfrac{n}{\sigma^2} & 0 \\ 0 & -\dfrac{n}{2\sigma^4} \end{pmatrix} = \begin{pmatrix} -2.275 & 0 \\ 0 & -0.2589 \end{pmatrix}$$

En effet, $\mathbb{E}\left(-\frac{n}{\sigma^2}\right) = -\frac{n}{\sigma^2}$;
$\mathbb{E}\left(-\frac{1}{\sigma^4}\sum_{i=1}^{n}(X_i - m)\right) = -\frac{1}{\sigma^4}\sum_{i=1}^{n}\mathbb{E}(X_i - m) = 0$;
$\mathbb{E}\left(\frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(X_i - m)^2\right) = \frac{n}{2\sigma^4} - \frac{2}{2\sigma^6}\sum_{i=1}^{n}\mathbb{E}(X_i - m)^2 = \frac{n}{2\sigma^4} - \frac{2n\sigma^2}{2\sigma^6}$.

Implémentation sous R:

```
y=c(4.634, 1.227, 2.337, 5.241, 3.597, 4.559, 5.912,
1.711,  -0.106, -0.195)
```

```
lNorm=function(theta,x){ n=length(x)
ss=sum((x-theta[1])^2)
n/2*(log(2*pi)+log(theta[2]))+1/(2*theta[2])*ss }
estim2=optim(fn=lNorm,par=c(0,0.01), x=y, hessian=TRUE,
lower=c(1,1),method="L-BFGS-B")
estim2[[1]]

[1] 2.891700 4.394126

round(estim2[[6]],4)

       [,1]  [,2]
[1,] 2.2758 0.000
[2,] 0.0000 0.259
```

Exercice 7

Soit la régression linéaire simple $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ *avec* $\varepsilon_i \overset{iid}{\sim} N(0, \sigma)$. *Donner une estimation ponctuelle des paramètres* $(\beta_0, \beta_1, \sigma)$ *par la méthode de vraisemblance.*

Application sous R:

$x = (0.1, 0.32, 0.52, 0.66, 0.41, 0.91, 0.29, 0.46, 0.33, 0.65)$

$y = (0.25, 0.81, 1.16, 1.31, 0.7, 1.68, 0.57, 1.2, 0.79, 1)$

Exercice 7

Soit la régression linéaire simple $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ *avec* $\varepsilon_i \overset{iid}{\sim} N(0, \sigma)$. *Donner une estimation ponctuelle des paramètres* $(\beta_0, \beta_1, \sigma)$ *par la méthode de vraisemblance.*

Application sous R:

$x = (0.1, 0.32, 0.52, 0.66, 0.41, 0.91, 0.29, 0.46, 0.33, 0.65)$

$y = (0.25, 0.81, 1.16, 1.31, 0.7, 1.68, 0.57, 1.2, 0.79, 1)$

$\varepsilon_i \overset{iid}{\sim} N(0, \sigma) \implies y_i \overset{iid}{\sim} N(\beta_0 + \beta_1 x_i, \sigma)$, donc

$$f(y_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2\right)$$

$$\implies logL = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

```r
x=c(0.1,0.32,0.52,0.66,0.41,0.91,0.29,0.46,0.33,0.65)
y=c(0.25,0.81,1.16,1.31,0.7,1.68,0.57, 1.2,0.79,1)
# VS
logL=function(x,y,theta){ # theta=(beta0,beta1,sigma)
n=length(x); b0=theta[1]
b1=theta[2]; sig=theta[3]
ll=-n/2*log(2*pi)-n*log(sig)-1/(2*sig^2)*sum((y-b0-b1*x)^2)
return(-ll)}
# Estimation
theta=optim(fn=logL,par=c(1,1,1), x=x,y=y,
lower=c(-Inf,-Inf,0.001), upper=c(Inf, Inf, Inf),
method="L-BFGS-B",hessian=TRUE)
theta[[1]]

[1] 0.1786281 1.6524133 0.1430038
```

Comparons le résultat par rapport l'estimation par la méthode des MCO

```
mco=lm(y~x)
mco[[1]] # coefficients

(Intercept)           x
  0.1786274   1.6524142

summary(mco)[[6]]  # sigma

[1] 0.1598768
```

La matrice des variances-covariances des paramètres:

```
vcov(mco)         # par MCO

              (Intercept)           x
(Intercept)  0.01408352 -0.02479024
x           -0.02479024  0.05331235

round(solve(theta[[6]]),6)  # MVS (Hessienne)

          [,1]      [,2]     [,3]
[1,]  0.011268 -0.019834 0.000000
[2,] -0.019834  0.042653 0.000000
[3,]  0.000000  0.000000 0.001022
```

1 Vocabulaires

2 Propriétés d'un estimateur

3 Méthodes d'estimation
   - La méthode de vraisemblance
   - Méthode des moments
   - Méthode des MCO

Définition 10

1. $\mathbb{E}(X^k)$ *est le* ***moment théorique non centré*** *d'ordre* k.

Définition 10

1. $\mathbb{E}(X^k)$ *est le **moment théorique non centré** d'ordre k.*
2. $\mathbb{E}\left[(x-m)^k\right]$ *est le **moment théorique centré** d'ordre k où* $m = \mathbb{E}(X)$.

Définition 10

1. $\mathbb{E}(X^k)$ *est le* ***moment théorique non centré*** *d'ordre k.*
2. $\mathbb{E}\left[(x-m)^k\right]$ *est le* ***moment théorique centré*** *d'ordre k où* $m = \mathbb{E}(X)$.
3. $M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ *est le* ***moment empirique non centré*** *d'ordre k.*

Définition 10

1. $\mathbb{E}(X^k)$ *est le* ***moment théorique non centré*** *d'ordre* k.
2. $\mathbb{E}\left[(x-m)^k\right]$ *est le* ***moment théorique centré*** *d'ordre* k *où* $m = \mathbb{E}(X)$.
3. $M_k = \dfrac{1}{n}\sum_{i=1}^{n} X_i^k$ *est le* ***moment empirique non centré*** *d'ordre* k.
4. $\mu_k = \dfrac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^k$ *est le* ***moment empirique centré*** *d'ordre* k.

Définition 10

1. $\mathbb{E}(X^k)$ *est le* ***moment théorique non centré*** *d'ordre* k.
2. $\mathbb{E}\left[(x-m)^k\right]$ *est le* ***moment théorique centré*** *d'ordre* k *où* $m = \mathbb{E}(X)$.
3. $M_k = \dfrac{1}{n}\sum_{i=1}^{n} X_i^k$ *est le* ***moment empirique non centré*** *d'ordre* k.
4. $\mu_k = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$ *est le* ***moment empirique centré*** *d'ordre* k.

Définition 10

1. $\mathbb{E}(X^k)$ *est le* ***moment théorique non centré*** *d'ordre* k.
2. $\mathbb{E}\left[(x-m)^k\right]$ *est le* ***moment théorique centré*** *d'ordre* k *où* $m = \mathbb{E}(X)$.
3. $M_k = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i^k$ *est le* ***moment empirique non centré*** *d'ordre* k.
4. $\mu_k = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^k$ *est le* ***moment empirique centré*** *d'ordre* k.

La méthode des moments consiste à ***égaliser*** *entre les moments théoriques et les moments empiriques.*

Exemple 3

Soit $(X_1, X_2, \dots, X_n)$ *une suite de v.a iid suivant la loi de* $\mathcal{B}(1, p)$ *.*
Déterminer l'estimateur de p par la méthode des moments.
On a $\mathbb{E}(X) = p$ *et* $M_1 = \overline{X}$*, d'où* $\mathbb{E}(X) = M_1 \Longrightarrow \widehat{p} = \overline{X}$*.*

Exemple 4

Soit $(X_1, X_2, \dots, X_n)$ *une suite de v.a telle que* $X_i \overset{iid}{\sim} N(m, \sigma)$*. Donner une estimation des paramètres* m *et* σ^2 *par la méthode des moments.*

$\mathbb{E}(X) = M_1 \Longrightarrow \widehat{m} = \overline{X}$.

$\mathbb{E}(X^2) = M_2 \Longrightarrow \sigma^2 + m^2 = \overline{X^2} \Longrightarrow \sigma^2 = \overline{X^2} - m^2$. *Soit*

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \widehat{m}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$$

Remarque: On peut, aussi, égaliser entre les moments centré théoriques et empiriques afin de déterminer l'estimateur de σ^2.

La méthode des MCO consiste à minimiser l'erreur quadratique (EQ). Soit le modèle linéaire simple $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$. L'erreur quadratique correspondant à ce modèle est donné par:

$EQ = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i\right)^2$. Les estimateurs de β_0 et β_1 sont solution de la fonction objective $\min_{\beta_0, \beta_1} EQ$.

La condition nécessaire est:

$$\frac{\partial EQ}{\partial \beta_0} = \frac{\partial EQ}{\partial \beta_1} = 0$$

Et la condition suffisante est: la matrice Hessienne est définie positive.

On peut montrer que les estimateurs des paramètres sont donnés par:

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} \quad \text{et} \quad \widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Dans le cas du modèle de régression multiple, on aura $\widehat{\beta} = (X'X)^{-1}X'Y$ et la matrice des variances-covariances des paramètres est estimée par $\widehat{\Sigma} = \widehat{\sigma}_{\varepsilon}^2 (X'X)^{-1}$.

Exercice 8

Reprenez l'exercice 7 et déterminer les estimateurs du modèle par la méthode des MCO.