Chapitre 3: Méthodes d'estimation

Rabah Messaci

Département de Probabilités-Statistique
USTHB

Octobre 2011

Introduction

Pour certaines quantités, moments, variances, covariances, probabilités il a été vu qu'il existe des estimateurs simples ayant de bonnes qualités : sans biais et convergents. Pour des paramètres quelconques, il est nécessaire d'avoir des méthodes d'estimation conduisant à des estimateurs de bonne qualité. Il existe plusieurs grandes méthodes d'estimation

- Méthode des moments
- Méthode du maximum de vraisemblance
- Recherche des estimateurs sans biais de variance minimale

Autres méthodes - estimation par les moindres carrés qui sera utilisée dans le contexte de la régression linéaire - estimation bayésienne qui ne sera pas abordée ici

Méthode des moments : exemples

Les moments empiriques sont des estimateurs sans biais et convergents des moments théoriques. La méthode des moments consiste à égaler les uns aux autres. Soit le modèle d'échantillon $(\mathfrak{X}, \mathfrak{B}, (\mathcal{P}_\theta)_{\theta\in\Theta})^{\prime n)}$ associé à n observations $(x_1, x_2, ..., x_n)$ d,'une v.a de loi P_θ.
On suppose $\theta = (\theta_1, \theta_2, ..., \theta_p)$. On veut estimer θ i.e p paramètres.

$$E(X) = \overline{X_n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$E(X^2) = \frac{\sum_{i=1}^{n} X_i^2}{n}$$

....

$$E(X^p) = \frac{\sum_{i=1}^{n} X_i^p}{n} :$$

Méthode des moments

Exemples :

- Soit le modèle d'échantillon poissonien $(N, \mathcal{P}(N), \{P(\lambda), \lambda \in R_+^*\})^{'n)}$ associé à n observations $(x_1, x_2, ..., x_n)$ d,'une v.a de loi $P(\lambda)$.
  Estimer λ par la méthode des moments.
  $E(X) = \lambda = \overline{x} \implies \widehat{\lambda} = \overline{x}$
- Soit le modèle d'échantillon gaussien $(R, \mathcal{P}(R), \{N(m, \sigma^2), m \in R\})^{'n)}$. On suppose σ^2 connu.
  Estimer m.
  $E(X) = m = \overline{x} \implies \widehat{m} = \overline{x}$
- Soit le modèle d'échantillon gaussien $(R, \mathcal{P}(R), \{N(m, \sigma^2), m \in R\})^{'n)}$
  Estimer m.et σ^2
  $E(X) = m = \overline{x}$
  $E(X^2) = \sigma^2 + m^2 = m_2 \implies$
  $\widehat{m} = \overline{x}$ et $\widehat{\sigma^2} = m_2 - \overline{x}^2 = S_n^2$

Méthode du maximum de vraisemblance

Considérons une pièce de monnaie pour laquelle la probabilité d'obtenir pile est inconnue. Cette pièce est lancée 10 fois et on obtient 6 piles. Une méthode d'estimation intuitive est de considérer que l'évènement réalisée, ici 6 piles et 4 faces, a une forte probabilité (puiqu'il s'est réalisé contrairemant aux autres). On estime alors p par la valeur $\widehat{p}$ de [0,1] qui attribue à cet évènement, la plus forte probabité, ou en d'autres termes par la valeur $\widehat{p}$ qui maximise la vraisemblance.

On a $L(p, \widetilde{x}) = p^6(1-p)^4$

$$\frac{\partial}{\partial p} L(p, \widetilde{x}) = 0 \iff \frac{\partial}{\partial p} \log L(p, \widetilde{x}) = 0$$

$$\iff \widehat{p} = \frac{6}{10}$$

Méthode du maximum de vraisemblance

Définition : Estimateurs du maximum de vraisemblance

Soit $\widetilde{x} = (x_1, x_2, ..., x_n)$ un n-échantillon d'une v.a X de loi $(f_\theta)_{\theta \in \Theta}$. On appelle estimateur du maximum de vraisemblance de θ, s'il existe, la statistique $\widehat{\theta}$ telle que

$$L(\widehat{\theta}, \widetilde{x}) = \max_{\theta \in \Theta} L(\theta, \widetilde{x})$$

La maximisation de la fonction de vraisemblance peut se faire de différentes manières

Méthode du maximum de vraisemblance

Cas 1 : θ réel et $L(\theta,\widetilde{x})$ dérivable par rapport à θ

$\widehat{\theta}$ est alors solution du système d'équations dit équations du maximum de vraisemblance

$$\begin{cases} \dfrac{\partial}{\partial\theta}L(\theta,\widetilde{x})=0 : \text{condition du } 1^{er} \text{ ordre} \\ \dfrac{\partial^2}{\partial\theta^2}L(\theta,\widetilde{x})<0 : \text{condition du } 2^{ème} \text{ ordre} \end{cases}$$

$$\Longleftrightarrow \begin{cases} \dfrac{\partial}{\partial\theta}\log L(\theta,\widetilde{x})=0 : \text{condition du } 1^{er} \text{ ordre} \\ \dfrac{\partial^2}{\partial\theta^2}\log L(\theta,\widetilde{x})<0 : \text{condition du } 2^{ème} \text{ ordre} \end{cases}$$

Car la fonction log est continue et strictement monotone.

Méthode du maximum de vraisemblance

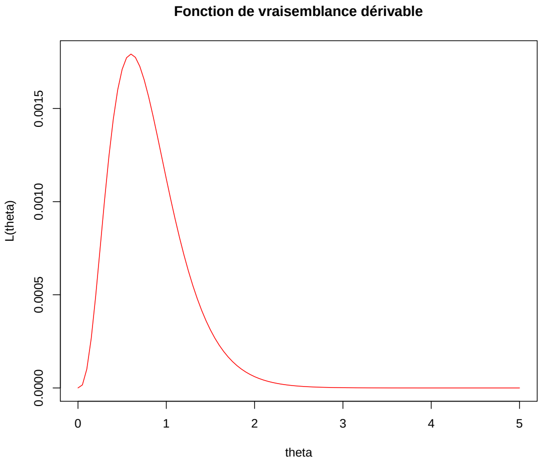

FIGURE: Fonction de vraisemblance

Méthode du maximum de vraisemblance

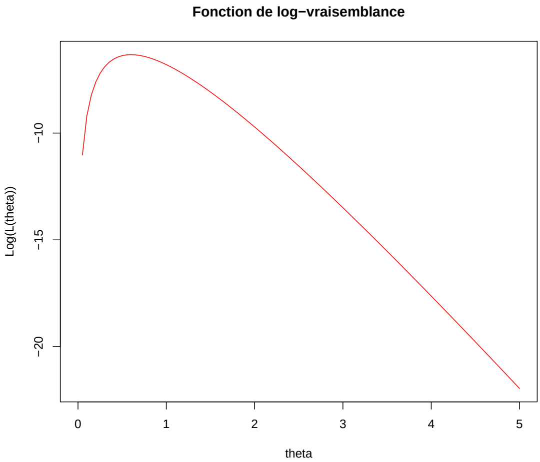

FIGURE: Fonction de log-vraisemblance

Méthode du maximum de vraisemblance

Cas 2 : θ réel et $L(\theta, \widetilde{x})$ non dérivable par rapport à θ
Il faut maximiser la vraisemblance directement

Cas 3 : $\theta = (\theta_1, \theta_2, .., \theta_p)$ vectoriel et $L(\theta, \widetilde{x})$ dérivable par rapport à θ

$$\begin{cases} \dfrac{\partial}{\partial \theta_1} \log L(\theta, \widetilde{x}) = 0 \\ \dfrac{\partial}{\partial \theta_2} \log L(\theta, \widetilde{x}) = 0 \\ ..... \\ \dfrac{\partial}{\partial \theta_p} \log L(\theta, \widetilde{x}) = 0 \end{cases} \quad \text{conditions du } 1^{er} \text{ ordre}$$

La condition du $2^{ème}$ ordre s'exprime à l'aide de la matrice Hessienne qui doit être définie négative.

Méthode du maximum de vraisemblance

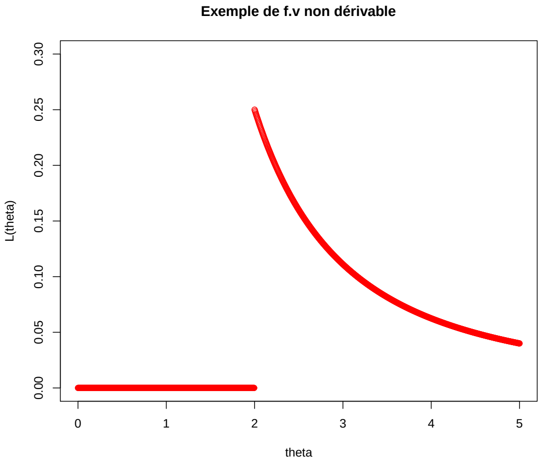

FIGURE: Fonction de vraisemblance

Méthode du maximum de vraisemblance

Théorème

Soit $\widetilde{x} = (x_1, x_2, ..., x_n)$ un n-échantillon d'une v.a X de loi $(f_\theta)_{\theta \in \Theta}$.
S'il existe une statistique exhaustive T pour θ alors l'estimateur du maximum de vraisemblance de θ est fonction de T.

Conséquence du théorème de factorème de factorisation

Théorème : invariance fonctionnelle

Soit $\widehat{\theta}$ l'estimateur du maximum de vraisemblance de θ alors l'estimateur du maximum de vraisemblance de $g(\theta)$ est $g(\widehat{\theta})$ pour toute fonction g.

Risque quadratique

Définition : Risque quadratique

Soit T un estimateur de $g(\theta)$
On appelle risque quadratique de l'estimateur T, l'erreur quadratique moyenne de T, i.e

$$R_\theta(T) = E_\theta((T(X) - g(\theta))^2)$$

Proposition

On a

$$R_\theta(T) = E_\theta((T(X) - E_\theta(T(X)))^2 + (E_\theta(T(X)) - g(\theta))^2$$

i.e

$$R_\theta(T) = Var(T) + biais(T)^2$$

Risque quadratique

Définition

Un estimateur T_1 est dit meilleur qu'un autre estimateur T_2 au sens du risque quadratique si

$$R_\theta(T_1) \leq R_\theta(T_2) \qquad \forall \theta \in \Theta$$

(uniformément en θ)

Proposition

Parmi tous les estimateurs sans biais le meilleur au sens du risque quadratique est celui qui a la plus petite variance.

$$Var_\theta(T_1) \leq Var_\theta(T_2) \qquad \forall \theta \in \Theta$$

pour tout T_2 autre estimateur sans biais.

Si un tel estimateur existe, il est alors dit estimateur sans biais de variance minimale (ou uniformément minimale) : ESBVUM.

Estimation sans biais

amélioration d'un estimateur

Soit T un estimateur sans biais de $g(\theta)$ et S une statistique exhaustive pour θ.
Alors
1) $E_\theta(T/S)$ est un estimateur sans biais de $g(\theta)$
2) $Var_\theta(E_\theta(T/S)) \leq Var_\theta(T)$

Conclusion : $E_\theta(T/S)$ est un meilleur estimateur sans biais de $g(\theta)$

Estimation sans biais

Démonstration :
$E_\theta(T/S)$ est un estimateur de $g(\theta)$ car ne dépend pas de θ.
1) $E_\theta(E_\theta(T/S)) = E_\theta(T) = g(\theta)$
(Théorème de l'espérance conditionnelle)
2) $Var_\theta(T) = Var_\theta(E_\theta(T/S)) + E_\theta(Var_\theta(T/S))$
(Théorème de la variance conditionnelle)
$\Longrightarrow Var_\theta(E_\theta(T/S)) = Var_\theta(T) - E_\theta(Var_\theta(T/S)) \leq Var_\theta(T)$
Remarque : Si $T = h(S)$ alors $E_\theta(h(S)/S) = h(S) = T$
Dans ce cas il n'y a pas d'amélioration.

Estimation sans biais

Théorème de Lehmann-Scheffé

Soit T un estimateur sans biais de $g(\theta)$ et S une statistique exhaustive et complète pour θ. Alors
$E_\theta(T/S)$ est l'unique estimateur sans biais de $g(\theta)$ dit estimateur sans biais de variance uniformément minimale (ESBVUM)

Démonstration

Estimation sans biais

Théorème de Rao-Cramer

Soit T un estimateur sans biais de $g(\theta)$. On suppose les conditions de régularité de Fischer vérifiées, alors

$$Var_\theta(T) \geq \frac{(g'(\theta))^2}{I(\theta)}$$

La quantité $\frac{(g'(\theta))^2}{I(\theta)}$ est dite borne de Rao-Cramer (BRC)

$$Var_\theta(T) = \frac{(g'(\theta))^2}{I(\theta)}$$

Estimation sans biais

Démonstration : Rappel :

$$|\rho_{X,Y}| = \left|\frac{Cov(X,Y)}{\sigma_X\sigma_Y}\right| \leq 1 \implies Cov(X,Y)^2 \leq \sigma_X^2\sigma_Y^2$$

$$E_\theta(T(X)) = \int_{\mathcal{X}} T(x)f_\theta(x)dx = \int_{\mathcal{X}} T(x)L(\theta,x)dx = g(\theta) \implies$$

$$\frac{\partial}{\partial\theta}(E_\theta(T(X)) = \int_{\mathcal{X}} T(x)\frac{\partial}{\partial\theta}(L(\theta,x))dx = g'(\theta)$$

$$= \int_{\mathcal{X}} T(x)\frac{\partial}{\partial\theta}(\log L(\theta,x))L(\theta,x)dx = Cov(T(X), \frac{\partial}{\partial\theta}(\log L(\theta,X))$$

$$\implies Cov^2(T(X), \frac{\partial}{\partial\theta}(\log L(\theta,X)) \leq Var_\theta(T(X))Var(\frac{\partial}{\partial\theta}(\log L(\theta,X))$$

$$\implies Var_\theta(T(X)) \geq \frac{Cov^2(T(X), \frac{\partial}{\partial\theta}(\log L(\theta,X))}{Var(\frac{\partial}{\partial\theta}(\log L(\theta,X))} = \frac{g'(\theta)^2}{I(\theta)}$$

Estimation sans biais

Cas particulier : $g(\theta) = \theta$

$BCR = \dfrac{1}{I(\theta)}$

Remarque :

Pour que $Var_\theta(T(X)) \geq \dfrac{Cov^2(T(X), \frac{\partial}{\partial\theta}(\log L(\theta, X))}{Var(\frac{\partial}{\partial\theta}(\log L(\theta, X))} = \dfrac{g'(\theta)^2}{I(\theta)}$

Il est nécessaire et suffisant que $\dfrac{\partial}{\partial\theta}(\log L(\theta, X) = k(\theta)T(X) + l(\theta)$

$$\implies E_\theta(\frac{\partial}{\partial\theta}(\log L(\theta, X)) = k(\theta)E_\theta(T(X)) + l(\theta)$$

$$\implies E_\theta(T(X)) = g(\theta) = -\frac{l(\theta)}{k(\theta)}$$

$$\iff \frac{\partial}{\partial\theta}(\log L(\theta, X) = k(\theta)(T(X) - g(\theta))$$

Estimation sans biais

Définition : estimateur efficace

Un estimateur sans biais T de $g(\theta)$ est dit efficace si

$$Var_\theta(T) = \frac{(g'(\theta))^2}{I(\theta)}$$

Information de Fischer : cas vectoriel

Cas d'un paramètre vectoriel $\theta = (\theta_1, \theta_2, ..., \theta_p)$
L'information de Fischer est alors donnée par la matrice des variances-covariances

$$\mathcal{I}(\theta) = \mathcal{V}(\mathit{Grad} \log L(\theta, X))$$

du vecteur $\mathit{Grad} \log L(\theta, X) = \begin{cases} \dfrac{\partial}{\partial \theta_1} \log L(\theta, X) \\ \dfrac{\partial}{\partial \theta_2} \log L(\theta, X) \\ . \\ . \\ \dfrac{\partial}{\partial \theta_p} \log L(\theta, X) \end{cases}$

Comme pour le cas réel on peut montrer que

$$\mathcal{I}(\theta) = -(E(\frac{\partial}{\partial \theta_i} \frac{\partial}{\partial \theta_j} \log L(\theta, X))_{1 \leq i \leq p, 1 \leq j \leq p},$$

Information de Fischer : cas vectoriel

Exemple

Soit x une observation de de $X \backsim N(m,\sigma^2)$

$$\mathcal{I}(m,\sigma^2) = -\begin{pmatrix} E(\frac{\partial^2}{\partial m^2}\log L(\theta,X) & E(\frac{\partial}{\partial m}\frac{\partial}{\partial \sigma^2}\log L(\theta,X) \\ E(\frac{\partial}{\partial m}\frac{\partial}{\partial \sigma^2}\log L(\theta,X) & E(\frac{\partial^2}{\partial (\sigma^2)^2}\log L(\theta,X) \end{pmatrix}$$

$$\mathcal{I}(m,\sigma^2) = -\begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

Théorème de Rao-Cramer : cas vectoriel

Soit deux matrices définies positives A et B, on dit que $A \geq B$ si
$Q_A(x) = x^t Ax \geq Q_B(x) = x^t Bx$
On rappelle que les matrices de variances-covariances sont définies positives.

Théorème de Rao-Cramer : cas vectoriel

Soit T un estimateur sans biais de $g(\theta)$ (à valeurs dans R^p). On suppose les conditions de régularité de Fischer vérifiées, alors

$$\mathcal{V}_\theta(T) \geq Grad(g(\theta)^t \mathcal{I}(\theta)^{-1} Grad(g(\theta)$$

$\mathcal{V}_\theta(T)$ est la matrice de variances-covariances de T.