Ch. 7 ESTIMATION PONCTUELLE

Despina Baskiotis

EISTI

12 décembre 2009

Introduction

Dans le Ch. précédent on a vu les propriétés (qualités) des *estimateurs :*

1. Biais : *Erreur systématique*
2. Risque Quadratique : La *précision de* l'estimateur
3. Efficacité relative : Comparaison des variances pour les *estimateurs sans biais*

Attention : Les différentes transformations ne conservent pas, en général, ces propriétés.

Pour obtenir des estimateurs meilleurs on doit :

1. Éliminer le biais
2. Minimiser le risque quadratique : c.a.d. minimiser sa variance car, $R_\theta(u) = V(U) + b^2$

On veut pouvoir définir dans la classe des estimateurs Sans Biais, l'estimateur qui aura *la plus petite variance.*

On obtiendra donc, un *Estimateur Sans Biais de Variance Minimale.*

Pour se faire on a besoin de :

1. Définir une borne inférieure pour la variance dans la classe des estimateurs sans biais
2. Mesurer, au sens statistique, la quantité de l'INFORMATION contenue dans un échantillon.

Pb : Quel sens donner au mot *Information*, au point de vue *statistique ?*

Fisher a proposé une définition.

Il a travaillé sur le $\mathrm{grad}\left[\ln(f(x;\theta))\right]$.

Cette définition nous permettra de définir l'efficacité absolue d'un estimateur.

Information de Fisher

On fait les hypothèses suivantes :

1. $f(x,\theta) > 0 \; \forall x \in S$ où S ensemble indépendant de $\theta \in \Theta \subset I\!R$.
2. Θ est un intervalle ouvert de $I\!R$.
3. $\frac{\partial}{\partial\theta} f(x,\theta)$ existe $\forall \theta \in \Theta$ et $\forall x \in S$.
4. $\sum_{x_1 \in S} \cdots \sum_{x_{n1} \in S} f(x_1,\theta) \cdots f(x_n,\theta)$ est différentiable.

Information de Fisher

Definition 1. *On appelle* **information de Fisher** *si elle existe la quantité :*

$$I(\theta) = E\left(\frac{\partial \ln f(\theta; X)}{\partial \theta}\right)^2$$

ou par la formule équivalente :

$$I(\theta) = -E\left(\frac{\partial^2 \ln f(\theta; X)}{\partial \theta^2}\right)$$

L'information $I(\theta)$ est donc toujours positive pour tout $\theta \in \Theta$.

Information contenue dans un échantillon

Definition 2. *Lorsque on considère un échantillon* $(X_1, \ldots, X_n)$ *,*

$I_n(\theta)$ *est l'information apportée par la réalisation* $(x_1, x_2, \ldots, x_n)$ *:*

$$I_n(\theta) = E\left(\frac{\partial \ln f(\theta; X_1, \ldots, X_n)}{\partial \theta}\right)^2$$

On montre que $\qquad I_n(\theta) = nI(\theta)$

Information contenue dans un échantillon

Preuve :

1. $I(\theta) = E\left(\frac{\partial \ln f(\theta;X)}{\partial\theta}\right)^2$ et pour un échantillon $\underline{X} = (X_1, X_2, ..., X_n)$,

   $I_n(\theta) = E\left(\frac{\partial \ln f(\theta;X_1,...,X_n)}{\partial\theta}\right)^2$

2. On pose $Y = I_n(\theta)$.
   Or $E(Y^2) = \mathrm{Var}(Y) - E^2(Y)$;

   On accepte sans démostration que $E^2(Y) = 0$,

   donc $E(Y^2) = \mathrm{Var}(Y)$.

Information contenue dans un échantillon

$$I_n(\theta) = \mathrm{Var}(Y) = Var(\frac{\partial \ln f\,(\theta; X_1, \ldots, X_n)}{\partial \theta}) = \mathrm{Var}\,\frac{\partial \ln}{\partial}\left(\prod_{i=1}^{n} f\,(\theta; X_i\ )\right)$$

$$= \mathrm{Var}\left(\sum_{i=1}^{n} \frac{\partial \ln}{\partial} f\,(\theta; X_i\ )\right) = nI(\theta).$$

Remarques

1. **Attention** cette quantité dépend **seulement de la loi**
2. $I(\theta)$ est toujours positive
3. $I(\theta)\ _{X,Y} = I(\theta)\ _X + I(\theta)\ _Y$

Il faut donc pouvoir définir une borne inférieure pour la variance d'un estimateur.

Borne inférieure de la variance

Le problème est résolu seulement pour les **estimateurs sans biais** par Fischer- Darmois- Cramer- Rao

Theorem 1. *Cramer-Rao : Sous certaines conditions,* **la variance minimale** *d'un estimateur* **sans biais,** *basé sur un échantillon de taille n :*

1. Si on veut estimer θ, est égale à :

$$\frac{1}{I_n(\theta)} = \frac{1}{nI(\theta)}$$

2. Si on veut estimer $\alpha(\theta)$ est égale à :

$$\frac{\left[\alpha'(\theta)\right]^2}{I_n(\theta)} = \frac{\left[\alpha'(\theta)\right]^2}{nI(\theta)}$$

Inégalité de Cramér – Rao

Pour tout estimateur **non biaisé** $U = U\left(X_1, \ldots, X_n\right)$ nous avons :

$$\sigma_\theta^2(U) \;\geq\; \frac{\left[\frac{d}{d\theta}\alpha(\theta)\right]^2}{n \cdot I(\theta)}$$

C'est *la Borne de Cramer-Rao*

Estimateur Efficace

Definition 3. *On dit qu'un estimateur* **sans biais** U *de* $\alpha(\theta)$ *(ou de* θ*) est* ***efficace***

s'il atteint la borne de Cramer-Rao.

C'est à dire si :

$$\sigma_\theta^2(U) = \frac{\left[\frac{d}{d\theta}\alpha(\theta)\right]^2}{n \cdot I(\theta)}$$

Construction des Estimateurs

Estimation par la méthode du maximum de vraisemblance

Fonction de Vraisemblance

Définition : On appelle **vraisemblance** de l'échantillon $(X_1, \ldots, X_n)$ la fonction

$$L(\ x_1, \ldots, x_n; \theta) = h(x_1, \ldots, x_n) \cdot P(X_1 = x_1, \ldots, X_n = x_n \mid \boldsymbol{\theta})$$

où $h(x_1, \ldots, x_n) > 0$; $\forall (x_1, \ldots, x_n) \in \mathbb{R}^n$
fonction qui ne dépend pas de $\boldsymbol{\theta}$, on a donc :

$$L\left(x_1, x_2, \ldots, x_n; \theta\right) = \prod_{i=1}^{n} P\left(X_i = x_i \mid \theta\right) \text{ si } X \text{ v.a. discrète}$$

et

$$L\left(x_1,x_2,...,x_n; \theta\right) = \prod_{i=1}^{n} f\left(x_i \mid \theta\right) \text{ si } X \text{ v.a. continue}$$

car les v.a.X_i sont indépendantes et de même loi que X.

Méthode du Maximum de Vraisemblance

Le principe :
On observe un échantillon :

Soit $(x_1, x_2, ..., x_n)$ une réalisation.

θ sera estimé par la valeur qui maximise $f(x_1, x_2, ..., x_n; \theta)$ (cas v.a. continue) ou bien

$P\left(X_1 = x_1, \ldots, X_n = x_n; \boldsymbol{\theta}\right)$ (cas v.a. discrète)
i.e
par la valeur de θ qui maximise la probabilité de la réalisation de l'événement $(x_1, x_2, ..., x_n)$**.**

Definition 4. *On appelle Estimateur du Maximum de Vraisemblance, E.M.V., la valeur* $\widehat{\theta}(x_1, x_2, ..., x_n)$ *de* θ *(si elle existe) qui maximise*

$$L\left(x_1, \ldots, x_n; \theta\right) = h\left(x_1, \ldots, x_n\right) \cdot P\left(X_1 = x_1, \ldots, X_n = x_n; \boldsymbol{\theta}\right)$$

si la v.a. X *est discrète et*

$$L\left(x_1, \ldots, x_n; \theta\right) = h\left(x_1, \ldots, x_n\right) \cdot f(x_1, x_2, ..., x_n; \theta)$$

si la v.a. est continue)

Étapes de calcul :

On a un problème de **maximisation d'une fonction**

1. Cas où θ est scalaire

   (a) On forme la fonction de vraisemblance

   $$L\left(x_1, x_2, \ldots, x_n; \theta\right) = \prod_{i=1}^{n} f\left(x_i \mid \theta\right) \text{ si } X \text{ v.a. continue} \quad (1)$$

   (b) On utiliser le logarithme de la fonction de vraisemblance :

   $$l\left(x_1, x_2, \ldots, x_n; \theta\right) = \ln(L\left(x_1, x_2, \ldots, x_n; \theta\right)) \qquad (2)$$

(c) On maximise cette dernière fonction :

Si les dérivées par rapport à θ existent :

$l'(\ x_1, \ldots, x_n; \theta) = \dfrac{\partial}{\partial \theta}\ l\,(x_1, \ldots, x_n; \theta)$

et

$l''(\boldsymbol{\theta}\ | x_1, \ldots, x_n) = \frac{\partial^2}{\partial \theta^2}\ l\,(x_1, \ldots, x_n; \theta)$

l'E.M.V. est calculé par les système des équations vraisemblance :

$$l'(x_1, \ldots, x_n; \theta) = 0$$

Une racine $\widehat{\theta}$ de cette équation pour laquelle nous avons

$$l''\left(\ x_1, \ldots, x_n; \widehat{\theta}\right) < 0$$

donne un maximum local.

On a :

$$\widehat{\theta} = \arg\max \left\{ l\left(\widehat{\theta}|x_1,\ldots,x_n\right) \ / \ l'\left(\widehat{\theta}|x_1,\ldots,x_n\right) = 0, l''\left(\widehat{\theta}|x_1,\ldots,x_n\right) < 0 \right\}$$

2. Cas où θ est un vecteur

   On travaille de la même façon :

   On maximise par rapport à un paramètre en fixant les autres

Theorem 2. *l'E.M.V est Invariant par rapport à une transformation bijective.*

Considérons la transformation bijective $\varphi : \Theta \to \Theta' \subset \mathbb{R}$.

Si $\widehat{\boldsymbol{\theta}}$ *est l'estimateur du maximum de vraisemblance de* $\boldsymbol{\theta}$, *alors* $\varphi\left(\widehat{\boldsymbol{\theta}}\right)$ *est l'estimateur du maximum de vraisemblance de* $\varphi(\boldsymbol{\theta})$.

Conséquences de la définition de l'E.M.V.

L'E.M.V

1. n'a aucune raison d'être *unique*.
2. peut avoir un *biais*
3. peut ne pas exister, car la fonction de vraisemblance n'est pas toujours dérivable.
   Dans ce cas on utilise des méthodes numériques pour calculer le max local de la fonction de vraisemblance

Exemple

Soit $X \sim B(p), \theta = p$ inconnue.

1.Calcul de la B.C.R. pour $B(p)$

On a : $P(X_i = x_i) = \theta^{x_i}(1-\theta)^{1-x_i}$

1. $I_1(\theta) = E\left(\dfrac{\partial \ln f\,(\theta; X)}{\partial \theta}\right)^2$
2. Or $\ln(P(X_i = x_i)) = x_i \ln \theta + (1-x_i)\ln(1-\theta)$
3. $\dfrac{\partial \ln f\,(x;\theta)}{\partial \theta} = \dfrac{x_i}{\theta} - \dfrac{1-x_i}{1-\theta} = \dfrac{x_i - \theta}{\theta(1-\theta)}$

4. $I_1(\theta) = E\left(\dfrac{\partial \ln f\,(\theta; X)}{\partial\theta}\right)^2 = \dfrac{1}{\theta^2(1-\theta)^2} E(X_i - \theta)^2 = \dfrac{1}{\theta(1-\theta)}$

5. Pour un $n-$échantillon on aura : $I_n(\theta) = \frac{n}{\theta(1-\theta)}$

6. B.C.R. : $\frac{\theta(1-\theta)}{n}$

Donc la variance d'un estimateur sans biais pour $\theta = p$ sur un $n-$échantillon, ne peut pas être inférieure à $\frac{\theta(1-\theta)}{n}$

2. E.M.V. pour $\theta = p$

1. On forme la fonction de vraisemblance
$$L\left(x_1, x_2, ..., x_n; \theta\right) = \prod_{i=1}^{n} f\left(x_i \mid \theta\right) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$
$$= \theta^{\sum\limits_{i=1}^{n} x_i}(1-\theta)^{n-\sum\limits_{i=1}^{n} x_i}$$
2. On utiliser le logarithme de la fonction de vraisemblance :
$$l\left(x_1, x_2, ..., x_n; \theta\right) = \ln(L\left(x_1, x_2, ..., x_n; \theta\right))$$
$$= \sum_{i=1}^{n} x_i \ln\theta + (n - \sum_{i=1}^{n} x_i)\ln(1-\theta)$$
3. On maximise cette dernière fonction :
$$l'\left(x_1, \ldots, x_n; \theta\right) = \frac{\sum\limits_{i=1}^{n} x_i}{\theta} - \frac{(n - \sum\limits_{i=1}^{n} x_i)}{1-\theta} = \frac{\sum\limits_{i=1}^{n} x_i - n\theta}{(1-\theta)\theta} \quad (I)$$

4. La valeur de θ qui annule (I) est : $\widehat{\theta} = \dfrac{\sum\limits_{i=1}^{n} x_i}{n}$

5. On vérifie que la dérivée seconde est négative pour $\theta = \widehat{\theta}$

6. L'E.M.V pour $\theta = p$ est $\widehat{\theta} = \dfrac{\sum\limits_{i=1}^{n} X_i}{n}$.

3. Propriétés de $f_n = \frac{\sum_{i=1}^{n} x_i}{n}$

1. On a déjà vu que $E(f_n)$ est sans biais et convergent.
2. On verra si f_n est *Efficace*
   On calcule $Var(f_n) = \frac{\theta(1-\theta)}{n}$; $\quad p = \theta$
3. La B.C.R. étant $\frac{\theta(1-\theta)}{n}$
   On obtient que $\widehat{\theta} = f_n$ est efficace.
4. On obtient que f_n est un estimateur efficace,
   Conclusion : f_n *sans biais et de variance minimale*