Stats

Qualité d'un estimateur :

L'erreur quadratique moyenne EQ (Tn) = $E_\theta((Tn-\theta)^2)$

Th : EQ(Tn) = V(Tn) + ${b_n}^2\ (\theta)$

Preuve

EQ (Tn) = $E_\theta((Tn-\theta)^2) = E_\theta([(Tn-E[Tn]) + (E[Tn]-\theta)]^2)$

$$= E[(T_n - E[T_n])^2] - E[(E[T_n]-\theta)^2] + 2E[(T_n - E[T_n])(E[T_n]-\theta)]$$

EQ(Tn) = Var(Tn) + $b_n^2(\theta) + 2(E[T_n]-\theta) - E[(T_n - E[T_n])] = E[T_n] - E\big[E[T_n]\big]$

$$= E[T_n] - E[T_n] = 0$$

Rq : Si tn est un estimateur sans biais de θ, alors EQ(Tn) = Var(Tn)

- ➔ Un estimateur sans biais est d'autant meilleur que sa variance est petite

Def : Soit Tn et Tn' deux estimateurs sans biais de θ

On dit que Tn est plus efficace que Tn' si Var(Tn) ≤ Var(Tn') pour tout $\theta \in \Theta$ et ceci à partir d'un certain ranf pour n

Borne de Cramer-Rao

La borne de Cramer-Rao est la valeur qui minimise la variance associée aux estimateurs sans biais

Def : On appelle information de Fisher la quantité,

$I_n\ (\theta) = E\left[\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right]$ où L est la vraisemblance associée à l'échantillon (X_1, …, X_n)

Rq : $I_n(\theta)$ est une quantité fixe qui en général depend de θ

Th : Sous certaines hypothèses (non détaillés dans leur totalité mais savoir qu'il y a un maximum Θ ouvert et f fonction dérivable par rapport à θ) et si $X(\Omega)$ est un domaine indépendant de θ, alors si Tn est un estimateur sans biais de θ, on a

$$Var(T_n) \geq \frac{1}{\underbrace{I_n(\theta)}_{borne\ de\ cramer\ rao}}$$

Rq : Si $X(\Omega)$ est indépendant de θ, alors $I_n\ (\theta) = E\left[\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right]$

Soit une va de loi $\mathcal{E}(\frac{1}{\theta})$

On veut estimer θ

On dispose de X_1, … X_n n va indépendantes et identiquement distribuées de loi $\mathcal{E}(\frac{1}{\theta})$

$$f(x) = \left| \begin{array}{l} \frac{1}{\theta} e^{-\frac{x}{\theta} si\ x\geq 0} \\ 0\ sinon \end{array} \right.$$

L(x_1, .., x_n) = $\prod_{i=1}^{n} f_\theta(xi) = \left| \begin{array}{c} \left(\frac{1}{\theta}\right)^n e^{-\sum\frac{xi}{\theta}}\ si\ pour\ tout\ i\ x_i \geq 0 \\ 0\ sinon \end{array} \right.$

X(Ω) = [0, +∞[(domaine indépendant de θ)

Soit x_1, … x_1 ∈ $X(\Omega)$

Ln L(x_1, …, x_n ; $\theta) = -n \ln \theta - \frac{\sum xi}{\theta}$

$$\frac{\eth \ln L}{\eth\ \theta} = -\frac{n}{\theta} + \frac{\sum xi}{\theta^2}$$

$$\frac{\eth^2 \ln L}{\eth\ \theta^2} = \frac{n}{\theta^2} - \frac{2\sum xi}{\theta^3}$$

Un estimateur du maximum de vraisemblance de θ est $\overline{Xn} = \frac{1}{n}\sum x_i$

$I_n(\theta) = E[-\frac{d^2 lnL}{d\theta^2}]$ = - n/$\theta + \frac{2}{\theta^3} E[\sum xi] = \sum E[xi] = \sum E[X] = nE[X] = n\theta$

$I_n(\theta) = -\frac{n}{\theta^2} + \frac{2n\theta}{\theta^3} = \frac{n}{\theta^2}$

$I_n\ (\theta) = E[(\frac{\eth lnL}{\eth\theta})^2]$ = E[($-\frac{n}{\theta} + \frac{\sum xi}{\theta^2}$) ^{2}] = E[($\frac{n^2}{\theta}$ + $\frac{(\sum xi)^2}{\theta^4} - \frac{2n\sum xi}{\theta^3}$]

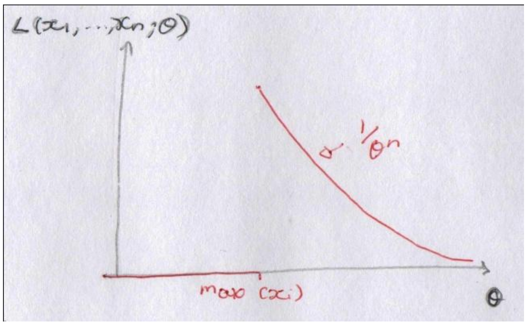
$I_n(\theta) = \frac{n^2}{\theta^2} - \frac{2n}{\theta^3}\ E[\sum xi] + \frac{1}{\theta^4}\ E[(\sum x)^2]$

$E[(\sum xi)^2] = var\ (\sum xi) + E^2[\sum xi]$ car Var(y) = E[y²] – E²[y]²

$= \sum var(xi) + n^2 E[(\sum x)^2]^2 = n\ var(x) + n^2\theta^2 = n\theta^2 + n^2\theta^2$

$$L(x1, .., xn;\ \theta) = \left| \begin{array}{c} \left(\frac{1}{\theta}\right)^n\ si\ pour\ i\ xi\ \in [0, \theta] \\ 0\ sinon \end{array} \right.$$

Dire Vi, xi ∈ $[0, \theta] \rightarrow \left| \begin{array}{l} \min(xi) \geq 0 \\ \max(xi) \leq \theta \end{array} \right.$



➔ Un estimateur du max de vraisemblance est max (X_i)

Def : Un estimateur sans biais de θ, noté Tn, est efficace si Var(Tn) = $\frac{1}{I_n\,(\theta)}$

Ex : X ~$\mathcal{E}(\frac{1}{\theta})$, X1, …, Xn n echantillon de loi $\mathcal{E}(\frac{1}{\theta})$ on a vu précédemment que

Tn = $\overline{Xn} = \frac{1}{n}\sum xi \rightarrow estimateur\ sans\ biais\ de\ \theta$

In$(\theta) = \frac{n}{\theta^2}$

Var(Tn) = Var($\frac{1}{n}\sum xi$) = $\frac{1}{n^2}$ Var($\sum xi) = \frac{n\theta^2}{n^2} = \frac{\theta^2}{n}$

Var(Tn) = $\frac{1}{I_n(\theta)}$

Tn est un estimateur efficace

Rq : un estimateur sans biais et efficace est le meilleur estimateur possible, au cou de l'erreur quadratique moyenne, où xi des estimateurs sans biais

Cependant il est parfois possible de trouver un estimateur biaisé qui sera meilleur que l'estimateur sans biais efficace

Ex : X~N(0, σ^2) : paramètre à estimateur est σ^2

X_1, … X_n n échantillon de mm loi que X

Soit $\hat{\sigma}_n^2 = \frac{1}{n}\sum x_i$ estimateur sans biais et efficace

Tn = $\frac{1}{n} + 2\ \sum {x_i}^2$: estimateur biaisé de σ^2

EQ(Tn) = $\frac{2}{n+2}\sigma^2$ < EQ($\hat{\sigma}_n^2$) =$\frac{2}{n}\sigma^2$

Retour aux probabilités

- loi de chi-deux
- loi de Student
- loi de Fisher

- la loi du chi-deux à n degré de liberté ($X^2(n)$)

On dit que $X \sim X^2(n)$ si

$X = Y_1^2 + Y_2^2 + \ldots + Y_n^2$ avec $Y_1, \ldots, Y_n$ n variable aléatoire indépendantes de loi N(0,1)

Prop : $X \sim X^2(n)$

$E[X] = n$

$V[X] = 2n$

degré de liberté	0,99	0,9	...
1	Seuil		
2			
⋮			
		P(X≤t)	

- La loi de Student

Soit $U \sim \mathcal{N}(0,1)$ et soit $V \sim \mathcal{X}^2(n)$ avec U et V indépendantes

$$T = \frac{U}{\sqrt{\frac{1}{n}}} \sim T(n)$$

$$S_n^2 = \frac{1}{n-1} \sum (Xi - \overline{Xn})^2 avec\ Xi \sim N(m, \sigma^2)$$

$\overline{Xn}$ estimateur sans biais de m

$\overline{X_n} \sim \mathcal{N}(0, \frac{\sigma^2}{n})$ -> $\overline{X_{n-m}} \sim \mathcal{N}(0, \frac{\sigma^2}{n})$

$\sqrt{n} * \frac{\overline{X_{n-m}}}{\sigma} \sim \mathcal{N}(0,1)$

$$\left.\frac{\overline{X_{n-m}}}{\sqrt{\frac{S_{n}2}{n}}} = \frac{\frac{\overline{X_{n-m}}}{\sigma} \sim \mathcal{N}(0,1)}{\sqrt{\frac{S_{n}2}{\sigma^2}\Big/ n \sim \mathcal{X}^2(n)}}\right\} \text{2 variables indépendantes}$$