MLE

20:19

consistent estimator

$$\lim_{n\to\infty} P_\theta(|W_n-\theta|\geq\epsilon)=0 \quad ①$$

by chebychev inequality:

$$P_\theta(|W_n-\theta|\geq\epsilon)\leq\frac{E_\theta(|W_n-\theta|^2)}{\epsilon^2}=0 \quad ②$$

思路: 证明 consistent, 将 ① 转化为 ②

关键在于凑 $E(\hat\theta)$

此时是 $\hat\theta$ 的 var / bias

(1) 凑 $E(\hat\theta)$ make sense

(2) $\hat\theta$ R.V. ; θ, $E\hat\theta$ scalar

$$E_\theta(|\hat\theta-\theta|^2)=E_\theta(\hat\theta-\underline{E(\hat\theta)+E(\hat\theta)}-\theta)^2$$
$$=E_\theta(\hat\theta-E(\hat\theta))^2+E_\theta(E(\hat\theta)-\theta)^2+2E_\theta(\hat\theta-E\hat\theta)(E\hat\theta-\theta)$$
$$=Var(\hat\theta)+[E(\hat\theta-\theta)]^2+2(E\hat\theta-\theta)\underbrace{E(\hat\theta-E\hat\theta)}_{=0}$$
$$=var(\hat\theta)+bias^2(\hat\theta)$$

(3) 了解 var, bias 定义

$Var\,X=E(X-EX)^2$

$bias\,\hat\theta=E(\hat\theta-\theta)$

$\Rightarrow$ (1) $\lim_{n\to\infty} Var(\hat\theta)=0$ (2) $\lim_{n\to\infty} bias(\hat\theta)=0$ $\Rightarrow\hat\theta$ consistent

Thm. consistency of MLE

$$L_n(\theta|x)=\prod_{i=1}^n f(x_i|\theta)$$

$$\ell_n(\theta|x)=\sum_{i=1}^n\log f(x_i|\theta)$$

proof.

sample mean $\xrightarrow{P}$ population mean

by WLLN. $\frac{1}{n}\sum_{i=1}^n\log f(x_i|\theta)\xrightarrow{P}E_{\theta_0}(\log f(x|\theta))$

思路: $\theta_{MLE}\xrightarrow{P}\theta'$

$$E_{\theta_0}(\log f(x|\theta))-E_{\theta_0}(\log f(x|\theta_0))=E\left(\log\frac{f(x|\theta)}{f(x|\theta_0)}\right)$$

by Jensen's inequality $\log x$ 求二阶导 $\frac{1}{x}\to-\frac{1}{x^2}<0$ concave

$$\therefore E_{\theta_0}\left(\log\frac{f(x|\theta)}{f(x|\theta_0)}\right)\leq\log E_{\theta_0}\frac{f(x|\theta)}{f(x|\theta_0)}=\log\int\frac{f(x|\theta)}{f(x|\theta_0)}f(x|\theta_0)dx$$
$$=\log 1=0$$

$\therefore\theta_{MLE}\xrightarrow{P}\theta_0$

Thm asymptotic normality of MLE

思路: score function $\ell'(\theta)$ 泰勒展开 (θ_0)

$$f(x)=f(x_0)+(x-x_0)f'(x_0)+(x-x_0)^2\frac{f''(x_0)}{2!}$$

$$\ell'(\theta)=\ell'(\theta_0)+(\theta-\theta_0)\cdot\ell''(\theta_0)$$

plugin $\hat\theta_{MLE}$

$$\ell'(\hat\theta_{MLE})=\ell'(\theta_0)+(\hat\theta_n-\theta_0)\ell''(\theta_0)$$

$\hat\theta_{MLE}$ 是由 $\ell'(\theta)=0$ 求得

0

$$\hat\theta_n-\theta_0=-\frac{\ell'(\theta_0)}{\ell''(\theta_0)}$$

看起来很像 asymptotical 的式子, 凑一个 $\sqrt{n}$

$$\sqrt{n}(\hat\theta_n-\theta_0)=-\frac{\sqrt{n}\ell'(\theta_0)}{\ell''(\theta_0)}$$

$\sqrt{n}\ell'(\theta_0)\to$ score function

$\ell''(\theta_0)\to$ FI

①

$$\Rightarrow\sqrt{n}(\hat\theta-\theta_0)\;\dot\sim\;N\left(0,\frac{1}{I(\theta_0)}\right)$$

$$\frac{\sqrt{n}(\hat\theta-\theta_0)}{\sqrt{\frac{1}{I(\theta_0)}}}\;\dot\sim\;N(0,1) \quad \text{Score test}$$

其中, ① $\ell'(\theta_0)=\frac{d\Sigma\log f(x_i|\theta_0)}{d\theta_0}$

$$=\sum_{i=1}^n\frac{d\log f(x_i|\theta_0)}{d\theta_0}$$

凑 sample mean

$$\therefore\frac{1}{n}\ell'(\theta_0)=\underbrace{\frac{1}{n}\sum_{i=1}^n\frac{d\log f(x_i|\theta_0)}{d\theta_0}}_{\text{define as } Z}$$

WLLN ✗

$$\xrightarrow{P}E_{\theta_0}\left(\frac{d\log f(x|\theta_0)}{d\theta_0}\right)$$

待补充: E(score), var(score)

CLT sample mean $\sim\mathcal{N}$

$$\sqrt{n}(Z-EZ)\sim N(0,Var Z)\Rightarrow\sqrt{n}Z\sim N(0,I(\theta_0))$$
$$\Rightarrow\frac{1}{\sqrt{n}}\cdot\sqrt{n}Z\sim N\left(0,\frac{1}{n}I(\theta_0)\right)$$

$$\therefore\frac{1}{n}\ell'(\theta_0)=Z$$

$$\frac{1}{n}\ell'(\theta_0)=Z\sim N\left(0,\frac{1}{n}I(\theta_0)\right)$$

$$\ell'(\theta_0)\;\dot\sim\;N(0,nI(\theta_0)) \quad ☆ ②$$

$$① \; -\sqrt{n}\frac{N(0,nI(\theta_0))}{n\cdot-I(\theta_0)}$$

$\to\ell'(\theta_0)$ ②

$\to\ell''(\theta_0)$ ③

$$=\frac{N(0,nI(\theta_0))}{\sqrt{n}I(\theta_0)}$$

$$\Rightarrow N\left(0,\frac{1}{I(\theta_0)}\right)$$

$n\to$ 这里稍微约了个 n.

continuous mapping

slutsky Thm.

$$\frac{n(N(0,I_n(\theta_0)))}{I_n(\theta_0)^2}=$$

② Denominator $\ell''(\theta_0)=\frac{d\sum_{i=1}^n\log f(x_i|\theta_0)}{d\theta_0}$

$$=\sum_{i=1}^n\frac{d\log f(x_i|\theta_0)}{d\theta_0}$$

$$\therefore\frac{1}{n}\ell''(\theta_0)=\frac{1}{n}\sum_{i=1}^n\frac{d^2\log f(x_i|\theta_0)}{d\theta_0^2}$$

写为 score function 的导数形式

$$=\frac{1}{n}\sum_{i=1}^n\left[\frac{d\log f(x_i|\theta_0)}{d\theta_0}\right]'$$

WLLN

$$\xrightarrow{P}E_{\theta_0}\left(\left[\frac{d\log f(x_i|\theta_0)}{d\theta_0}\right]'\right)=-I(\theta_0)$$

待补充

对 score 求导

对 likelihood 求二阶导

⇓

取 expectation, 取负

可得到 $I(\theta)$

$$\ell''(\theta_0)\xrightarrow{P}n\cdot-I(\theta_0) \quad ☆ ③$$

求 score function 的 mean & variance

$$\ell_n(\theta|x)=\sum_{i=1}^n\log f(x_i|\theta)$$

$$\frac{\partial\ell_n(\theta|x)}{\partial\theta}=\sum_{i=1}^n\frac{\partial\log f(x_i|\theta)}{\partial\theta}\Rightarrow\text{score function}$$

$$E\left(\Sigma\frac{\partial\log f(x_i|\theta)}{\partial\theta}\right)=\sum_{i=1}^n E\left(\frac{\partial\log f(x_i|\theta)}{\partial\theta}\right)$$
$$=\sum_{i=1}^n\int_x\frac{\partial\log f(x_i|\theta)}{\partial\theta}f(x_i|\theta)dx$$
$$=\sum_{i=1}^n\int_x\frac{1}{f(x_i|\theta)}\cdot\frac{\partial f(x_i|\theta)}{\partial\theta}\cdot f(x_i|\theta)dx$$
$$=\sum_{i=1}^n\int_x\frac{\partial f(x_i|\theta)}{\partial\theta}dx$$
$$=\sum_{i=1}^n\frac{\partial}{d\theta}\int f(x_i|\theta)dx$$
$$=0$$

对 E(score) 求导 =0

$$\frac{\partial E\left(\sum_{i=1}^n\frac{\partial\log f(x_i|\theta)}{\partial\theta}\right)}{\partial\theta}=\frac{\partial}{\partial\theta}\int_x\sum_{i=1}^n\frac{\partial\log f(x_i|\theta)}{\partial\theta}\cdot f(x_i|\theta)dx$$
$$=\int\frac{\partial}{\partial\theta}\Big(\underbrace{\Sigma\frac{\partial\log f(x_i|\theta)}{\partial\theta}}_{A}\cdot\underbrace{f(x_i|\theta)}_{B}\Big)dx$$

$(A\cdot B)'=A'B+B'\cdot A$

$$=\int\sum_{i=1}^n\left[\frac{\partial^2\log f(x_i|\theta)}{\partial\theta^2}\cdot f(x_i|\theta)+\frac{\partial\log f(x_i|\theta)}{\partial\theta}\cdot\frac{\partial f(x_i|\theta)}{\partial\theta}\right]dx$$
$$=\int\sum_{i=1}^n\frac{\partial^2\log f(x_i|\theta)}{\partial\theta^2}\cdot f(x_i|\theta)dx+\int\Sigma\frac{\partial\log f(x_i|\theta)}{\partial\theta}\cdot\frac{\partial f(x_i|\theta)}{\partial\theta}dx$$

可以凑为 $\frac{\partial\log f(x_i|\theta)}{\partial\theta}$

$$=E\left(\Sigma\frac{\partial^2\log f(x_i|\theta)}{\partial\theta^2}\right)+\int\Sigma\frac{\partial\log f(x_i|\theta)}{\partial\theta}\cdot\frac{\partial f(x_i|\theta)}{\partial\theta}f(x_i|\theta)\cdot\frac{1}{f(x_i|\theta)}dx$$
$$=E\left(\Sigma\frac{\partial^2\log f(x_i|\theta)}{\partial\theta^2}\right)+\int\Sigma\left(\frac{\partial\log f(x_i|\theta)}{\partial\theta}\right)^2\cdot f(x_i|\theta)dx$$
$$=E\left(\Sigma\frac{\partial^2\log f(x_i|\theta)}{\partial\theta^2}\right)+\sum_{i=1}^n E\left(\frac{\partial\log f(x_i|\theta)}{\partial\theta}\right)^2$$

$\to$ Var(score) as E(score) = 0

$=E(score^2)-[E(score)]^2$

$=E(score^2)-0$

Var(score)

$$\therefore \textbf{Var}(score)\overset{②}{=}-E\left(\Sigma\frac{\partial^2\log f(x_i|\theta)}{\partial\theta^2}\right)$$

$=-E$ (score 求导) $=-E$ (likelihood 求二阶导)

$$=E\left(\Sigma\left(\frac{\partial\log f(x_i|\theta)}{\partial\theta}\right)^2\right)\overset{①}{=}\mathcal{I}(\theta)$$

$I(\theta)$: Fisher information is { the variance of the score / the expected value of observed information }

$$I=-\Sigma\frac{\partial^2\log f(x_i|\theta)}{\partial\theta^2}$$

likelihood 求二阶导, 取负号

$$Var(\hat u)=\frac{1}{FI}=\frac{1}{E(-\ell''(\hat u))}$$

$$Var(\hat u)=\frac{1}{V(score)}=\frac{1}{var(\ell'(\hat u))}$$

(→ if unbiased.) check CRLB.

$$\Rightarrow E(-\ell''(u))=var(\ell'(\hat u))$$

$$I(\theta)=-\ell''(\theta)>0 \quad \text{information function}$$

$$\mathcal{I}(\theta)=E(I(\theta)) \quad \text{Expected information function}$$

estimator W_n is asymptotically efficient for param $\tau(\theta)$ if

$$\sqrt{n}[W_n-\tau(\theta)]\xrightarrow{d}N(0,var(\theta))$$

and

$$var(\theta)=\frac{[\tau'(\theta)]^2}{E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)^2\right)}$$

cramér-rao lower bound:

$$Var_\theta(W(\mathbf{X}))\geq\frac{\left(\frac{d}{d\theta}E_\theta W\mathbf{X}\right)^2}{E_\theta\left(\left[\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)\right]^2\right)}$$

MLE is asymptotic efficient: achieve CRLB.

→ defined only when estimator is asymp. normal.

asym. normality $\xrightarrow{\text{imply}}$ consistency

estimator W_n. V_n. satisfy

$$\sqrt{n}(W_n-\tau(\theta))\xrightarrow{d}N(0,\sigma_W^2)$$

$$\sqrt{n}(V_n-\tau(\theta))\xrightarrow{d}N(0,\sigma_V^2)$$

asymptotic relative efficiency (ARE) of V_n, w.r.t. W_n is

$$ARE(V_n,W_n)=\frac{\sigma_W^2}{\sigma_V^2}$$