# 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

关键步：凑 $E(\hat\theta)$

(1) 凑 $E(\hat\theta)$ make sense 时是否的 var/bias

(2) $\hat\theta$ R.V.

$\theta$, $E\hat\theta$ scalar

$$E_\theta(|\hat\theta-\theta|^2)=E_\theta|\hat\theta-\underline{E(\hat\theta)+E(\hat\theta)}-\theta|^2$$

$$=E_\theta(\hat\theta-E(\hat\theta))^2+E_\theta(E(\hat\theta)-\theta)^2+2E_\theta(\underset{\text{R.V.}}{\hat\theta-E\hat\theta})(E\hat\theta-\theta)$$

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

$$=var(\hat\theta)+bias^2(\hat\theta)$$

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

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

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

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

$$= \log 1 = 0$$

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

Thm asymptotic normality of MLE

思路: score function $l'(\theta)$ 泰勒展开 $(\theta_0)$

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

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

plugin $\hat{\theta}_{MLE}$

$$l'(\hat{\theta}_{MLE}) = l'(\theta_0) + (\hat{\theta}_n - \theta_0) l''(\theta_0)$$

$\hat{\theta}_{MLE}$ 是由 $l'(\theta) = 0$ 求得

0

$$\hat{\theta}_n - \theta_0 = -\frac{l'(\theta_0)}{l''(\theta_0)}$$

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

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\frac{\sqrt{n}\, l'(\theta_0)}{l''(\theta_0)}$$

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

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

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

$$\frac{\sqrt{n}(\hat{\theta} - \theta_0)}{\sqrt{\frac{1}{I(\theta_0)}}} \sim N(0, 1)$$

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

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

凑 sample mean

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

taking 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 ~ $\mathcal{N}$

$\sqrt{n}(Z - EZ) \sim N(0, \mathrm{Var} Z) \Rightarrow \sqrt{n}\, Z \sim N(0, I(\theta_0))$

$\Rightarrow \frac{1}{\sqrt{n}} \cdot \sqrt{n} Z \sim N(0, \frac{1}{n} I(\theta_0))$

$\because \frac{1}{n} l'(\theta_0) = Z$

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

$l'(\theta_0) \dot\sim N(0, n I(\theta_0))$ ☆ ②

① $-\sqrt{n}\, \dfrac{N(0, n I(\theta_0))}{n \cdot -I(\theta_0)}$ → $l'(\theta_0)$ ② , $l''(\theta_0)$ ③

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

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

continuous mapping

Slutsky Thm.

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

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

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

写为 score function 的导数形式

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

待补充

WLLN

WLLN

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

对 score 求导

对 likelihood 求二阶导

⇓

取 expectation，取负

可得到 $I(\theta)$

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

求 score function 的 mean & variance

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

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

$$E\left(\sum\frac{\partial \log f(x_i|\theta)}{\partial\theta}\right) = \sum_{i=1}^{n} E\left(\frac{\partial \log f(x_i|\theta)}{\partial\theta}\right)$$

$$= \sum_{i=1}^{n}\int_x \frac{\partial \log f(x_i|\theta)}{\partial\theta} f(x_i|\theta)\,dx$$

$$= \sum_{i=1}^{n}\int_x \frac{1}{f(x_i|\theta)}\cdot\frac{\partial f(x_i|\theta)}{\partial\theta}\cdot f(x_i|\theta)\,dx$$

$$= \sum_{i=1}^{n}\int_x \frac{\partial f(x_i|\theta)}{\partial\theta}\,dx$$

$$= \sum_{i=1}^{n}\frac{\partial}{d\theta}\int f(x_i|\theta)\,dx$$

$$= 0$$

at E(score) 求导 = 0

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

$$= \int \frac{\partial}{\partial\theta}\Big(\underbrace{\sum \frac{\partial \log f(x_i|\theta)}{\partial\theta}}_{A}\cdot \underbrace{f(x_i|\theta)}_{B}\Big)\,dx$$

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

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

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

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

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

$$= E\left(\sum \frac{\partial^2 \log f(x_i|\theta)}{\partial\theta^2}\right) + \int \sum\left(\frac{\partial \log f(x_i|\theta)}{\partial\theta}\right)^2\cdot f(x_i|\theta)\,dx$$

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

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

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

$= E(score^2) - 0$

$Var(score)$

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

$= E(\text{score 求导}) = -E(\text{Likelihood 求二阶导})$

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

$I(\theta)$ : Fisher information is $\begin{cases} \text{the variance of the score} \\ \text{the expected value of observed information} \end{cases}$

: $I = -\sum \dfrac{\partial^2 \log f(x_i|\theta)}{\partial \theta^2}$

Likelihood 求二阶导, 取负号

$$\left.\begin{aligned} Var(\hat{u}) &= \frac{1}{FI} = \frac{1}{E(-l''(\hat{u}))} \\ Var(\hat{u}) &= \frac{1}{V(score)} = \frac{1}{var(l'(\hat{u}))} \end{aligned}\right\} \Rightarrow E(-l''(u)) = var(l'(u))$$