Lecture 6. Rao-Cramer bound.

1

~~fo~~ finite sample || Before unbiasedness and efficiency,

Asymptotic properties: consistency

$\cdot$ CLT

$\rightarrow$ use LLN & cont. mapping theor & Delta method

non-parametric est.

We say that $\hat{\theta}_n$ is *consistent* for θ if $\hat{\theta}_n \rightarrow_p \theta$.

$X_1, \dots, X_n \sim \text{iid } F \quad \mu = EX_i$

$$\hat{\mu} = \bar{x}_n \xrightarrow{P} \mu \quad \text{as } n \to \infty$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^{n} x_i^2 - \frac{n}{n-1} (\bar{x})^2$$

$$\frac{1}{n} \sum_{i=1}^{n} X_i^2 \xrightarrow{P} E X_i^2$$

$$\bar{x} \xrightarrow{P} EX_i = u$$

$$\downarrow P$$

$$E X_i^2 - (E X_i)^2$$

$$= Var(X_i)$$

CMT: $\frac{n}{n-1} \to 1$

Asymptotic normality

We say that $\hat{\theta}$ is *asymptotically normal* if there are sequences $\{a_n\}_{n=1}^{\infty}$ and $\{r_n\}_{n=1}^{\infty}$ and constant σ^2 such that $r_n(\hat{\theta} - a_n) \Rightarrow N(0, \sigma^2)$.
Then r_n is called the *rate of convergence*, a_n - the *asymptotic mean*, and σ^2 - the *asymptotic variance*.
In many cases, one can choose $a_n = \theta$ and $r_n = \sqrt{n}$.

$\sqrt{n}(\hat{\theta} - \theta) \Rightarrow N(0, \sigma^2_{\theta})$

$\bar{X} = \hat{\mu} \quad EX_i = \mu$

$y_i = X_i - \mu$

$E Y_i = 0$

$E Y_i^2 = \sigma^2$

By CLT: $\sqrt{n}(\hat{\mu} - \mu) \Rightarrow N(0, \sigma^2) \quad \sigma^2 = \mathrm{Var}(X_i)$

$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 = \underbrace{\frac{n}{n-1} \frac{1}{n} \sum_i y_i^2 - \frac{n}{n-1} (\bar{y})^2}_{\text{from day 2 ex.}}$

$y_i - \bar{Y} = X_i - \bar{X}$

$\sqrt{n}\left[\hat{\sigma}^2 - \sigma^2\right] = \frac{n}{n-1} \cdot \sqrt{n}\left[\frac{1}{n}\sum y_i - \sigma^2\right] + \frac{\sqrt{n}}{n-1}\sigma^2 - \frac{n}{n-1}\sqrt{n}(\bar{y})^2$

$$\sqrt{n}\Big[\underbrace{\tfrac{1}{n}\textstyle\sum Y_i^2}_{\text{avg}} - \underbrace{\sigma^2}_{\text{var}}\Big]$$ is the CLT

$$\Longrightarrow N(0,V) \qquad V = E\big[(Y_i^2 - EY_i^2)^2\big]$$

σ^2

- $\frac{n}{n-1} \to 1 \qquad \frac{\sqrt{n}}{n-1} \to 0$

$\bar{Y} \to EY_i = 0$

$\sqrt{n}[\bar{Y} - 0] \to N(0,\sigma^2)$

- $\sqrt{n}(\bar{Y})^2 = \underbrace{\sqrt{n}\cdot\bar{Y}}_{\Downarrow\, N(0,\sigma^2)} \cdot \bar{Y} \xrightarrow{P} 0$

d

$\therefore \sqrt{n}[S^2 - \sigma^2] \longrightarrow N(0,V)$

Methods for constructing estimators: Maximum Likelihood

f fn of x: density fn (pdf) f fn of realised x & θ : likelihood.

- Let $f(x|\theta) = f(x_1, ..., x_n|\theta)$ be joint pdf of $X = (X_1, ..., X_n)$ $\sim$ iid $f(x|\theta)$
- for i.i.d. sample $f(x|\theta) = \prod_{i=1}^n f_1(x_i|\theta)$, where $f_1(x_i|\theta)$ is the pdf of one observation

$$\hat{\theta}_{ML} = \arg\max_{\theta\in\Theta} f(x_1, ..., x_n|\theta).$$

Eg 1) $X_1, \dots, X_n \sim iid \ N(\mu, \sigma^2)$ $\qquad \theta = (\mu, \sigma^2)$

1) (pdf) $f(x_i|\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x_i-\mu)^2\right\}$

plug in true obs

$f(x|\theta) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_i (X_i-\mu)^2\right\} \longrightarrow \max_{\mu,\sigma^2}$

$l(\theta) = n \log(\sqrt{2\pi}) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_i (X_i-\mu)^2 \longrightarrow \max_{\mu,\sigma^2}$

σ^2 as an object

FOC $\frac{\partial l}{\partial \mu} = \frac{1}{2\sigma^2} \sum_i (X_i - \hat{\mu}) = 0 \rightarrow \sum (X_i - \mu) = 0$ $\underline{\underline{\hat{\mu} = \bar{X}}}$

$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2} \cdot \frac{1}{\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum (X_i - \mu)^2 = 0$ when ~~...~~ plug in optimal $\hat{\sigma}$ & $\hat{\mu}$

This is the biased MLE as not correct for degrees of freedom $\rightarrow$ $\hat{\sigma}^2 = \frac{1}{n} \sum_i (X_i - \bar{X})^2$

Eg 2) $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$; discrete in x but cont. in p.

PMF for one obs $f(X_i = x_i \mid \theta) = p^{x_i} (1-p)^{1-x_i}$

log likeli for one obs $\ell_1(\theta) = x_i \log p + (1-x_i) \log(1-p)$

$\underline{\underline{\ell_n(p)}} = \log p \cdot \sum x_i + \log(1-p)\,(n - \sum x_i) \rightarrow$ max wrt p.

FOC $\quad \dfrac{\sum x_i}{\hat{p}} - \dfrac{n - \sum x_i}{1-\hat{p}} = 0$

$$(1-\hat{p}) \cdot \sum x_i = \hat{p}\,(n - \sum x_i)$$

$$\sum x_i = \hat{p} \cdot n - \sum x_i)$$

$$\sum x_i = \hat{p} \cdot n$$

$$\hat{p} = \frac{\sum x_i}{n} = \bar{X}$$

Fisher Information

Parametric family:

- $\mathcal{X} \overset{\text{joint pdf}}{\sim} f(x|\theta)$ with $\theta \in \Theta$, $f(\cdot|\theta)$ is pdf or pmf

  unknown

- $Supp_\theta = \{x : f(x|\theta) > 0\}$ is support of distribution $f(x|\theta)$ for $\theta \in \Theta$
- $\ell(\theta|X) = \log f(X|\theta)$ is log-likelihood function
- $s(\theta) = \frac{\partial \ell(\theta|X)}{\partial \theta}$ is called score — depends on θ but is random bec plug in sample

  so you may think about of mean, variance.

$f(\cancel{x}|\theta) = \mathcal{L}(\theta|x) \longrightarrow \max \theta$

$\ell(\theta|x) \longrightarrow \max \theta$

FOC: $S(\hat{\theta}) = 0$

Fisher Information

Assume that

- ☐ $Supp_\theta$ does not depend on θ
- ☐ $\ell(\theta|x)$ is twice continuously differentiable in θ for all $x \in Supp$
- ☐ $\left|\frac{\partial^2 \ell(\theta|x)}{\partial \theta^2}\right| \leq g(x)$ such that $\mathbb{E}g(X) < \infty$

Definition 1.

$I(\theta_0) = \mathbb{E}_\theta\left[\left(\frac{\partial \ell(\theta|X)}{\partial \theta}\right)^2\right]$ is called Fisher information.

$var(s(\theta_0))$ =

Information equalities

Theorem 1.

In the setting above,

(1) $\underbrace{\mathbb{E}_\theta\left[\frac{\partial \ell(\theta|X)}{\partial \theta}\right]} = \mathbb{E}_{\theta_0}[s(\theta_0)] = 0$ Information equality $\underline{E\, s(\theta_0) = 0}$

(2) $I(\theta) = -\mathbb{E}_\theta\left[\frac{\partial^2 \ell(\theta|X)}{\partial \theta^2}\right]$.

- Fisher info is : $I = \text{var}(s(\theta)) = E[S^2(\theta)] = E\left[\left(\frac{\partial \ell}{\partial \theta}\right)^2\right]$

- you take two derivation of log. lik $= E\left[\frac{\partial^2 \ell}{\partial \theta^2}\right]$ ← also calculate info eg1) For one observation

$X \sim N(\mu, \sigma^2)$

$$l_1(\theta|x) = \log(\sqrt{2\pi}) - \frac{1}{2}\log(\sigma^2) - \frac{(X-\mu)^2}{2\sigma^2}$$

$$\frac{\partial l}{\partial \mu} = \frac{2(X-\mu)}{2\sigma^2} = \frac{x-\mu}{\sigma^2}$$

random

not random

$$E\left(\frac{X-\mu}{\sigma^2}\right) = 0$$

$$\frac{\partial l}{\partial \sigma^2} = \underset{\downarrow N.R}{\frac{-1}{2\sigma^2}} + \frac{(X-\mu)^2}{2\sigma^4}$$

$$= \underbrace{\qquad}_{= 0}$$

$$S(\theta) = \begin{pmatrix} \frac{x-\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2\sigma^4} \end{pmatrix}$$

Two parameters, info be matrix:

$$I_1 = E\, S(\theta) S^T(\theta) = \begin{pmatrix} \frac{E(x-u)^2}{6^4} & E\left[\frac{x-u}{6^2}\cdot\left(-\frac{1}{26^2}+\frac{(x-u)^2}{26^4}\right)\right] \\ \parallel & E\left[\left(-\frac{1}{26^2}+\frac{(x-u)^2}{26^4}\right)^2\right] \end{pmatrix}$$

$$\frac{6^2}{6^4} = \frac{1}{6^2}$$

$$\frac{1}{46^4} + \frac{\overbrace{E(x-u)^4}^{=36^4}}{46^8} - \frac{2(x-\mu)^2}{46^6}$$

$$= \frac{1}{26^4}$$

Fisher infor. be $I_1 = \begin{pmatrix} \frac{1}{6^2} & 0 \\ 0 & \frac{1}{26^4} \end{pmatrix}$

$$S(\theta) = \begin{pmatrix} \frac{x-\mu}{6^2} \\ \frac{-1}{26^2} + \frac{(x-u)^2}{26^2} \end{pmatrix}$$

$$\frac{\partial^2 l}{\partial \theta^2} = \begin{pmatrix} \frac{\partial^2 l}{\partial \mu \partial \mu} & \frac{\partial^2 l}{\partial u \partial 6^2} \\ & \frac{\partial^2 l}{\partial 6^2 \partial 6^2} \end{pmatrix}$$

Examples

$$E\frac{\partial^2 l}{\partial\mu\partial\mu} = -\frac{1}{\sigma} \; ; \; E\frac{\partial^2 l}{\partial\mu\partial\sigma^2} = \frac{-E(x-\mu)}{\sigma^4} \; ; \; \frac{\partial^2 l}{\partial\sigma^2\partial\sigma^2} = \frac{1}{2\sigma^4} - \frac{2(x-\mu)^2}{2\sigma^6}$$

$$= 0$$

$$= \frac{1}{2\sigma^4} \frac{-\sigma^2}{\sigma^6} = \frac{-1}{2\sigma^4}$$

$$I_2 = -E\frac{\partial^2 l}{\partial\theta^2} = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

Eg 2) Bernoulli (for 1 obs)

$$f(x|\theta) = p^x (1-p)^{1-x}$$

1) $l(p) = \log p \cdot x - \log(1-p)\cdot(1-x)$ — log

2) $s(p) = \frac{\partial l}{\partial p} = \frac{x}{p} - \frac{1-x}{1-p}$ — Score $\quad Es = \frac{Ex}{p} - \frac{1-Ex}{1-p} = 0$

Examples

3) Now to info!

$$I_1 = E\left(\left[\frac{x}{p} - \frac{1-x}{1-p}\right]^2\right) = E\left[\frac{x - \cancel{px} - p + \cancel{px}}{p(1-p)}\right]^2$$

$$= \frac{E[(X-p)^2]}{p^2(1-p)^2} \leftarrow \text{var}$$

$$= \frac{\text{Var}(X)}{p^2(1-p)^2} = \frac{p(1-p)}{p^2(1-p)} = \frac{1}{p(1-p)} \leftarrow \text{info. of one draw.}$$

Now in second way:

$$\frac{\partial^2 l}{\partial p \partial p} = \frac{\partial}{\partial p}\left[\frac{X}{p} - \frac{1-x}{1-p}\right] = \frac{-X}{p^2} - \frac{(1-x)}{(1-p)^2}$$

Now taking Exp.

$$E\frac{\partial^2 l}{\partial p \partial p} = -\frac{E\,x}{p^2} - \frac{E(1-x)}{(1-p)^2} = \frac{p}{p^2} - \frac{1-p}{(1-p)^2}$$

$$= \frac{-1}{p} - \frac{1}{1-p} = -\frac{1}{p(\cancel{t}-p)}$$

$$I_2 = -E\frac{\partial^2 l}{\partial p \partial p} = \frac{1}{p(1-p)}$$

Information for a random sample

- Let $X = (X_1, ..., X_n)$ be i.i.d. sample from $f_1(x_i|\theta)$.

  full sample

  pdf for one obs. so gaussian,
- Fisher information for the full sample X is

$$I_n(\theta) = nI_1(\theta).$$

where $I_1(\theta)$ is Fisher information for one random draw from $f_1(x_i|\theta)$.

$$f_n(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f_1(x_i|\theta)$$

$$l_n(\theta) = \sum_{i=1}^{n} l_1(\theta|x_i)$$

$$E\left[-\frac{\partial^2 l_n}{\partial\theta\partial\theta}\right] = E\left[\sum_{i=1}^{n} -\frac{\partial^2 l_1}{\partial\theta\partial\theta}(\theta|x_i)\right] = \sum_{i=1}^{n}\left[-E\frac{\partial^2 l_1}{\partial\theta\partial\theta}(\theta|x_i)\right]$$

$$= n\,I_1(\theta)$$

 : bound on variance : help found efficient est. wi class biasedness

Theorem 2 (Rao-Cramer bound).

all data here

try to est.

Let X be a random data with distribution $f(x|\theta)$ and information $I(\theta)$. Let $W(X)$ be an estimator of θ such that

$\hat{\theta} = W(X)$

if $\hat{\theta}$ is unbiased $E\hat{\theta} = \theta$

(1) $\frac{d}{d\theta}E_\theta[W(X)] = \int W(x)\frac{\partial f(x|\theta)}{\partial\theta}dx$ || technical ass.

(2) $Var(W) < \infty$.

$\frac{d}{d\theta} E\hat{\theta}$

$\frac{d}{d\theta} E\hat{\theta} = 1$

Then

$$Var(W) \geq \left(\underbrace{\frac{d}{d\theta}E_\theta[W(X)]}\right)^2 \frac{1}{I(\theta)}.$$

Info.

In particular, if $X = (X_1, ..., X_n)$ is an i.i.d. random sample and W is unbiased for θ, then
$Var(W) \geq \frac{1}{I(\theta)} = \frac{1}{nI_1(\theta)}$

for any unbiased est. $\hat{\theta}$

$Var(\hat{\theta}) \geq \frac{1}{I_n(\theta)} = \frac{1}{nI_1(\theta)}$

eg 1) $X_i \sim$ iid Bernoulli(p)

MLE est. for $\hat{p} = \bar{X}$ $\quad E\hat{p} = p$ unbiased

$$Var(\hat{p}) = \frac{p(1-p)}{n}$$

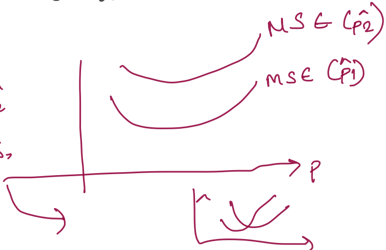
$$\underline{\underline{Var(\text{unbiased est})}} \geq \frac{1}{nI_1} = \frac{1}{n\frac{1}{p(1-p)}} = \underline{\underline{\frac{p(1-p)}{n}}}$$

Thus, est. is unbiased one.

$\hat{p}_1$ $\quad \hat{p}_2$

more efficient than $\hat{p}_2$

But mse curves cross,



efficient est. not be defeated