Lecture 20

2024 Spring

July 23, 2024

Last Lecture

{ ① support depends on θ
② ... does not depend on θ.

- How to derive the maximum likelihood estimator (or estimate).
- Preparations for the properties of the maximum likelihood estimator.

① Score function: $S(\theta) = S(\theta; x_1, \cdots, x_n) = \frac{d}{d\theta} \ell(\theta)$

② Information function: $I(\theta) = I(\theta; x_1 \cdots x_n) = -\frac{d}{d\theta} S(\theta) = -\frac{d^2}{d\theta^2} \ell(\theta)$

③ Fisher Information: $J(\theta) = E\{I(\theta; X_1, \cdots, X_n)\}$

This Lecture

- Properties of maximum likelihood estimator.

(Cont') Fisher Information

1). $X_1 \cdots X_n$ iid. with pdf $f(x;\theta)$

$$J(\theta) = nJ_1(\theta), \quad J_1(\theta) = -E\left\{\frac{d^2 \log f(X_1;\theta)}{d\theta^2}\right\}$$

2). Alterative form of the Fisher Information.

$$J_1(\theta) = -E\left\{\frac{d^2 \log f(X_1;\theta)}{d\theta^2}\right\} = E\left[\left\{\frac{d \log f(X_1;\theta)}{d\theta}\right\}^2\right]$$

$$\frac{d^2 \log f(X_1;\theta)}{d\theta^2} = \boxed{\frac{\frac{d^2 f(X_1;\theta)}{d\theta^2}}{f(X_1;\theta)}} - \left\{\frac{\frac{df(X_1;\theta)}{d\theta}}{f(X_1;\theta)}\right\}^2 \qquad \frac{d \log f(X_1;\theta)}{d\theta} = \frac{1}{f(X_1;\theta)} \cdot \frac{df(X_1;\theta)}{d\theta}$$

$$E\left\{\frac{d^2 f(X_1;\theta)/d\theta^2}{f(X_1;\theta)}\right\} = \int \frac{d^2 f(x_1;\theta)}{d\theta^2} dx_1 = \frac{d^2}{d\theta^2}\int f(x_1;\theta)dx_1 = 0$$

Examples:

① $X_1 \cdots X_n \overset{iid}{\sim} \mathrm{Poi}(\theta)$

$$\log f(x_1;\theta) = x_1 \log\theta - \theta - \log(x_1!)$$

$$l(\theta) = \sum_{i=1}^n \log f(x_i;\theta) = \log\theta\left(\sum_{i=1}^n x_i\right) - n\theta - \sum_{i=1}^n \log(x_i!)$$

$$S(\theta) = \frac{dl(\theta)}{d\theta} = \frac{\sum_{i=1}^n x_i}{\theta} - n \quad \Rightarrow \quad \hat\theta_{ML} = \bar{X}_n$$

$$I(\theta) = -\frac{dS(\theta)}{d\theta} = \frac{\sum_{i=1}^n x_i}{\theta^2}, \quad J(\theta) = E\{I(\theta)\} = E\left(\frac{\sum_{i=1}^n X_i}{\theta^2}\right) = \frac{n\theta}{\theta^2} = \frac{n}{\theta}$$

$$\mathrm{var}(\hat\theta_{ML}) = \frac{\theta}{n}$$

② $X_1 \cdots X_n \overset{iid}{\sim} \mathrm{Exp}(\theta)$. ($\theta$ is the mean). $f(x;\theta) = \frac{1}{\theta}\cdot\exp(-\frac{x}{\theta})$. $x>0, \theta>0$

$$\log f(x_1;\theta) = -\log\theta - \frac{x_1}{\theta}. \quad l(\theta) = -n\log\theta - \frac{\sum_{i=1}^n x_i}{\theta}$$

$$S(\theta) = \frac{dl(\theta)}{d\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\cdot\sum_{i=1}^n x_i. \quad \Rightarrow \quad \hat\theta_{ML} = \bar{X}_n$$

$$I(\theta) = -\frac{dS(\theta)}{d\theta} = -\frac{n}{\theta^2} + \frac{2\sum_{i=1}^n x_i}{\theta^3},$$

$$J(\theta) = E\{I(\theta; X_1 \cdots X_n)\} = -\frac{n}{\theta^2} + \frac{2n\cdot\theta}{\theta^3} = \frac{n}{\theta^2}$$

$$\mathrm{var}(\hat\theta_{ML}) = \frac{\theta^2}{n}$$

③ $X_1 \cdots X_n \overset{iid}{\sim} \mathrm{Norm}(\theta, \sigma^2)$, suppose σ^2 is known

$$f(x;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\}$$

$$\log f(x_1;\theta) = -\frac{1}{2}\cdot\log(2\pi\sigma^2) - \frac{(x_1-\theta)^2}{2\sigma^2}$$

$$l(\theta) = \sum_{i=1}^n \log f(x_i;\theta) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}$$

$$S(\theta) = \frac{\sum_{i=1}^n (x_i-\theta)}{\sigma^2} \quad \Rightarrow \quad \hat\theta_{ML} = \bar{X}_n$$

$$I(\theta) = \frac{n}{\sigma^2}. \quad J(\theta) = E\{I(\theta)\} = \frac{n}{\theta^2}$$

$$\mathrm{var}(\hat\theta_{ML}) = \frac{\sigma^2}{n}$$

The Cramer-Rao Lower Bound The Cramer-Rao bound says that for any **unbiased** estimator $T = T(X_1, \ldots, X_n)$ for $\tau(\theta)$, the lower bound of its variance is given by $\tau(\cdot)$ is known.

$$\text{Var}(T) \geq \frac{\{\tau'(\theta)\}^2}{\text{J}(\theta)}.$$

Comments:

- Conditions: Some regularity conditions (e.g. support does not depend on θ)

- If $\tau(\theta) = \theta$: $\tau'(\theta) = 1$, $\frac{\{\tau'(\theta)\}^2}{J(\theta)} = \frac{1}{J(\theta)}$

- If iid data: $\frac{\{\tau'(\theta)\}^2}{J(\theta)} = \frac{\{\tau'(\theta)\}^2}{nJ_1(\theta)}$

  $\hookrightarrow$ Fisher Information of X_1

- Attainable?

  ① Not always attainable

  ② If $\hat{\theta}$ is an unbiased estimator that satisfies $\text{Var}(\hat{\theta}) = \frac{1}{J(\theta)}$, we say $\hat{\theta}$ is the UMVUE

  uniformly variance

  minimal unbiased

Examples:

① $\text{Poi}(\theta)$ $\hat{\theta} = \bar{X}_n$. $\text{Var}(\hat{\theta}) = \frac{1}{J(\theta)}$. $\hat{\theta}$ is the UMVUE

② $\text{Exp}(\theta)$. $\hat{\theta} = \bar{X}_n$. $\text{Var}(\hat{\theta}) = \frac{1}{J(\theta)}$. $\hat{\theta}$ is the UMVUE

③ $\text{Norm}(\theta, \sigma^2)$, σ^2 is known. $\hat{\theta} = \bar{X}_n$. $\text{Var}(\hat{\theta}) = \frac{1}{J(\theta)}$. $\hat{\theta}$ is the UMVUE

Optional : Inference using MLE

$H_0: \theta = \theta_0$

$Z_n = \sqrt{n J_1(\theta_0)}\,(\hat{\theta} - \theta_0) \xrightarrow{d} \text{Norm}(0,1)$

Z_n is a test statistic.

$p = 1 - \Phi(Z_n)$

standard normal cdf.

ML Estimator Property Under regularity conditions (e.g., the support does not depend on θ), with iid $X_1, \ldots, X_n$, the ML estimator for θ satisfies

1. $\widehat{\theta} \xrightarrow{p} \theta$ as $n \to \infty$. ML estimator is consistent.
2. $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \text{Norm}\left(0, \frac{1}{J_1(\theta)}\right)$ $J_1(\theta) = -E\left\{\frac{d^2 \log f(X_i;\theta)}{d\theta^2}\right\}$
3. $\sqrt{n}(\tau(\widehat{\theta}) - \tau(\theta)) \xrightarrow{d} \text{Norm}\left(0, \frac{\{\tau'(\theta)\}^2}{J_1(\theta)}\right)$ using the delta method + (2)

   $\tau(\cdot)$ is known.

Comments:

From 2. $(\hat{\theta} - \theta) \overset{\text{approx.}}{\sim} \text{Norm}\left(0, \frac{1}{nJ_1(\theta)}\right)$

$\hat{\theta} \overset{\text{approx}}{\sim} \text{Norm}\left(\theta, \frac{1}{nJ_1(\theta)}\right)$ $\frac{1}{J(\theta)}$

$\Rightarrow E(\hat{\theta}) \to \theta$, $n \to \infty$

$\text{Var}(\hat{\theta}) \to \frac{1}{J(\theta)}$ (CR Bound), $n \to \infty$

$\Rightarrow \hat{\theta}$ is asymptotically unbiased.

$\Rightarrow \text{Var}(\hat{\theta})$ is asymptotically converging the CR bound

Best among unbiased estimators

Example: $X_1, \ldots, X_n$ are iid Poi(θ).

1. Find the ML estimator of θ, denoted by $\widehat{\theta}$.

   $\hat{\theta} = \bar{X}_n$

2. Find the ML estimator of $\psi \equiv \Pr(X_1 = 0)$, denoted by $\widehat{\psi}$.

   $\psi = Pr(X_1 = 0) = e^{-\theta}$

   $\hat{\psi} = \exp(-\hat{\theta}) = \exp(-\bar{X}_n)$

3. Find the limiting distribution of $\sqrt{n}(\widehat{\theta} - \theta)$.

Method 1: CLT $\hat{\theta} = \overline{X}_n$

Method 2: By the property of ML estimator:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \text{Norm}\left(0, \frac{1}{J_1(\theta)}\right)$$

$$J_1(\theta) = \frac{1}{\theta}$$

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \text{Norm}(0, \theta)$$

4. Find the limiting distribution of $\sqrt{n}(\widehat{\psi} - \psi)$.

Method 1: CLT + Delta method.

Method 2: By the property of ML estimator.

$$\sqrt{n}(\hat{\psi} - \psi) \xrightarrow{d} \text{Norm}\left(0, \frac{\{(e^{-\theta})'\}^2}{J_1(\theta)}\right)$$

$$\xrightarrow{d} \text{Norm}(0, \theta \cdot e^{-2\theta})$$

5. Find the mean of $\widehat{\theta}$ and $\widehat{\psi}$.

$E(\hat{\theta}) = E(\overline{X}_n) = \theta$. , $\hat{\theta}$ is unbiased.

$$E(\hat{\psi}) = E(e^{-\overline{X}_n}) = E(e^{-\frac{1}{n}\sum_{i=1}^{n} X_i}) = E(e^{-\frac{1}{n} Y_n})$$

$Y_n = \sum_{i=1}^{n} X_i \sim \text{Poi}(n\theta)$, $E(\hat{\psi}) = M_{Y_n}(-\frac{1}{n})$

$$= \exp\{\theta n (e^{-\frac{1}{n}} - 1)\}$$

$$\psi = \exp(-\theta) \neq E(\hat{\psi})$$

not unbiased.

$$\lim_{n\to\infty} E(\hat{\psi}) = \lim_{n\to\infty} \exp\{\theta \cdot n(e^{-\frac{1}{n}} - 1)\} = \exp(-\theta)$$

$$n(e^{-\frac{1}{n}} - 1) = \boxed{\frac{e^{-\frac{1}{n}} - e^0}{-\frac{1}{n} - 0}} \cdot (-1) \longrightarrow -1$$

$$\to (e^x)'\big|_{x=0} = 1$$