STAT 5430

Lecture 05, F, Jan 31

- Homework 1 is assigned in Canvas (submit/due by next Monday, Feb 3, by midnight)

practice on point estimation (method of moments & likeihood estimation)

- Office hours to be announced
  Mine: FM, 12-1 PM & by appointment
  TA (Min-Yi): WR 11-12 in Snedecor 2404

Point Estimation

Finding Maximum Likelihood Estimators (MLEs)/Multiparameter Case

Suppose $X_1, X_2, \ldots, X_n$ have joint pmf/pdf $f(x_1, x_2, \ldots, x_n | \underset{\sim}{\theta})$ where $\underset{\sim}{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)' \in \Theta \subset \mathbb{R}^k$ (i.e., k parameters).

Want to find MLEs $\hat{\underset{\sim}{\theta}} \equiv (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)'$ of $\underset{\sim}{\theta}$, which solve

$$L(\hat{\underset{\sim}{\theta}}) = \max_{\underset{\sim}{\theta} \in \Theta} L(\underset{\sim}{\theta}), \quad \text{where} \quad L(\underset{\sim}{\theta}) \equiv f(x_1, x_2, \ldots, x_n | \underset{\sim}{\theta})$$

can depend on more than 1 parameter

.

Result: If $\Theta \subset \mathbb{R}^k$ is open and $L(\underset{\sim}{\theta}) \equiv f(x_1, x_2, \ldots, x_n | \underset{\sim}{\theta})$ has 2nd order partial derivatives on Θ, then $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ are MLEs of $\theta_1, \theta_2, \ldots, \theta_k$ provided

(review of using calculus to find maximum)

1. for each $i = 1, \ldots, k$

   → partial derivatives $$\frac{\partial \log L(\underset{\sim}{\theta})}{\partial \theta_i}\Big|_{\hat{\underset{\sim}{\theta}}} = 0;$$

2. denote the $k \times k$ Hessian matrix at $\hat{\underset{\sim}{\theta}}$ as

   $$H = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1k} \\ h_{21} & h_{22} & \cdots & h_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ h_{k1} & h_{k2} & \cdots & h_{kk} \end{pmatrix}$$

   2nd partial derivatives matrix

   where

   $$h_{ij} = \frac{\partial^2 \log L(\underset{\sim}{\theta})}{\partial \theta_i \partial \theta_j}\Big|_{\hat{\underset{\sim}{\theta}}} \quad \text{for } i, j = 1, \ldots k,$$

   and let

   $$\Delta_i = \det \begin{pmatrix} \boxed{h_{11}} & \cdots & h_{1i} \\ \vdots & \ddots & \vdots \\ h_{i1} & \cdots & h_{ii} \end{pmatrix} \quad \text{for } i = 1, \ldots k,$$

   be the determinant of the $i \times i$ submatrix of H consisting of the first i rows. Then, we need $\Delta_1 < 0, \Delta_2 > 0, \Delta_3 < 0, \cdots$ and so on. (Must compute k determinants $\Delta_1, \Delta_2, \ldots, \Delta_k$ to see if they alternate in positive/negative.)

$\underset{1\times 1}{\Delta_1} = h_{11} < 0 \qquad \underset{2\times 2}{\Delta_2} = \det\begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} = h_{11}h_{22} - h_{21}h_{12} > 0$

Point Estimation

Finding Maximum Likelihood Estimators/Example in Multiparameter Case

Example: Let $X_1, X_2, \ldots, X_n$ be iid $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$. Find MLEs of $\mu \& \sigma^2$.

Solution: Write $\theta_1 \equiv \mu$ and $\theta_2 \equiv \sigma^2$ and

$$\begin{aligned} L(\theta_1, \theta_2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{(x_i-\theta_1)^2}{2\theta_2}} = (2\pi)^{-n/2}\theta_2^{-n/2} e^{-\sum_{i=1}^{n} \frac{(x_i-\theta_1)^2}{2\theta_2}} \quad \leftarrow \text{joint density} \\ \log L(\theta_1, \theta_2) &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\theta_2 - \sum_{i=1}^{n} \frac{(x_i-\theta_1)^2}{2\theta_2} \end{aligned}$$

Then, setting

$$\frac{\partial \log L(\underset{\sim}{\theta})}{\partial \theta_i}\Big|_{\hat{\underset{\sim}{\theta}}} = 0 \quad \text{for } i = 1, 2,$$

we see that the MLEs $\hat{\underset{\sim}{\theta}} \equiv (\hat\theta_1, \hat\theta_2)'$ satisfy

$$\sum_{i=1}^{n} \frac{(x_i - \hat\theta_1)}{\hat\theta_2} = 0 \qquad \& \qquad -\frac{n}{2\hat\theta_2} - \sum_{i=1}^{n} \frac{(x_i-\hat\theta_1)^2}{2(\hat\theta_2)^2} = 0,$$

implying that

$$\hat\mu = \hat\theta_1 = \bar{x}_n = \sum_{i=1}^{n} x_i/n, \qquad \hat\theta_2 = \sum_{i=1}^{n}(x_i - \hat\theta_1)^2/n = \sum_{i=1}^{n}(x_i - \bar{x}_n)^2/n, = \hat\sigma^2$$

though need to check 2nd partials conditions too. Note

$$\begin{aligned} \frac{\partial^2 \log L(\theta_1,\theta_2)}{\partial\theta_1^2} &= \frac{\partial}{\partial\theta_1}\left[\frac{\partial \log L(\theta_1,\theta_2)}{\partial\theta_1}\right] = -\frac{n}{\theta_2}, \\ \frac{\partial^2 \log L(\theta_1,\theta_2)}{\partial\theta_2^2} &= \frac{\partial}{\partial\theta_2}\left[\frac{\partial \log L(\theta_1,\theta_2)}{\partial\theta_2}\right] = \frac{n}{2(\theta_2)^2} - \frac{\sum_{i=1}^{n}(x_i-\theta_1)^2}{(\theta_2)^3} \\ \frac{\partial^2 \log L(\theta_1,\theta_2)}{\partial\theta_2\partial\theta_1} &= \frac{\partial}{\partial\theta_2}\left[\frac{\partial \log L(\theta_1,\theta_2)}{\partial\theta_1}\right] = -\frac{\sum_{i=1}^{n}(x_i-\theta_1)}{(\theta_2)^2} = \frac{\partial}{\partial\theta_1}\left[\frac{\partial \log L(\theta_1,\theta_2)}{\partial\theta_2}\right] = \frac{\partial^2 \log L(\theta_1,\theta_2)}{\partial\theta_1\partial\theta_2} \end{aligned}$$

and hence

$$h_{11} \equiv \frac{\partial^2 \log L(\theta_1,\theta_2)}{\partial\theta_1^2}\Big|_{\hat{\underset{\sim}{\theta}}} = -\frac{n}{\hat\theta_2}, \quad h_{22} \equiv \frac{\partial^2 \log L(\theta_1,\theta_2)}{\partial\theta_2^2}\Big|_{\hat{\underset{\sim}{\theta}}} = \frac{n}{2(\hat\theta_2)^2} - \frac{\sum_{i=1}^{n}(x_i-\hat\theta_1)^2}{(\hat\theta_2)^3} = -\frac{n}{2(\hat\theta_2)^2}$$

$$h_{12} \equiv \frac{\partial^2 \log L(\theta_1,\theta_2)}{\partial\theta_1\partial\theta_2}\Big|_{\hat{\underset{\sim}{\theta}}} = -\frac{\sum_{i=1}^{n}(x_i-\bar{x}_n)}{(\hat\theta_2)^2} = 0 = \frac{\partial^2 \log L(\theta_1,\theta_2)}{\partial\theta_2\partial\theta_1}\Big|_{\hat{\underset{\sim}{\theta}}} \equiv h_{21}$$

$$H \equiv \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} = \begin{bmatrix} -n/\hat\theta_2 & 0 \\ 0 & -n/(2\hat\theta_2^2) \end{bmatrix} \quad \Rightarrow \quad \Delta_1 \equiv -\frac{n}{\hat\theta_2} < 0 \;\&\; \Delta_2 \equiv \det(H) = \frac{n^2}{2\hat\theta_2^3} > 0$$

$\Rightarrow$ $\hat\mu$ & $\hat\sigma^2$ are indeed maximums (MLEs)

Point Estimation

Maximum Likelihood Estimators (MLEs) of Parametric Functions

Definition: For a parametric function $\gamma(\theta_1, \theta_2, \ldots, \theta_k)$, we define $\gamma(\hat{\theta}_1, \ldots, \hat{\theta}_k)$ as the MLE of $\gamma(\theta_1, \theta_2, \ldots, \theta_k)$, where $\hat{\theta}_1, \ldots, \hat{\theta}_k$ are the MLEs of $\theta_1, \ldots, \theta_k$.

Last Example: Let $X_1, X_2, \ldots, X_n$ be iid $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$. Find the MLE of $\log(\mathrm{E}X_1^2) = \log(\mu^2 + \sigma^2)$.

Here MLEs are $\hat{\mu} = \overline{X}_n$ & $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$

$$= \frac{1}{n}\sum_{i=1}^n X_i^2 - (\overline{X}_n)^2$$

Find the MLE of $\log(\mu^2 + \sigma^2)$

is $\log(\hat{\mu}^2 + \hat{\sigma}^2) = \log\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right)$

STAT 5430: Summary to date
Where we have been
& where we are headed

- Completed
  - Introduction to Statistical Inference
  - Point Estimation
    - MME as strategy for point estimation
    - MLE as strategy for point estimation
    - Finding MLEs: examples using/without calculus, multivariate case
    - MME/MLE of parametric functions

- Next: Criteria for Evaluating Point Estimators

  (Section 7.3.1 & 7.3.2)

  looking for good properties of estimators
  - bias
  - variance
  - UMVUE
  - CRLB
  - relative efficiency
  - MSE

Criteria for Evaluating Point Estimators

Bias

↙ data function $T = h(X_1, \dots, X_n)$

Definition: An estimator T of a parametric function $\gamma(\theta)$ is called **unbiased** if

"$\forall$" $\equiv$ for all

$$E_\theta(T) = \mathrm{E}(T) = \gamma(\theta), \quad \forall \theta \in \Theta.$$

↑ data $X_1, \dots X_n$ come from some distribution depending on θ

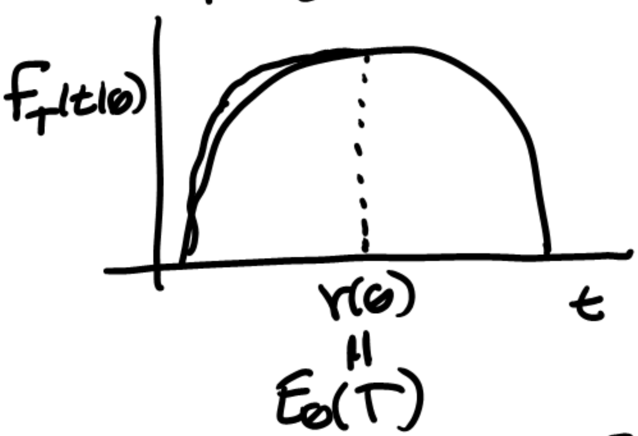
Definition: T is **biased** if T is NOT unbiased.

Definition: The **bias** of T is denoted as: $b_\theta(T) \equiv \mathrm{E}_\theta(T) - \gamma(\theta)$.

(if T is unbiased, $b_\theta(T) = 0$, $\forall \theta \in \Theta$)

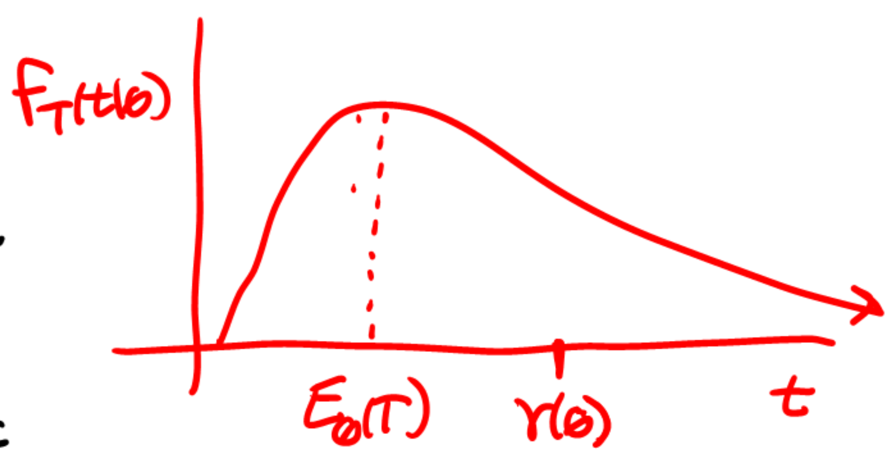
Picture: Suppose T is a continuous r.v. with pdf $f_T(t|\theta)$.

T is unbiased for $\gamma(\theta)$



↗ "Typical value of T is $\gamma(\theta)$"

T is biased



bias is $b_\theta(T)$
$= E_\theta(T) - \gamma(\theta)$

Criteria for Evaluating Point Estimators

Bias, cont'd

Examples: Let $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}, \sigma^2 > 0$ & consider estimators $\bar{X}_n$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

$E_{\mu,\sigma^2}(\bar{X}_n) = \mu, \quad \forall \mu, \sigma^2$

$E_{\mu,\sigma^2}(S^2) = \sigma^2, \quad \forall \mu, \sigma^2$

Hence, $\bar{X}_n$ is U.E of μ & S^2 is U.E of σ^2.

Next, $S = \sqrt{S^2}$ sample standard deviation is estimator of σ

$$E_{\mu,\sigma^2}(S) = E_{\mu,\sigma^2}(\sqrt{S^2}) \leftarrow Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$= E_{\mu,\sigma^2}\left(\sqrt{\frac{\sigma^2 Y}{n-1}}\right) \qquad E_{\mu,\sigma^2}(\sqrt{Y}) = \int_0^\infty y^{1/2} \underbrace{f(y)}_{\chi^2_{n-1}\text{ density}} dy$$

$$= \frac{\sigma}{\sqrt{n-1}} \sqrt{2} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \neq \sigma$$

gamma functions

ie S is NOT U.E. of σ

Notes on Unbiasedness: (Below, let "U.E." $\equiv$ "unbiased estimator.")

1. If T is an U.E. of θ, then $\gamma(T)$ need NOT be an U.E. of $\gamma(\theta)$.

   Example: See last example

   S^2 is UE of σ^2

   $\sqrt{S^2}$ is NOT UE of $\sqrt{\sigma^2} = \sigma$

Criteria for Evaluating Point Estimators

Bias, cont'd

2. It is NOT always possible to find an U.E. of $\gamma(\theta)$

   Example: Let X $\leftarrow$ data be Binomial(n, p), $0 < p < 1$. Show there is no U.E. of $\gamma(p) = 1/p$.

Solution: If possible, suppose $h(X)$ is U.E of $1/p$

$$\Rightarrow E_p h(X) = \sum_{x=0}^{n} h(x)\binom{n}{x} p^x (1-p)^{n-x} = 1/p, \ \forall\, 0<p<1$$

multiply both sides by p & let $p \downarrow 0$

$$\underbrace{\sum_{x=0}^{n} h(x)\binom{n}{x} p^{x+1}(1-p)^{n-x}}_{\downarrow 0 \text{ as } p \downarrow 0} = 1, \ \forall\, 0<p<1$$

a contradiction!

Note: $\frac{X}{n}$ is estimator of p (U.E of p) $\Rightarrow \frac{n}{X}$ is estimator of $1/p$ but not U.E.

3. Unbiasedness, while good, is not everything!

   Example: Let $X_1, \ldots, X_n$ be iid Bernoulli(p), $0 < p < 1$, and consider two estimators of p given by

$$T_1 = X_1, \qquad T_2 = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$