Biostatistics 602 - Statistical Inference
Lecture 11
Evaluation of Point Estimators

Hyun Min Kang

February 14th, 2013

Some News

- Homework 3 is posted.
  - Due is Tuesday, February 26th.
- Next Thursday (Feb 21) is the midterm day.
  - We will start sharply at 1:10pm.
  - It would be better to solve homework 3 yourself to get prepared.
  - The exam is closed book, covering all the material from Lecture 1 to 12.
  - Last year's midterm is posted on the web page.

Last Lecture

1. What is a maximum likelihood estimator (MLE)?
2. How can you find an MLE?
3. Does an ML estimate always fall into a valid parameter space?
4. If you know MLE of θ, can you also know MLE of $\tau(\theta)$?

Recap - Maximum Likelihood Estimator

Definition

- For a given sample point $\mathbf{x} = (x_1, \cdots, x_n)$,
- let $\hat{\theta}(\mathbf{x})$ be the value such that
- $L(\theta|\mathbf{x})$ attains its maximum.
- More formally, $L(\hat{\theta}(\mathbf{x})|\mathbf{x}) \geq L(\theta|\mathbf{x})\ \forall \theta \in \Omega$ where $\hat{\theta}(\mathbf{x}) \in \Omega$.
- $\hat{\theta}(\mathbf{x})$ is called the *maximum likelihood estimate* of θ based on data $\mathbf{x}$,
- and $\hat{\theta}(\mathbf{X})$ is the *maximum likelihood estimator (MLE)* of θ.

Recap - Invariance Property of MLE

Question

If $\hat{\theta}$ is the MLE of θ, what is the MLE of $\tau(\theta)$?

Example

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$ where $0 < p < 1$.

1. What is the MLE of p?
2. What is the MLE of odds, defined by $\eta = p/(1-p)$?

Getting MLE of $\eta = \frac{p}{1-p}$ from $\hat{p}$

$$L^*(\eta|\mathbf{x}) = \frac{\eta^{\sum x_i}}{(1+\eta)^n}$$

- From MLE of $\hat{p}$, we know $L^*(\eta|\mathbf{x})$ is maximized when $p = \eta/(1+\eta) = \hat{p}$.
- Equivalently, $L^*(\eta|\mathbf{x})$ is maximized when $\eta = \hat{p}/(1-\hat{p}) = \tau(\hat{p})$, because τ is a one-to-one function.
- Therefore $\hat{\eta} = \tau(\hat{p})$.

Invariance Property of MLE

Fact

Denote the MLE of θ by $\hat{\theta}$. If $\tau(\theta)$ is an one-to-one function of θ, then MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$.

Proof

The likelihood function in terms of $\tau(\theta) = \eta$ is

$$\begin{aligned} L^*(\tau(\theta)|\mathbf{x}) &= \prod_{i=1}^n f_X(x_i|\theta) = \prod_{i=1}^n f(x_i|\tau^{-1}(\eta)) \\ &= L(\tau^{-1}(\eta)|\mathbf{x}) \end{aligned}$$

We know this function is maximized when $\tau^{-1}(\eta) = \hat{\theta}$, or equivalently, when $\eta = \tau(\hat{\theta})$. Therefore, MLE of $\eta = \tau(\theta)$ is $\tau(\hat{\theta})$.

Induced Likelihood Function

Definition

- Let $L(\theta|\mathbf{x})$ be the likelihood function for a given data $x_1, \cdots, x_n$,
- and let $\eta = \tau(\theta)$ be a (possibly not a one-to-one) function of θ.

We define the *induced likelihood function* L^* by

$$L^*(\eta|\mathbf{x}) = \sup_{\theta \in \tau^{-1}(\eta)} L(\theta|\mathbf{x})$$

where $\tau^{-1}(\eta) = \{\theta : \tau(\theta) = \eta, \ \theta \in \Omega\}$.

- The value of η that maximize $L^*(\eta|\mathbf{x})$ is called the MLE of $\eta = \tau(\theta)$.

Invariance Property of MLE

Theorem 7.2.10

If θ is the MLE of $\hat{\theta}$, then the MLE of $\eta = \tau(\theta)$ is $\tau(\hat{\theta})$, where $\tau(\theta)$ is any function of θ.

Proof - Using Induced Likelihood Function

$$\begin{aligned} L^*(\hat{\eta}|\mathbf{x}) &= \sup_{\eta} L^*(\eta|\mathbf{x}) = \sup_{\eta} \sup_{\theta \in \tau^{-1}(\eta)} L(\theta|\mathbf{x}) \\ &= \sup_{\theta} L(\theta|\mathbf{x}) = L(\hat{\theta}|\mathbf{x}) \\ L(\hat{\theta}|\mathbf{x}) &= \sup_{\theta \in \tau^{-1}(\tau(\hat{\theta}))} L(\theta|\mathbf{x}) = L^*[\tau(\hat{\theta})|\mathbf{x}] \end{aligned}$$

Hence, $L^*(\hat{\eta}|\mathbf{x}) = L^*[\tau(\hat{\theta})|\mathbf{x}]$ and $\tau(\hat{\theta})$ is the MLE of $\tau(\theta)$.

Properties of MLE

1. Optimal in some sense : We will study this later
2. By definition, MLE will always fall into the range of the parameter space.
3. Not always easy to obtain; may be hard to find the global maximum.
4. Heavily depends on the underlying distributional assumptions (i.e. not robust).

Method of Evaluating Estimators

Definition : Unbiasedness

Suppose $\hat{\theta}$ is an estimator for θ, then the bias of θ is defined as

$$\text{Bias}(\theta) = E(\hat{\theta}) - \theta$$

If the bias is equal to 0, then $\hat{\theta}$ is an unbiased estimator for θ.

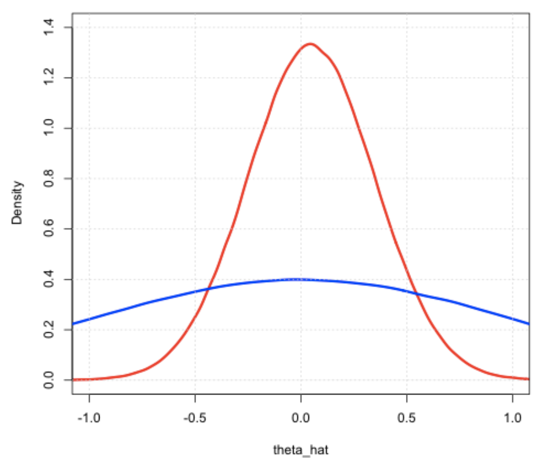
Example

$X_1, \cdots, X_n$ are iid samples from a distribution with mean μ. Let $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ is an estimator of μ. The bias is

$$\begin{aligned} \text{Bias}(\mu) &= E(\overline{X}) - \mu \\ &= E\left(\frac{1}{n}\sum_{i=1}^n X_i\right) - \mu = \frac{1}{n}\sum_{i=1}^n E(X_i) - \mu = \mu - \mu = 0 \end{aligned}$$

Therefore $\overline{X}$ is an unbiased estimator for μ.

How important is unbiased?



- $\hat{\theta}_1$ (blue) is unbiased but has a chance to be very far away from $\theta = 0$.
- $\hat{\theta}_2$ (red) is biased but more likely to be closer to the true θ than $\hat{\theta}_1$.

Mean Squared Error

Definition

Mean Squared Error (MSE) of an estimator $\hat{\theta}$ is defined as

$$\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)]^2$$

Property of MSE

$$\begin{aligned} \text{MSE}(\hat{\theta}) &= E[(\hat{\theta} - E\hat{\theta} + E\hat{\theta} - \theta)]^2 \\ &= E[(\hat{\theta} - E\hat{\theta})^2] + E[(E\hat{\theta} - \theta)^2] + 2E[(\hat{\theta} - E\hat{\theta})]E[(E\hat{\theta} - \theta)] \\ &= E[(\hat{\theta} - E\hat{\theta})^2] + (E\hat{\theta} - \theta)^2 + 2(E\hat{\theta} - E\hat{\theta})E[(E\hat{\theta} - \theta)] \\ &= \text{Var}(\hat{\theta}) + \text{Bias}^2(\theta) \end{aligned}$$

Example

- $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1)$
- $\mu_1 = 1,\ \mu_2 = \overline{X}$.

$$\begin{aligned} \text{MSE}(\hat{\mu}_1) &= E(\hat{\mu}_1 - \mu)^2 = (1 - \mu)^2 \\ \text{MSE}(\hat{\mu}_2) &= E(\overline{X} - \mu)^2 = \text{Var}(\overline{X}) = \frac{1}{n} \end{aligned}$$

- Suppose that the true $\mu = 1$, then $\text{MSE}(\mu_1) = 0 < \text{MSE}(\mu_2)$, and no estimator can beat μ_1 in terms of MSE when true $\mu = 1$.
- Therefore, we cannot find an estimator that is uniformly the best in terms of MSE across all $\theta \in \Omega$ among all estimators
- Restrict the class of estimators, and find the "best" estimator within the small class.

Uniformly Minimum Variance Unbiased Estimator

Definition

$W^*(\mathbf{X})$ is the *best unbiased estimator*, or *uniformly minimum variance unbiased estimator (UMVUE)* of $\tau(\theta)$ if,

1. $E[W^*(\mathbf{X})|\theta] = \tau(\theta)$ for all θ (unbiased)
2. and $\text{Var}[W^*(\mathbf{X})|\theta] \leq \text{Var}[W(\mathbf{X})|\theta]$ for all θ, where W is any other unbiased estimator of $\tau(\theta)$ (minimum variance).

How to find the Best Unbiased Estimator

- Find the lower bound of variances of any unbiased estimator of $\tau(\theta)$, say $B(\theta)$.
- If W^* is an unbiased estimator of $\tau(\theta)$ and satisfies $\text{Var}[W^*(\mathbf{X})|\theta] = B(\theta)$, then W^* is the best unbiased estimator.

Cramer-Rao inequality

Theorem 7.3.9 : Cramer-Rao Theorem

Let $X_1, \cdots, X_n$ be a sample with joint pdf/pmf of $f_{\mathbf{X}}(\mathbf{x}|\theta)$. Suppose $W(\mathbf{X})$ is an estimator satisfying

1. $E[W(\mathbf{X})|\theta] = \tau(\theta),\ \forall\theta \in \Omega$.
2. $\text{Var}[W(\mathbf{X})|\theta] < \infty$.

For $h(\mathbf{x}) = 1$ and $h(\mathbf{x}) = W(\mathbf{x})$, if the differentiation and integrations are interchangeable, i.e.

$$\frac{d}{d\theta} E[h(\mathbf{x})|\theta] = \frac{d}{d\theta} \int_{x \in \mathcal{X}} h(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}|\theta) d\mathbf{x} = \int_{x \in \mathcal{X}} h(\mathbf{x}) \frac{\partial}{\partial\theta} f_{\mathbf{X}}(\mathbf{x}|\theta) d\mathbf{x}$$

Then, a lower bound of $\text{Var}[W(\mathbf{X})|\theta]$ is

$$\text{Var}[W(\mathbf{X})] \geq \frac{[\tau'(\theta)]^2}{E\left[\{\frac{\partial}{\partial\theta} \log f_{\mathbf{X}}(\mathbf{x}|\theta)\}^2\right]}$$

Proving Cramer-Rao Theorem (1/4)

By Cauchy-Schwarz inequality,

$$[\text{Cov}(X, Y)]^2 \leq \text{Var}(X)\text{Var}(Y)$$

Replacing X and Y,

$$\left[\text{Cov}\{W(\mathbf{X}), \frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\}\right]^2 \leq \text{Var}[W(\mathbf{X})]\text{Var}\left[\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right]$$

$$\text{Var}[W(\mathbf{X})] \geq \frac{\left[\text{Cov}\{W(\mathbf{X}), \frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\}\right]^2}{\text{Var}\left[\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right]}$$

Using $\text{Var}(X) = EX^2 - (EX)^2$,

$$\text{Var}\left[\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right] = E\left[\left\{\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right\}^2\right] - E\left[\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right]^2$$

Proving Cramer-Rao Theorem (2/4)

$$\begin{aligned}
E\left[\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right] &= \int_{\mathbf{x}\in\mathcal{X}}\left[\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{x}|\theta)\right] f_{\mathbf{X}}(\mathbf{x}|\theta)\,d\mathbf{x} \\
&= \int_{\mathbf{x}\in\mathcal{X}} \frac{\frac{\partial}{\partial\theta}f_{\mathbf{X}}(\mathbf{x}|\theta)}{f_{\mathbf{X}}(\mathbf{x}|\theta)} f_{\mathbf{X}}(\mathbf{x}|\theta)\,d\mathbf{x} \\
&= \int_{\mathbf{x}\in\mathcal{X}} \frac{\partial}{\partial\theta}f_{\mathbf{X}}(\mathbf{x}|\theta)\,d\mathbf{x} \\
&= \frac{d}{d\theta}\int_{\mathbf{x}\in\mathcal{X}} f_{\mathbf{X}}(\mathbf{x}|\theta)\,d\mathbf{x} \quad \text{(by assumption)} \\
&= \frac{d}{d\theta}1 = 0 \\
\text{Var}\left[\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right] &= E\left[\left\{\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right\}^2\right]
\end{aligned}$$

Proving Cramer-Rao Theorem (3/4)

$$\begin{aligned}
& \text{Cov}\left[W(\mathbf{X}), \frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right] \\
&= E\left[W(\mathbf{X}) \cdot \frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right] - E[W(\mathbf{X})]\, E\left[\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right] \\
&= E\left[W(\mathbf{X}) \cdot \frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right] = \int_{\mathbf{x}\in\mathcal{X}} W(\mathbf{x})\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{x}|\theta) f(\mathbf{x}|\theta)\,d\mathbf{x} \\
&= \int_{\mathbf{x}\in\mathcal{X}} W(\mathbf{x})\frac{\frac{\partial}{\partial\theta}f_{\mathbf{X}}(\mathbf{x}|\theta)}{f(\mathbf{x}|\theta)} f(\mathbf{x}|\theta)\,d\mathbf{x} = \int_{\mathbf{x}\in\mathcal{X}} W(\mathbf{x})\frac{\partial}{\partial\theta}f_{\mathbf{X}}(\mathbf{x}|\theta) \\
&= \frac{d}{d\theta}\int_{\mathbf{x}\in\mathcal{X}} W(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}|\theta) \quad \text{(by assumption)} \\
&= \frac{d}{d\theta}E[W(\mathbf{X})] = \frac{d}{d\theta}\tau(\theta) = \tau'(\theta)
\end{aligned}$$

Proving Cramer-Rao Theorem (4/4)

From the previous results

$$\text{Var}\left[\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right] = E\left[\left\{\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right\}^2\right]$$

$$\text{Cov}\left[W(\mathbf{X}), \frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right] = \tau'(\theta)$$

Therefore, Cramer-Rao lower bound is

$$\begin{aligned}
\text{Var}[W(\mathbf{X})] &\geq \frac{\left[\text{Cov}\{W(\mathbf{X}), \frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\}\right]^2}{\text{Var}\left[\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right]} \\
&= \frac{[\tau'(\theta)]^2}{E\left[\{\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\}^2\right]}
\end{aligned}$$

Cramer-Rao bound in iid case

Corollary 7.3.10

If $X_1, \cdots, X_n$ are iid samples from pdf/pmf $f_X(x|\theta)$, and the assumptions in the above Cramer-Rao theorem hold, then the lower-bound of $\text{Var}[W(\mathbf{X})|\theta]$ becomes

$$\text{Var}[W(\mathbf{X})] \geq \frac{[\tau'(\theta)]^2}{nE\left[\{\frac{\partial}{\partial\theta}\log f_X(X|\theta)\}^2\right]}$$

Proof

We need to show that

$$E\left[\left\{\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right\}^2\right] = nE\left[\left\{\frac{\partial}{\partial\theta}\log f_X(X|\theta)\right\}^2\right]$$

Proving Corollary 7.3.10

$$\begin{aligned}
E\left[\left\{\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right\}^2\right] &= E\left[\left\{\frac{\partial}{\partial\theta}\log\prod_{i=1}^n f_X(X_i|\theta)\right\}^2\right] \\
&= E\left[\left\{\frac{\partial}{\partial\theta}\sum_{i=1}^n\log f_X(X_i|\theta)\right\}^2\right] \\
&= E\left[\left\{\sum_{i=1}^n\frac{\partial}{\partial\theta}\log f_X(X_i|\theta)\right\}^2\right] \\
&= E\Big[\textstyle\sum_{i=1}^n\left\{\frac{\partial}{\partial\theta}\log f_X(X_i|\theta)\right\}^2 + \\
&\quad \textstyle\sum_{i\neq j}\frac{\partial}{\partial\theta}\log f_X(X_i|\theta)\frac{\partial}{\partial\theta}\log f_X(X_j|\theta)\Big]
\end{aligned}$$

Proving Corollary 7.3.10

Because $X_1, \cdots, X_n$ are independent,

$$E\left[\sum_{i\neq j}\frac{\partial}{\partial\theta}\log f_X(X_i|\theta)\frac{\partial}{\partial\theta}\log f_X(X_j|\theta)\right]$$

$$= \sum_{i\neq j}E\left[\frac{\partial}{\partial\theta}\log f_X(X_i|\theta)\right]E\left[\frac{\partial}{\partial\theta}\log f_X(X_j|\theta)\right] = 0$$

$$\begin{aligned}
E\left[\left\{\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right\}^2\right] &= E\left[\sum_{i=1}^n\left\{\frac{\partial}{\partial\theta}\log f_X(X_i|\theta)\right\}^2\right] \\
&= \sum_{i=1}^n E\left[\left\{\frac{\partial}{\partial\theta}\log f_X(X_i|\theta)\right\}^2\right] \\
&= nE\left[\left\{\frac{\partial}{\partial\theta}\log f_X(X|\theta)\right\}^2\right]
\end{aligned}$$

Remark from Corollary 7.3.10

In iid case, Cramer-Rao lower bound for an unbiased estimator of θ is

$$\text{Var}[W(\mathbf{X})] \geq \frac{1}{nE\left[\{\frac{\partial}{\partial\theta}\log f_X(X|\theta)\}^2\right]}$$

Because $\tau(\theta) = \theta$ and $\tau'(\theta) = 1$.

Score Function

Definition: Score or Score Function for X

$$\begin{aligned} X_1, \cdots, X_n &\overset{\text{i.i.d.}}{\sim} f_X(x|\theta) \\ S(X|\theta) &= \frac{\partial}{\partial\theta} \log f_X(X|\theta) \\ E[S(X|\theta)] &= 0 \\ S_n(X|\theta) &= \frac{\partial}{\partial\theta} \log f_{\mathbf{X}}(\mathbf{X}|\theta) \end{aligned}$$

Fisher Information Number

Definition: Fisher Information Number

$$\begin{aligned} I(\theta) &= E\left[\left\{\frac{\partial}{\partial\theta}\log f_X(X|\theta)\right\}^2\right] = E\left[S^2(X|\theta)\right] \\ I_n(\theta) &= E\left[\left\{\frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{X}|\theta)\right\}^2\right] \\ &= nE\left[\left\{\frac{\partial}{\partial\theta}\log f_X(X|\theta)\right\}^2\right] = nI(\theta) \end{aligned}$$

The bigger the information number, the more information we have about θ, the smaller bound on the variance of unbiased estimates.

Simplified Fisher Information

Lemma 7.3.11

If $f_X(x|\theta)$ satisfies the two interchangeability conditions

$$\begin{aligned} \frac{d}{d\theta}\int_{x\in\mathcal{X}} f_X(x|\theta)\,dx &= \int_{x\in\mathcal{X}} \frac{\partial}{\partial\theta} f_X(x|\theta)\,dx \\ \frac{d}{d\theta}\int_{x\in\mathcal{X}} \frac{\partial}{\partial\theta} f_X(x|\theta)\,dx &= \int_{x\in\mathcal{X}} \frac{\partial^2}{\partial\theta^2} f_X(x|\theta)\,dx \end{aligned}$$

which are true for exponential family, then

$$I(\theta) = E\left[\left\{\frac{\partial}{\partial\theta}\log f_X(X|\theta)\right\}^2\right] = -E\left[\frac{\partial^2}{\partial\theta^2}\log f_X(X|\theta)\right]$$

Example - Poisson Distribution

- $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Poisson}(\lambda)$
- $\lambda_1 = \overline{X}$
- $\lambda_2 = s^2_{\mathbf{X}}$
- $E[\lambda_1] = E(\overline{X}) = \lambda.$

Cramer-Rao lower bound is $I_n^{-1}(\lambda) = [nI(\lambda)]^{-1}$.

$$\begin{aligned} I(\lambda) &= E\left[\left\{\frac{\partial}{\partial\lambda}\log f_X(X|\lambda)\right\}^2\right] = -E\left[\frac{\partial^2}{\partial\lambda^2}\log f_X(X|\lambda)\right] \\ &= -E\left[\frac{\partial^2}{\partial\lambda^2}\log\frac{e^{-\lambda}\lambda^X}{X!}\right] = -E\left[\frac{\partial^2}{\partial\lambda^2}(-\lambda + X\log\lambda - \log X!)\right] \\ &= E\left[\frac{X}{\lambda^2}\right] = \frac{1}{\lambda^2}E(X) = \frac{1}{\lambda} \end{aligned}$$

Example - Poisson Distribution (cont'd)

Therefore, the Cramer-Rao lower bound is

$$\mathrm{Var}[W(\mathbf{X})] \geq \frac{1}{nI(\lambda)} = \frac{\lambda}{n}$$

where W is any unbiased estimator.

$$\mathrm{Var}(\hat{\lambda}_1) = \mathrm{Var}(\overline{X}) = \frac{\mathrm{Var}(X)}{n} = \frac{\lambda}{n}$$

Therefore, $\lambda_1 = \overline{X}$ is the best unbiased estimator of λ.

$$\mathrm{Var}(\hat{\lambda}_2) > \frac{\lambda}{n}$$

(details is omitted), so $\hat{\lambda}_2$ is not the best unbiased estimator.

With and without Lemma 7.3.11

With Lemma 7.3.11

$$I(\lambda) = -E\left[\frac{\partial^2}{\partial\lambda^2}\log f_X(X|\lambda)\right] = -E\left[\frac{\partial^2}{\partial\lambda^2}(-\lambda + X\log\lambda - \log X!)\right] = \frac{1}{\lambda}$$

Without Lemma 7.3.11

$$\begin{aligned} I(\lambda) &= E\left[\left\{\frac{\partial}{\partial\lambda}\log f_X(X|\lambda)\right\}^2\right] = E\left[\left\{\frac{\partial}{\partial\lambda}(-\lambda + X\log\lambda - \log X!)\right\}^2\right] \\ &= E\left[\left\{-1 + \frac{X}{\lambda}\right\}^2\right] = E\left[1 - 2\frac{X}{\lambda} + \frac{X^2}{\lambda^2}\right] = 1 - 2\frac{E(X)}{\lambda} + \frac{E(X^2)}{\lambda^2} \\ &= 1 - 2\frac{E(X)}{\lambda} + \frac{\mathrm{Var}(X) + [E(X)]^2}{\lambda^2} = 1 - 2\frac{\lambda}{\lambda} + \frac{\lambda + \lambda^2}{\lambda^2} = \frac{1}{\lambda} \end{aligned}$$

Example - Normal Distribution

- $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, where σ^2 is known.
- The Cramer-Rao bound for μ is $[nI(\mu)]^{-1}$.

$$\begin{aligned} I(\mu) &= -E\left[\frac{\partial^2}{\partial\mu^2}\log f_X(X|\mu)\right] \\ &= -E\left[\frac{\partial^2}{\partial\mu^2}\log\left\{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(X-\mu)^2}{2\sigma^2}\right)\right\}\right] \\ &= -E\left[\frac{\partial^2}{\partial\mu^2}\left\{-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(X-\mu)^2}{2\sigma^2}\right\}\right] \\ &= -E\left[\frac{\partial}{\partial\mu}\left\{\frac{2(X-\mu)}{2\sigma^2}\right\}\right] = \frac{1}{\sigma^2} \end{aligned}$$

Applying Lemma 7.3.11

Question

When can we interchange the order of differentiation and integration?

Answer

- For exponential family, always yes.
- Not always yes for non-exponential family. Will have to check the individual case.

Example

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathrm{Uniform}(0, \theta)$

$$\frac{d}{d\theta}\int_0^\theta h(x) f_X(x|\theta)\,dx \neq \int_0^\theta h(x)\frac{\partial}{\partial\theta} f_X(x|\theta)\,dx$$

Summary

Today

- Invariance Property
- Mean Squared Error
- Unbiased Estimator
- Cramer-Rao inequality

Next Lecture

- More on Cramer-Rao inequality