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.
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.
1. What is a maximum likelihood estimator (MLE)?
Last Lecture

1. What is a maximum likelihood estimator (MLE)?
2. How can you find an MLE?
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?
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 $\theta$, can you also know MLE of $\tau(\theta)$?
Recap - Maximum Likelihood Estimator

Definition

- For a given sample point $\mathbf{x} = (x_1, \ldots, x_n)$,
- let $\hat{\theta}(\mathbf{x})$ be the value such that $L(\theta|\mathbf{x})$ attains its maximum.
- $L(\theta|\mathbf{x})$ 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 $\theta$ based on data $\mathbf{x}$,
- and $\hat{\theta}(\mathbf{X})$ is the maximum likelihood estimator (MLE) of $\theta$. 
Recap - Invariance Property of MLE

Question

If $\hat{\theta}$ is the MLE of $\theta$, 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 $\tau$ is a one-to-one function.
- Therefore $\hat{\eta} = \tau(\hat{p})$. 
Invariance Property of MLE

Fact

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

Proof.

The likelihood function in terms of $\tau(\theta)$ is $L(\tau(\theta)|x) = \prod_{i=1}^{n} f(X(x_i|\tau(\theta))) = L(\tau^{-1}(x)|\theta)$. We know this function is maximized when $\tau^{-1}(x) = \hat{\theta}$, or equivalently, when $\theta = \tau(\hat{\theta})$. Therefore, MLE of $\theta = \tau(\theta)$ is $\tau(\hat{\theta})$. 

Hyun Min Kang
Invariance Property of MLE

Fact

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

Proof

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

$$L^*(\tau(\theta)|x) = \prod_{i=1}^{n} f_X(x_i|\theta)$$
Invariance Property of MLE

Fact

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

Proof

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

$$L^*(\tau(\theta) | \mathbf{x}) = \prod_{i=1}^{n} f_X(x_i | \theta) = \prod_{i=1}^{n} f(x_i | \tau^{-1}(\eta))$$
Invariance Property of MLE

Fact

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

Proof

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

$$L^*(\tau(\theta)|x) = \prod_{i=1}^{n} f_{X}(x_i|\theta) = \prod_{i=1}^{n} f(x_i|\tau^{-1}(\eta))$$

$$= L(\tau^{-1}(\eta)|x)$$
Invariance Property of MLE

Fact

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

Proof

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

\[
L^*(\tau(\theta) | \mathbf{x}) = \prod_{i=1}^{n} f_X(x_i | \theta) = \prod_{i=1}^{n} f(x_i | \tau^{-1}(\eta))
\]

\[
= L(\tau^{-1}(\eta) | \mathbf{x})
\]

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

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

Proof

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

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

$$= L(\tau^{-1}(\eta)|x)$$

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 $L(\theta|x)$ be the likelihood function for a given data $x_1, \cdots, x_n$. 

\[
\text{Induced Likelihood Function}
\]
Induced Likelihood Function

Definition

- Let $L(\theta|x)$ 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 $\theta$. 
Induced Likelihood Function

**Definition**

- Let $L(\theta|x)$ 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 $\theta$.

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

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

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

Definition

- Let $L(\theta|x)$ 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 $\theta$.

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

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

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

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

Invariance Property of MLE

Theorem 7.2.10

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

**Theorem 7.2.10**

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

**Proof - Using Induced Likelihood Function**

\[
L^*(\hat{\eta}|x) = \sup_{\eta} L^*(\eta|x)
\]
Invariance Property of MLE

**Theorem 7.2.10**

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

**Proof - Using Induced Likelihood Function**

\[
L^*(\hat{\eta}|\mathbf{x}) = \sup_{\eta} L^*(\eta|\mathbf{x}) = \sup_{\eta} \sup_{\theta \in \tau^{-1}(\eta)} L(\theta|\mathbf{x})
\]
Invariance Property of MLE

**Theorem 7.2.10**

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

**Proof - Using Induced Likelihood Function**

\[
L^*(\hat{\eta}|x) = \sup_{\eta} L^*(\eta|x) = \sup_{\eta} \sup_{\theta \in \tau^{-1}(\eta)} L(\theta|x) = \sup_{\theta} L(\theta|x)
\]
Invariance Property of MLE

Theorem 7.2.10

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

Proof - Using Induced Likelihood Function

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

$$= \sup_{\theta} L(\theta|\mathbf{x}) = L(\hat{\theta}|\mathbf{x})$$
Invariance Property of MLE

**Theorem 7.2.10**

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

**Proof - Using Induced Likelihood Function**

\[
L^*(\hat{\eta}|\mathbf{x}) = \sup_{\eta} L^*(\eta|x) = \sup_{\eta} \sup_{\theta \in \tau^{-1}(\eta)} L(\theta|x) \\
= \sup_{\theta} L(\theta|x) = L(\hat{\theta}|\mathbf{x}) \\
L(\hat{\theta}|\mathbf{x}) = \sup_{\theta \in \tau^{-1}(\tau(\hat{\theta}))} L(\theta|x)
\]
Invariance Property of MLE

Theorem 7.2.10

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

Proof - Using Induced Likelihood Function

\[
L^*(\hat{\eta}|x) = \sup_{\eta} L^*(\eta|x) = \sup_{\eta} \sup_{\theta \in \tau^{-1}(\eta)} L(\theta|x)
\]

\[
= \sup_{\theta} L(\theta|x) = L(\hat{\theta}|x)
\]

\[
L(\hat{\theta}|x) = \sup_{\theta \in \tau^{-1}(\tau(\hat{\theta}))} L(\theta|x) = L^*[\tau(\hat{\theta})|x]
\]
Theorem 7.2.10

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

Proof - Using Induced Likelihood Function

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

$$= \sup_{\theta} L(\theta|\mathbf{x}) = L(\hat{\theta}|\mathbf{x})$$

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

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.
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.
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.
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 \( \theta \), then the bias of \( \theta \) is defined as

\[
\text{Bias}(\theta) = E(\hat{\theta}) - \theta
\]
Method of Evaluating Estimators

Definition: Unbiasedness

Suppose \( \hat{\theta} \) is an estimator for \( \theta \), then the bias of \( \theta \) 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 \( \theta \).
Method of Evaluating Estimators

**Definition : Unbiasedness**

Suppose \( \hat{\theta} \) is an estimator for \( \theta \), then the bias of \( \theta \) 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 \( \theta \).

**Example**

\( X_1, \ldots, X_n \) are iid samples from a distribution with mean \( \mu \). Let

\[
\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i
\]

is an estimator of \( \mu \).
Method of Evaluating Estimators

Definition: Unbiasedness

Suppose \( \hat{\theta} \) is an estimator for \( \theta \), then the bias of \( \theta \) 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 \( \theta \).

Example

\( X_1, \ldots, X_n \) are iid samples from a distribution with mean \( \mu \). Let

\[
\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i
\]

is an estimator of \( \mu \). The bias is

\[
\text{Bias}(\mu) = E(\bar{X}) - \mu
\]
Definition: Unbiasedness

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

Example

$X_1, \cdots, X_n$ are iid samples from a distribution with mean $\mu$. Let

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

is an estimator of $\mu$. The bias is

$$\text{Bias}(\mu) = E(\bar{X}) - \mu$$

$$= E \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) - \mu$$
Method of Evaluating Estimators

Definition: Unbiasedness

Suppose \( \hat{\theta} \) is an estimator for \( \theta \), then the bias of \( \theta \) 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 \( \theta \).

Example

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

\[
\text{Bias}(\mu) = E(\bar{X}) - \mu
\]

\[
= E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) - \mu
\]
Method of Evaluating Estimators

Definition: Unbiasedness

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

Example

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

$$\text{Bias}(\mu) = E(\overline{X}) - \mu$$

$$= E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) - \mu = \frac{1}{n} \sum_{i=1}^{n} E(X_i) - \mu$$
Method of Evaluating Estimators

Definition: Unbiasedness

Suppose \( \hat{\theta} \) is an estimator for \( \theta \), then the bias of \( \theta \) 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 \( \theta \).

Example

\( X_1, \ldots, X_n \) are iid samples from a distribution with mean \( \mu \). Let \( \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \) is an estimator of \( \mu \). The bias is

\[
\text{Bias}(\mu) = E(\bar{X}) - \mu
\]

\[
= E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) - \mu
= \frac{1}{n} \sum_{i=1}^{n} E(X_i) - \mu
= \mu - \mu = 0
\]
Method of Evaluating Estimators

Definition: Unbiasedness

Suppose \( \hat{\theta} \) is an estimator for \( \theta \), then the bias of \( \theta \) 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 \( \theta \).

Example

\( X_1, \ldots, X_n \) are iid samples from a distribution with mean \( \mu \). Let

\[
\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i
\]

is an estimator of \( \mu \). The bias is

\[
\text{Bias}(\mu) = E(\overline{X}) - \mu
\]

\[
= E \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) - \mu = \frac{1}{n} \sum_{i=1}^{n} E(X_i) - \mu = \mu - \mu = 0
\]

Therefore \( \overline{X} \) is an unbiased estimator for \( \mu \).
How important is unbiased?

- $^1$ (blue) is unbiased but has a chance to be very far away from $\theta = 0$.
- $^2$ (red) is biased but more likely to be closer to the true $\theta$ than $^1$.
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 $\theta$ than $\hat{\theta}_1$. 
Mean Squared Error

**Definition**

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

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

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

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

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

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

$$= E[(\hat{\theta} - E\hat{\theta})^2] + (E\hat{\theta} - \theta)^2 + 2(E\hat{\theta} - E\hat{\theta}) E[(E\hat{\theta} - \theta)]$$
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**

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

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

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

$$= \text{Var}(\hat{\theta}) + \text{Bias}^2(\theta)$$
Example

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

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

$$\text{MSE}(\hat{\mu}_1) = E(\hat{\mu}_1 - \mu)^2 = (1 - \mu)^2$$
Example

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

\[
\text{MSE}(\hat{\mu}_1) = E(\hat{\mu}_1 - \mu)^2 = (1 - \mu)^2
\]
\[
\text{MSE}(\hat{\mu}_2) = E(\overline{X} - \mu)^2 = \text{Var}(\overline{X}) = \frac{1}{n}
\]
Example

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

\[
\begin{align*}
\text{MSE}(\hat{\mu}_1) &= E(\hat{\mu}_1 - \mu)^2 = (1 - \mu)^2 \\
\text{MSE}(\hat{\mu}_2) &= E(\bar{X} - \mu)^2 = \text{Var}(\bar{X}) = \frac{1}{n}
\end{align*}
\]

- Suppose that the true $\mu = 1$, then $\text{MSE}(\mu_1) = 0 < \text{MSE}(\mu_2)$, and no estimator can beat $\mu_1$ in terms of MSE when true $\mu = 1$. 
Example

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

\[
\begin{align*}
\text{MSE}(\hat{\mu}_1) &= E(\hat{\mu}_1 - \mu)^2 = (1 - \mu)^2 \\
\text{MSE}(\hat{\mu}_2) &= E(\bar{X} - \mu)^2 = \text{Var}(\bar{X}) = \frac{1}{n}
\end{align*}
\]

- Suppose that the true $\mu = 1$, then $\text{MSE}(\mu_1) = 0 < \text{MSE}(\mu_2)$, and no estimator can beat $\mu_1$ 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.
Example

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

\[
\text{MSE}(\hat{\mu}_1) = E(\hat{\mu}_1 - \mu)^2 = (1 - \mu)^2
\]
\[
\text{MSE}(\hat{\mu}_2) = E(\bar{X} - \mu)^2 = \text{Var}(\bar{X}) = \frac{1}{n}
\]

- Suppose that the true $\mu = 1$, then $\text{MSE}(\mu_1) = 0 < \text{MSE}(\mu_2)$, and no estimator can beat $\mu_1$ 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**

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

\[
\begin{align*}
\mathbb{E}[W^*(X)] &= \tau(\theta) \\
\text{Var}[W^*(X)] &= \text{Var}[W(X)]
\end{align*}
\] for all \( \theta \), where \( W \) is any other unbiased estimator of \( \tau(\theta) \) (minimum variance).
Uniformly Minimum Variance Unbiased Estimator

**Definition**

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

1. \( E[W^*(X)|\theta] = \tau(\theta) \) for all \( \theta \) (unbiased)
Uniformly Minimum Variance Unbiased Estimator

**Definition**

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

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

**Definition**

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

1. $E[W^*(X)|\theta] = \tau(\theta)$ for all $\theta$ (unbiased)
2. and $\text{Var}[W^*(X)|\theta] \leq \text{Var}[W(X)|\theta]$ for all $\theta$, 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 $B(\theta)$.
Uniformly Minimum Variance Unbiased Estimator

**Definition**

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

1. \( E[W^*(X)|\theta] = \tau(\theta) \) for all \( \theta \) (unbiased)
2. and \( \text{Var}[W^*(X)|\theta] \leq \text{Var}[W(X)|\theta] \) for all \( \theta \), 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 \( \text{Var}[W^*(X)|\theta] = B(\theta) \), then \( W^* \) is the best unbiased estimator.
Cramer-Rao inequality

Theorem 7.3.9 : Cramer-Rao Theorem

Let $X_1, \ldots, X_n$ be a sample with joint pdf/pmf of $f_X(x|\theta)$. Suppose $W(X)$ is an estimator satisfying
Cramer-Rao inequality

**Theorem 7.3.9 : Cramer-Rao Theorem**

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

1. $E[W(X)|\theta] = \tau(\theta), \ \forall \theta \in \Omega$. 

Let $X_1, \cdots, X_n$ be a sample with joint pdf/pmf of $f_X(x|\theta)$. Suppose $W(X)$ is an estimator satisfying
Cramer-Rao inequality

**Theorem 7.3.9 : Cramer-Rao Theorem**

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

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

---

Hyun Min Kang  
Biostatistics 602 - Lecture 11  
February 14th, 2013  
16 / 33
Cramer-Rao inequality

Theorem 7.3.9 : Cramer-Rao Theorem

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

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

For $h(x) = 1$ and $h(x) = W(x)$, if the differentiation and integrations are interchangeable, i.e.
Theorem 7.3.9 : Cramer-Rao Theorem

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

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

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

$$
\frac{d}{d\theta} E[h(x)|\theta] = \frac{d}{d\theta} \int_{x \in X} h(x)f_X(x|\theta) \, dx = \int_{x \in X} h(x) \frac{\partial}{\partial \theta} f_X(x|\theta) \, dx
$$
Cramer-Rao inequality

**Theorem 7.3.9 : Cramer-Rao Theorem**

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

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

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

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

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

$$
\text{Var}[W(X)] \geq \frac{[\tau'(\theta)]^2}{E \left[ \left\{ \frac{\partial}{\partial \theta} \log f_X(x|\theta) \right\}^2 \right]}
$$
Proving Cramer-Rao Theorem (1/4)

By Cauchy-Schwarz inequality,

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

\[
\left[ \text{Cov}\{ W(\mathbf{X}), \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}\mid \theta) \} \right]^2 \leq \text{Var}[W(\mathbf{X})] \text{Var} \left[ \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}\mid \theta) \right]
\]
Proving Cramer-Rao Theorem (1/4)

By Cauchy-Schwarz inequality,

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

Replacing $X$ and $Y$,

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

$$\text{Var}[W(X)] \geq \frac{\left[ \text{Cov}\{ W(X), \frac{\partial}{\partial \theta} \log f_X(X|\theta) \} \right]^2}{\text{Var} \left[ \frac{\partial}{\partial \theta} \log f_X(X|\theta) \right]}$$
Proving Cramer-Rao Theorem (1/4)

By Cauchy-Schwarz inequality,

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

Replacing \(X\) and \(Y\),

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

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

Using \(\text{Var}(X) = E[X^2] - (E[X])^2\),

\[
\text{Var} \left[\frac{\partial}{\partial \theta} \log f_X(X|\theta)\right] = E \left[\left\{\frac{\partial}{\partial \theta} \log f_X(X|\theta)\right\}^2\right] - E \left[\frac{\partial}{\partial \theta} \log f_X(X|\theta)\right]^2
\]
Proving Cramer-Rao Theorem (2/4)

\[ E \left[ \frac{\partial}{\partial \theta} \log f_X(x|\theta) \right] = \int_{x \in \mathcal{X}} \left[ \frac{\partial}{\partial \theta} \log f_X(x|\theta) \right] f_X(x|\theta) \, dx \]
Proving Cramer-Rao Theorem (2/4)

\[
E \left[ \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}|\theta) \right] = \int_{\mathbf{x} \in \mathcal{X}} \left[ \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{x}|\theta) \right] f_{\mathbf{X}}(\mathbf{x}|\theta) \, d\mathbf{x} \\
= \int_{\mathbf{x} \in \mathcal{X}} \frac{\partial f_{\mathbf{X}}(\mathbf{x}|\theta)}{\partial \theta} \frac{f_{\mathbf{X}}(\mathbf{x}|\theta)}{f_{\mathbf{X}}(\mathbf{x}|\theta)} f_{\mathbf{X}}(\mathbf{x}|\theta) \, d\mathbf{x} \\
= \frac{\partial}{\partial \theta} \int_{\mathbf{x} \in \mathcal{X}} f_{\mathbf{X}}(\mathbf{x}|\theta) \, d\mathbf{x} \\
= \frac{\partial}{\partial \theta} 1 = 0
\]
Proving Cramer-Rao Theorem (2/4)

\[
E \left[ \frac{\partial}{\partial \theta} \log f_X(X|\theta) \right] = \int_{x \in \mathcal{X}} \left[ \frac{\partial}{\partial \theta} \log f_X(x|\theta) \right] f_X(x|\theta) \, dx
\]

\[
= \int_{x \in \mathcal{X}} \frac{\partial f_X(x|\theta)}{f_X(x|\theta)} f_X(x|\theta) \, dx
\]

\[
= \int_{x \in \mathcal{X}} \frac{\partial}{\partial \theta} f_X(x|\theta) \, dx
\]
Proving Cramer-Rao Theorem (2/4)

\[
E \left[ \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}\mid \theta) \right] = \int_{\mathbf{x} \in \mathcal{X}} \left[ \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{x}\mid \theta) \right] f_{\mathbf{X}}(\mathbf{x}\mid \theta) \, d\mathbf{x} \\
= \int_{\mathbf{x} \in \mathcal{X}} \frac{\partial}{\partial \theta} f_{\mathbf{X}}(\mathbf{x}\mid \theta) \frac{f_{\mathbf{X}}(\mathbf{x}\mid \theta)}{f_{\mathbf{X}}(\mathbf{x}\mid \theta)} \, d\mathbf{x} \\
= \int_{\mathbf{x} \in \mathcal{X}} \frac{\partial}{\partial \theta} f_{\mathbf{X}}(\mathbf{x}\mid \theta) \, d\mathbf{x} \\
= \frac{d}{d\theta} \int_{\mathbf{x} \in \mathcal{X}} f_{\mathbf{X}}(\mathbf{x}\mid \theta) \, d\mathbf{x} \quad \text{(by assumption)}
\]
Proving Cramer-Rao Theorem (2/4)

\[
E \left[ \frac{\partial}{\partial \theta} \log f_X(x|\theta) \right] = \int_{x \in \mathcal{X}} \left[ \frac{\partial}{\partial \theta} \log f_X(x|\theta) \right] f_X(x|\theta) \, dx \\
= \int_{x \in \mathcal{X}} \frac{\partial}{\partial \theta} f_X(x|\theta) f_X(x|\theta) \, dx \\
= \int_{x \in \mathcal{X}} \frac{\partial}{\partial \theta} f_X(x|\theta) f_X(x|\theta) \, dx \\
= \frac{d}{d\theta} \int_{x \in \mathcal{X}} f_X(x|\theta) \, dx \\
= \frac{d}{d\theta} \int_{x \in \mathcal{X}} f_X(x|\theta) \, dx \quad \text{(by assumption)} \\
= \frac{d}{d\theta} 1 = 0
\]
Proving Cramer-Rao Theorem (2/4)

\[
E \left[ \frac{\partial}{\partial \theta} \log f_X(x|\theta) \right] = \int_{x \in \mathcal{X}} \left[ \frac{\partial}{\partial \theta} \log f_X(x|\theta) \right] f_X(x|\theta) \, dx
\]

\[
= \int_{x \in \mathcal{X}} \frac{\partial}{\partial \theta} f_X(x|\theta) f_X(x|\theta) \, dx
\]

\[
= \int_{x \in \mathcal{X}} \frac{\partial}{\partial \theta} f_X(x|\theta) \, dx
\]

\[
= \frac{d}{d\theta} \int_{x \in \mathcal{X}} f_X(x|\theta) \, dx \quad \text{(by assumption)}
\]

\[
= \frac{d}{d\theta} 1 = 0
\]

\[
\text{Var} \left[ \frac{\partial}{\partial \theta} \log f_X(x|\theta) \right] = E \left[ \left\{ \frac{\partial}{\partial \theta} \log f_X(x|\theta) \right\}^2 \right]
\]
Proving Cramer-Rao Theorem (3/4)

\[ \text{Cov} \left[ W(X), \frac{\partial}{\partial \theta} \log f_X(X|\theta) \right] \]
Proving Cramer-Rao Theorem (3/4)

\[
\text{Cov} \left[ W(\mathbf{X}), \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}|\theta) \right] \\
= E \left[ W(\mathbf{X}) \cdot \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}|\theta) \right] - E[ W(\mathbf{X})] \ E \left[ \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}|\theta) \right]
\]
Proving Cramer-Rao Theorem (3/4)

\[
\text{Cov} \left[ W(\mathbf{X}), \frac{\partial}{\partial \theta} \log f_\mathbf{X}(\mathbf{X}|\theta) \right] \\
= E \left[ W(\mathbf{X}) \cdot \frac{\partial}{\partial \theta} \log f_\mathbf{X}(\mathbf{X}|\theta) \right] - E[ W(\mathbf{X})] E \left[ \frac{\partial}{\partial \theta} \log f_\mathbf{X}(\mathbf{X}|\theta) \right] \\
= E \left[ W(\mathbf{X}) \cdot \frac{\partial}{\partial \theta} \log f_\mathbf{X}(\mathbf{X}|\theta) \right]
\]
Proving Cramer-Rao Theorem (3/4)

\[
\text{Cov} \left[ W(\mathbf{X}), \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X} | \theta) \right] \\
= E \left[ W(\mathbf{X}) \cdot \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X} | \theta) \right] - E[ W(\mathbf{X})] E \left[ \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X} | \theta) \right] \\
= E \left[ W(\mathbf{X}) \cdot \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X} | \theta) \right] = \int_{\mathbf{x} \in \mathcal{X}} W(\mathbf{x}) \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{x} | \theta) f(\mathbf{x} | \theta) d\mathbf{x}
\]
Proving Cramer-Rao Theorem (3/4)

\[
\text{Cov} \left[ W(\mathbf{X}), \frac{\partial}{\partial \theta} \log f_X(\mathbf{X}|\theta) \right]
\]

\[
= E \left[ W(\mathbf{X}) \cdot \frac{\partial}{\partial \theta} \log f_X(\mathbf{X}|\theta) \right] - E[W(\mathbf{X})] E \left[ \frac{\partial}{\partial \theta} \log f_X(\mathbf{X}|\theta) \right]
\]

\[
= E \left[ W(\mathbf{X}) \cdot \frac{\partial}{\partial \theta} \log f_X(\mathbf{X}|\theta) \right] = \int_{\mathbf{x} \in \mathcal{X}} W(\mathbf{x}) \frac{\partial}{\partial \theta} \log f_X(\mathbf{x}|\theta) f(\mathbf{x}|\theta) d\mathbf{x}
\]

\[
= \int_{\mathbf{x} \in \mathcal{X}} W(\mathbf{x}) \frac{\partial}{\partial \theta} f_X(\mathbf{x}|\theta) \frac{f(\mathbf{x}|\theta)}{f(\mathbf{x}|\theta)} d\mathbf{x}
\]
Proving Cramer-Rao Theorem (3/4)

\[
\text{Cov} \left[ W(\mathbf{X}), \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}|\theta) \right] \\
= E \left[ W(\mathbf{X}) \cdot \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}|\theta) \right] - E[ W(\mathbf{X})] \cdot E \left[ \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}|\theta) \right] \\
= E \left[ W(\mathbf{X}) \cdot \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}|\theta) \right] = \int_{\mathbf{x} \in \mathcal{X}} W(\mathbf{x}) \frac{\partial}{\partial \theta} \log f_{\mathbf{x}}(\mathbf{x}|\theta) f(\mathbf{x}|\theta) d\mathbf{x} \\
= \int_{\mathbf{x} \in \mathcal{X}} W(\mathbf{x}) \frac{\partial}{\partial \theta} \frac{f_{\mathbf{x}}(\mathbf{x}|\theta)}{f(\mathbf{x}|\theta)} f(\mathbf{x}|\theta) d\mathbf{x} = \int_{\mathbf{x} \in \mathcal{X}} W(\mathbf{x}) \frac{\partial}{\partial \theta} f_{\mathbf{x}}(\mathbf{x}|\theta)
\]
Proving Cramer-Rao Theorem (3/4)

\[
\text{Cov} \left[ W(\mathbf{X}), \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}|\theta) \right] = E \left[ W(\mathbf{X}) \cdot \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}|\theta) \right] - E[ W(\mathbf{X})] E \left[ \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}|\theta) \right] \\
= E \left[ W(\mathbf{X}) \cdot \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}|\theta) \right] = \int_{\mathbf{x} \in \mathcal{X}} W(\mathbf{x}) \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{x}|\theta) f(\mathbf{x}|\theta) d\mathbf{x} \\
= \int_{\mathbf{x} \in \mathcal{X}} W(\mathbf{x}) \frac{\partial}{\partial \theta} f_{\mathbf{X}}(\mathbf{x}|\theta) f(\mathbf{x}|\theta) d\mathbf{x} = \int_{\mathbf{x} \in \mathcal{X}} W(\mathbf{x}) \frac{\partial}{\partial \theta} f_{\mathbf{X}}(\mathbf{x}|\theta) \\
= \frac{d}{d\theta} \int_{\mathbf{x} \in \mathcal{X}} W(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}|\theta) \quad \text{(by assumption)}
\]
Proving Cramer-Rao Theorem (3/4)

\[
\text{Cov} \left[ W(\mathbf{X}), \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}|\theta) \right] \\
= E \left[ W(\mathbf{X}) \cdot \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}|\theta) \right] - E[ W(\mathbf{X})] E \left[ \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}|\theta) \right] \\
= E \left[ W(\mathbf{X}) \cdot \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X}|\theta) \right] = \int_{\mathbf{x} \in \mathcal{X}} W(\mathbf{x}) \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{x}|\theta) f(\mathbf{x}|\theta) \, d\mathbf{x} \\
= \int_{\mathbf{x} \in \mathcal{X}} W(\mathbf{x}) \frac{\partial}{\partial \theta} f_{\mathbf{X}}(\mathbf{x}|\theta) f(\mathbf{x}|\theta) \, d\mathbf{x} = \int_{\mathbf{x} \in \mathcal{X}} W(\mathbf{x}) \frac{\partial}{\partial \theta} f_{\mathbf{X}}(\mathbf{x}|\theta) \\
= \frac{d}{d\theta} \int_{\mathbf{x} \in \mathcal{X}} W(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}|\theta) \quad \text{(by assumption)} \\
= \frac{d}{d\theta} E[ W(\mathbf{X})] = \frac{d}{d\theta} \tau(\theta) = \tau'(\theta)
\]
Proving Cramer-Rao Theorem (4/4)

From the previous results

\[
\text{Var} \left[ \frac{\partial}{\partial \theta} \log f_X(X | \theta) \right] = E \left[ \left\{ \frac{\partial}{\partial \theta} \log f_X(X | \theta) \right\}^2 \right]
\]
Proving Cramer-Rao Theorem (4/4)

From the previous results

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

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

Therefore, Cramer-Rao lower bound is

\[
\text{Var}[W(X)] \geq \frac{\left[ \text{Cov}\{ W(X), \frac{\partial}{\partial \theta} \log f_X(X|\theta) \} \right]^2}{\text{Var} \left[ \frac{\partial}{\partial \theta} \log f_X(X|\theta) \right]}
\]
Proving Cramer-Rao Theorem (4/4)

From the previous results

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

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

Therefore, Cramer-Rao lower bound is

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

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

\[
\text{Var}[W(X)|\theta] \leq \frac{1}{n \mathbb{E}\left[\left(\frac{\partial}{\partial \theta} \log f_X(X_j|\theta)\right)^2\right]}
\]
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 $\text{Var}[W(X)|\theta]$ becomes

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

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

Proof

We need to show that

$$E \left[ \left\{ \frac{\partial}{\partial \theta} \log f_X(X|\theta) \right\}^2 \right] = nE \left[ \left\{ \frac{\partial}{\partial \theta} \log f_X(X|\theta) \right\}^2 \right]$$
Proving Corollary 7.3.10

\[ E \left[ \left\{ \frac{\partial}{\partial \theta} \log f_X(x|\theta) \right\}^2 \right] = E \left[ \left\{ \frac{\partial}{\partial \theta} \log \prod_{i=1}^{n} f_X(X_i|\theta) \right\}^2 \right] \]
Proving Corollary 7.3.10

\[
E \left[ \left\{ \frac{\partial}{\partial \theta} \log f_{X}(X|\theta) \right\}^2 \right] = E \left[ \left\{ \frac{\partial}{\partial \theta} \log \prod_{i=1}^{n} f_{X}(X_{i}|\theta) \right\}^2 \right] = E \left[ \left\{ \frac{\partial}{\partial \theta} \sum_{i=1}^{n} \log f_{X}(X_{i}|\theta) \right\}^2 \right]
\]
Proving Corollary 7.3.10

\[
E \left[ \left\{ \frac{\partial}{\partial \theta} \log f_{\mathbf{X}}(\mathbf{X} | \theta) \right\}^2 \right] = E \left[ \left\{ \frac{\partial}{\partial \theta} \log \prod_{i=1}^{n} f_{X_i}(X_i | \theta) \right\}^2 \right] = E \left[ \left\{ \frac{\partial}{\partial \theta} \sum_{i=1}^{n} \log f_{X_i}(X_i | \theta) \right\}^2 \right] = E \left[ \sum_{i=1}^{n} \left( \frac{\partial}{\partial \theta} \log f_{X_i}(X_i | \theta) \right)^2 \right]
\]
Proving Corollary 7.3.10

\[
E \left[ \left\{ \frac{\partial}{\partial \theta} \log f_X(X|\theta) \right\}^2 \right] = E \left[ \left\{ \frac{\partial}{\partial \theta} \log \prod_{i=1}^{n} f_X(X_i|\theta) \right\}^2 \right] = E \left[ \left\{ \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f_X(X_i|\theta) \right\}^2 \right] = E \left[ \sum_{i=1}^{n} \left\{ \frac{\partial}{\partial \theta} \log f_X(X_i|\theta) \right\}^2 + \sum_{i \neq j} \frac{\partial}{\partial \theta} \log f_X(X_i|\theta) \frac{\partial}{\partial \theta} \log f_X(X_j|\theta) \right] \]
Proving Corollary 7.3.10

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

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

$$= \sum_{i \neq j} E \left[ \frac{\partial}{\partial \theta} \log f_X(X_i|\theta) \right] E \left[ \frac{\partial}{\partial \theta} \log f_X(X_j|\theta) \right] = 0$$
Proving Corollary 7.3.10

Because \( X_1, \cdots, X_n \) are independent,

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

\[
E \left[ \left\{ \frac{\partial}{\partial \theta} \log f_X(X | \theta) \right\}^2 \right] = E \left[ \sum_{i=1}^{n} \left\{ \frac{\partial}{\partial \theta} \log f_X(X_i | \theta) \right\}^2 \right]
\]
Proving Corollary 7.3.10

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

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

\[
E \left[ \left\{ \frac{\partial}{\partial \theta} \log f_X(X|\theta) \right\}^2 \right] = E \left[ \sum_{i=1}^{n} \left\{ \frac{\partial}{\partial \theta} \log f_X(X_i|\theta) \right\}^2 \right] = \sum_{i=1}^{n} E \left[ \left\{ \frac{\partial}{\partial \theta} \log f_X(X_i|\theta) \right\}^2 \right]
\]
Proving Corollary 7.3.10

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

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

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

$$E \left[ \left\{ \frac{\partial}{\partial \theta} \log f_X(X|\theta) \right\}^2 \right] = E \left[ \sum_{i=1}^n \left\{ \frac{\partial}{\partial \theta} \log f_X(X_i|\theta) \right\}^2 \right]$$

$$= \sum_{i=1}^n E \left[ \left\{ \frac{\partial}{\partial \theta} \log f_X(X_i|\theta) \right\}^2 \right]$$

$$= nE \left[ \left\{ \frac{\partial}{\partial \theta} \log f_X(X_i|\theta) \right\}^2 \right]$$
Remark from Corollary 7.3.10

In iid case, Cramer-Rao lower bound for an unbiased estimator of $\theta$ is
Remark from Corollary 7.3.10

In iid case, Cramer-Rao lower bound for an unbiased estimator of $\theta$ is

$$\text{Var}[W(X)] \geq \frac{1}{nE\left[\left\{\frac{\partial}{\partial \theta} \log f_X(X|\theta)\right\}^2\right]}$$
Remark from Corollary 7.3.10

In iid case, Cramer-Rao lower bound for an unbiased estimator of \( \theta \) is

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

Because \( \tau(\theta) = \theta \) and \( \tau'(\theta) = 1 \).
Score Function

Definition: Score or Score Function for $X$

$$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} f_X(x|\theta)$$
Score Function

Definition: Score or Score Function for $X$

$$X_1, \ldots, X_n \overset{i.i.d.}{\sim} f_X(x|\theta)$$

$$S(X|\theta) = \frac{\partial}{\partial \theta} \log f_X(X|\theta)$$
Score Function

Definition: Score or Score Function for $X$

\begin{align*}
  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
\end{align*}
Definition: Score or Score Function for $X$

\[ X_1, \ldots, X_n \overset{i.i.d.}{\sim} f_X(x|\theta) \]

\[
S(X|\theta) = \frac{\partial}{\partial \theta} \log f_X(X|\theta)
\]

\[
E[S(X|\theta)] = 0
\]

\[
S_n(X|\theta) = \frac{\partial}{\partial \theta} \log f_X(X|\theta)
\]
Definition: Fisher Information Number

\[
I(\theta) = E \left[ \left\{ \frac{\partial}{\partial \theta} \log f_X(X|\theta) \right\}^2 \right] = E \left[ S^2(X|\theta) \right]
\]
Fisher Information Number

**Definition: Fisher Information Number**

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

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

The bigger the information number, the more information we have about \( \theta \), the smaller bound on the variance of unbiased estimates.
Fisher Information Number

**Definition: Fisher Information Number**

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

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

The bigger the information number, the more information we have about \( \theta \), the smaller bound on the variance of unbiased estimates.
Fisher Information Number

Definition: Fisher Information Number

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

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

The bigger the information number, the more information we have about \( \theta \), the smaller bound on the variance of unbiased estimates.
Lemma 7.3.11

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

\[
\frac{d}{d\theta} \int_{x \in X} f_X(x|\theta) \, dx \quad = \quad \int_{x \in X} \frac{\partial}{\partial \theta} f_X(x|\theta) \, dx
\]
Simplified Fisher Information

Lemma 7.3.11

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

$$\frac{d}{d\theta} \int_{x \in X} f_X(x|\theta) \, dx \quad = \quad \int_{x \in X} \frac{\partial}{\partial \theta} f_X(x|\theta) \, dx$$

$$\frac{d}{d\theta} \int_{x \in X} \frac{\partial}{\partial \theta} f_X(x|\theta) \, dx \quad = \quad \int_{x \in X} \frac{\partial^2}{\partial \theta^2} f_X(x|\theta) \, dx$$
Lemma 7.3.11

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

\[
\frac{d}{d\theta} \int_{x \in X} f_X(x|\theta) \, dx = \int_{x \in X} \frac{\partial}{\partial \theta} f_X(x|\theta) \, dx
\]

\[
\frac{d}{d\theta} \int_{x \in X} \frac{\partial}{\partial \theta} f_X(x|\theta) \, dx = \int_{x \in X} \frac{\partial^2}{\partial \theta^2} f_X(x|\theta) \, dx
\]

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

- $X_1, \cdots, X_n \stackrel{i.i.d.}{\sim} \text{Poisson}(\lambda)$
- $\lambda_1 = \bar{X}$
- $\lambda_2 = s^2_X$
- $E[\lambda_1] = E(\bar{X}) = \lambda.$
Example - Poisson Distribution

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

Cramer-Rao lower bound is $I^{-1}_n(\lambda) = [nI(\lambda)]^{-1}.$
Example - Poisson Distribution

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

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

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

- \( X_1, \cdots, X_n \) i.i.d. Poisson(\( \lambda \))
- \( \lambda_1 = \bar{X} \)
- \( \lambda_2 = s_X^2 \)
- \( E[\lambda_1] = E(\bar{X}) = \lambda. \)

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

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

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

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

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

\[
= -E \left[ \frac{\partial^2}{\partial \lambda^2} \log \frac{e^{-\lambda} \lambda^X}{X!} \right] = -E \left[ \frac{\partial^2}{\partial \lambda^2} (-\lambda + X \log \lambda - \log X!) \right]
\]

\[
= E \left[ \frac{X}{\lambda^2} \right] = \frac{1}{\lambda^2} E(X) = \frac{1}{\lambda}
\]
Example - Poisson Distribution (cont’d)

Therefore, the Cramer-Rao lower bound is

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

where $W$ is any unbiased estimator.
Example - Poisson Distribution (cont’d)

Therefore, the Cramer-Rao lower bound is

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

where \( W \) is any unbiased estimator.

\[ \text{Var}(\hat{\lambda}_1) = \text{Var}(\bar{X}) = \frac{\text{Var}(X)}{n} = \frac{\lambda}{n} \]
Example - Poisson Distribution (cont’d)

Therefore, the Cramer-Rao lower bound is

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

where $W$ is any unbiased estimator.

$$\text{Var}(\hat{\lambda}_1) = \text{Var}(\bar{X}) = \frac{\text{Var}(X)}{n} = \frac{\lambda}{n}$$

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

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

(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**

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

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

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

\[ = E \left[ \left( -1 + \frac{X}{\lambda} \right)^2 \right] = E \left[ 1 - 2 \frac{X}{\lambda} + \frac{X^2}{\lambda^2} \right] = 1 - 2 \frac{E(X)}{\lambda} + \frac{E(X^2)}{\lambda^2} \]
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**

\[
\begin{align*}
I(\lambda) & = E \left[ \left\{ \frac{\partial}{\partial \lambda} \log f_X(X|\lambda) \right\}^2 \right] = E \left[ \left\{ \frac{\partial}{\partial \lambda} (-\lambda + X \log \lambda - \log X!) \right\}^2 \right] \\
& = E \left[ \left\{ -1 + \frac{X}{\lambda} \right\}^2 \right] = E \left[ 1 - 2 \frac{X}{\lambda} + \frac{X^2}{\lambda^2} \right] = 1 - 2 \frac{E(X)}{\lambda} + \frac{E(X^2)}{\lambda^2} \\
& = 1 - 2 \frac{E(X)}{\lambda} + \frac{\text{Var}(X) + [E(X)]^2}{\lambda^2} = 1 - 2 \frac{\lambda}{\lambda} + \frac{\lambda + \lambda^2}{\lambda^2} = \frac{1}{\lambda}
\end{align*}
\]
Example - Normal Distribution

- $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, where $\sigma^2$ is known.
Example - Normal Distribution

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

\[ I(\mu) = \mathbb{E} \left[ \frac{\partial^2}{\partial \mu^2} \log f_{X_j}(X_j) \right] = \mathbb{E} \left[ \frac{\partial^2}{\partial \mu^2} \left\{ \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left( -\frac{(X_j - \mu)^2}{2\sigma^2} \right) \right\} \right] = \mathbb{E} \left[ \frac{1}{\sqrt{2\pi \sigma^2}} \frac{1}{2\sigma^2} \right] = \frac{1}{2\sigma^2}. \]
Example - Normal Distribution

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

\[
I(\mu) = -E \left[ \frac{\partial^2}{\partial \mu^2} \log f_X(X|\mu) \right]
\]
Example - Normal Distribution

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

\[
I(\mu) = -E \left[ \frac{\partial^2}{\partial \mu^2} \log f_X(X|\mu) \right] \\
= -E \left[ \frac{\partial^2}{\partial \mu^2} \log \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(X - \mu)^2}{2\sigma^2} \right) \right\} \right]
\]
Example - Normal Distribution

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

\[
I(\mu) = -E \left[ \frac{\partial^2}{\partial \mu^2} \log f_X(X|\mu) \right]
\]
\[
= -E \left[ \frac{\partial^2}{\partial \mu^2} \log \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(X - \mu)^2}{2\sigma^2} \right) \right\} \right]
\]
\[
= -E \left[ \frac{\partial^2}{\partial \mu^2} \left\{ -\frac{1}{2} \log(2\pi\sigma^2) - \frac{(X - \mu)^2}{2\sigma^2} \right\} \right]
\]
Example - Normal Distribution

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

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

\[
= -E \left[ \frac{\partial^2}{\partial \mu^2} \log \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(X-\mu)^2}{2\sigma^2} \right) \right\} \right]
\]

\[
= -E \left[ \frac{\partial}{\partial \mu} \left\{ -\frac{1}{2} \log(2\pi\sigma^2) - \frac{(X-\mu)^2}{2\sigma^2} \right\} \right]
\]

\[
= -E \left[ \frac{\partial}{\partial \mu} \left\{ \frac{2(X-\mu)}{2\sigma^2} \right\} \right] = \frac{1}{\sigma^2}
\]
Applying Lemma 7.3.11

Question

When can we interchange the order of differentiation and integration?

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

Example

\[ X_1; X_2; \ldots; X_n \text{i.i.d.} \text{Uniform}(0; \theta) \]

\[
\int_0^\theta h(x) f_{X_j}(x) \, dx \neq \int_0^\theta h(x) \frac{d}{dx} f_{X_j}(x) \, dx
\]
Applying Lemma 7.3.11

Question
When can we interchange the order of differentiation and integration?

Answer
- For exponential family, always yes.
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.
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
\[ X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Uniform}(0, \theta) \]
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

\[ X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Uniform}(0, \theta) \]

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

Today

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

Today

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

Next Lecture

- More on Cramer-Rao inequality