Likelihood Function

**Definition**

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

The join distribution of \( \mathbf{X} = (X_1, \cdots, X_n) \) is

\[
f_{\mathbf{X}}(\mathbf{x}|\theta) = \prod_{i=1}^{n} f_{X}(x_i|\theta)
\]

Given that \( \mathbf{X} = \mathbf{x} \) is observed, the function of \( \theta \) defined by \( L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta) \) is called the likelihood function.
Examples of Likelihood Function - 1/3

- \( X_1, X_2, X_3, X_4 \) \( \sim \) Bernoulli\((p)\), \( 0 < p < 1 \).
Examples of Likelihood Function - 1/3

- $X_1, X_2, X_3, X_4 \overset{i.i.d.}\sim \text{Bernoulli}(p)$, $0 < p < 1$.
- $\mathbf{x} = (1, 1, 1, 1)^T$
Examples of Likelihood Function - 1/3

- $X_1, X_2, X_3, X_4 \overset{i.i.d.}{\sim} \text{Bernoulli}(p), \ 0 < p < 1.$
- $x = (1, 1, 1, 1)^T$
- Intuitively, it is more likely that $p$ is larger than smaller.
Examples of Likelihood Function - 1/3

- \( X_1, X_2, X_3, X_4 \) i.i.d. \( \text{Bernoulli}(p) \), \( 0 < p < 1 \).
- \( \mathbf{x} = (1, 1, 1, 1)^T \)
- Intuitively, it is more likely that \( p \) is larger than smaller.
- \( L(p|\mathbf{x}) = f(\mathbf{x}|p) = \prod_{i=1}^{4} p^{x_i} (1 - p)^{1-x_i} = p^4 \).
Examples of Likelihood Function - 1/3

- $X_1, X_2, X_3, X_4 \overset{i.i.d.}{\sim} \text{Bernoulli}(p), \ 0 < p < 1$.
- $x = (1, 1, 1, 1)^T$
- Intuitively, it is more likely that $p$ is larger than smaller.
- $L(p|x) = f(x|p) = \prod_{i=1}^{4} p^{x_i} (1 - p)^{1-x_1} = p^4$. 

![Graph showing likelihood function](image)
Examples of Likelihood Function - 2/3

- $X_1, X_2, X_3, X_4 \sim \text{Bernoulli}(p), \ 0 < p < 1.$
Examples of Likelihood Function - 2/3

- \( X_1, X_2, X_3, X_4 \overset{i.i.d.}{\sim} \text{Bernoulli}(p), \, 0 < p < 1. \)
- \( x = (0, 0, 0, 0)^T \)
Examples of Likelihood Function - 2/3

- $X_1, X_2, X_3, X_4 \overset{i.i.d.}{\sim} \text{Bernoulli}(p), \; 0 < p < 1.$
- $x = (0, 0, 0, 0)^T$
- Intuitively, it is more likely that $p$ is smaller than larger.
Examples of Likelihood Function - 2/3

- $X_1, X_2, X_3, X_4 \overset{i.i.d.}{\sim} \text{Bernoulli}(p), \; 0 < p < 1.$
- $x = (0, 0, 0, 0)^T$
- Intuitively, it is more likely that $p$ is smaller than larger.
- $L(p|x) = f(x|p) = \prod_{i=1}^{4} p^{x_i} (1 - p)^{1-x_1} = (1 - p)^4.$
Examples of Likelihood Function - 2/3

- $X_1, X_2, X_3, X_4 \overset{i.i.d.}{\sim} \text{Bernoulli}(p), \ 0 < p < 1.$
- $\mathbf{x} = (0, 0, 0, 0)^T$
- Intuitively, it is more likely that $p$ is smaller than larger.
- $L(p|\mathbf{x}) = f(\mathbf{x}|p) = \prod_{i=1}^{4} p^{x_i}(1 - p)^{1-x_i} = (1 - p)^4.$
Examples of Likelihood Function - 3/3

- $X_1, X_2, X_3, X_4 \overset{i.i.d.}{\sim} \text{Bernoulli}(p), \ 0 < p < 1.$
Examples of Likelihood Function - 3/3

- $X_1, X_2, X_3, X_4 \overset{i.i.d.}{\sim} \text{Bernoulli}(p), \ 0 < p < 1.$
- $\mathbf{x} = (1, 1, 0, 0)^T$

Intuitively, it is more likely that $p$ is somewhere in the middle than in the extremes.

$$L(p|\mathbf{x}) = f(\mathbf{x}|p) = \prod_{i=1}^{4} p^{x_i} (1-p)^{1-x_i}.$$
Examples of Likelihood Function - 3/3

- $X_1, X_2, X_3, X_4 \overset{i.i.d.}{\sim} \text{Bernoulli}(p)$, $0 < p < 1$.
- $x = (1, 1, 0, 0)^T$
- Intuitively, it is more likely that $p$ is somewhere in the middle than in the extremes.
Examples of Likelihood Function - 3/3

- $X_1, X_2, X_3, X_4 \overset{i.i.d.}{\sim} \text{Bernoulli}(p), \ 0 < p < 1.$
- $\mathbf{x} = (1, 1, 0, 0)^T$
- Intuitively, it is more likely that $p$ is somewhere in the middle than in the extremes.
- $L(p|\mathbf{x}) = f(\mathbf{x}|p) = \prod_{i=1}^{4} p^{x_i} (1 - p)^{1-x_1} = p^2(1 - p)^2.$
Examples of Likelihood Function - 3/3

- \( X_1, X_2, X_3, X_4 \) i.i.d. Bernoulli\((p)\), \( 0 < p < 1 \).
- \( x = (1, 1, 0, 0)^T \)
- Intuitively, it is more likely that \( p \) is somewhere in the middle than in the extremes.
- \( L(p|x) = f(x|p) = \prod_{i=1}^{4} p^{x_i} (1 - p)^{1-x_i} = p^2 (1 - p)^2. \)
Point Estimation: Ingredients

- **Data**: $\mathbf{x} = (x_1, \cdots, x_n)$ - realizations of random variables $(X_1, \cdots, X_n)$. 
Point Estimation: Ingredients

- **Data:** \( \mathbf{x} = (x_1, \cdots, x_n) \) - realizations of random variables \((X_1, \cdots, X_n)\).

- \( X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} f_X(x|\theta) \).
Point Estimation: Ingredients

- **Data:** \( \mathbf{x} = (x_1, \cdots, x_n) \) - realizations of random variables \((X_1, \cdots, X_n)\).
- \( X_1, \cdots, X_n \) i.i.d. \( f_X(x|\theta) \).
- Assume a model \( \mathcal{P} = \{f_X(x|\theta) : \theta \in \Omega \subset \mathbb{R}^p\} \) where the functional form of \( f_X(x|\theta) \) is known, but \( \theta \) is unknown.
Point Estimation: Ingredients

- **Data**: $\mathbf{x} = (x_1, \cdots, x_n)$ - realizations of random variables $(X_1, \cdots, X_n)$.
- $X_1, \cdots, X_n \overset{i.i.d.}{\sim} f_X(x|\theta)$.
- Assume a model $\mathcal{P} = \{f_X(x|\theta) : \theta \in \Omega \subset \mathbb{R}^p\}$ where the functional form of $f_X(x|\theta)$ is known, but $\theta$ is unknown.
- Task is to use data $\mathbf{x}$ to make inference on $\theta$
Definition

If we use a function of sample \( w(X_1, \cdots, X_n) \) as a "guess" of \( \tau(\theta) \), where \( \tau(\theta) \) is a function of true parameter \( \theta \).
Point Estimation

Definition

If we use a function of sample $w(X_1, \cdots, X_n)$ as a "guess" of $\tau(\theta)$, where $\tau(\theta)$ is a function of true parameter $\theta$. Then $w(X) = w(X_1, \cdots, X_n)$ is called a point estimator of $\tau(\theta)$.
Point Estimation

Definition

If we use a function of sample \( w(X_1, \cdots, X_n) \) as a "guess" of \( \tau(\theta) \), where \( \tau(\theta) \) is a function of true parameter \( \theta \). Then \( w(X) = w(X_1, \cdots, X_n) \) is called a point estimator of \( \tau(\theta) \). The realization of the estimation, \( w(x) = w(x_1, \cdots, x_n) \) is called the estimate of \( \tau(\theta) \).

Example

- \( X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, 1) \), where \( \theta \in \Omega \in \mathbb{R} \).
Point Estimation

Definition

If we use a function of sample $w(X_1, \cdots, X_n)$ as a "guess" of $\tau(\theta)$, where $\tau(\theta)$ is a function of true parameter $\theta$. Then $w(X) = w(X_1, \cdots, X_n)$ is called a point estimator of $\tau(\theta)$. The realization of the estimation, $w(x) = w(x_1, \cdots, x_n)$ is called the estimate of $\tau(\theta)$.

Example

- $X_1, \cdots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$, where $\theta \in \Omega \in \mathbb{R}$.
- Suppose $n = 6$, and $(x_1, \cdots, x_6) = (2.0, 2.1, 2.9, 2.6, 1.2, 1.8)$.
Point Estimation

**Definition**

If we use a function of sample $w(X_1, \cdots, X_n)$ as a "guess" of $\tau(\theta)$, where $\tau(\theta)$ is a function of true parameter $\theta$. Then $w(X) = w(X_1, \cdots, X_n)$ is called a point estimator of $\tau(\theta)$. The realization of the estimation, $w(x) = w(x_1, \cdots, x_n)$ is called the estimate of $\tau(\theta)$.

**Example**

- $X_1, \cdots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$, where $\theta \in \Omega \in \mathbb{R}$.
- Suppose $n = 6$, and $(x_1, \cdots, x_6) = (2.0, 2.1, 2.9, 2.6, 1.2, 1.8)$.
- Define $w_1(X_1, \cdots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X} = 2.1$. 
Point Estimation

Definition

If we use a function of sample $w(X_1, \cdots, X_n)$ as a "guess" of $\tau(\theta)$, where $\tau(\theta)$ is a function of true parameter $\theta$. Then $w(X) = w(X_1, \cdots, X_n)$ is called a point estimator of $\tau(\theta)$. The realization of the estimation, $w(x) = w(x_1, \cdots, x_n)$ is called the estimate of $\tau(\theta)$.

Example

- $X_1, \cdots, X_n \overset{	ext{i.i.d.}}{\sim} \mathcal{N}(\theta, 1)$, where $\theta \in \Omega \subset \mathbb{R}$.
- Suppose $n = 6$, and $(x_1, \cdots, x_6) = (2.0, 2.1, 2.9, 2.6, 1.2, 1.8)$.
- Define $w_1(X_1, \cdots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X} = 2.1$.
- Define $w_2(X_1, \cdots, X_n) = X_{(1)} = 1.2.$
Method of Moments

A method to equate sample moments to population moments and solve equations.
Method of Moments

A method to equate sample moments to population moments and solve equations.

<table>
<thead>
<tr>
<th>Sample moments</th>
<th>Population moments</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_1 = \frac{1}{n} \sum_{i=1}^{n} X_i$</td>
<td>$\mu'_1 = E[X</td>
</tr>
<tr>
<td>$m_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2$</td>
<td>$\mu'_2 = E[X</td>
</tr>
<tr>
<td>$m_3 = \frac{1}{n} \sum_{i=1}^{n} X_i^3$</td>
<td>$\mu'_3 = E[X</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Method of Moments

A method to equate sample moments to population moments and solve equations.

<table>
<thead>
<tr>
<th>Sample moments</th>
<th>Population moments</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_1 = \frac{1}{n} \sum_{i=1}^{n} X_i$</td>
<td>$\mu'_1 = E[X</td>
</tr>
<tr>
<td>$m_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2$</td>
<td>$\mu'_2 = E[X</td>
</tr>
<tr>
<td>$m_3 = \frac{1}{n} \sum_{i=1}^{n} X_i^3$</td>
<td>$\mu'_3 = E[X</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Point estimator of $T(\theta)$ is obtained by solving equations like this.
Method of Moments

A method to equate sample moments to population moments and solve equations.

<table>
<thead>
<tr>
<th>Sample moments</th>
<th>Population moments</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_1 = \frac{1}{n} \sum_{i=1}^{n} X_i$</td>
<td>$\mu_1' = E[X</td>
</tr>
<tr>
<td>$m_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2$</td>
<td>$\mu_2' = E[X</td>
</tr>
<tr>
<td>$m_3 = \frac{1}{n} \sum_{i=1}^{n} X_i^3$</td>
<td>$\mu_3' = E[X</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Point estimator of $T(\theta)$ is obtained by solving equations like this.

\[
\begin{align*}
m_1 &= \mu_1'(\theta) \\
m_2 &= \mu_2'(\theta) \\
\vdots &= \vdots \\
m_k &= \mu_k'(\theta)
\end{align*}
\]
Examples of method of moments estimator

Problem

\( X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2) \). Find estimator for \( \mu, \sigma^2 \).
Examples of method of moments estimator

Problem

\( X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2) \). Find estimator for \( \mu, \sigma^2 \).

Solution

\[
\mu_1' = E X = \mu = \bar{X}
\]
Examples of method of moments estimator

Problem

\[ X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2). \] Find estimator for \( \mu, \sigma^2 \).

Solution

\[
\begin{align*}
\mu'_1 &= E[X] = \mu = \bar{X} \\
\mu'_2 &= E[X^2] = [E[X]]^2 + \text{Var}(X) = \mu^2 + \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2
\end{align*}
\]
Examples of method of moments estimator

**Problem**

\( X_1, \ldots, X_n \overset{	ext{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2). \) Find estimator for \( \mu, \sigma^2. \)

**Solution**

\[
\begin{align*}
\mu_1' &= EX = \mu = \bar{X} \\
\mu_2' &= EX^2 = [EX]^2 + \text{Var}(X) = \mu^2 + \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 \\
& \quad \left\{ \hat{\mu} = \bar{X} \right\}
\end{align*}
\]
Examples of method of moments estimator

Problem

\(X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)\). Find estimator for \(\mu, \sigma^2\).

Solution

\[
\begin{align*}
\mu_1' &= E[X] = \mu = \bar{X} \\
\mu_2' &= E[X^2] = [E[X]]^2 + \text{Var}(X) = \mu^2 + \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2
\end{align*}
\]

\[
\begin{cases}
\hat{\mu} = \bar{X} \\
\hat{\mu}^2 + \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2
\end{cases}
\]
Examples of method of moments estimator

Problem

\[ X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2). \] Find estimator for \( \mu, \sigma^2 \).

Solution

\[ \mu_1' = E \mathbf{X} = \mu = \overline{X} \]

\[ \mu_2' = E \mathbf{X}^2 = (E \mathbf{X})^2 + \text{Var}(\mathbf{X}) = \mu^2 + \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 \]

\[
\begin{cases}
\hat{\mu} = \overline{X} \\
\hat{\mu}^2 + \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2
\end{cases}
\]

Solving the two equations above, \( \hat{\mu} = \overline{X}, \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n} \).
Method of moments estimator - Binomial

Problem

\[ X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Binomial}(k, p). \text{ Find an estimator for } k, p. \]
Method of moments estimator - Binomial

Problem

\( X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Binomial}(k, p) \). Find an estimator for \( k, p \).

Solution

\[
f_X(x | k, p) = \binom{k}{x} p^x (1 - p)^{k-x} \quad x \in \{0, 1, \cdots, k\}
\]
Method of moments estimator - Binomial

Problem

\( X_1, \cdots, X_n \overset{i.i.d.}{\sim} \text{Binomial}(k, p) \). Find an estimator for \( k, p \).

Solution

\[
f_X(x|k, p) = \binom{k}{x} p^x (1 - p)^{k-x} \quad x \in \{0, 1, \cdots, k\}
\]

Equating first two sample moments,

\[
\frac{1}{n} \sum_{i=1}^{n} X_i = \bar{x} = \mu_1' = E X = k p
\]
Method of moments estimator - Binomial

**Problem**

\(X_1, \cdots, X_n \sim \text{i.i.d. Binomial}(k, p)\). Find an estimator for \(k, p\).

**Solution**

\[
f_X(x|k, p) = \binom{k}{x} p^x (1 - p)^{k-x} \quad x \in \{0, 1, \cdots, k\}
\]

Equating first two sample moments,

\[
\frac{1}{n} \sum_{i=1}^{n} X_i = \bar{x} = \mu_1' = E[X] = kp
\]

\[
\frac{1}{n} \sum_{i=1}^{n} X_i^2 = \mu_2' = E[X^2] = (E[X])^2 + \text{Var}(X) = k^2 p^2 + kp(1 - p)
\]
The method of moments estimators are

\[ \hat{k} = \frac{\overline{X}^2}{\overline{X} - \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2} \]
The method of moments estimators are

\[ \hat{k} = \frac{\overline{X}^2}{\overline{X} - \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2} \]

\[ \hat{p} = \frac{\overline{X}}{\hat{k}} \]
The method of moments estimators are

\[ \hat{k} = \frac{\bar{X}^2}{\bar{X} - \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2} \]

\[ \hat{p} = \frac{\bar{X}}{\hat{k}} \]

These are not the best estimators. It is possible to get negative estimates of \( k \) and \( p \).
Examples of MoM estimator - Negative Binomial

Problem

\[ X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Negative Binomial}(r, p). \text{ Find estimator for } (r, p). \]
Examples of MoM estimator - Negative Binomial

Problem

\[ X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Negative Binomial}(r, p). \text{ Find estimator for } (r, p). \]

Solution

\[
m_1 = \frac{1}{n} \sum_{i=1}^{n} X_i = E X = \frac{r(1 - p)}{p}
\]
Examples of MoM estimator - Negative Binomial

Problem

\( X_1, \ldots, X_n \sim \text{i.i.d. Negative Binomial}(r, p) \). Find estimator for \((r, p)\).

Solution

\[
\begin{align*}
  m_1 & = \frac{1}{n} \sum_{i=1}^{n} X_i = E(X) = \frac{r(1 - p)}{p} \\
  m_2 & = \frac{1}{n} \sum_{i=1}^{n} X_i^2 = E(X^2) = \left( \frac{r(1 - p)}{p} \right)^2 + \frac{r(1 - p)}{p^2}
\end{align*}
\]
Examples of MoM estimator - Negative Binomial

Problem

\( X_1, \cdots, X_n \overset{i.i.d.}{\sim} \text{Negative Binomial}(r, p). \) Find estimator for \((r, p)\).

Solution

\[
\begin{align*}
m_1 &= \frac{1}{n} \sum_{i=1}^{n} X_i = E(X) = \frac{r(1 - p)}{p} \\
m_2 &= \frac{1}{n} \sum_{i=1}^{n} X_i^2 = E(X^2) = \left( \frac{r(1 - p)}{p} \right)^2 + \frac{r(1 - p)}{p^2} \\
\hat{p} &= \frac{m_1}{m_2 - m_1} = \frac{\bar{X}}{\frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}^2}
\end{align*}
\]
Examples of MoM estimator - Negative Binomial

**Problem**

\( X_1, \ldots, X_n \text{i.i.d.} \sim \text{Negative Binomial}(r, p) \). Find estimator for \((r, p)\).

**Solution**

\[
\begin{align*}
m_1 &= \frac{1}{n} \sum_{i=1}^{n} X_i = E(X) = \frac{r(1 - p)}{p} \\
m_2 &= \frac{1}{n} \sum_{i=1}^{n} X_i^2 = E(X^2) = \left( \frac{r(1 - p)}{p} \right)^2 + \frac{r(1 - p)}{p^2} \\
\hat{p} &= \frac{m_1}{m_2 - m_1^2} = \frac{\bar{X}}{1 - \bar{X}} \\
\hat{r} &= \frac{m_1 \hat{p}}{1 - \hat{p}} = \frac{\bar{X} \hat{p}}{1 - \hat{p}}
\end{align*}
\]
Satterthwaite Approximation

Problem

Let $Y_1, \cdots, Y_k$ are independently (but not identically) distributed random variables from $\chi^2_{r_1}, \cdots, \chi^2_{r_k}$, respectively. We know that the distribution $\sum_{i=1}^{n} Y_i$ is also chi-squared with degrees of freedom equal to $\sum_{i=1}^{k} r_i$. 
Satterthwaite Approximation

Problem

Let $Y_1, \ldots, Y_k$ are independently (but not identically) distributed random variables from $\chi^2_{r_1}, \ldots, \chi^2_{r_k}$, respectively. We know that the distribution $\sum_{i=1}^{n} Y_i$ is also chi-squared with degrees of freedom equal to $\sum_{i=1}^{k} r_i$.

However, the distribution of $\sum_{i=1}^{k} a_i Y_i$, where $a_i$'s are known constants with $\sum_{i=1}^{n} a_i r_i = 1$, in general, the distribution is hard to obtain.
Satterthwaite Approximation

Problem

Let $Y_1, \cdots, Y_k$ are independently (but not identically) distributed random variables from $\chi^2_{r_1}, \cdots, \chi^2_{r_k}$, respectively. We know that the distribution $\sum_{i=1}^{n} Y_i$ is also chi-squared with degrees of freedom equal to $\sum_{i=1}^{k} r_i$.

However, the distribution of $\sum_{i=1}^{k} a_i Y_i$, where $a_i$s are known constants with $\sum_{i=1}^{n} a_i r_i = 1$, in general, the distribution is hard to obtain.

It is often reasonable to assume that the distribution of $\sum_{i=1}^{k} a_i Y_i$ follows $\frac{1}{\nu} \chi^2_{\nu}$ approximately. Find a moment-based estimator of $\nu$. 
A Naive Solution

To match the first moment, let $X \sim \chi^2_{\nu}/\nu$. Then $E(X) = 1$, and $\text{Var}(X) = 2/\nu$. 

Note that $\nu$ can be negative, which is not desirable.
A Naive Solution

To match the first moment, let $X \sim \chi^2_{\nu}/\nu$. Then $E(X) = 1$, and $\text{Var}(X) = 2/\nu$.

$$E \left( \sum_{i=1}^{k} a_i Y_i \right) = \sum_{i=1}^{k} a_i EY_i = \sum_{i=1}^{k} a_i r_i = 1 = E(X)$$
A Naive Solution

To match the first moment, let $X \sim \chi^2_{\nu}/\nu$. Then $E(X) = 1$, and $\text{Var}(X) = 2/\nu$.

$$E \left( \sum_{i=1}^{k} a_i Y_i \right) = \sum_{i=1}^{k} a_i E(Y_i) = \sum_{i=1}^{k} a_i r_i = 1 = E(X)$$

To match the second moments,

$$E \left( \left( \sum_{i=1}^{k} a_i Y_i \right)^2 \right) = E(X^2) = \frac{2}{\nu} + 1$$
A Naive Solution

To match the first moment, let $X \sim \chi^2_{\nu}/\nu$. Then $E(X) = 1$, and $\text{Var}(X) = 2/\nu$.

$$E\left(\sum_{i=1}^{k} a_i Y_i\right) = \sum_{i=1}^{k} a_i E(Y_i) = \sum_{i=1}^{k} a_i r_i = 1 = E(X)$$

To match the second moments,

$$E\left(\sum_{i=1}^{k} a_i Y_i\right)^2 = E(X^2) = \frac{2}{\nu} + 1$$

Therefore, the method of moment estimator of $\nu$ is

$$\hat{\nu} = \frac{2}{\left(\sum_{i=1}^{k} a_i Y_i\right)^2 - 1}$$

Note that $\nu$ can be negative, which is not desirable.
An alternative Solution

To match the second moments,

\[
E\left(\sum_{i=1}^{k} a_i Y_i\right)^2 = \text{Var}\left(\sum_{i=1}^{k} a_i Y_i\right) + \left[ E\left(\sum_{i=1}^{k} a_i Y_i\right) \right]^2
\]
An alternative Solution

To match the second moments,

\[
E \left( \sum_{i=1}^{k} a_i Y_i \right)^2 = \text{Var} \left( \sum_{i=1}^{k} a_i Y_i \right) + \left[ E \left( \sum_{i=1}^{k} a_i Y_i \right) \right]^2
\]

\[
= \left[ E \left( \sum_{i=1}^{k} a_i Y_i \right) \right]^2 \left[ \frac{\text{Var} \left( \sum_{i=1}^{k} a_i Y_i \right)}{E \left( \sum_{i=1}^{k} a_i Y_i \right)^2} + 1 \right]
\]
An alternative Solution

To match the second moments,

\[
E\left(\sum_{i=1}^{k} a_i Y_i\right)^2 = \text{Var}\left(\sum_{i=1}^{k} a_i Y_i\right) + \left[E\left(\sum_{i=1}^{k} a_i Y_i\right)\right]^2
\]

\[
= \left[E\left(\sum_{i=1}^{k} a_i Y_i\right)\right]^2 \left[\frac{\text{Var}\left(\sum_{i=1}^{k} a_i Y_i\right)}{\left[E\left(\sum_{i=1}^{k} a_i Y_i\right)\right]^2} + 1\right]
\]

\[
= \left[\frac{\text{Var}\left(\sum_{i=1}^{k} a_i Y_i\right)}{\left[E\left(\sum_{i=1}^{k} a_i Y_i\right)\right]^2} + 1\right] = \frac{2}{\nu} + 1
\]
An alternative Solution

To match the second moments,

\[
E \left( \sum_{i=1}^{k} a_i Y_i \right)^2 = \text{Var} \left( \sum_{i=1}^{k} a_i Y_i \right) + \left[ E(\sum_{i=1}^{k} a_i Y_i) \right]^2 
\]

\[
= \left[ E(\sum_{i=1}^{k} a_i Y_i) \right]^2 \left[ \frac{\text{Var}(\sum_{i=1}^{k} a_i Y_i)}{E(\sum_{i=1}^{k} a_i Y_i)^2} \right] + 1
\]

\[
= \left[ \frac{\text{Var}(\sum_{i=1}^{k} a_i Y_i)}{E(\sum_{i=1}^{k} a_i Y_i)^2} \right]^2 + 1 = \frac{2}{\nu} + 1
\]

\[
\nu = \frac{2 \left[ E(\sum_{i=1}^{k} a_i Y_i) \right]^2}{\text{Var}(\sum_{i=1}^{k} a_i Y_i)}
\]
Alternative Solution (cont’d)

To match the second moments, Finally, use the fact that $Y_1, \cdots, Y_k$ are independent chi-squared random variables.
Alternative Solution (cont’d)

To match the second moments, finally, use the fact that $Y_1, \cdots, Y_k$ are independent chi-squared random variables.

\[
\text{Var}\left(\sum_{i=1}^{n} a_i Y_i\right) = \sum_{i=1}^{k} a_i \text{Var}(Y_i)
\]
Alternative Solution (cont’d)

To match the second moments, finally, use the fact that $Y_1, \cdots, Y_k$ are independent chi-squared random variables.

\[
\text{Var}\left( \sum_{i=1}^{n} a_i Y_i \right) = \sum_{i=1}^{k} a_i \text{Var}(Y_i) \\
= 2 \sum_{i=1}^{n} \frac{a_i^2 (EY_i)^2}{r_i}
\]
Alternative Solution (cont’d)

To match the second moments, finally, use the fact that $Y_1, \cdots, Y_k$ are independent chi-squared random variables.

$$\text{Var}(\sum_{i=1}^{n} a_i Y_i) = \sum_{i=1}^{k} a_i \text{Var}(Y_i)$$

$$= 2 \sum_{i=1}^{n} \frac{a_i^2 (EY_i)^2}{r_i}$$

Substituting this expression for the variance and removing expectations, we obtain Satterthwaite’s estimator.
Alternative Solution (cont’d)

To match the second moments, Finally, use the fact that $Y_1, \cdots, Y_k$ are independent chi-squared random variables.

\[
\text{Var}\left(\sum_{i=1}^{n} a_i Y_i \right) = \sum_{i=1}^{k} a_i \text{Var}(Y_i)
\]

\[
= 2 \sum_{i=1}^{n} \frac{a_i^2 (EY_i)^2}{r_i}
\]

Substituting this expression for the variance and removing expectations, we obtain Satterthwaite’s estimator

\[
\hat{v} = \frac{\sum_{i=1}^{n} a_i Y_i}{\sum_{i=1}^{n} \frac{a_i^2}{r_i} Y_i^2}
\]
Definition

- For a given sample point \( x = (x_1, \cdots, x_n) \),
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

\[
L(\hat{\theta}(\mathbf{x})) = \max_{\theta \in \Theta} L(\theta | \mathbf{x})
\]

\( \hat{\theta}(\mathbf{x}) \) 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 \( \theta \).
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.
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 \( \hat{\theta}(\mathbf{x}) \in \Omega \).
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 $\theta$ based on data $\mathbf{x}$,
Maximum Likelihood Estimator

Definition

- For a given sample point \( \mathbf{x} \equiv (x_1, \cdots, x_n) \),
- 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 \textit{maximum likelihood estimate} of \( \theta \) based on data \( \mathbf{x} \),
- and \( \hat{\theta}(\mathbf{X}) \) is the \textit{maximum likelihood estimator (MLE)} of \( \theta \).
Example of MLE - Exponential Distribution

Problem

Let $X_1, \cdots, X_n \overset{i.i.d.}{\sim} \text{Exponential}(\beta)$. Find MLE of $\beta$. 

\[
L(j; x) = f_{\text{X}}(x_j) = n \prod_{i=1}^{n} f_{\text{X}}(x_i) = \frac{1}{n} \exp\left(-\frac{\sum_{i=1}^{n} x_i}{\beta}\right)
\]

where $\beta > 0$. 

Hyun Min Kang

Biostatistics 602 - Lecture 09

February 7th, 2013
Example of MLE - Exponential Distribution

Problem

Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Exponential}(\beta)$. Find MLE of $\beta$.

Solution

$$L(\beta|x) = f_X(x|\theta) = \prod_{i=1}^{n} f_X(x_i|\theta)$$
Example of MLE - Exponential Distribution

Problem

Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Exponential}(\beta)$. Find MLE of $\beta$.

Solution

$$L(\beta|x) = f_X(x|\theta) = \prod_{i=1}^{n} f_X(x_i|\theta)$$

$$= \prod_{i=1}^{n} \left[ \frac{1}{\beta} e^{-x_i/\beta} \right] = \frac{1}{\beta^n} \exp \left( -\sum_{i=1}^{n} \frac{x_i}{\beta} \right)$$

where $\beta > 0$. 
Use the derivative to find potential MLE

To maximize the likelihood function $L(\beta|x)$ is equivalent to maximize the log-likelihood function.
Use the derivative to find potential MLE

To maximize the likelihood function $L(\beta|x)$ is equivalent to maximize the log-likelihood function

$$l(\beta|x) = \log L(\beta|x) = \log \left[ \frac{1}{\beta^n} \exp \left( - \sum_{i=1}^{n} \frac{x_i}{\beta} \right) \right]$$
Use the derivative to find potential MLE

To maximize the likelihood function $L(\beta|x)$ is equivalent to maximize the log-likelihood function

$$l(\beta|x) = \log L(\beta|x) = \log \left( \frac{1}{\beta^n} \exp \left( - \sum_{i=1}^{n} \frac{x_i}{\beta} \right) \right)$$

$$= - \frac{\sum_{i=1}^{n} x_i}{\beta} - n \log \beta$$
Use the derivative to find potential MLE

To maximize the likelihood function $L(\beta|x)$ is equivalent to maximize the log-likelihood function

$$l(\beta|x) = \log L(\beta|x) = \log \left[ \frac{1}{\beta^n} \exp \left( -\sum_{i=1}^{n} \frac{x_i}{\beta} \right) \right]$$

$$= -\frac{\sum_{i=1}^{n} x_i}{\beta} - n \log \beta$$

$$\frac{\partial l}{\partial \beta} = \frac{\sum_{i=1}^{n} x_i}{\beta^2} - \frac{n}{\beta} = 0$$
Use the derivative to find potential MLE

To maximize the likelihood function \( L(\beta|x) \) is equivalent to maximize the log-likelihood function

\[
l(\beta|x) = \log L(\beta|x) = \log \left[ \frac{1}{\beta^n} \exp \left( -\sum_{i=1}^{n} \frac{x_i}{\beta} \right) \right]
\]

\[
= -\frac{\sum_{i=1}^{n} x_i}{\beta} - n \log(\beta)
\]

\[
\frac{\partial l}{\partial \beta} = \sum_{i=1}^{n} \frac{x_i}{\beta^2} - \frac{n}{\beta} = 0
\]

\[
\sum_{i=1}^{n} x_i = n\beta
\]
Use the derivative to find potential MLE

To maximize the likelihood function $L(\beta|x)$ is equivalent to maximize the log-likelihood function

$$ l(\beta|x) = \log L(\beta|x) = \log \left[ \frac{1}{\beta^n} \exp \left( -\sum_{i=1}^{n} \frac{x_i}{\beta} \right) \right] $$

$$ = -\frac{\sum_{i=1}^{n} x_i}{\beta} - n \log \beta $$

$$ \frac{\partial l}{\partial \beta} = \frac{\sum_{i=1}^{n} x_i}{\beta^2} - \frac{n}{\beta} = 0 $$

$$ \sum_{i=1}^{n} x_i = n \beta $$

$$ \hat{\beta} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} $$
Use the double derivative to confirm local maximum

\[ \frac{\partial^2 l}{\partial \beta^2} \bigg|_{\beta=\bar{x}} = -2 \sum_{i=1}^{n} \frac{x_i}{\beta^3} + \frac{n}{\beta^2} \bigg|_{\beta=\bar{x}} \]
Use the double derivative to confirm local maximum

\[ \frac{\partial^2 l}{\partial \beta^2} \bigg|_{\beta=\bar{x}} = -2 \frac{\sum_{i=1}^{n} x_i}{\beta^3} + \frac{n}{\beta^2} \bigg|_{\beta=\bar{x}} \]

\[ = \frac{1}{\beta^2} \left( -2 \frac{\sum_{i=1}^{n} x_i}{\beta} + n \right) \bigg|_{\beta=\bar{x}} \]
Use the double derivative to confirm local maximum

\[
\left. \frac{\partial^2 l}{\partial \beta^2} \right|_{\beta=\bar{x}} = -2 \frac{\sum_{i=1}^{n} x_i}{\beta^3} + \frac{n}{\beta^2} \left|_{\beta=\bar{x}} \right.
\]

\[
= \frac{1}{\beta^2} \left( - \frac{2 \sum_{i=1}^{n} x_i}{\beta} + n \right) \left|_{\beta=\bar{x}} \right.
\]

\[
= \frac{1}{\bar{x}^2} \left( - \frac{2n\bar{x}}{\bar{x}} + n \right)
\]
Use the double derivative to confirm local maximum

\[
\frac{\partial^2 l}{\partial \beta^2} \bigg|_{\beta = \bar{x}} = -2 \sum_{i=1}^{n} \frac{x_i}{\beta^3} + \frac{n}{\beta^2} \bigg|_{\beta = \bar{x}}
\]

\[
= \frac{1}{\beta^2} \left( -2 \sum_{i=1}^{n} \frac{x_i}{\beta} + n \right) \bigg|_{\beta = \bar{x}}
\]

\[
= \frac{1}{\bar{x}^2} \left( -2 \frac{n\bar{x}}{\bar{x}} + n \right)
\]

\[
= \frac{1}{\bar{x}^2} (-n) < 0
\]
Use the double derivative to confirm local maximum

\[
\frac{\partial^2 l}{\partial \beta^2} \bigg|_{\beta=\bar{x}} = -2 \sum_{i=1}^{n} \frac{x_i}{\beta^3} + \frac{n}{\beta^2} \bigg|_{\beta=\bar{x}}
\]

\[
= \frac{1}{\beta^2} \left( - \frac{2 \sum_{i=1}^{n} x_i}{\beta} + n \right) \bigg|_{\beta=\bar{x}}
\]

\[
= \frac{1}{\bar{x}^2} \left( - \frac{2n\bar{x}}{\bar{x}} + n \right)
\]

\[
= \frac{1}{\bar{x}^2} (-n) < 0
\]

Therefore, we can conclude that \( \hat{\beta}(X) = \bar{X} \) is unique local maximum on the interval.
Check boundary and confirm global maximum

\[ \beta \in (0, \infty). \text{ If } \beta \to \infty \]
Check boundary and confirm global maximum

\( \beta \in (0, \infty) \). If \( \beta \to \infty \)

\[
l(\beta|\mathbf{x}) = -\frac{\sum_{i=1}^{n} x_i}{\beta} - n \log \beta \to -\infty
\]
Check boundary and confirm global maximum

\[ \beta \in (0, \infty). \text{ If } \beta \to \infty \]

\[
\begin{align*}
l(\beta|x) & = -\frac{\sum_{i=1}^{n} x_i}{\beta} - n \log \beta \to -\infty \\
L(\beta|x) & \to 0
\end{align*}
\]
Check boundary and confirm global maximum

\( \beta \in (0, \infty) \). If \( \beta \to \infty \)

\[
\ell(\beta|x) = -\sum_{i=1}^{n} \frac{x_i}{\beta} - n \log \beta \to -\infty
\]

\[
L(\beta|x) \to 0
\]

If \( \beta \to 0 \), use \( \log(x) = \lim_{\beta \to 0} \frac{1}{\beta} (x^\beta - 1) \)
Check boundary and confirm global maximum

\[ \beta \in (0, \infty). \text{ If } \beta \to \infty \]

\[
l(\beta | x) = - \frac{\sum_{i=1}^{n} x_i}{\beta} - n \log \beta \to -\infty
\]

\[ L(\beta | x) \to 0 \]

If \( \beta \to 0 \), use \( \log(x) = \lim_{\beta \to 0} \frac{1}{\beta} (x^\beta - 1) \)

\[
l(\beta | x) = - \frac{\sum_{i=1}^{n} x_i}{\beta} - n \log \beta
\]
Check boundary and confirm global maximum

\( \beta \in (0, \infty) \). If \( \beta \to \infty \)

\[
l(\beta|x) = -\sum_{i=1}^{n} \frac{x_i}{\beta} - n \log \beta \to -\infty
\]

\[
L(\beta|x) \to 0
\]

If \( \beta \to 0 \), use \( \log(x) = \lim_{\beta \to 0} \frac{1}{\beta}(x^\beta - 1) \)

\[
l(\beta|x) = -\sum_{i=1}^{n} \frac{x_i}{\beta} - n \log \beta
\]

\[
= -\sum_{i=1}^{n} \frac{x_i}{\beta} - n \left( \frac{1}{\beta} \beta^\beta - 1 \right)
\]
Check boundary and confirm global maximum

$\beta \in (0, \infty)$. If $\beta \to \infty$

$$l(\beta|x) = -\sum_{i=1}^{n} \frac{x_i}{\beta} - n \log \beta \to -\infty$$

$L(\beta|x) \to 0$

If $\beta \to 0$, use $\log(x) = \lim_{\beta \to 0} \frac{1}{\beta} (x^\beta - 1)$

$$l(\beta|x) = -\sum_{i=1}^{n} \frac{x_i}{\beta} - n \log \beta$$

$$= -\sum_{i=1}^{n} \frac{x_i}{\beta} - n \left( \frac{1}{\beta} \beta^\beta - 1 \right)$$

$$= -\sum_{i=1}^{n} \frac{x_i}{\beta} - n(\beta^\beta - 1) \to -\infty$$
Check boundary and confirm global maximum

\( \beta \in (0, \infty) \). If \( \beta \to \infty \)

\[
l(\beta | x) = -\sum_{i=1}^{n} \frac{x_i}{\beta} - n \log \beta \to -\infty
\]

\( L(\beta | x) \to 0 \)

If \( \beta \to 0 \), use \( \log(x) = \lim_{\beta \to 0} \frac{1}{\beta}(x^\beta - 1) \)

\[
l(\beta | x) = -\sum_{i=1}^{n} \frac{x_i}{\beta} - n \log \beta
\]

\[
= -\sum_{i=1}^{n} \frac{x_i}{\beta} - n \left( \frac{1}{\beta} \beta^\beta - 1 \right)
\]

\[
= -\sum_{i=1}^{n} \frac{x_i}{\beta} - n(\beta^\beta - 1) \to -\infty
\]

\( L(\beta | x) \to 0 \)
Putting Things Together

1. $\frac{\partial l}{\partial \beta} = 0$ at $\hat{\beta} = \bar{x}$
Putting Things Together

1. \( \frac{\partial l}{\partial \beta} = 0 \) at \( \hat{\beta} = \bar{x} \)

2. \( \frac{\partial^2 l}{\partial \beta^2} < 0 \) at \( \hat{\beta} = \bar{x} \)
Putting Things Together

1. \( \frac{\partial l}{\partial \beta} = 0 \) at \( \hat{\beta} = \bar{x} \)
2. \( \frac{\partial^2 l}{\partial \beta^2} < 0 \) at \( \hat{\beta} = \bar{x} \)
3. \( L(\beta|x) \rightarrow 0 \) (lowest bound) when \( \beta \) approaches the boundary
Putting Things Together

1. \[ \frac{\partial l}{\partial \beta} = 0 \text{ at } \hat{\beta} = \bar{x} \]

2. \[ \frac{\partial^2 l}{\partial \beta^2} < 0 \text{ at } \hat{\beta} = \bar{x} \]

3. \[ L(\beta|x) \rightarrow 0 \text{ (lowest bound) when } \beta \text{ approaches the boundary} \]

Therefore \( l(\beta|x) \) and \( L(\beta|x) \) attains the global maximum when \( \hat{\beta} = \bar{x} \)

\( \hat{\beta}(X) = \bar{X} \) is the MLE of \( \beta \).
How do we find MLE?

If the function is differentiable with respect to $\theta$. 
How do we find MLE?

If the function is differentiable with respect to $\theta$.

1. Find candidates that makes first order derivative to be zero

2. Check second-order derivative to check local maximum.
   - For one-dimensional parameter, negative second order derivative implies local maximum.
   - For two-dimensional parameter, suppose $L(\theta_1; \theta_2)$ is the likelihood function. Then we need to show
     \[
     \frac{\partial^2}{\partial \theta_1^2} L(\theta_1; \theta_2) < 0 \quad \text{or} \quad \frac{\partial^2}{\partial \theta_2^2} L(\theta_1; \theta_2) < 0.
     \]

   2. Check boundary points to see whether boundary gives global maximum.

If the function is NOT differentiable with respect to $\theta$.

• Use numerical methods
• Or perform directly maximization, using inequalities, or properties of the function.
How do we find MLE?

If the function is differentiable with respect to $\theta$.

1. Find candidates that makes first order derivative to be zero
2. Check second-order derivative to check local maximum.
How do we find MLE?

If the function is differentiable with respect to $\theta$.

1. Find candidates that makes first order derivative to be zero
2. Check second-order derivative to check local maximum.
   - For one-dimensional parameter, negative second order derivative implies local maximum.
How do we find MLE?

If the function is differentiable with respect to $\theta$.

1. Find candidates that makes first order derivative to be zero
2. Check second-order derivative to check local maximum.
   - For one-dimensional parameter, negative second order derivative implies local maximum.
   - For two-dimensional parameter, suppose $L(\theta_1, \theta_2)$ is the likelihood function. Then we need to show

How do we find MLE?

If the function is differentiable with respect to $\theta$.

1. Find candidates that makes first order derivative to be zero.
2. Check second-order derivative to check local maximum.

- For one-dimensional parameter, negative second order derivative implies local maximum.
- For two-dimensional parameter, suppose $L(\theta_1, \theta_2)$ is the likelihood function. Then we need to show
  
  \[ \frac{\partial^2 L(\theta_1, \theta_2)^2}{\partial \theta_1^2} < 0 \text{ or } \frac{\partial^2 L(\theta_1, \theta_2)^2}{\partial \theta_2^2} < 0. \]
How do we find MLE?

If the function is differentiable with respect to $\theta$.

1. Find candidates that makes first order derivative to be zero
2. Check second-order derivative to check local maximum.
   - For one-dimensional parameter, negative second order derivative implies local maximum.
   - For two-dimensional parameter, suppose $L(\theta_1, \theta_2)$ is the likelihood function. Then we need to show
     (a) $\partial^2 L(\theta_1, \theta_2)^2 / \partial \theta_1^2 < 0$ or $\partial^2 L(\theta_1, \theta_2)^2 / \partial \theta_2^2 < 0$.
     (b) Determinant of second-order derivative is positive
How do we find MLE?

If the function is differentiable with respect to $\theta$.

1. Find candidates that makes first order derivative to be zero
2. Check second-order derivative to check local maximum.
   - For one-dimensional parameter, negative second order derivative implies local maximum.
   - For two-dimensional parameter, suppose $L(\theta_1, \theta_2)$ is the likelihood function. Then we need to show
     (a) $\frac{\partial^2 L(\theta_1, \theta_2)^2}{\partial \theta_1^2} < 0$ or $\frac{\partial^2 L(\theta_1, \theta_2)^2}{\partial \theta_2^2} < 0$.
     (b) Determinant of second-order derivative is positive
   - Check boundary points to see whether boundary gives global maximum.
How do we find MLE?

If the function is differentiable with respect to $\theta$.

1. Find candidates that makes first order derivative to be zero
2. Check second-order derivative to check local maximum.
   - For one-dimensional parameter, negative second order derivative implies local maximum.
   - For two-dimensional parameter, suppose $L(\theta_1, \theta_2)$ is the likelihood function. Then we need to show
     (a) $\frac{\partial^2 L(\theta_1, \theta_2)^2}{\partial \theta_1^2} < 0$ or $\frac{\partial^2 L(\theta_1, \theta_2)^2}{\partial \theta_2^2} < 0$.
     (b) Determinant of second-order derivative is positive
   - Check boundary points to see whether boundary gives global maximum.

If the function is NOT differentiable with respect to $\theta$.

- Use numerical methods
How do we find MLE?

If the function is differentiable with respect to $\theta$.

1. Find candidates that makes first order derivative to be zero
2. Check second-order derivative to check local maximum.
   - For one-dimensional parameter, negative second order derivative implies local maximum.
   - For two-dimensional parameter, suppose $L(\theta_1, \theta_2)$ is the likelihood function. Then we need to show
     (a) $\frac{\partial^2 L(\theta_1, \theta_2)}{\partial \theta_1^2} < 0$ or $\frac{\partial^2 L(\theta_1, \theta_2)}{\partial \theta_2^2} < 0$.
     (b) Determinant of second-order derivative is positive
   - Check boundary points to see whether boundary gives global maximum.

If the function is NOT differentiable with respect to $\theta$.

- Use numerical methods
- Or perform directly maximization, using inequalities, or properties of the function.
Summary

Today

- Likelihood Function
- Point Estimator
- Method of Moments Estimator
- Maximum Likelihood Estimator
Summary

Today

- Likelihood Function
- Point Estimator
- Method of Moments Estimator
- Maximum Likelihood Estimator

Next Lecture

- Maximum Likelihood Estimator