IEOR165 Discussion
Week 3

Sheng Liu
University of California, Berkeley

Feb 5, 2016
Outline

1 Weighted Least Squares

2 Maximum Likelihood Estimation (MLE)
Heteroskedasticity

Weighted Least Squares

In simple linear regression

\[ Y = \beta_0 + \beta_1 X + \epsilon \]

We assume the error \( \epsilon \) is independent with \( X \). This implies that \( \text{Var}(\epsilon|X) = \sigma^2 \). When this assumption is violated, we may get poor estimates using ordinary least squares. This motivates us to apply weighted least squares.

(IPSES E9.8b) The following data present travel times in the Berkeley downtown area with different travel distances.

<table>
<thead>
<tr>
<th>Distance</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>8</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>15</td>
<td>15.1</td>
<td>16.5</td>
<td>19.9</td>
<td>27.7</td>
<td>29.7</td>
<td>26.7</td>
<td>35.9</td>
<td>42</td>
<td>49.4</td>
</tr>
</tbody>
</table>

Suppose you are planning a trip that will take 2.5 miles in this area, how to establish a linear model to estimate the travel time?
Response: Time
Predictor: Distance
Linear Model:

\[ \text{Distance} = \beta_0 + \beta_1 \times \text{Distance} + \epsilon \]

From ordinary least squares, we have

\[ \hat{\beta}_0 = 12.722 \quad \hat{\beta}_1 = 3.675 \]

What about the residuals?
Imagine that longer distance involves more blocks and thus we can assume

\[ \text{Var}(Y|X = x) = \text{Var}(\epsilon|X = x) = x\sigma^2 \]

Thus we can choose weight as \( w_i = x_i \) or \( w_i = x_i\sigma^2 \). Then the weighted least squares aims to minimize

\[
\sum_{i=1}^{n} \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{x_i}
\]

We have

\[ \hat{\beta}_0^{\text{wls}} = 12.554 \quad \hat{\beta}_1^{\text{wls}} = 3.716 \]
Maxmum Likelihood Estimation

MLE:

\[ \hat{\theta}_{MLE} = \arg \max_{\theta} L(\theta) \]

When \( X_1, \ldots, X_n \) are an iid sample, the likelihood function is

\[ L(\theta) = \prod_{i=1}^{n} f_{\theta}(x) \]

Maximization can be simplified if you deal with the log-likelihood function instead of the likelihood function:

\[ l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f_{\theta}(x) \]

Intuition: find the parameter for which the sample is mostly likely.
MLE: examples

(SI 7.1) Suppose we are given one observation on a discrete random variable $X$ with pmf $f_\theta(x)$, where $\theta \in \{1, 2, 3\}$. Find the MLE of $\theta$.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$f_1(x)$</th>
<th>$f_2(x)$</th>
<th>$f_3(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{4}$</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{4}$</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{1}{6}$</td>
<td>0</td>
<td>$\frac{1}{4}$</td>
</tr>
</tbody>
</table>

1. If the only observation is $x = 0$, what is the MLE of $\theta$?
2. If the only observation is $x = 2$, what is the MLE of $\theta$?
3. If the only observation is $x = 4$, what is the MLE of $\theta$?
MLE: examples

(SI 7.12) Let $X_1, \ldots, X_n$ be a random sample from a population with pmf

$$P_\theta(X = x) = \theta^x (1 - \theta)^{1-x}, \quad x \in \{0, 1\}, \quad 0 \leq \theta \leq 1/2$$

Find the MLE estimate of $\theta$. 
(SI 7.6) Let $X_1, \ldots, X_n$ be a random sample from the pdf

$$f(x|\theta) = \theta x^{-2}, \theta \leq x < \infty$$

1. Use the method of moments to estimate $\theta$.
2. Find the MLE estimate of $\theta$. 
(PSES 7.6) We are given a random sample:
And we know they are drawn from a population with pdf

\[ f_\theta(x) = \begin{cases} 
\frac{\theta}{x^{\theta+1}} & x > 1 \\
0 & o.w. 
\end{cases} \]

Find the MLE estimate of \( \theta \).