# IEOR165 Discussion Week 3

Sheng Liu

University of California, Berkeley

Feb 5, 2016

## Outline



1 Weighted Least Squares

2 Maxmum Likelihood Estimation (MLE)

#### Heteroskedasticity



Reference: James, Gareth, et al. An introduction to statistical learning. New York: springer, 2013.



### 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  $Var(\epsilon|X) = \sigma^2$ . 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.

Distance	0.5	1	1.5	2	3	4	5	6	8	10
Time	15	15.1	16.5	19.9	27.7	29.7	26.7	35.9	42	49.4

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?

Weighted Least Squares Maxmum Likelihood Estimation (MLE)

## Weighted Least Squares (Contd.)

- Response: Time
- Predictor: Distance
- Linear Model:

$$\mathsf{Distance} = \beta_0 + \beta_1 \times \mathsf{Distance} + \epsilon$$

From ordinary least squares, we have

$$\hat{\beta}_0 = 12.722$$
  $\hat{\beta}_1 = 3.675$ 

What about the residuals?



Weighted Least Squares Maxmum Likelihood Estimation (MLE)

# Weighted Least Squares (Contd.)



Berkelev

## Weighted Least Squares (Contd.)



$$Var(Y|X = x) = Var(\epsilon|X = x) = x\sigma^2$$

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

$$\hat{\beta}_0^{wls} = 12.554 \quad \hat{\beta}_1^{wls} = 3.716$$



## 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.



(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  $\theta$ .

х	$f_1(x)$	$f_2(x)$	$f_3(x)$
0	1/3	1/4	0
1	1/3	1/4	0
2	0	1/4	1/4
3	1/6	1/4	1/2
4	1/6	0	1/4

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  $\theta$ ?



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

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

Find the MLE estimate of  $\theta$ .



(SI 7.6) Let  $X_1, \ldots, X_n$  be a random sample from the pdf

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

- 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: 12, 11.2, 13.5, 12.3, 13.8, 11.9. 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  $\theta$ .

