Department of Industrial Engineering & Operations Research

IEOR 165 (Spring 2016)

Homework 1

Due: Friday, Feb 12

Question 1. Let $X_1, \ldots, X_n$ be iid from the pdf

$$f(x) = \theta x^{\theta-1}, \quad 0 \le x \le 1,\ 0 < \theta < \infty$$

Find the method of moments estimator of θ. (*hint:* $E(X) = \int_0^1 \theta x^{\theta-1} \cdot x dx$)

Solution. The first moment is

$$E(X) = \int_0^1 \theta x^{\theta-1} \cdot x dx = \frac{\theta}{\theta+1}$$

Replace $E(X)$ by $\hat{\mu}_1$, we have

$$\hat{\theta} = \frac{\hat{\mu}_1}{1-\hat{\mu}_1} = \frac{\sum_{i=1}^n X_i}{n - \sum_{i=1}^n X_i}$$

Question 2. Let $X_1, \ldots, X_n$ be iid from the uniform distribution on (θ_1, θ_2), where $\theta_1 < \theta_2 < \infty$. Use the method of moments to estimate θ_1 and θ_2.

Solution. The first two moments are

$$E(X) = \frac{\theta_1 + \theta_2}{2}$$

$$E(X^2) = \frac{\theta_1^2 + \theta_2^2 + \theta_1\theta_2}{3}$$

Replace $E(X)$ and $E(X^2)$ by $\hat{\mu}_1$ and $\hat{\mu}_2$, then solve these two equations for θ_1 and θ_2:

$$\hat{\theta}_2 = \hat{\mu}_1 + \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)}$$

$$\hat{\theta}_1 = \hat{\mu}_1 - \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)}$$

where $\hat{\mu}_1 = \sum_{i=1}^n X_i/n$ and $\hat{\mu}_2 = \sum_{i=1}^n X_i^2/n$.
Note: since the sample follows a uniform distribution on (θ_1, θ_2), our estimate of θ_2 must be greater than the sample mean, i.e. $\hat{\theta}_2 > \hat{\mu}_1$. Thus we drop one solution in solving the quadratic equation for θ_2.

Question 3. The following data set specifies the number of units of a good ordered and the price of the good at six different locations. Suppose we would like to build a linear model that predicts the number ordered as a function of price.

Number Ordered	88	112	123	136	158	172
Price	50	40	35	30	20	15

a) State the predictor and the response.

Solution. The response is Number ordered and the predictor is Price.

b) What is the linear model?

Solution. The linear model is

$$\text{Number Ordered} = \beta_0 + \beta_1 \text{Price} + \epsilon$$

c) Estimate the parameters of the linear model using least squares.

Solution. Let x be the price and y be the number ordered, from the formula of least squares we have

$$\hat{\beta}_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2} = -2.376$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 206.74$$

Question 4. The corrosion of a certain metallic substance has been studied in dry oxygen at 500 degrees Centigrade. In this experiment, the gain in weight after various periods of exposure was used as a measure of the amount of oxygen that had reacted with the sample. Here are the data:

Hour	Percent Gain
1.0	0.020
2.0	0.030
2.5	0.035
3.0	0.042
3.5	0.050
4.0	0.054

Suppose we would like to build a linear model that predicts the percent weight gain as a function of time of exposure.
a) State the predictor and the response.

Solution. The response is Percent Gain and the predictor is Hour.

b) What is the linear model?

Solution. The linear model is

$$\text{Percent Gain} = \beta_0 + \beta_1 \text{Hour} + \epsilon$$

c) Estimate the parameters of the linear model using least squares.

Solution. Let x be the hour and y be the percent gain, from the formula of least squares we have

$$\hat{\beta}_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2} = 0.0117$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 0.0073$$

d) Predict the percent weight gain when the metal is exposed for 4.5 hours.

Solution. From the linear model

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot 4.5 = 0.060$$

So the percent weight gain after 4.5 hours is predicted as 0.060.

Question 5. Imagine you are a consultant for the Bay Area Bike Share system and the following information is available:

- Average bike demand per day
- Average wind speed
- Average temperature
- Weekday $\in \{Sun, Mon, \ldots, Sat\}$

Suppose you would like to build a linear model to predict the demand based on wind speed, temperature and the weekday information. Please specify the response and predictor variables.

Solution. The response: Average bike demand per day. The predictors: Average wind speed, Average temperature, and Weekday. And for the Weekday variable, we need to code it by dummy variables. In total, we need 6 dummy variables.