ECON 5350 Solutions to the Midterm Exam – Fall 2017

1. **Probability and Statistics (50 pts)**. Let X have a Pareto cdf where $F(x;\theta) = 1-(1/x)^{\theta}$ for $x \geq 1$ and zero elsewhere; $\theta > 3$.

   (a) Find the pdf for X, $f(x)$, and verify it is a valid pdf.

   Solution. The pdf is

$$f(x) = \frac{dF(x)}{dx} = \theta x^{-(\theta+1)} \text{ for } x \geq 1$$

   and zero otherwise. Integrating $f(x)$ over the range $x \geq 1$ gives

$$\int_1^{\infty} f(x)dx = \int_1^{\infty} \theta x^{-(\theta+1)}dx = \theta[-\frac{1}{\theta}x^{-\theta}]|_{x=1}^{\infty} = 1.$$

   (b) Find the mean and variance of X.

   Solution. The mean is given by

$$\mu_X = E(X) = \int_1^{\infty} \theta x^{-\theta}dx = \frac{\theta}{(\theta-1)}.$$

   To calculate the variance, first find

$$E(X^2) = \int_1^{\infty} \theta x^{-\theta+1}dx = \frac{\theta}{(\theta-2)}.$$

   The variance is then

$$\sigma_X^2 = var(X) = E(X^2) - E(X)^2 = \frac{\theta}{(\theta-1)^2(\theta-2)}.$$

   (c) Let $\theta = 4$. Find the pdf for $Y = X^2$, $g(y)$. Find the mean of Y and verify that $g(y)$ is a valid pdf.

   Solution. Using the change-of-variable technique, $X = \sqrt{Y}$ and $J = 0.5Y^{-0.5}$. The pdf is

$$g(y) = 4y^{-5/2}(0.5y^{-0.5}) = 2y^{-3} \text{ for } y \geq 1$$

   and zero otherwise. The mean of Y is given by

$$E(Y) = \int_1^{\infty} 2y^{-2}dy = -2y^{-1}|_{y=1}^{\infty} = 2.$$

Integrating $g(y)$ over the range $y \geq 1$ gives

$$\int_1^\infty g(y)dy = \int_1^\infty 2y^{-3}dy = -y^{-2}]|_{y=1}^\infty = 1.$$

(d) Outline two different procedures for estimating θ from a random sample $\{X_1, X_2, ..., X_n\}$.

Solution. The first is the method of moments. Since there is only one unknown parameter, only one moment is needed. Set the estimated mean $\bar{X}$ equal to the population mean from part (b); then solve for θ. A second possibility is maximum likelihood. Use $f(x)$ from part (a) and independence to form the likelihood function (joint probability). Then choose the θ that maximizes the likelihood function.

(e) Find the pdf for the smallest value from a random sample of size $n = 2$, $\{X_1, X_2\}$.

The pdf for the first-order statistic when $n = 2$ is

$$\begin{aligned} f_1(y_1) &= n(1 - F(y_1))^{n-1} f(y_1) = 2(1 - F(y_1))f(y_1) \\ &= 2y_1^{-\theta}\theta y_1^{-(\theta+1)} = 2\theta y_1^{-(2\theta+1)}, y_1 \geq 1 \end{aligned}$$

and zero otherwise.

2. **Classical Linear Regression Model (50 pts)**. Consider the following model: $Y_i = \beta_1 + \beta_2 X_i + \epsilon_i$ for $i = 1, ..., n$.

(a) Without using matrices, derive the least squares estimator for the intercept, β_1.

Solution. The least squares objective is

$$\min \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - b_1 - b_2 X_i)^2.$$

The first-order condition for the intercept is

$$\frac{\partial \sum_{i=1}^n e_i^2}{\partial b_1} = -2\sum_{i=1}^n (Y_i - b_1 - b_2 X_i) = 0 \Rightarrow b_1 = \bar{Y} - b_2 \bar{X}.$$

(b) Without using matrices, derive the least squares estimator for the slope, β_2.

Solution. The first-order condition for the intercept is

$$\frac{\partial \sum_{i=1}^n e_i^2}{\partial b_2} = -2\sum_{i=1}^n (Y_i - b_1 - b_2 X_i)X_i = 0 \Rightarrow b_2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}.$$

(c) Show that b_1 and b_2 are unbiased, make sure to highlight only the necessary Classical assumptions as you go.

Solution. To show b_1 is unbiased, we find

$$\begin{aligned} E(b_1) &= E(\bar{Y}) - E(b_2)\bar{X} \\ &= E(\beta_1 + \beta_2\bar{X} + \bar{\epsilon}) - \beta_2\bar{X} \\ &= \beta_1. \end{aligned}$$

To show b_2 is unbiased, we find

$$\begin{aligned} E(b_2) &= E\left[\frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}\right] \\ &= \frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2} E\left[\sum_{i=1}^n \left(\beta_2 (X_i - \bar{X})^2 + (\epsilon_i - \bar{\epsilon})(X_i - \bar{X})\right)\right] \\ &= \beta_2 + \frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2} E\left[\sum_{i=1}^n (\epsilon_i - \bar{\epsilon})(X_i - \bar{X})\right] \\ &= \beta_2. \end{aligned}$$

Each proof requires that X is fixed in repeated sampling and the errors are mean zero.

(d) Now consider the alternative model $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i$, where $\bar{X}_2 = \bar{X}_3 = 0$ and $corr(X_{2i}, X_{3i}) = 0$. Use matrix algebra to find the least squares estimates of β_1, β_2 and β_3.

Solution. The least squares estimate is

$$\begin{aligned} b &= \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} n & 0 & 0 \\ 0 & \sum x_{2i}^2 & 0 \\ 0 & 0 & \sum x_{3i}^2 \end{bmatrix}^{-1} \begin{bmatrix} n\overline{y} \\ \sum x_{2i} y_i \\ \sum x_{3i} y_i \end{bmatrix} \\ &= \begin{bmatrix} 1/n & 0 & 0 \\ 0 & 1/\sum x_{2i}^2 & 0 \\ 0 & 0 & 1/\sum x_{3i}^2 \end{bmatrix} \begin{bmatrix} n\overline{y} \\ \sum x_{2i} y_i \\ \sum x_{3i} y_i \end{bmatrix} = \begin{bmatrix} \overline{y} \\ \sum x_{2i} y_i / \sum x_{2i}^2 \\ \sum x_{3i} y_i / \sum x_{3i}^2 \end{bmatrix}. \end{aligned}$$

(e) Assume the model in part (d) is the true population regression model, but you mistakenly estimate the following model: $Y_i = \beta_1 + \beta_2 X_{2i} + \epsilon_i$. Is the OLS estimate of β_2 biased or unbiased? Defend your answer.

Solution. Unbiased. The expectation of b_2 is

$$\begin{aligned}
E(b_2) &= E\left[\frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_{2i} - \bar{X}_2)}{\sum_{i=1}^n (X_{2i} - \bar{X}_2)^2}\right] \\
&= \frac{1}{\sum_{i=1}^n X_{2i}^2} E\left[\sum\nolimits_{i=1}^n \left(\beta_2 X_{2i}^2 + \beta_3 X_{2i} X_{3i} + (\epsilon_i - \overline{\epsilon})(X_{2i})\right)\right] \\
&= \beta_2 + \frac{1}{\sum_{i=1}^n X_{2i}^2} E\left[\sum\nolimits_{i=1}^n (\epsilon_i - \overline{\epsilon})(X_{2i})\right] \\
&= \beta_2.
\end{aligned}$$