Economics 217

Homework #1

Due Thursday, Jan 22nd

Problem 1

The Pareto distribution is a tractable and extremely convenient distribution that matches a wide variety of economic phenomena. For example, the Pareto distribution matches the relationship between the rank of city population and the log of population. Specifically, if one regresses the log city rank (largest having a rank of 1, second largest having a rank of 2) on the log of actual population, the one typically gets a straight line. A similar exercise holds when you look at the rank of firm-size and firm revenues.

The Pareto distribution itself can be characterized by the following cumulative density function

$$F(y) = \Pr(Y < y) = \left(\frac{y}{y_{max}} \right)^k \tag{1}$$

where $Y \in [0; y_{max}]$ and $k > 0$. Using (1) the probability distribution function is derived as:

$$\frac{dF(y)}{dy} = f(y) = k \frac{y^{k-1}}{y_{max}^k} \tag{2}$$

a. Using (2), please show that the Pareto distribution is a member of the exponential family. You may take y_{max} to be the nuisance parameter, and thus, please indicate the functions $b(k)$ and $c(k)$ in your answer (10 points).

b. Please derive the mean and variance of the Pareto distribution (10 points).

c. Using R, assume $y_{max} = 10$ and please plot the PDF and CDF of the Pareto distribution using the shape parameters $k = 0.5$, $k = 1$, and $k = 2$. Although you may plot each on separate graphs, I recommend experimenting with the function lines and plotting all PDFs on one plot and all CDFs on a separate plot (10 points).

Problem 2

For this question, we will use the "org_example" data from the website to estimate the relationship between education and the probability of being unemployed during the great recession. We will step through this process to get familiar with R.

a. Load the data and restrict the sample to be only include data from 2008 (5 points).

b. To be considered unemployed, you must be in the labor force. Please restrict the sample to only include individuals who are in the labor force (5 points).

c. Run a probit model predicting unemployment as a function of the education categories. Please interpret each coefficient and comment on its statistical significance (10 points).

d. Please compute the marginal effect of getting and advanced degree, relative to receiving a college degree, on the probability of being unemployed. Remember for this question that this is a marginal effect of a categorical variable. To do this, I want you to program the following difference in R:

$$\Pr(Unemployed|Advanced) - \Pr(Unemployed|College)$$

The CDF of the standard normal distribution in R is given by the function dnorm(), which will help you with these probabilities (10 points).