Problem Set 2 **Solutions**

EC 303: Empirical Economic Analysis

Due October 8, 2019 at 3:30pm.

1 Theoretical Problems

Problem 1.1: Random Variables.** Working with discrete random variables.

1. Complete Exercise 3.23 from the textbook.
2. Complete Exercise 3.24 from the textbook.

Problem 1.2: Expectations.** Working with the expectation operator.

1. Complete Exercise 3.32 from the textbook.
2. Use the formula on the bottom of page 115 of the textbook to prove that $\mathbb{E}[X - c] = \mathbb{E}[X] - c$, where $c \in \mathbb{R}$ is any constant. What does this imply is true about $\mathbb{E}[X - c]$ when $c = \mu_X$?

Problem 1.3: Continuous RVs and Moments.** This is a compilation of 4.6 and 4.22 in the text.

The grade point averages (GPA's) for graduating seniors at a college are distributed as a continuous random variable X with pdf

$$f(x) = \begin{cases} k[1 - (x-3)^2] & x \in [2, 4] \\ 0 & \text{otherwise.} \end{cases}$$

a. Sketch the graph of $f(x)$.
b. What is the value of k needed to make this a valid pdf?
c. What is the probability that a GPA is greater than 3?
d. What is the probability that a GPA is differs from 3 by more than 0.5?
e. Compute $\mathbb{E}[X]$ and $\mathbb{V}[X]$.

Problem 1.4: Moment Generating Functions.** Exploring mgf's.

1. Prove the following: If X is a *discrete* random variable that has a mgf $M_X(t)$ and $Y = aX + b$, then $M_Y(t) = e^{bt} M_X(at)$. How does this proof change if X is continuous? (*Hint:* There's no need to rewrite the proof for continuous case, just explain the key difference.)
2. Show that $g(t) = te^t$ cannot be a moment generating function.
3. Complete Exercise 4.32 from the textbook.

Problem 1.5: Popular Distributions. Let's zoom in on the Poisson and Pareto distributions.

1. Complete Exercise 3.100 from the textbook.
2. Complete Exercise 4.28 parts (a) and (b) from the textbook. The Pareto distribution is frequently used to model the distribution of incomes.

Problem 1.6: Transformations of a Normal Distribution.. If $X \sim \mathcal{N}(\mu, \sigma^2)$, prove that the following transformations are legitimate:

a. $Y = e^X$ has a lognormal pdf (see Exercise 4.112 in the text).

b. $Z = (X - \mu)^{-2}$ has the pdf given by $\frac{e^{-1/(2y\sigma^2)}}{\sqrt{2y^3}\pi\sigma}$. This is referred to as the Lévy distribution, used for modelling Brownian motion and—occasionally—income flows. (*Hint:* Recall that the standardization does nothing but change this to a centered normal distribution, where $\mu = 0$.)

2 Stata Exercises

Problem 2.1: Probability Plots. Use the bwght2 data set for this problem.

a. Create a (good-looking) histogram of the bwght variable. Does this appear to be normally distributed?

b. Transform this variable to be approximately a standard normal. Verify through another histogram that the new variable is centered around 0. Do you have any concerns about this as a standard normal?

c. Now construct and interpret a probability plot for the transformed variable. How does it compare to the plot for the untransformed variable?

d. Make a probability plot for lbwght, a variable created by log(bwght). What does the concavity of the plot indicate? Verify this in the histogram.

Problem 2.2: Working with Normal Distributions. Stata can be quite useful in calculating probabilities from a normal distribution. To see this, we will use the openness data set, which contains data on openness to foreign trade for various countries.

a. Look at a histogram and a probability plot for open (no need to report them). Does this variable appear normally distributed? What about lopen? What does this imply about the approximate distribution of open?

b. What are the mean and standard deviation of lopen?

c. Use this website to find current data on a country's openness to trade (the link should have the data show immediately; if not, select the variable "Openness at constant prices" from the drop down menu). Select a country of interest and report their (i) openness score and (ii) log openness score[1]. Then calculate (iii) the z-score using the measures you calculated in part (b). How many standard deviations above/below the mean is the country you selected?

d. Back in Stata, transform lopen into a new variable that is its z-score (if you're curious, check that the new variable appears to be a standard normal). What proportion of observations in the data are below the z-score you reported in part (c)?

e. Stata can also calculate probabilities from a true normal distribution. Use the display, normprob and/or invnorm commands to report the probability that a randomly selected country has (i) an openness score lower than the country you chose, and (ii) an openness score between your country and the average openness score. How does your answer to (i) compare to your answer to part (d) above? Why are the numbers different?

Problem 2.3: Binomial Random Variables. An investor considers the use of a Binomial random variable to model the number X of days (per week) on which the price of S&P500 is higher than that of the previous business day. In the model, X is assumed to follow a Binomial distribution with $n = 5$ (business days), but the probability p of S&P500's achieving a higher price than the previous trading day is unknown. The investor conjectures that p is one of the following values: $\theta \in \{0.25, 0.3, 0.5, 0.7, 0.85\}$.

[1] *Note:* When taking logs here, do not convert the percentage into a decimal, but use percentage points. For example, if the openness score is 72%, the log openness score should be ln(72), not ln(0.72).

a. Calculate the PMF $p(x)$ of a Binomial random variable for each of the parameter values $(\theta; n)$. Plot the calculated probability mass functions. (You may use hand calculations, Stata, or any other software e.g. Matlab, Excel.) Comment on any symmetries you see.

b. Calculate the expected value and variance of a Binomial random variable for each of the parameter values $(\theta; n)$.

Download UpDowns.dta from the course website. This file contains the number of days X (per week) on which the close price of S&P500 was higher than the close price of the previous business day.

c. Create a (nice) histogram of X based on the relative frequency distribution of X. Which value of θ do you think would be the best for modeling S&P500's up and down movements? Explain.

d. Calculate the sample mean and variance of X. Which value of θ do you think would be the best for modeling S&P500's up and down movements? Explain.

3 References

- Federal Reserve Bank of St. Louis and U.S. Office of Management and Budget, Openness to Trade, retrieved from FRED, Federal Reserve Bank of St. Louis, https://geofred.stlouisfed.org, August 6, 2019.