ECOM20001: Econometrics 1

Tutorial 3: Distributions, Law of Large Numbers, Central Limit Theorem, Conditional Probability Practice Problems

---

Part 1: Normal, Chi-Square, t, F, Distributions; LLN, and CLT in R

A. Getting Started

Please create a Tutorial3 folder on your computer, and then go to the LMS site for ECOM 20001 and download the following files into the Tutorial3 folder:

- tute3.R

There is no dataset for this week's tutorial.

B. Go to the Code

With the R file downloaded into your Tutorial3 folder, you are ready to proceed with the tutorial. Please go to the tute3.R file to continue with the tutorial.

C. Questions

Having worked through the tute3.R code and graphs, please answer the following:

1. Section 2.A from the code discusses probability densities and cumulative densities for the standard normal distribution. Explain why the following equality holds from the code

   ```
   pnorm(-1.65,mean=0,sd=1)=1-pnorm(1.65,mean=0,sd=1)
   ```

2. Explain the relationship between the output from the following two lines of code:

   - `pnorm(1.96,mean=0,sd=1)`
   - `qnorm(0.975,mean=0,sd=1)`

3. Comment on the shape of the Normal (mean=0, sd=1), Chi-Square (df=3), t (df=12), and F (df1=5, df2=2) distributions produced by tute3.R. Discuss whether each distribution is symmetric, right or left skewed, and whether you would expect the mean of the distribution to equal the median, be smaller than the median, or larger than the median.

4. Using Sections 3.A and 3.B from the code, compute the sampling distribution of the mean from an underlying sample that is Chi-Square with df=3 for sample sizes of nobs=10, nobs=50, nobs=100, and nobs=1000.

   - What is the variance of the sampling distribution of the mean for nobs=10, nobs=50, nobs=100, and nobs=1000?
   - For nobs=10, nobs=50, nobs=100, and nobs=1000, what percentage of sample means lies within 0.3 of the true population mean of 3?
   - Explain what the Law of Large Numbers (LLN) is, and how your findings from these two bullet points sheds light on how the LLN works in this example.

5. Using Section 3.C and the PDF graphs of the distribution of sample means produced at the bottom of tute3.R, carefully explain how the results highlight how the Central Limit Theorem (CLT) works as nobs increases from nobs=10 to nobs=1000.

Note: For questions 4 and 5, to obtain the required results for nobs=10, nobs=50, nobs=100, and nobs=1000, follow these 3 steps:

- Go to the line in tute3.R that says:

  "# ***** ALTER FOR ASSIGNMENT HERE ******"

- Change the value of nobs in tute3.R to the necessary value just under this line, to either nobs=10, nobs=50, nobs=100, and nobs=1000
- Re-run the code from the "# ***** ALTER FOR ASSIGNMENT HERE ******" line to the bottom of tute3.R to obtain the results for a given value of nobs.
- LLN results: The output in the console screen of R for a given nobs value can be used to discuss the first two bullet points for the LLN questions above. Perhaps write down the results on a piece of paper for each nobs value.
- CLT results: Change the PDF file name on the pdf() line at the bottom of the tute3.R code to save the PDF graph of the distribution of sample means for nobs=10, nobs=50, nobs=100, and nobs=1000. The pdf() line should either be:
  - pdf("means_nobs_10.pdf")
  - pdf("means_nobs_50.pdf")
  - pdf("means_nobs_100.pdf")
  - pdf("means_nobs_1000.pdf")

Part 2: Conditional Distribution Practice Problems

1. Consider the following table which describes the joint probability distribution for all combinations of studying and performance. The outcome space for Studying (Y) and Performance (X) is:
   - Y - Studying: Study Hard, Study Sometimes, Study Never
   - X - Performance: High Grade, Medium Grade, Low Grade

	High Grade	Medium Grade	Low Grade	Total
Study Hard	0.20	0.10	0.02	0.32
Study Sometimes	0.07	0.30	0.10	0.47
Study Never	0.01	0.05	0.15	0.21
Total	0.28	0.45	0.27	1

   a. What is the marginal distribution for studying?

   b. What is the marginal distribution for performance?

   c. What is the probability distribution of Performance, conditional on Studying hard?

   d. What is the probability distribution of Performance, conditional on Studying Sometimes?

   e. What is the probability distribution of Studying, conditional on Medium Grade?

   f. What is the probability distribution of Studying, conditional on Low Grade?

   g. Using an example from the table above, show that Studying and Performance are not independently distributed.

2. Suppose you have a random variable X that is i.i.d distributed from a N(mu**x**,1) distribution, and a separate random variable Y that is defined as follows:

Y=2+2X

a. What is the distribution of Y?

b. Graphically plot the conditional distribution of Y if mu**x**=2, if mu**x**=5, and if mu**x**=10. What is happening to the conditional distribution of Y for these different X values?

c. Suppose Y was instead distributed as

Y=2+4X.

- What is the distribution of Y now?
- Again, graphically plot the conditional distribution of Y if mu**x**=2, if mu**x**=5, and if mu**x**=10 and compare your results to what you found in part b.
- What can you conclude about the magnitude of how shifts in the conditional distribution of Y as a function of different X values as the magnitude of the slope in the linear function that defines Y increases?

Note: In answering this question, you may simply draw the conditional distributions of Y by hand, or you may use R to plot the different conditional distributions of Y given X.