Econ 114: Assignment 1

DUE DATE: Tuesday 1/18 at 9:50am.

1. **Question 1: Law of Large Numbers**
   In this question we will take a closer look at the law of large numbers (LLN) and investigate situations in which it holds and in which it fails to hold. The law of large numbers states that if X_i are *i.i.d.* with $E[|X_i|] < \infty$, and common mean $\mu = E[X_i]$ then

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu. \tag{1}$$

   One way to interpret the statement is that as n increases, the mean of repeated samples will tend to be closer and closer to μ, the true expected value. This means that if we take S different samples of the size n, compute S different means, then the histogram will be close to μ. The goal of this exercise is to show that by simulation.

   Recall[1] that if X_i is has a uniform distribution in (a, b), then $\mu = E[X_i] = \frac{a+b}{2}$.

   (a) Generate four different samples of sizes $n_1 = 10$, $n_2 = 100$, $n_3 = 1,000$ and $n_4 = 10,000$ all from a uniform distribution with $a = 2$ and $b = 4$. What is μ in this case? Compute the four sample means. Do you get closer to μ?

   (b) For each sample size $n_1 = 10$, $n_2 = 100$, $n_3 = 1,000$ and $n_4 = 10,000$, generate $S = 10,000$ different samples all from a uniform distribution with $a = 2$ and $b = 4$. For each sample size, compute the $10,000$ sample means. For each sample size compute the standard deviation of the sample means. Why is the standard deviation getting close to 0 as the sample size increases?

   (c) Plot four densities with the sample means for each of the sample sizes. What do you observe?

   (d) Now we look at a counterexample. We are going to draw samples from Student's t-distribution with 1 degree of freedom. It is well known that in this case the mean is undefined. Not surprisingly, the law of large numbers does not apply. To see this, repeat (a), (b), and (c) this time using Student's t-distribution with 1 degree of freedom.

2. **Question 2: Central Limit Theorem**
   In this question we are going to take a look at the Central Limit Theorem. The CLT says that if X_i are *i.i.d.* with $E[|X_i|] < \infty$ and $Var[X_i] < \infty$, then

$$\sqrt{n}\frac{\bar{X} - E[X_i]}{\sqrt{Var[X_i]}} \overset{a}{\sim} N(0, 1) \tag{2}$$

   For this exercise, you need to know that if $Z_1, ..., Z_k$ are independent standard normals, then

$$X = Z_1^2 + Z_2^2 + ... + Z_k^2 \tag{3}$$

[1] For example see `https://en.wikipedia.org/wiki/Continuous_uniform_distribution`.

has a chi-squared distribution with k degrees of freedom, and with $E[X] = k$ and $Var[X] = 2k$.

(a) Generate $S = 1,000$ different samples of size $n = 100$ from a chi-squared distribution with $k = 4$ degrees of freedom. Plot the density of $\sqrt{n}\frac{\bar{X}-E[X]}{\sqrt{Var[X]}}$. Here: $E[X] = 4$, and $\sqrt{Var[X]} = \sqrt{8}$.

(b) Incorrect rate: Using the same samples, instead of using the $\sqrt{n}$ rate, we use $n^{1/4}$. Plot the density of $n^{1/4}\frac{\bar{X}-4}{\sqrt{8}}$ for $E[X] = k$. What do you observe?

(c) Incorrect centering: Using the same samples, instead of centering by subtracting $E[X] = 4$, we center (incorrectly) by subtracting 3. Plot the density of $\sqrt{n}\frac{\bar{X}-3}{\sqrt{8}}$. What do you observe?

3. **Question 3: Estimating π**
It is well known[2] that $\pi = 3.14...$ equals

$$\pi = \frac{\text{Circumference}}{\text{Diameter}} \tag{4}$$

It is also related to the area of a circle which equals πR^2, where R is the radius of the circle. If you only have access to a random number generator, how would you provide an estimate of π?

4. **Question 4: Reverse regression**
Suppose your data is generated according to the following model:

$$Y_i = 1 + 2X_i + U_i, \tag{5}$$

where $X_i \sim N(0,1)$, $U_i \sim N(0,1)$, and X_i and U_i are independent. As we saw in class, the OLS estimator $\hat{\beta}$ is unbiased and consistent: $E[\hat{\beta}] = 2$ and $\hat{\beta}$ will get arbitrarily close 2. In this question we study the reverse regression:

$$X_i = \gamma_0 + \gamma_1 Y_i + V_i, \tag{6}$$

We will focus on the OLS estimator of γ_1.

(a) According the model in (5), what is the value of γ_1?

(b) Fix the sample size in $n = 500$ and generate $S = 100$ different samples of $\{X_i, Y_i\}_{i=1}^{100}$ according to (5). Obtain 100 values of $\hat{\gamma}_1$ and plot the histogram or density. What is the average value of $\hat{\gamma}_1$ across samples?

(c) Is $\hat{\gamma}_1$ consistent? Plot its density with sample sizes $n = 100$, $n = 500$, and $n = 1,000$.

5. **Question 5: Exercise 4.4 in Introductory Econometrics**
Are rent rates influenced by the student population in a college town? Let *rent* be the average

[2] See `https://en.wikipedia.org/wiki/Pi`.

monthly rent paid on rental units in a college town in the United States. Let *pop* denote the total city population, *avginc* the average city income, and *pctstu* the student population as a percent of the total population. One model to test for a relationship is

$$\log(rent) = \beta_0 + \beta_1 \times \log(pop) + \beta_2 \times \log(avginc) + \beta_3 \times pctstu + U \tag{7}$$

(a) State the null hypothesis that size of the student body relative to the population has no ceteris paribus effect on monthly rents. State the alternative that there is an effect.

(b) What signs do you expect for β_1 and β_2?

(c) Using the dataset[3] *rental*, estimate equation (7) for the year 1990. What is wrong with the statement: "A 10% increase in population is associated with about a 6.6% increase in rent"?

(d) Test the hypothesis stated in part (a) at the 1% level.

[3]To access the dataset you need the package 'wooldridge': `https://cran.r-project.org/web/packages/wooldridge/wooldridge.pdf`.