Problem Set 1
MFE 402: Econometrics
Professor Rossi

This is designed to review material on summation, covariance, and the normal distribution.

Question 1

Review the basics of summation notation and covariance formulas. Show that:

a. $\sum_{i=1}^{N}(Y_i - \bar{Y}) = 0$
b. $\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{N}(X_i - \bar{X})Y_i$

Question 2

Define both (and explain the difference between) the expectation of a random variable and the sample average?

Question 3

Review the normal distribution and the mean and variance of a linear combination of two normally distributed random variables. Let $X \sim \mathcal{N}(1, 2)$ and $Y \sim \mathcal{N}(2, 3)$. Note that the second parameter is variance. X and Y are independent. Compute:

a. $\mathbb{E}[3X]$
b. $\text{Var}(3X)$
c. $\text{Var}(2X - 2Y)$ and $\text{Var}(2X + 2Y)$
d. Explain why in part (c) you get the same answer no matter whether you add or subtract. (Your answer should discuss both the coefficient on Y and why independence between X and Y is important.)

Question 4

a. Describe the Central Limit Theorem as simply as you can.
b. Let $X \sim \text{Gamma}(\alpha = 2,\ \beta = 2)$. For the Gamma distribution, α is often called the "shape" parameter, β is often called the "scale" parameter, and the $\mathbb{E}[X] = \alpha\beta$. Plot the density of X. You may find the functions `dgamma()` or `curve()` to be helpful.
c. Let n be the number of draws from that distribution in one sample and r be the number of times we repeat the process of sampling from that distribution. Draw an iid sample of size $n = 10$ from the Gamma(2,2) distribution and calculate the sample average; call this $\bar{X}_n^{(1)}$. Repeat this process r times where $r = 1000$ so that you have $\bar{X}_n^{(1)}, \ldots, \bar{X}_n^{(r)}$. Plot a histogram of these r values and describe what you see. This is the sampling distribution of $\bar{X}_{(n)}$.
d. Repeat part (c) but with $n = 100$. Be sure to produce and describe the histogram.
e. Let's say you were given a dataset for 2,000 people with 2 variables: each person's height and weight. What are the values for n and r in this "real world" example?

Question 5

The normal distribution is often said to have "thin tails" relative to other distributions like the t-distribution. Use random number generation in R to illustrate that a $\mathcal{N}(0, 1)$ distribution has much thinner tails than a t-distribution with 5 degrees of freedom. (Note that `rnorm()` and `rt()` are the functions in R to draw from a normal distribution and a t-distribution.)

Question 6

a. From the Vanguard dataset, compute the standard error of the mean for the `VFIAX` index fund return.
b. For this fund, the mean and the standard error of the mean are almost exactly the same. Why is this a problem for a financial analyst who wants to assess the performance of this fund?
c. Calculate the size of the sample which would be required to reduce the standard error of the mean to 1/10th of the size of the mean return.

Question 7

a. Plot the `VFIAX` index fund return against the `ewretd` (equal-weighted market return) and add the fitted regression line to the plot. You might find the function `abline()` to be helpful.
b. Provide the regression output using the `lmSumm()` function from the `DataAnalytics` package.