POL 345: Quantitative Analysis and Politics

Precept 6

Week 7 (Verzani Chapter 7: 7.1-7.3)

In this precept, we cover the following new command:

- Using `matrix()` to generate a data matrix.

1 Law of Large Numbers

The **Law of Large Numbers** states that the sample mean approaches the population mean as we increase the sample size. We demonstrate this important law by randomly drawing samples of increasing sizes from the normal distribution with a specified mean and standard deviation. According to the Law of Large Numbers, as we increase the sample size, we should see the sample mean and standard deviation approach the specified mean and standard deviation of the distribution.

1. We begin by illustrating the law of large numbers with a simple numerical example. Using similuation, begin with a sample size of one draw from the normal distribution with a mean of 1 and standard deviation of 3. Increase the sample size by an increment to one until you reach 1200 draws. For each sample size, calculate and store the mean and standard deviation. Plot the results.

2 Central Limit Theorem

The **Central Limit Theorem** (CLT) states that the sample mean of identically distributed independent random variables is approximately normally distributed if the sample size is large. This is true for virtually any distribution! We illustrate the Central Limit Theorem using draws from a skewed Binominal population distribution.

1. We begin by using the following formula from the lecture notes: $Z \equiv \frac{\bar{X}-\mathbb{E}(X)}{\sqrt{\text{Var}(X)/n_s}} \sim N(0,1)$, where n_s is the sample size. Also, recall that if $X \sim B(n_t, p)$, then $\mathbb{E}(X) = n_t p$ and $\text{Var}(X) = n_t p(1-p)$, where n_t is the number of trials (population size). Illustrate the Central Limit Theorem by simulating 4000 times an experiment in which you take random draws from a Binomial distribution with a success rate of 0.05 and a size of three for three sample sizes: 20, 100, and 800. Calculate and store the Z-score for the mean of each sample size. **Note:** The command `matrix(data, nrow=X, ncol=Y)` will create a matrix with `X` rows and `Y` columns, and fill those spaces with a vector of `data`.

2. Next, create three stacked histogram of the Z-scores, one for each sample size. Add the density curve from the standard normal distribution to each histogram.