Statistical Inference Course Project

Part 1: Simulation Exercise

Lief Pagalan

January 3, 2023

Overview

In this project, I investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution is simulated in R with `rexp(n, lambda)` where λ is the rate parameter. The mean of the exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. I've set $\lambda = 0.2$ for all of the simulations, and I investigate the distribution of averages of 40 exponentials across 1,000 simulations.

Simulations

Load libraries and set a seed for reproducibility.

```
library(tidyverse)
library(kableExtra)
set.seed(12345)
```

Define the rate parameter as $\lambda = 0.2$. Then set the number of exponentials to 40 and the number of simulations to 1,000.

```
lambda <-  0.2
n      <-   40
sim    <- 1000
```

Simulate the 40 exponentials across 1,000 simulations.

```
means <- NULL
for (i in 1:sim) means[i] <- mean(rexp(n, lambda))
```

Mean of sampling distribution of sample means vs. mean of the exponential distribution.

```
tbl_summary <- tibble(Distribution = c("Sampling Distribution", "Exponential Distribution"),
                      Mean         = c(mean(means),             1/lambda))
```

Variance of sampling distribution of sample means vs. variance of the exponential distribution.

```
tbl_summary <- left_join(tbl_summary,
  tibble(Distribution         = c("Sampling Distribution",  "Exponential Distribution"),
         Variance             = c(sd(means)^2,              ((1/lambda)/sqrt(n))^2),
         `Standard Deviation` = c(sd(means),                 (1/lambda)/sqrt(n))))
```

Table 1: Mean and Variance of Sampling Distribution vs. Exponential Distribution

Distribution	Mean	Variance	Standard Deviation
Sampling Distribution	4.9720	0.5954	0.7716
Exponential Distribution	5.0000	0.6250	0.7906

Sample Mean versus Theoretical Mean

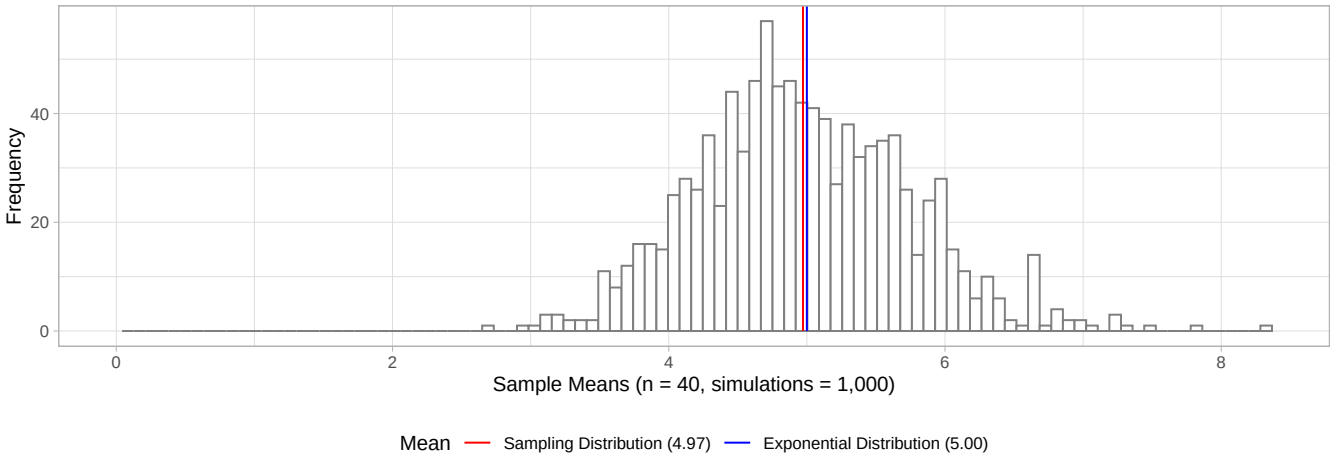

Figure 1: Sampling Distribution Mean vs. Exponential Distribution Mean

Figure 1 shows the sampling distribution of the sample means of 40 exponentials across 1,000 simulations. The average of the sample means was 4.97, which approximates the theoretical mean of the exponential distribution, 5.00 (see Table 1). As the number of simulations increase, the average of the sample means will tend towards the theoretical distribution mean.

Sample Variance versus Theoretical Variance

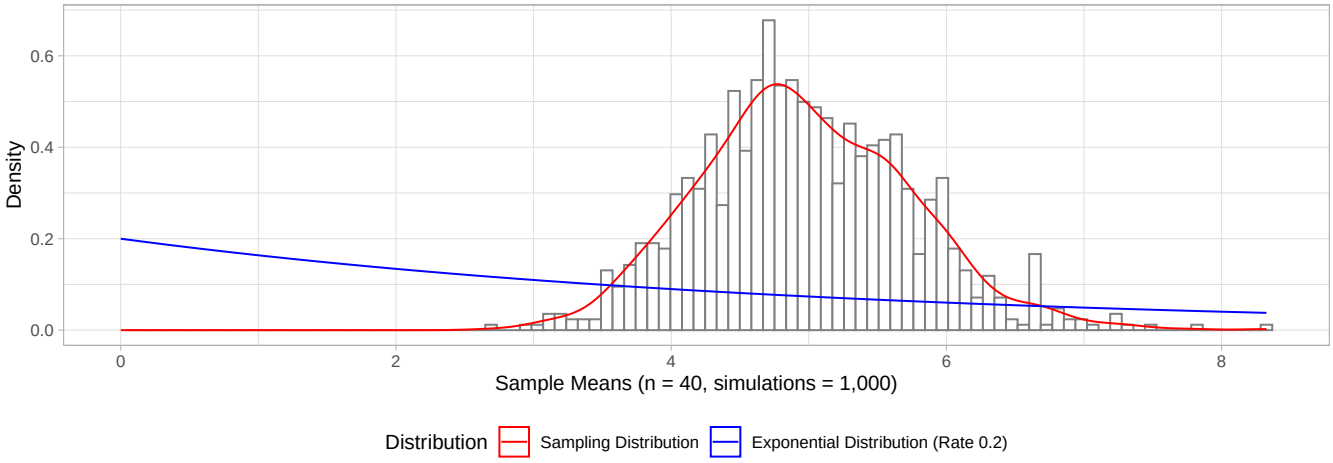

Figure 2: Sampling Distribution vs. Exponential Distribution

Figure 2 shows the difference in shape between the sampling distribution (red) and the theoretical exponential distribution (blue). Despite the difference in distribution shapes, the variance of the sampling distribution of sample

means is 0.60 with standard deviation 0.77, which closely approximates the theoretical exponential distribution variance 0.62 with standard deviation 0.79 (see see Table 1). As the number of simulations increase, the variance of the sampling distribution will tend towards the theoretical distribution variance.

Sampling, Exponential, and Normal Distributions

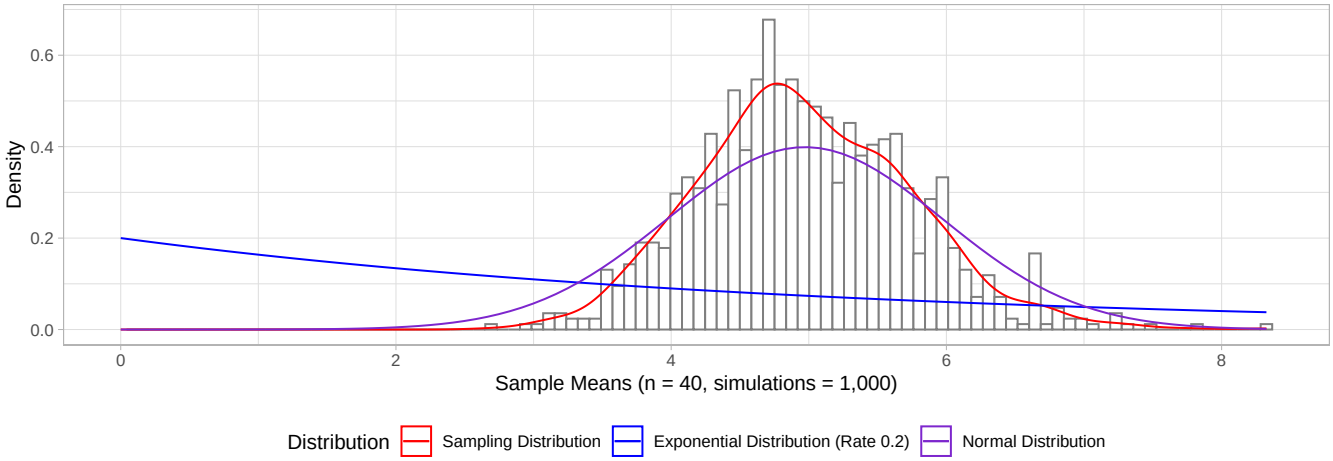

Figure 3: Sampling Distribution vs. Exponential Distribution vs. Normal Distribution

Figure 3 shows that the sampling distribution of sample means (red) approximates the normal distribution (purple) with $n \geq 30$. This demonstrate the central limit theorem, which states that if you have a population with defined mean μ and standard deviation σ, and you take *sufficiently large random samples from the population with replacement multiple times*, then the sampling distribution of the *sample means* will be approximately normally distributed regardless of the shape of the original population distribution. As sample size n increases, the sampling distribution of sample means will tend towards the normal distribution.