

Exponential Distribution Project and Tooth Growth Analysis

Raymond Dineen

9/21/2020

Part 1:

Overview

In this first section, we will do a bit of exploration of the exponential distribution. We will run simulations of sampling from the exponential distribution and show how the results of these simulations demonstrate the Central Limit Theorem.

Simulations

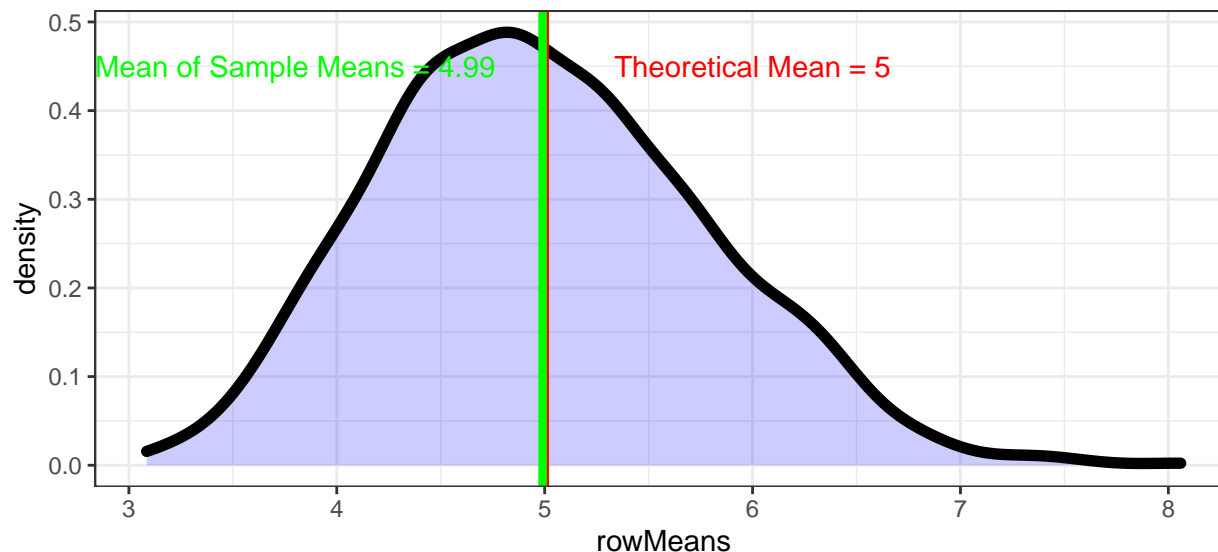
The first thing we need to do to is to simulate some random data from the exponential distribution. We will create 1000 samples of 40 random exponentials using the rate of $\lambda = 0.2$

```
B <- 1000
n <- 40
lambda <- 0.2
set.seed(1)
# Creating a matrix of 1000 rows of 40 random exponentials
sim <- matrix(rexp(n * B, rate = lambda), B, n)
```

Sample Mean versus Theoretical Mean

The theoretical mean of the exponential distribution is $\frac{1}{\lambda}$. In our case, our theoretical mean is $\frac{1}{(1/5)} = 5$ and we expect the distribution of our sampled means to be centered around this value. We also expect the mean of our sampled means to be very close to our theoretical mean. The density plot below shows both of our expectations to be true.

```
rowMeans <- apply(sim, 1, mean)
theoMean <- 5
sampleMean <- mean(rowMeans)
library(dplyr)
library(ggplot2)
data.frame(rowMeans = rowMeans) %>% ggplot(aes(x = rowMeans)) +
  geom_density(size = 2, alpha = 0.2, fill = "blue", size = 2) +
  geom_vline(xintercept = theoMean, color = "red", size = 1.5) +
  geom_vline(xintercept = sampleMean, color = "green", size = 1.5) +
  annotate("text", 6, 0.45, label = "Theoretical Mean = 5", color = "red") +
  annotate("text", 3.8, 0.45, label = "Mean of Sample Means = 4.99", color = "green") +
  theme_bw()
```



Sample Variance versus Theoretical Variance

The sample variance of our observed means can be calculated with $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$. The `var()` function in R does this for us. Our theoretical variance is σ^2/n which in our case is $5^2/1000 = 0.625$. We expect our sample variance and theoretical variance to be close to each other due to the the high amount of simulations we did. Our expectations are met as shown in the code below.

```
theoVar <- 0.625
sampleVar <- var(rowMeans)
sampleVar
```

```
## [1] 0.6177072
```

Distribution

The Central Limit Theorem tells us that the the distribution of our properly normalized sample means becomes standard normal as the sample size increases. To show this we first must normalize our data (subtract the mean divide by the standard deviation). The `scale()` function in R does this for us. We then plot the the normalized data against a normal distribution curve, and notice how similar the two distributions are. It is also worth noticing how different the distribution of 1000 random exponentials is compared to the distribution of 1000 means of 40 random exponentials. Even though the means are taken the from exponential an exponentially distributed population, they follow a normal distribution due to the Central Limit Theorem.

```
library(ggpubr)
normalized <- scale(rowMeans)
cltdemo <- data.frame(normalized = normalized) %>% ggplot(aes(normalized)) +
  geom_density(size = 2, alpha = 0.2, fill = "blue") +
  stat_function(fun = dnorm, n = 1000, color = "red", size = 1.5) +
  annotate("text", 2.5, 0.35, label = "Normal Distribution", color = "red", size = 3) +
  labs(x = "distribution", title = "Normalized means vs. Normal Distribution") +
  theme_bw() +
```

```

theme(plot.title = element_text(size = 10))

rexpdemo <- data.frame(rexps = rexp(1000, rate = 0.2)) %>% ggplot(aes(rexps)) +
  geom_density(size = 2, fill = "blue", alpha = 0.2) +
  labs(x = "", title = "1000 Random Exponentials") +
  theme_bw()

ggarrange(cltdemo, rexpdemo, nrow = 1)

```

