Statistical Inference Assignment 1 - Question 1

Lachlan Glascott

April 2020

Question 1

Overview

This project is to investigate the exponential distribution in R and compare it with the Central Limit Theorem. The key questions are to:

```
 1.  Show the sample mean and compare it to the theoretical mean of the distribution.
 2.  Show how variable the sample is (via variance) and compare it to the theoretical variance of the
 3.  Show that the distribution is approximately normal.
```

Simulation

As per the exercise instructions, the exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter: - The theoretical mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. - Set lambda = 0.2 for all of the simulations. - The distribution of averages of 40 exponentials is investigated over 1000 simulations.

The code chunk below performs the simulation and calculates the sample mean across each of the simulations.

```
simulations <- 1000
sample_size <- 40
lambda <- 0.2

exp_means = NULL
for (i in 1 : simulations) exp_means = c(exp_means, mean(rexp(sample_size, lambda)))

exp_means_df <- as.data.frame(exp_means)
```

Sample Mean versus Theoretical Mean

The sample mean of the distrubution is very similar to the theoretical mean of 5. This is shown in the plot below wher the center of the distibution for the sample is blue and the theoretical centre is red.

```
sample_mean <- mean(exp_means)
theoretical_mean <- 1/lambda

sample_mean
```

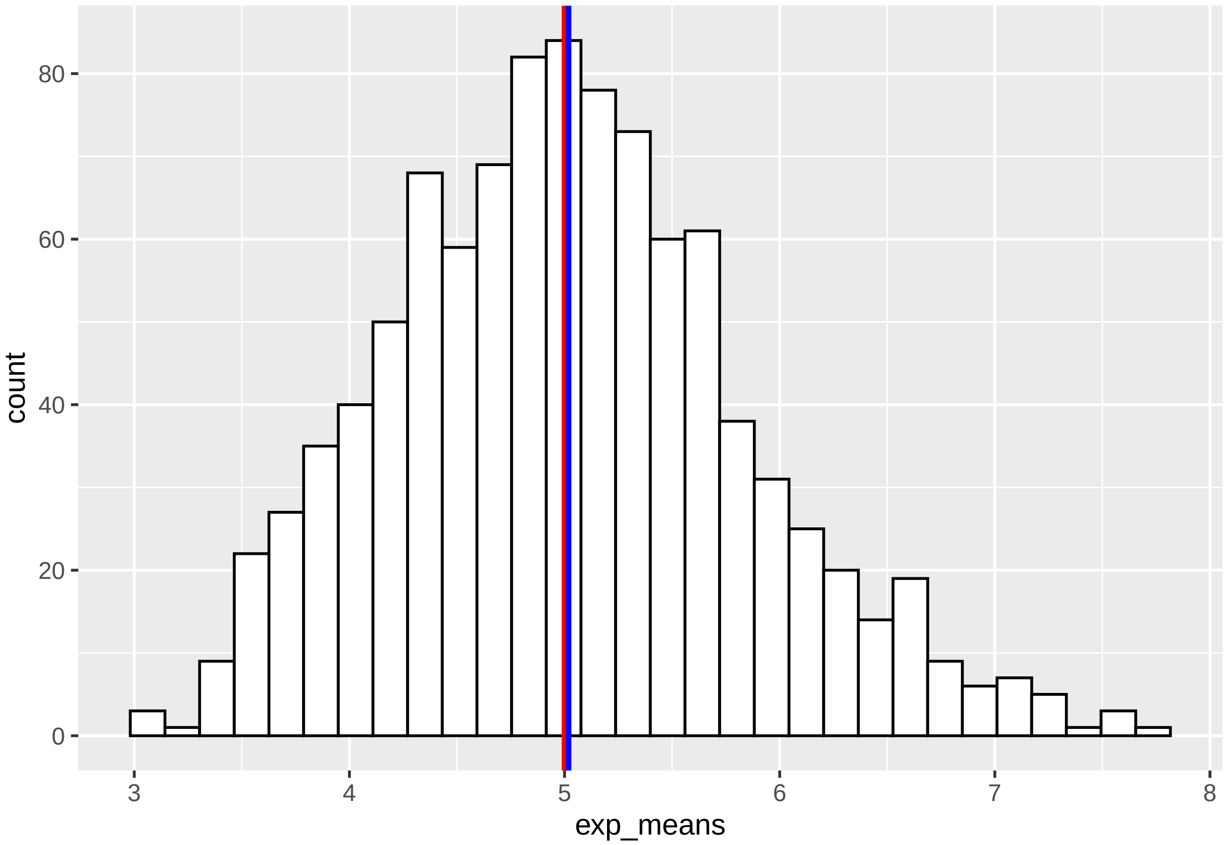
```
## [1] 5.018351
```

```
theoretical_mean
```

```
## [1] 5
```

```
ggplot(exp_means_df, aes(x=exp_means)) +
  geom_histogram(color="black", fill="white") +
  geom_vline(xintercept=theoretical_mean, color="red", size=1) +
  geom_vline(xintercept=sample_mean, color="blue", size=1);
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Sample Variance versus Theoretical Variance

The theoretical variance is calculated as the standard deviation squared devided by the sample size.

The sample variance is the variance between the expected means generated from the simulations.

```
theoretical_variance = ((1/lambda)^2/sample_size)
sample_variance <- var(exp_means)

theoretical_variance
```

```
## [1] 0.625
```

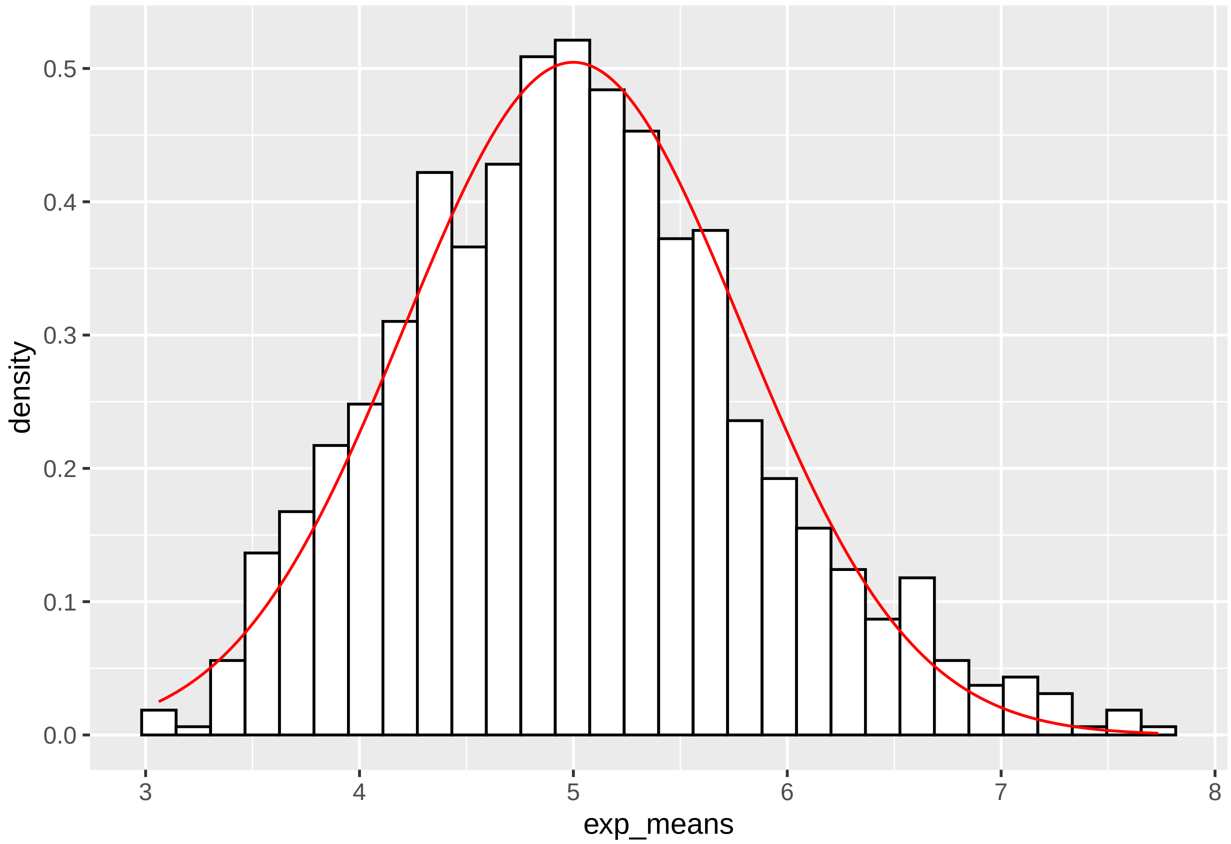
```
sample_variance
```

```
## [1] 0.6782607
```

The variance of the sample distribution is very similar to the theoretical variance is 0.625.

```
ggplot(exp_means_df, aes(x = exp_means)) +
  geom_histogram(color="black", fill="white", aes(y = ..density..)) +
  stat_function(fun = dnorm, n = 1000, args = list(theoretical_mean, sqrt(theoretical_variance)), color
```

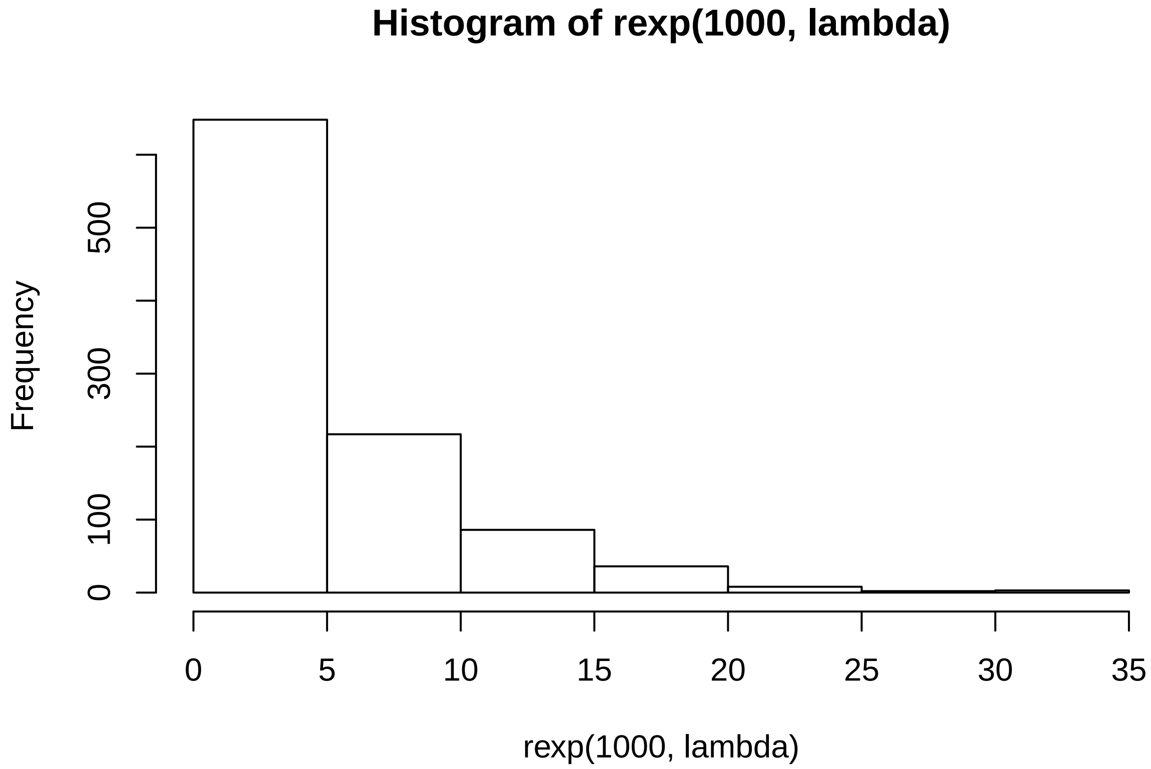
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



A large collection of random exponentials is not Gaussian (refer to Appendix), however and the distribution of a large collection of averages of 40 exponentials is Gaussian as demonstrated by the bell shaped distibution of the histogram of averages. The shape of the distribution is very similar to a a distribution of values randonly generated from a normal distibution - see red line in chart above. This shows that the distribution is approximately normal.

Appendix

```
hist(rexp(1000, lambda))
```


Histogram of rexp(1000, lambda)
Frequency
rexp(1000, lambda)