

Statistical Inference: Week 4 Project Part I

Author: Alexander Johannes

Date: 10 August 2021

Link to my Statistical Inference repository: [My Repo](#)

Overview

This report details the process of carrying out a simulation using the exponential distribution.

Part I: Simulation Exercise

First, we define some useful variables and load a plotting library.

```
lambda <- .2
samples <- 40
simulations <- 1000
set.seed(7)
library(ggplot2)
```

Next, let's take a sample of 40 exponential distributions of `rate = lambda`, using 1000 simulations, in order to determine the sampling distribution of the mean.

```
simulatedData <- replicate(simulations, rexp(samples, rate = lambda))
```

So, we have a 40 x 1000 matrix containing our simulated data. Let's compute the mean of the sampling distribution and its variability:

```
meansData <- apply(simulatedData, 2, mean)
sampleMean <- mean(meansData)
sampleMean
```

```
## [1] 4.983294
```

```
theoreticalMean <- 1/lambda
```

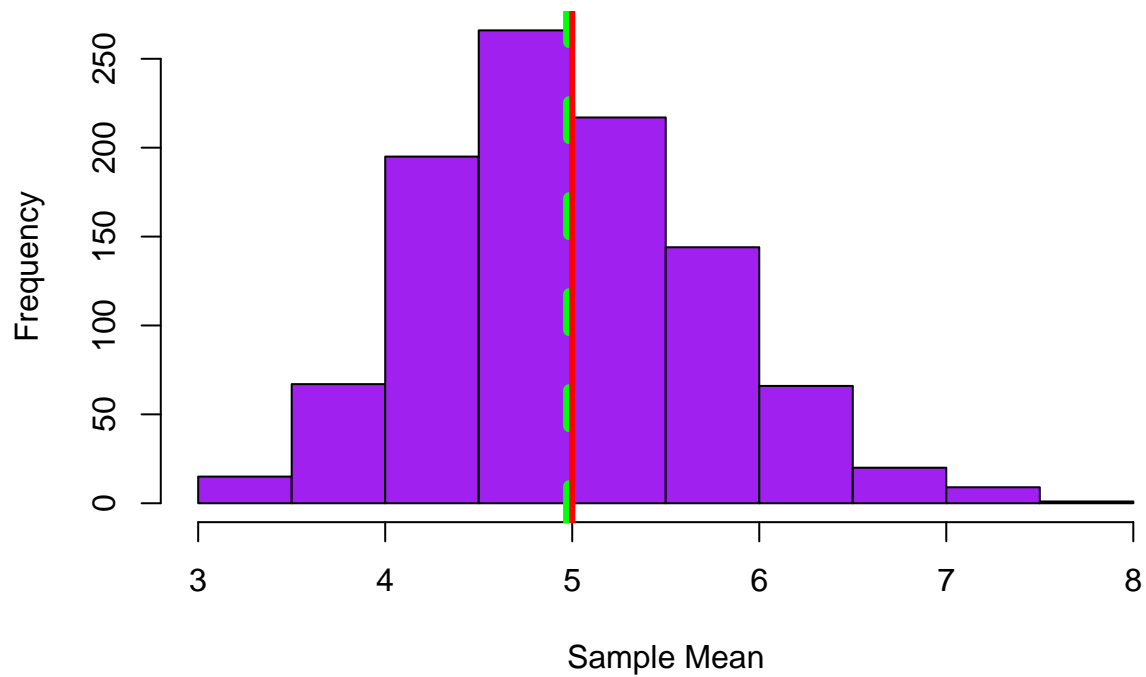
```
sampleVar <- var(meansData)
sampleVar
```

```
## [1] 0.5792547
```

```
theoreticalVar <- (1/lambda)^2 / samples
```

Visualizing our results yield:

Histogram of 1000 means of 40 sample exponentials

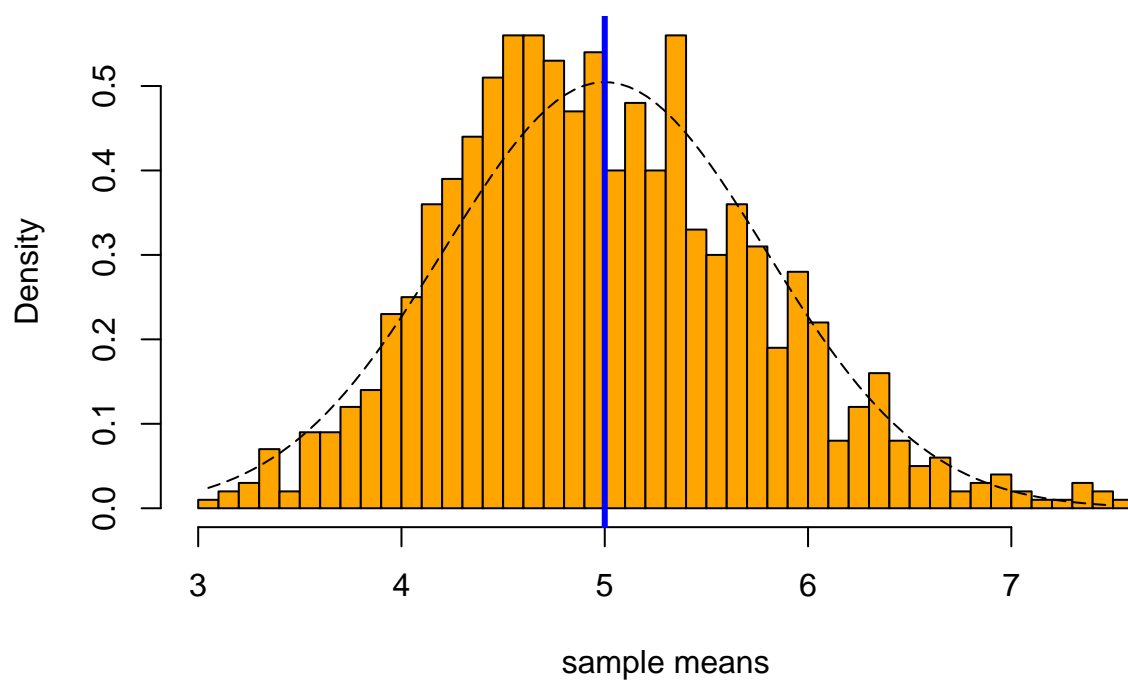


Clearly, the sample mean (4.98) and theoretical mean (5) from the simulated exponential distributions are approximately equal.

The sample variance is .58, with the theoretical variance being .625. We expect some variability in our random sample and these values are still fairly similar.

Finally, let's visualize the sampling distribution.

Distribution of the sample mean



Due to the Central Limit theorem, we see that the distribution of the sample mean is approximately normal. The bell-shaped curve superimposed on the histogram illustrates how the distribution roughly resembles the density function of a normal random variable, centered around the mean (5).