Course-Project-1

Part 1: Simulation Exercise

Overview

This section contains simulations of the exponential distribution, measurement and comparison of sample parameters to theoretical expectations, and a demonstration that the distribution of these sample parameters is roughly Gaussian using both visual and statistical methods.

Simulations

As instructed in the assignment, simulate the exponential distribution using R's rexp() function, using samples of 40 exponentials, and simulating 1000 samples.

```
data <- apply(matrix(1:1000), 1, function(x) rexp(40, 0.2))
```

This creates a 40x1000 matrix, where each of the 1000 columns contains a sample of 40 observations taken from an exponential distribution with lambda=0.2. Next, take the mean of each sample of 40 by applying the mean function column-wise.

```
meansDist <- apply(data, 2, mean)
```

Sample Mean vs Theoretical Mean

Each of these 40-observation samples has its own sample mean. To get an idea of what that sample mean tends to be, find the mean of the sample means:

```
mean(meansDist)
```

```
## [1] 4.999938
```

In theory, the sample mean should be centered around 1/lambda = 1/0.2 = 5. Thus, the mean observed here closely matches our theoretical expectations. Some small difference is to be expected, but this difference will approach zero as the number of observations per sample or the number of samples approach infinity.

Sample Variance vs Theoretical Variance

Each sample also has a sample standard deviation. Calculate the mean sample standard deviation:

```
var(meansDist)
```

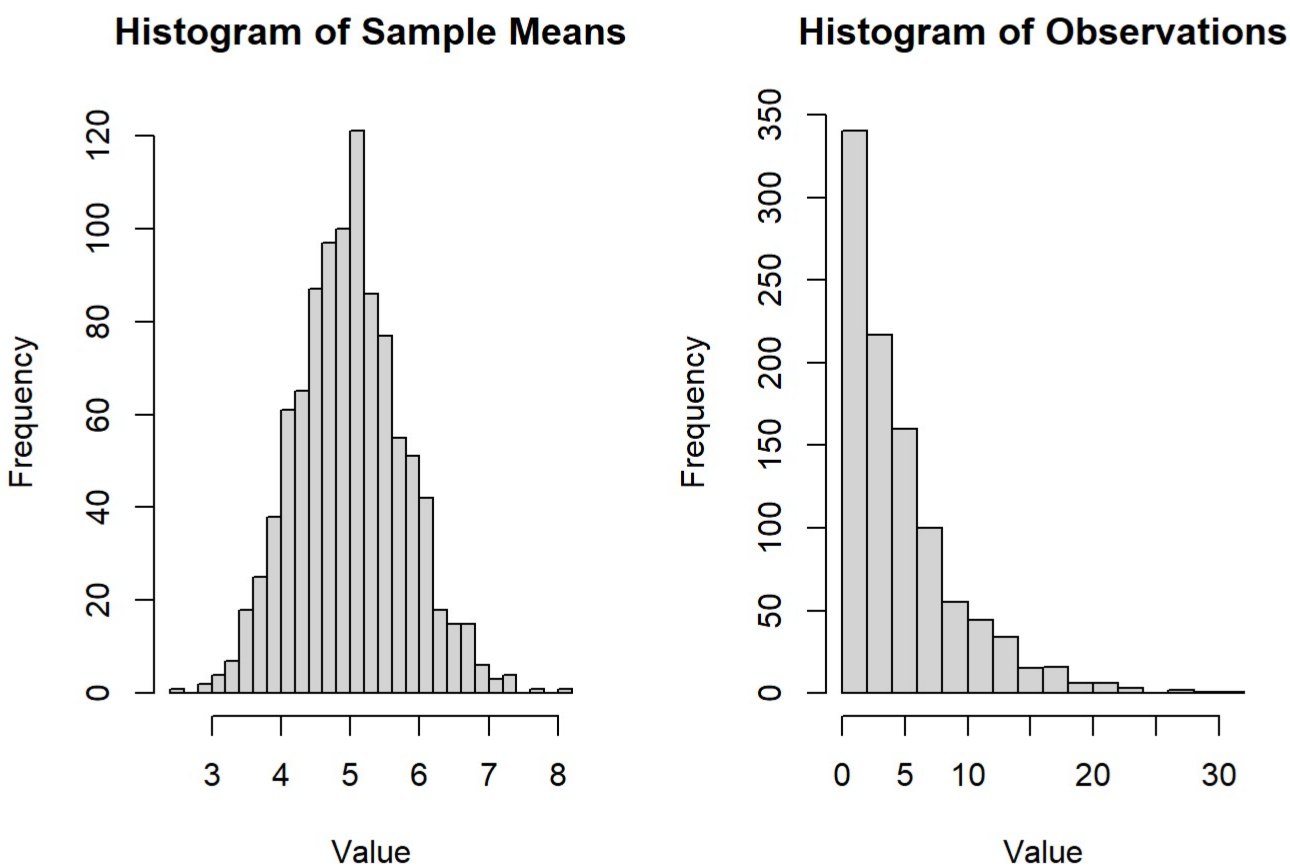
```
## [1] 0.6131281
```

This distribution should have a variance close to s^2/sqrt(n). Since the sample variance is expected to be 1/lambda = 5, and each sample has n=40, the variance of meansDist should be about 5^2/40 = 25/40 = 0.625. Thus, the result observed here matches theoretical expectations. Some difference is expected, but this will go to zero as n goes to infinity.

Distribution

Visualize the distribution of sample means and check if it is roughly normal:

```
par(mfrow=c(1,2))
hist(apply(data, 2, mean), breaks=20, xlab="Value", ylab="Frequency", main="Histogram of Sample Means")
hist(data[1,], breaks=20, xlab="Value", ylab="Frequency", main="Histogram of Observations")
```



While the histogram of original observations on the right is exponential, the Central Limit Theorem means that the distribution of sample means on the left is roughly normal. Besides this visual test, further check normality using the Shapiro-Wilk Normality Test, with the R function shapiro.test().

```
shapiro.test(data[1,])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data[1, ]
## W = 0.83685, p-value < 2.2e-16
```

This returns an extremely low p-value: thus, the distribution is almost certainly normal.