Statistical Inference Course Project

Chompasaurus for Coursera

Overview

In this project we are using R to explore the exponential distribution and compare it with the expectations of the Central Limit Theorem (CLT), which expects the distribution of sample means from the exponential distribution to approach the normal distribution as n increases.

Simulations

The samples we're comparing to CLT expected values are the means of 1000 simulated sets of 40 values, using the rate parameter lambda=0.2 and R's rexp() function.

```r
set.seed(9) # because good habits and reproducability
lambda <- 0.2
n_rexp <- 40 # number of points in each simulated set
n_sims <- 1000 # number of simulations
mean_theoretical <- 1/lambda # theoretical mean according to CLT
variance_theoretical <- 1/(n_rexp*lambda^2) # theoretical variance according to CLT
standardDeviation_theoretical <- sqrt(variance_theoretical)
```

```r
means_exponential <- NULL # create variable. using only means so no need to store all sim data
for(i in 1:n_sims){ means_exponential <- rbind(means_exponential, mean(rexp(n_rexp, lambda)))}
mean_exponential <- mean(means_exponential)
variance_exponential <- var(means_exponential)
standardDeviation_exponential <- sqrt(variance_exponential)
```

Findings

1. Sample Mean versus Theoretical Mean

Comparison between the center of the simulated distribution data and the theoretical mean 1/lambda show close similarity.

```r
print(mean_theoretical)
```

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

```r
print(mean_exponential)
```

```
## [1] 4.993253
```

2. Sample Variance versus Theoretical Variance

Again, comparing the variance of the simulated vector of means to the theorized variance 1/(n*lambda^2) shows that our data approaches the theoretical value.

```
print(variance_exponential)
```

```
##          [,1]
## [1,] 0.666388
```

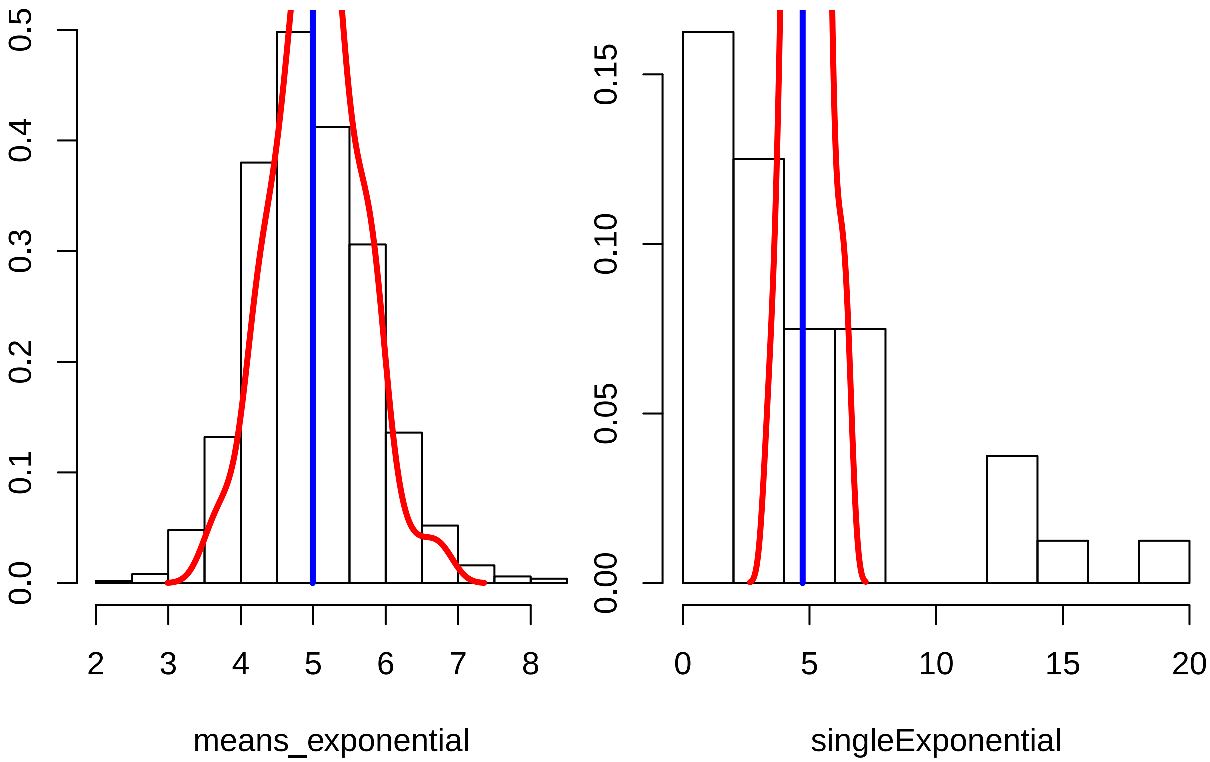
```
print(variance_theoretical)
```

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

3. Distribution

Compare the distributions of: one set of 40 exponentially distributed variables, and 1000 means. We see that the distribution of means is clustered around their mean(blue line), where the distribution of one sample is scattered about. The red normal distribution curve fits the distribution of means.

```
singleExponential <- rexp(40, 0.2)
par(mfrow = c(1,2), mar = c(4,3,3,0))
hist(means_exponential, main = "", freq = FALSE)
plotLineX <- matrix(mean_exponential, 10,1)
lines(plotLineX, y = c(seq(0,0.9,0.1)), col = "blue", lwd = "3")
lines(density(rnorm(100,5,0.625)), col = "red", lwd = "3")
par(mar = c(4,2,3,0))
hist(singleExponential, main = "", ylab = "", freq = FALSE)
plotLineX <- matrix(mean(singleExponential),10,1)
lines(plotLineX, y = c(seq(0,0.9,0.1)), col = "blue", lwd = "3")
lines(density(rnorm(100,5,0.625)), col = "red", lwd = "3")
```



0.0
0.1
0.2
0.3
0.4
0.5
2
3
4
5
6
7
8
means_exponential
0.00
0.05
0.10
0.15
0
5
10
15
20
singleExponential