Statistical Inference Course Project. Part 1: Simulation Exercise

Vitalii Zhukov

Project Description

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter.

1. Simulation

Initializations, Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. (Note that you will need to do a thousand simulations.)

```
lambda <- 0.2
set.seed(1234)
sim.data <- data.frame(ncol = 2,nrow = 1000)
names(sim.data) <- c("simulation.run","mean")

for (i in 1:1000) {
  sim.data[i,1] <- i
  sim.data[i,2] <- mean(rexp(40,lambda))
}
```

The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda

```
t.mean <- 1 / lambda
s.mean <- mean(sim.data$mean)
paste("Theoretical mean = ", t.mean)
```

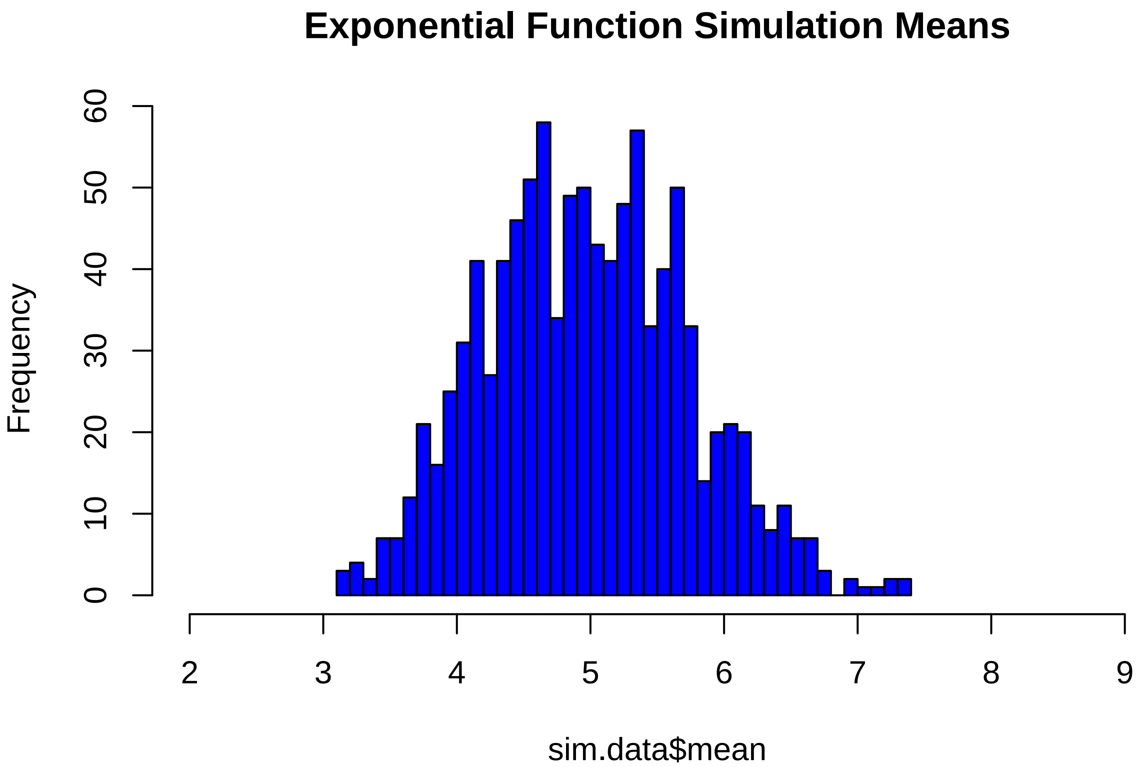
```
## [1] "Theoretical mean =  5"
```

```
paste("Simulated mean =  ", s.mean)
```

```
## [1] "Simulated mean =   4.97423877125153"
```

Histogram Exponential function simulation means

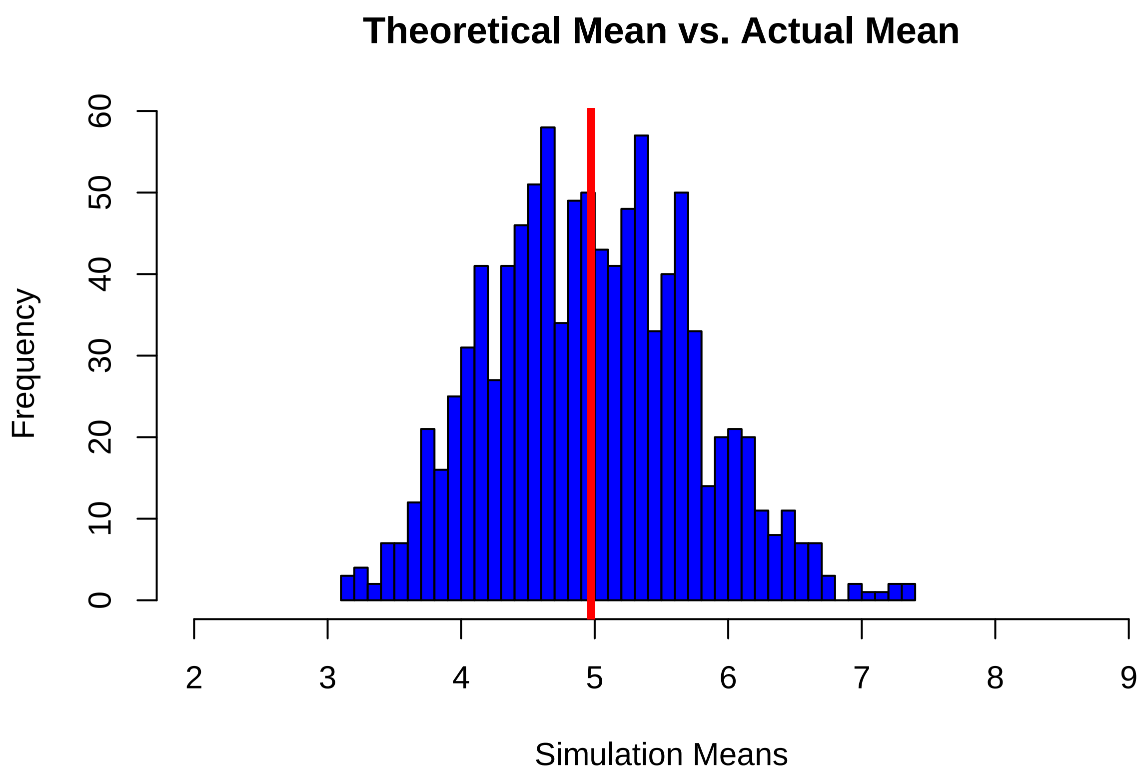
```
hist(sim.data$mean, breaks=40, xlim = c(2,9), main="Exponential Function Simulation Means", col = "blue
```



2. Sample Mean vs Theoretical Mean

The mean of the exponential distribution is 1/lambda. In this case, lambda is 0.2. Therefore, the theoretical mean should result as 5 (i.e. 1 / 0.2). Lets see if that holds true. (plot histogram of the sample means)

```
hist(sim.data$mean, col="blue", main="Theoretical Mean vs. Actual Mean",
     xlim = c(2,9),breaks=40, xlab = "Simulation Means")
abline(v=mean(sim.data$mean), lwd="4", col="red")
```




```r
mean(sim.data$mean)
```

```
## [1] 4.974239
```

3. Sample Variance vs Theoretical Variance

The standard deviation of the exponential distribution is (1/lambda) / sqrt(n). Next, we'll see if this matches our simulations. Theoretical standard deviation vs. simulation standard deviation.

```r
paste("Theoretical standard deviation: ", round( (1/lambda)/sqrt(40) ,4))
```

```
## [1] "Theoretical standard deviation:  0.7906"
```

```r
paste("Practical standard deviation: ", round(sd(sim.data$mean) ,4))
```

```
## [1] "Practical standard deviation:  0.7554"
```

```r
paste("Theoretical variance: ", round( ((1/lambda)/sqrt(40))^2 ,4))
```

```
## [1] "Theoretical variance:  0.625"
```

```r
paste("Practical variance: ", round(sd(sim.data$mean)^2 ,4))
```

```
## [1] "Practical variance:  0.5707"
```

4. Distribution

Finally, we'll investigate whether the exponential distribution is approximately normal. Due to the Central Limit Theorem, the means of the sample simulations should follow a normal distribution. - General Plot with ditribution curve drawn

```r
hist(sim.data$mean, prob=TRUE, col="blue", main="Exponential Function Simulation Means",
     breaks=40, xlim=c(2,9), xlab = "Simulation Means")
lines(density(sim.data$mean), lwd=3, col="yellow")
x <- seq(min(sim.data$mean), max(sim.data$mean), length=2*40)
y <- dnorm(x, mean=1/lambda, sd=sqrt(((1/lambda)/sqrt(40))^2))
lines(x, y, pch=22, col="green", lwd=2, lty = 2)
```