Statistical Inference Part 1

Simon Grasdal

2024-12-11

Part 1: Simulation Exercise

In this project I will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution will be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. I will investigate the distribution of averages of 40 exponentials. Note that I will do a thousand simulations to complete this.

1. Mean Comparision Sample mean in comparison with the theoretical

mean of distributions.

```
# Sample Mean
sampleMean <- mean(mean_sim_data) # Mean of sample means
print (paste("Sample Mean = ", sampleMean))
```

```
## [1] "Sample Mean =  5.02010698674351"
```

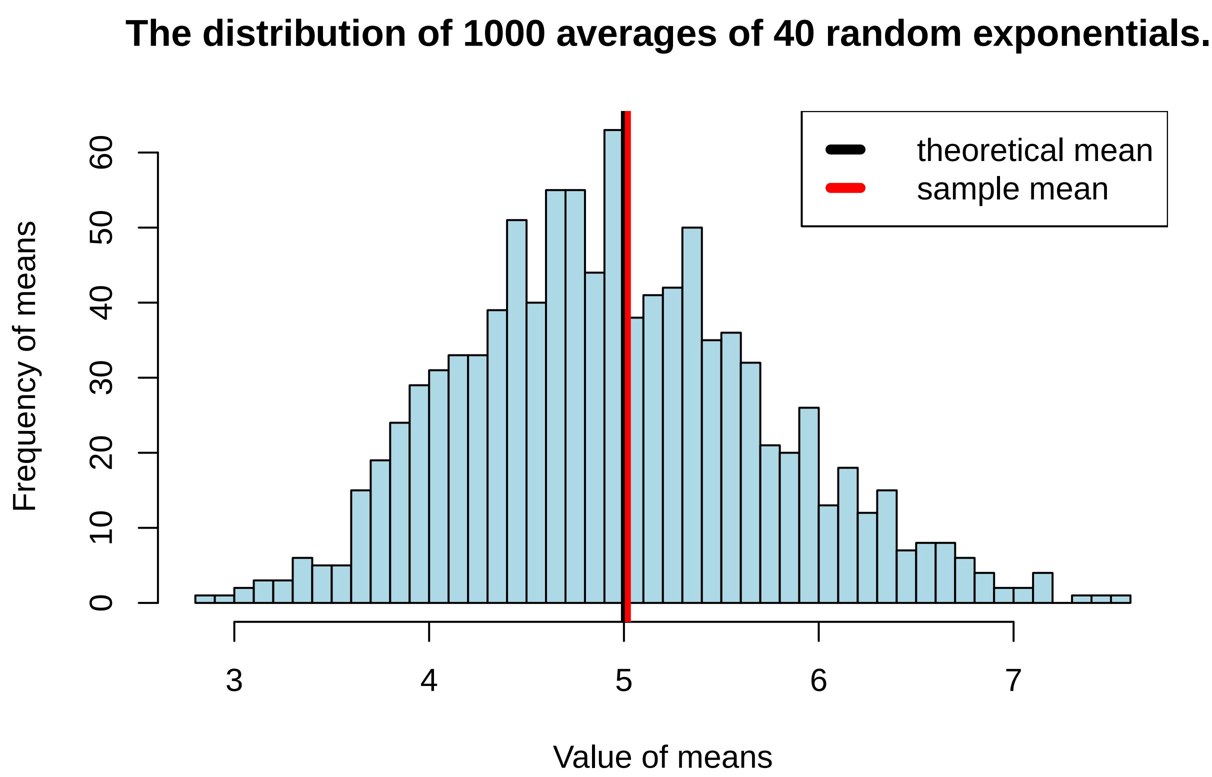
```
# Theoretical Mean
# the expected mean of the exponential distribution of rate = 1/lambda
theoretical_mean <- (1/lambda)
print (paste("Theoretical Mean = ", theoretical_mean))
```

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

Visualization of comparison between theoretical and sample means across 1000 replications.

```
lambda <- 0.2
sim_data <- matrix(rexp(1000*40, lambda),
                   nrow = 1000, ncol = 40)

dist_mean <- apply(sim_data, 1, mean)

# Histogram shows differences
hist(dist_mean, breaks = 40,
     main = " The distribution of 1000 averages of 40 random exponentials.",
     xlab = "Value of means",
     ylab = "Frequency of means", col = "lightblue")
```

```r
abline(v=1/lambda, lty = 1,
       lwd=3, col = "black")
abline(v=sampleMean, lty = 1,
       lwd=3, col = "red")
legend("topright", lty = 2,
       lwd = 5, col = c("black", "red"), legend = c("theoretical mean", "sample mean"))
```



2. Sample Variance and Theoriteical Variance The Sample mean in

comparison with the theoretical variance of the distribution.

Calculating the theoretical and sample variance

```r
sample_d <- sd(mean_sim_data)
sample_v <- sample_d^2
theoretical_d <- (1/lambda)/sqrt(n)
theoretical_v <- ((1/lambda)*(1/sqrt(n)))^2
sample_v
```

```
## [1] 0.6261326
```

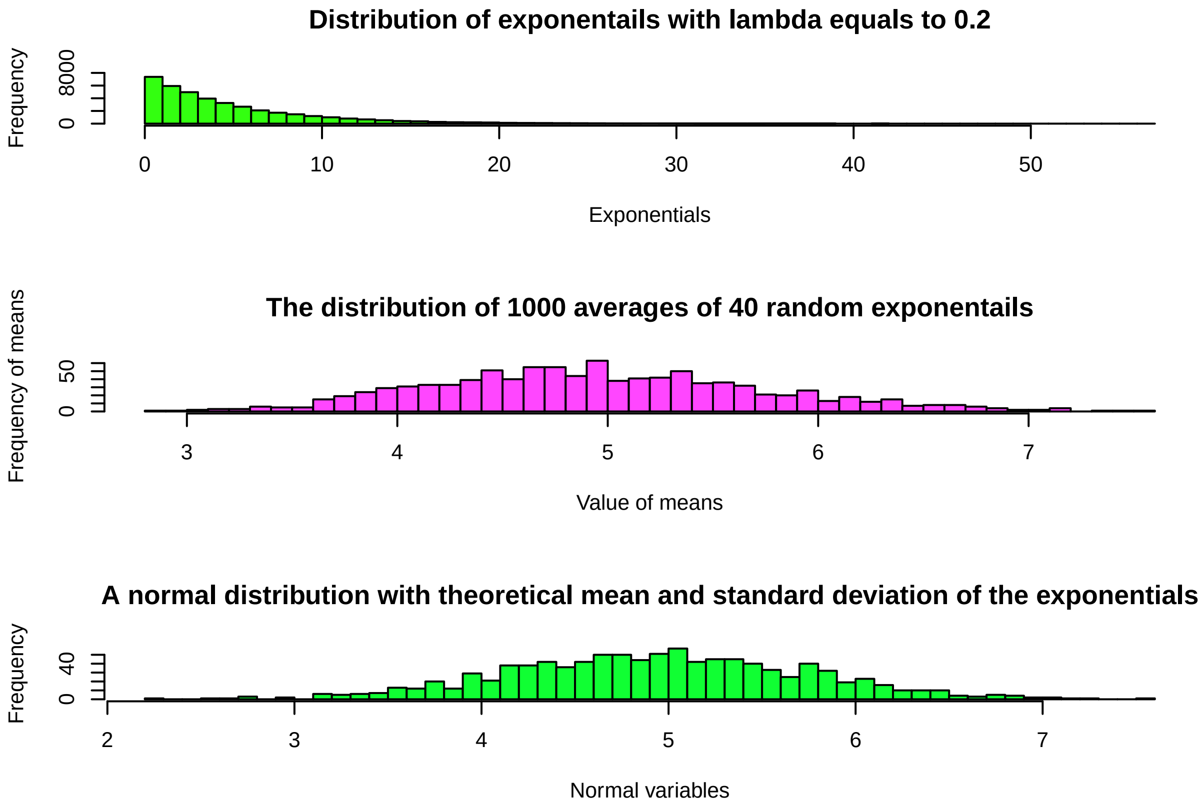
```
theoretical_v
```

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

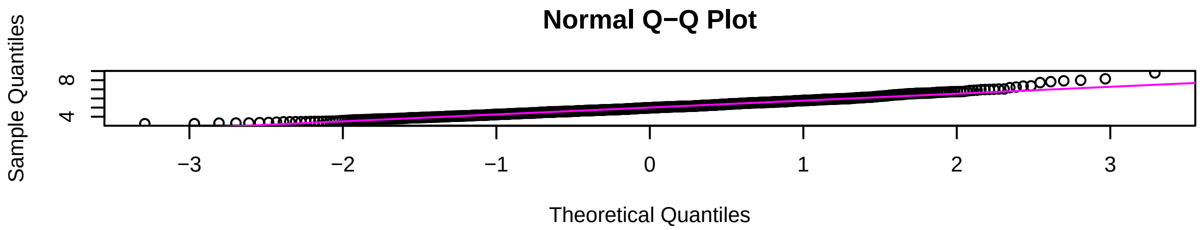
The variance for the sample is .626 vs the theoretical variance which is .625. They are very close but not quite the same. ## 3. Distribution Normality Show that the distribution is approximately normal

Histogram with Density and sample means:

```
par(mfrow = c(3,1))

hist(sim_data, breaks = 50,
     main = "Distribution of exponentails with lambda equals to 0.2",
     xlab = "Exponentials",
     ylim = c(0,8000),
     col = "#33ff10")

hist(dist_mean, breaks = 40,
     main = "The distribution of 1000 averages of 40 random exponentails",
     xlab = "Value of means", ylab = "Frequency of means", col="#ff46ff")

normal_sim <- rnorm(1000, mean = mean(dist_mean), sd = sd(dist_mean))
hist(normal_sim, breaks = 40,
     main = "A normal distribution with theoretical mean and standard deviation of the exponentials",
     xlab = "Normal variables",
     col = "#10ff33")
```




```
qqnorm(mean_sim_data)
 qqline(mean_sim_data, col = "magenta")
```



The first histogram is the distribution of exponential with lambda equals to 0.2. The second histogram is the distribution of 100 averages of 40 random exponential. The third histogram is a real normal distribution with a mean and standard deviation equals to the second histograms. Comparing the first with the second histogram, we can see the distribution becomes normal as the means were taken from each groups. It is a result of the Central Limit Theorem. Comparing the second and the third histogram, we can see the distribution with the same mean and standard deviation.The qqplot also indicates a normal distribution within the sample data.