Statistical Inference Course Project Part 1

Part 1. Simulation Exercise

Overview:

This report describes an investigation into the exponential distribution and the Central Limit Theorem. Simulations will be carried out to illustrate the properties of the distribution.

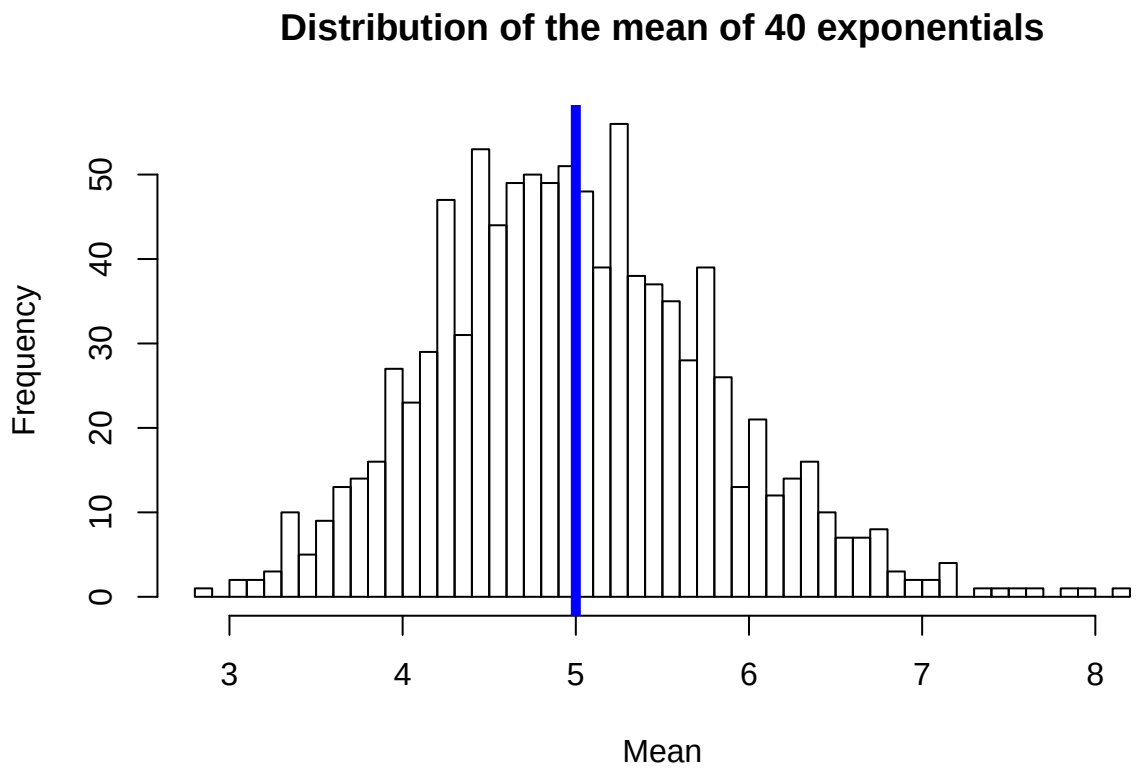
Simulations:

Set up to run 1000 simulations of 40 randomly generated exponentials:

```
lambda <- 0.2
sim <- rexp(40000,lambda)
mat <- matrix(sim, nrow = 1000, ncol = 40)
```

Calculate the mean of each exponential set and compare it to the theoretical mean of the distribution:

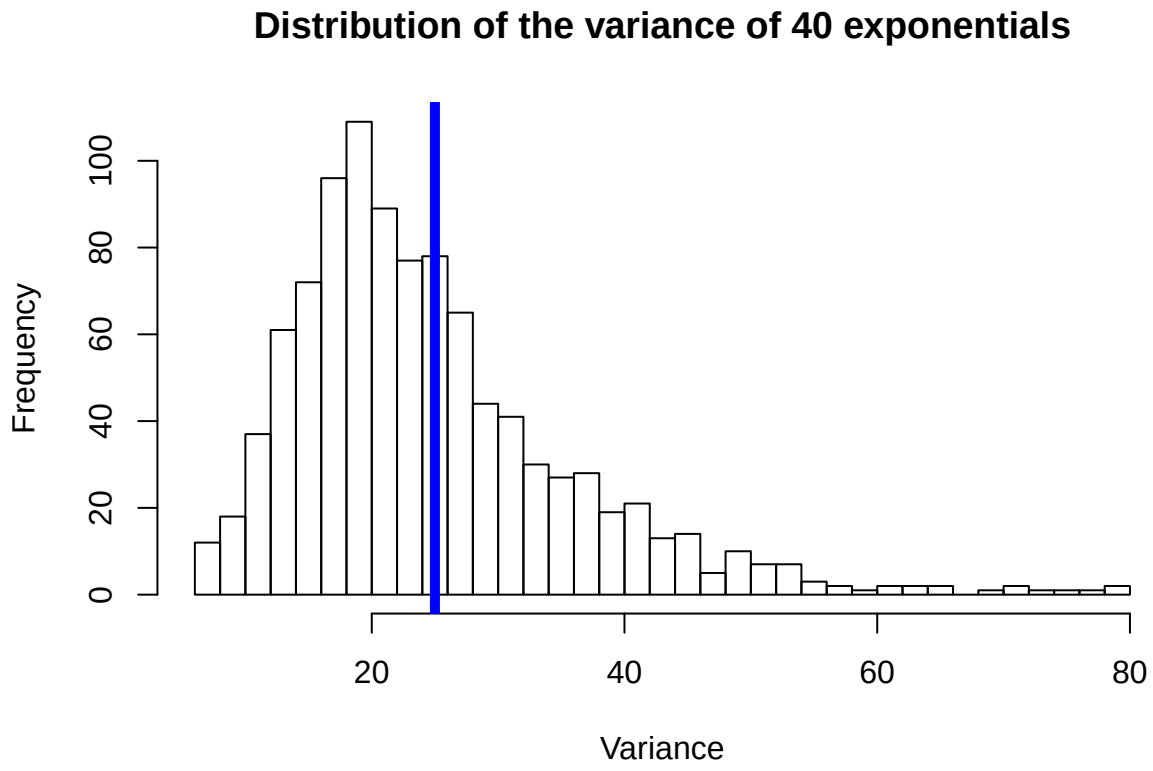
```
means <- apply(mat, 1, mean)
hist(means, breaks = 40, main = "Distribution of the mean of 40 exponentials",
     xlab = "Mean", ylab = "Frequency")
abline(v = 1/lambda, lty = 1, lwd = 5, col = "blue")
```



Calculate the variance of each exponential set and compare it to the theoretical variance of the distribution:

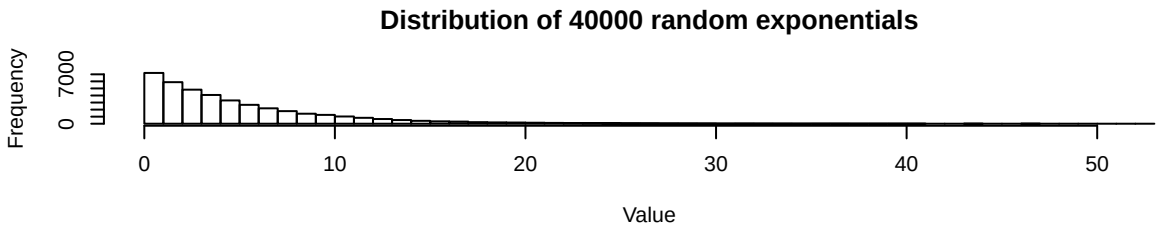
```
vars <- apply(mat, 1, var)
hist(vars, breaks = 40, main = "Distribution of the variance of 40 exponentials",
```

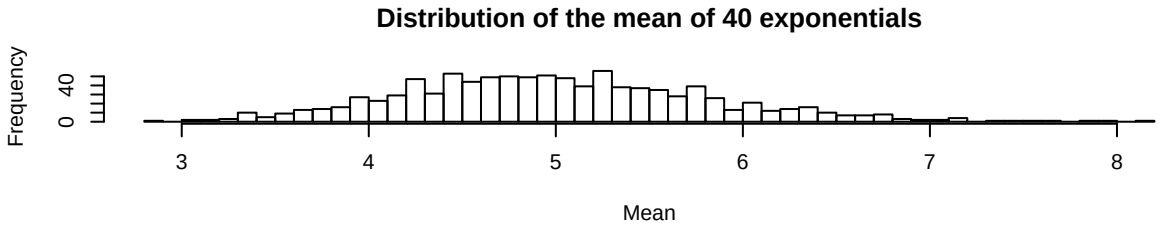
```
     xlab = "Variance", ylab = "Frequency")
abline(v = 1/lambda^2, lty = 1, lwd = 5, col = "blue")
```

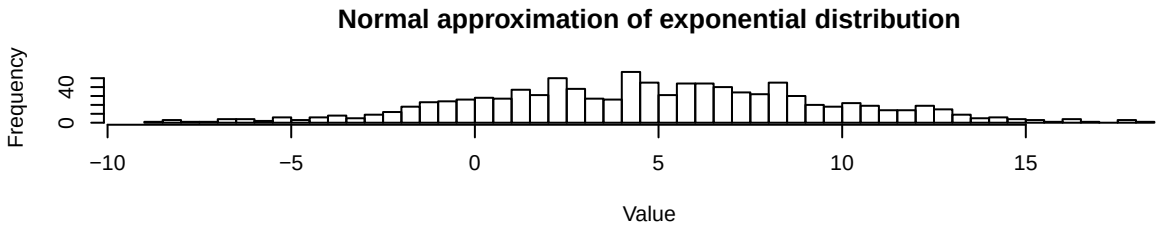


Compare large collection of random exponentials with large collection of averages of 40 exponentials and normal distribution:

```
par(mfrow = c(3, 1))
hist(sim, breaks = 40, main = "Distribution of 40000 random exponentials",
     xlab = "Value", ylab = "Frequency")
hist(means, breaks = 40, main = "Distribution of the mean of 40 exponentials",
     xlab = "Mean", ylab = "Frequency")
normshape <- rnorm(1000, 1/lambda, 1/lambda)
hist(normshape, breaks = 40, main = "Normal approximation of exponential distribution",
     xlab = "Value", ylab = "Frequency")
```







Result plots:

- The first plot is the unprocessed full set of randomly generated exponentials
- The second plot is the sample means of the generated exponentials set
- The third plot is a normal distribution with the exponential distribution's mean and sd

This supports the Central Limit Theorem which states that the sample mean of a population is approximately normally distributed when a large sample is taken.