Assignment: Statistical Inference Course Project

Part 1: Exponential distribution and CLT

Overview

In this project we 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. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. We will investigate the distribution of averages of 40 exponentials.

TodoList

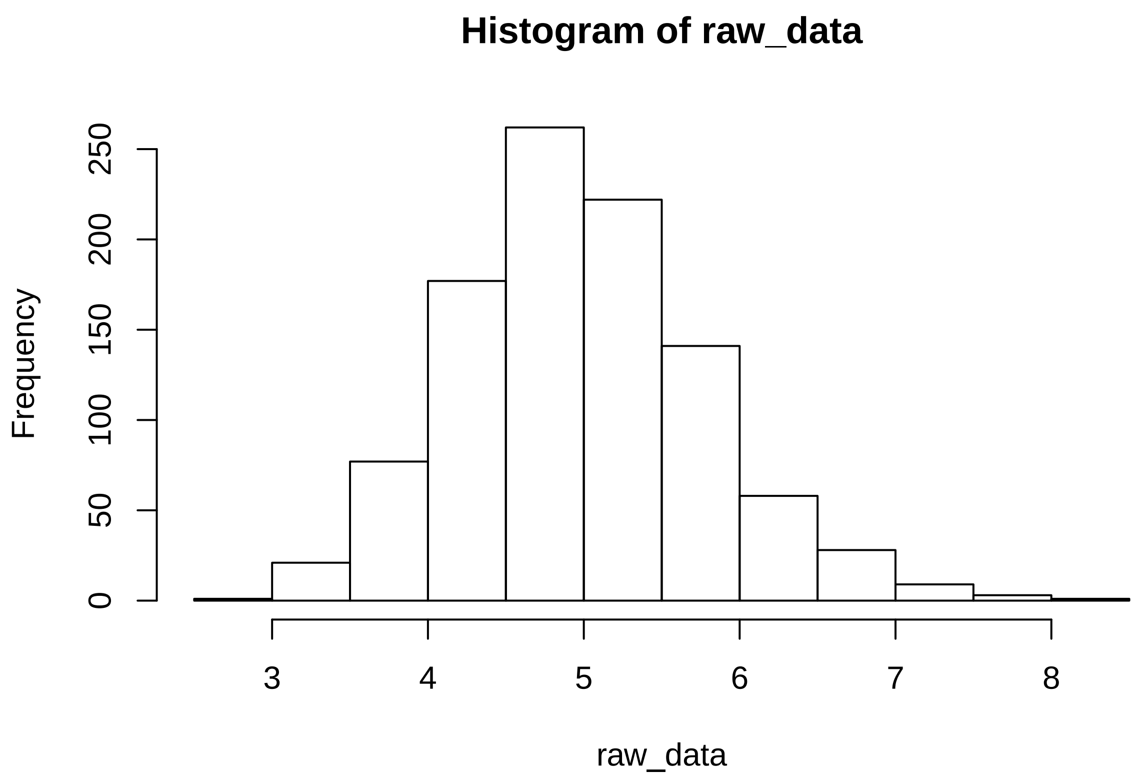
Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Part 0

Let's create simulation data.

```
set.seed(42)
lambda <- 0.2
n <- 40
numberOfSimulations <- 1000
raw_data <- rep(NA, numberOfSimulations)
for (i in 1:numberOfSimulations) {
  raw_data[i] <- mean(rexp(n, lambda))
}
hist(raw_data)
```




```
data <- data.frame(value = raw_data, distribution = "simulation")
```

Part 1

Task 1 - Show the sample mean and compare it to the theoretical mean of the distribution.

```
simulationMean <- mean(data$value)
theoreticalMean <- 1/lambda
meanData <- data.frame(c(simulationMean, theoreticalMean),
                       c("simulation","theoretical"))
colnames(meanData) <- c("mean","distribution")
meanData
```

```
##       mean distribution
## 1 4.986508   simulation
## 2 5.000000  theoretical
```

Which is very close to the expected theoretical center of the distribution:

Part 2

Task 2 - Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
simulationSD <- sd(data$value)
theoreticalSD <- (1/lambda)/sqrt(n)
simulationVariance <- simulationSD^2
theoreticalVariance <- theoreticalSD^2
```

```r
varianceData <- data.frame(c(simulationVariance, theoreticalVariance),
                           c("simulation","theoretical"))
colnames(varianceData) <- c("variance","distribution")
varianceData
```

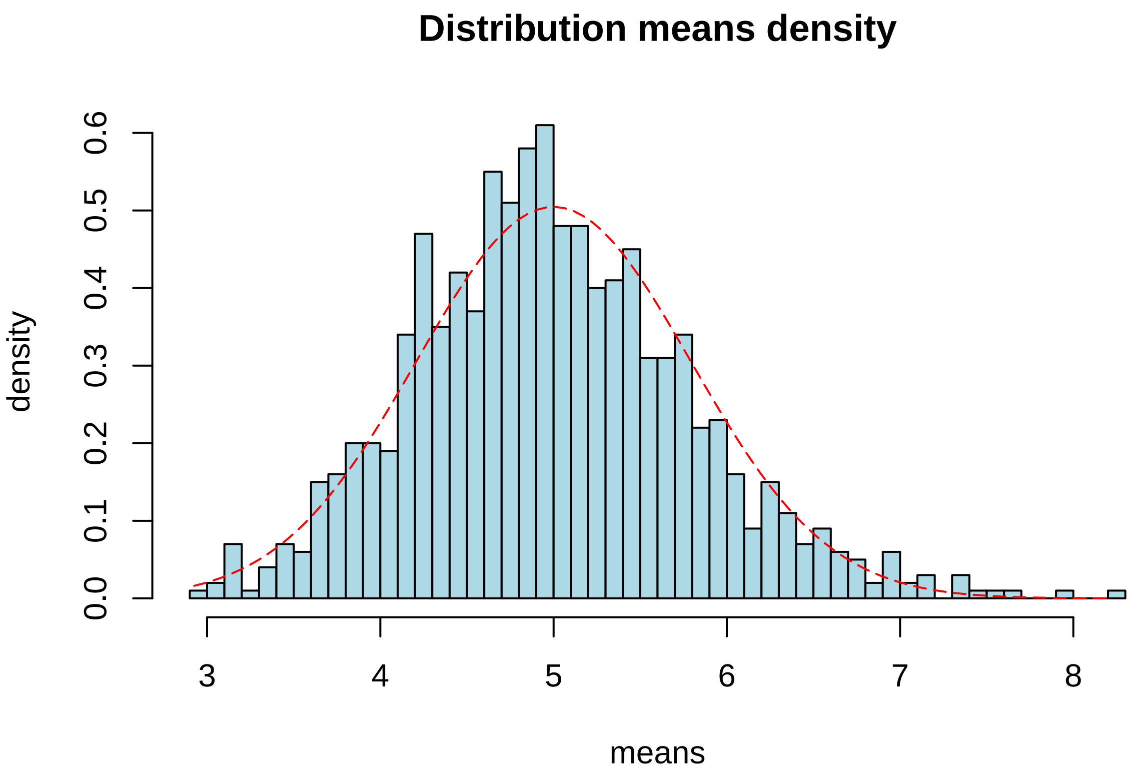
```
##    variance distribution
## 1 0.6344405   simulation
## 2 0.6250000  theoretical
```

We can safely conclude that the differences are minimal, as expected.

Part 3

Task 3 - Show that the distribution is approximately normal

```r
xfit <- seq(min(data$value), max(data$value), length=100)
yfit <- dnorm(xfit, mean = theoreticalMean, sd = theoreticalSD)
hist(data$value,
     breaks = n,
     probability = T,
     col = "lightblue",
     xlab = "means",
     ylab = "density",
     main = "Distribution means density")
lines(xfit, yfit, pch = 42, col = "red", lty = 5)
```



This helps us conclude that the function appears to aproximate to nearly Normal.