C6W4 Assignment Part 1

Nic Neo

9 July 2020

C6W4 Assignment Part 1

Simulation Exercise

Overview

This project investigates the exponential distribution and compares 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.

Instructions

- Set lambda = 0.2 for all of the simulations.
- Investigate the distribution of averages of 40 exponentials.
- Run 1000 simulations.

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

In point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

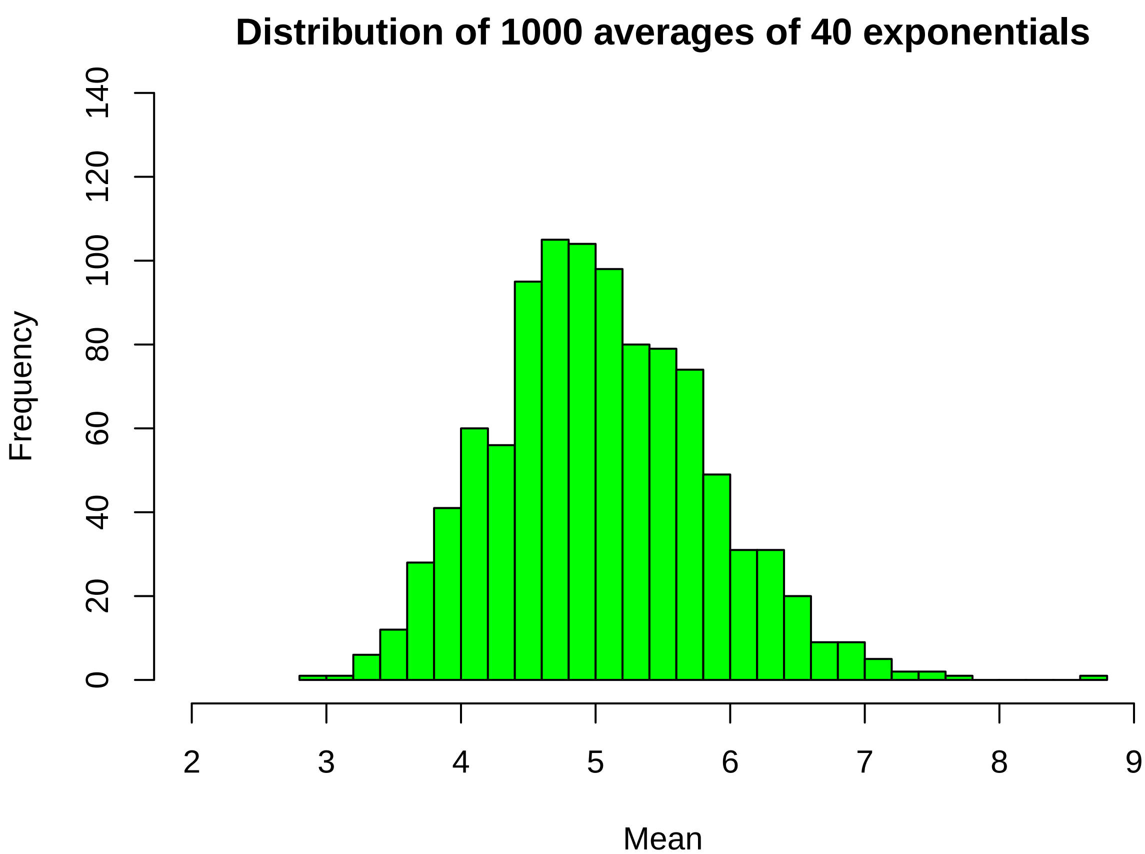
Simulations

Setup the parameter values

```
set.seed(7777)
# Set parameters
# Lambda value
lambda <- 0.2
# Number of exponentials
n <- 40
# Number of simulations
nsims <- 1000
# Simulation with 1000 replications
sim <- replicate(nsims, mean(rexp(n, lambda)))
```

Obtain the distribution of mean of 40 exponentials.

```r
par(mar = c(4, 4, 2, 2))
hist(sim,
     breaks = 40,
     xlim = range(2, 9),
     ylim = range(0, 140),
     col = "green",
     main = "Distribution of 1000 averages of 40 exponentials",
     xlab = "Mean"
     )
```



Sample Mean versus Theoretical Mean

Obtain the sample mean of the distribution (as illustrated above) and compare it to the theoretical mean.

```r
theoretical.mean <- 1/lambda
sample.mean <- mean(sim)
cbind(theoretical.mean, sample.mean)
```

```
##      theoretical.mean sample.mean
## [1,]                5    5.035271
```

The sample mean is similar to the theoretical mean as illustrated above.

Sample Variance versus Theoretical Variance

Obtain the sample variance of the distribution (as illustrated above) and compare it to the theoretical variance.

```
theoretical.variance <- ((1/lambda)/sqrt(n))^2
sample.variance <- var(sim)
cbind(theoretical.variance, sample.variance)
```

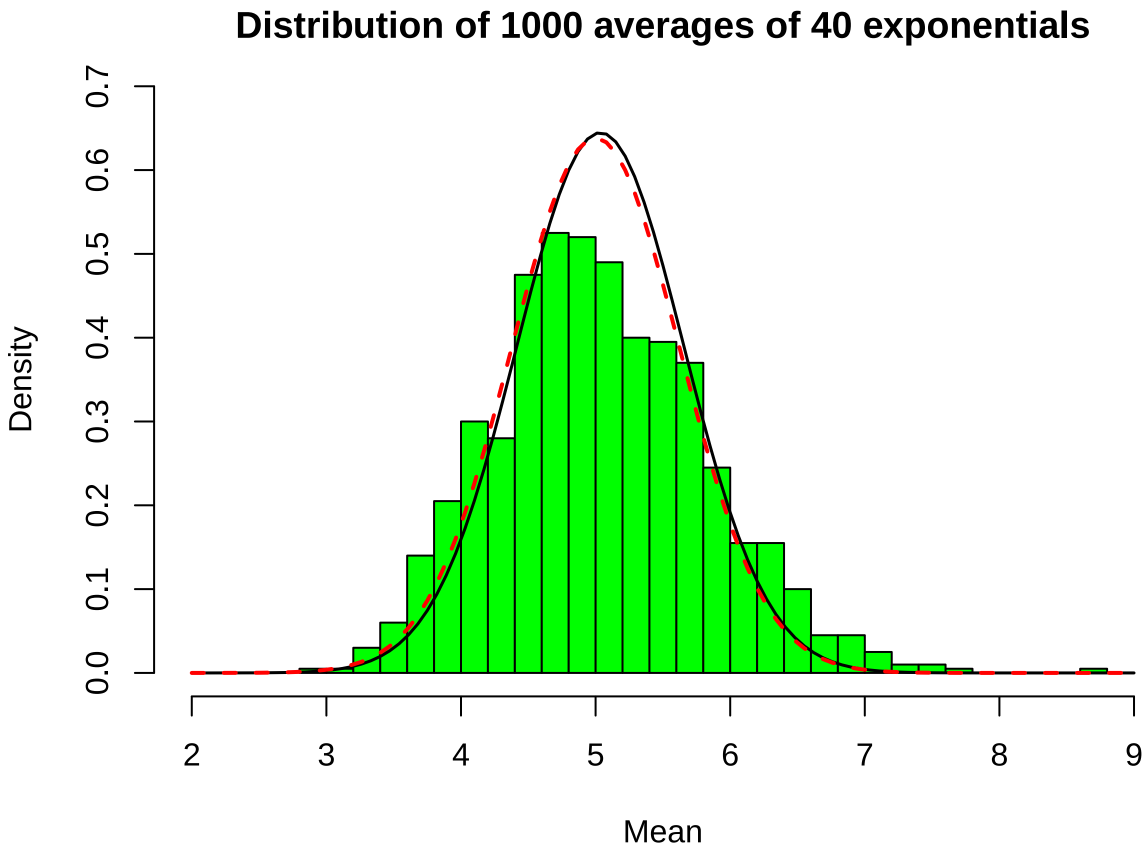
```
##      theoretical.variance sample.variance
## [1,]                0.625       0.6188305
```

The sample variance is somehow close to the theoretical variance.

Distribution

Prove that the distribution of the averages is approximately normal.

```
# Plot density distribution
par(mar = c(4, 4, 2, 2))
hist(sim,
     probability = TRUE,
     breaks = 40,
     xlim = range(2, 9),
     ylim = range(0.0, 0.7),
     col = "green",
     main = "Distribution of 1000 averages of 40 exponentials",
     xlab = "Mean"
     )
# Add sample normal curve
curve(dnorm(x,
            mean = sample.mean,
            sd = sample.variance),
      col = "black",
      lwd = 1.5,
      add = TRUE
      )
# Add theoretical normal curve
curve(dnorm(x,
            mean = theoretical.mean,
            sd = theoretical.variance),
      col = "red",
      lwd = 2,
      lty = 2,
      add = TRUE
      )
```



The black solid line plots the normal curve based on our sample distribution, while the dotted red line plots the true normal curve. Our sample distribution is roughly normal as illustrated in the chart above.