Sama Issa Simulate Exercise

Overview

The purpose of this analysis is to 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. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that I will need to do a thousand simulations.

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.

Simulation

Setting seed and parameter values first, this code run a 1000 simulations to generates a exponential distribution of averages of 40 observation.

```r
library(ggplot2)
library(knitr)

set.seed(1)
no_simulation <- 1000   # number of simulations
lambda <-  0.2
n <- 40             # sample size

sampleMean <- c()
for( i in 1:no_simulation)
    sampleMean <- c(sampleMean,mean(rexp(n,lambda)))
```

Sample Mean Vs. Theoretical Mean

```r
theoMean <- 1/lambda
sampMean <- mean(sampleMean)
print(paste("Theoretical mean of distribution: ",theoMean))
```

```
## [1] "Theoretical mean of distribution:  5"
```

```r
print(paste("Sample mean of distribution: ", sampMean))
```

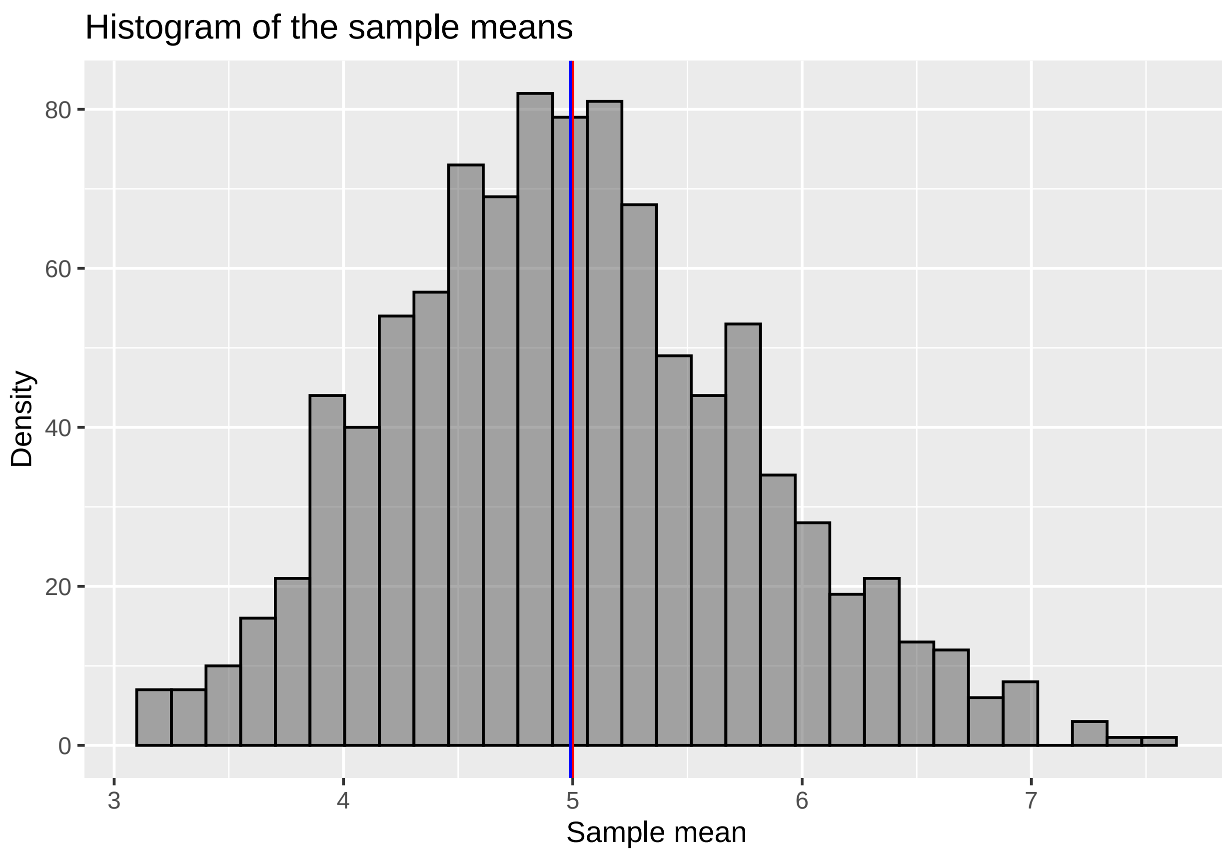
```
## [1] "Sample mean of distribution:  4.99002520077716"
```

The simulation mean is so close to the theoretical mean

```r
ggplot(as.data.frame(sampleMean), aes(sampleMean)) +
    geom_histogram(alpha=.5, position="identity", col="black") +
    geom_vline(xintercept = theoMean, colour="red",show.legend=TRUE) +
    geom_vline(xintercept = sampMean, colour="blue", show.legend=TRUE) +
    ggtitle ("Histogram of the sample means ") +
    xlab("Sample mean") +
    ylab("Density")
```

Histogram

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Sample Variance Vs. Theoretical Variance

```
theoVar <- ((1/lambda)^2)/n
sampVar <- var(sampleMean)
print(paste("Theoretical variance of distribution: ", theoVar))
```

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

```
print(paste("Sample variance of distribution: ", sampVar))
```

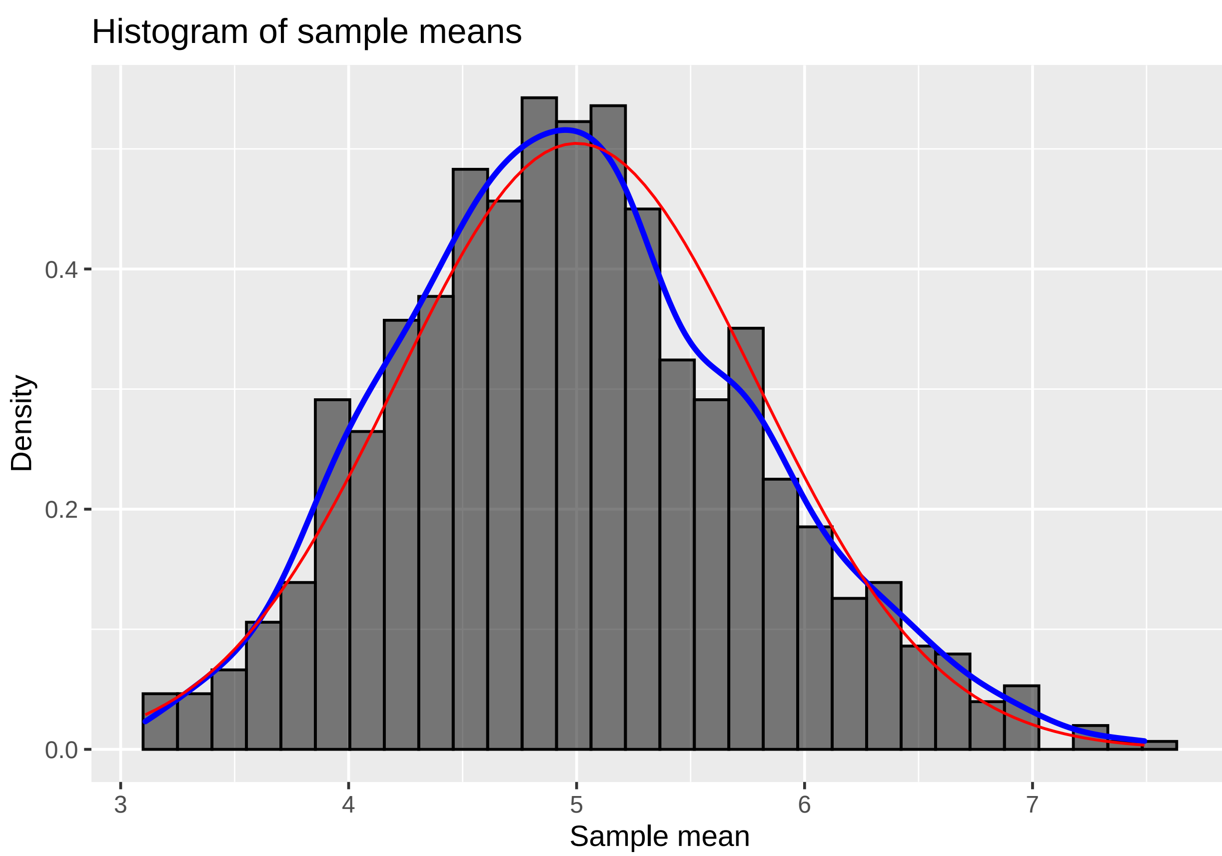
```
## [1] "Sample variance of distribution:  0.611116466559575"
```

Distribution

According to the central limit theorem (CLT), the sample mean follow normal distribution.

```
ggplot(as.data.frame(sampleMean), aes(sampleMean))+
    geom_histogram(aes(y=..density..), alpha=.5, position="identity", fill="black", col="black")+
    geom_density(colour="blue", size=1)+
    stat_function(fun = dnorm, colour = "red", args = list(mean = theoMean, sd = sqrt(theoVar)))+
    ggtitle ("Histogram of sample means ")+
    xlab("Sample mean")+
    ylab("Density")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



As the plot above show, the normal density distribution fits the density histograms well, and hence shows that the distribution of sample mean of a large sample size of exponential is normal.