Statistical Inference Course Project 1

Ben White

26 July 2015

The aim of this project is to investigate the exponential distribution and the Central Limit Theorem in R. The exponential distibution is simulated using `rexp(n, lambda)` where n the number of results, and lambda is the rate parameter. The mean μ and standard deviation σ are both 1/lambda.

Simulations

Set the constants

```
lambda <- 0.2
n <- 40
noSimulations <- 1000
```

Set the seed so that the test can be reproduced

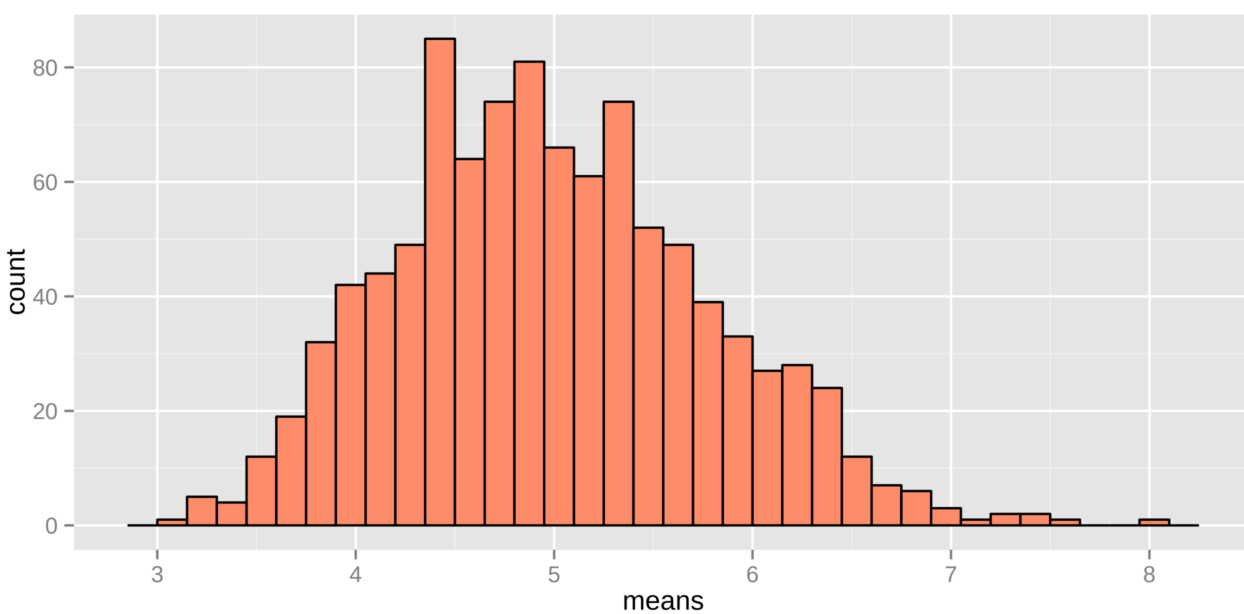
```
set.seed(1)
```

Run the 1000 simulations using rexp

```
simulationResults <- matrix(data=rexp(n * noSimulations, lambda), nrow=noSimulations)
simulationMeans <- data.frame(means=rowMeans(simulationResults))
```

Plot of the Means

```
g <- ggplot(data = simulationMeans, aes(x = means))
g <- g + geom_histogram(fill = "salmon1",
                        binwidth = 0.15,
                        color = "black")
g
```



Question 1. Sample Mean versus Theoretical Mean

The mean μ of a exponential distribution is:

$\mu = \frac{1}{\lambda}$

```
mu <- 1/lambda
mu
```

```
## [1] 5
```

The mean of the sample means is as follows:

```
meanOfMeans <- mean(simulationMeans$means)
meanOfMeans
```

```
## [1] 4.990025
```

It can be seen the theoretical mean of 5 and the mean of the sample means 4.9900252 are close.

Question 2. Sample Variance versus Theoretical Variance

The expected standard deviation σ is:

$\sigma = \frac{1/\lambda}{\sqrt{n}}$

```
sigma <- 1/lambda/sqrt(n)
sigma
```

```
## [1] 0.7905694
```

The variance Var is the standard deviation σ squared:

$Var = \sigma^2$

```
sigma2 <- sigma^2
sigma2
```

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

Calculating the standard deviation and variances of the simulations

```
sdOfSimulation <- sd(simulationMeans$means)
sdOfSimulation
```

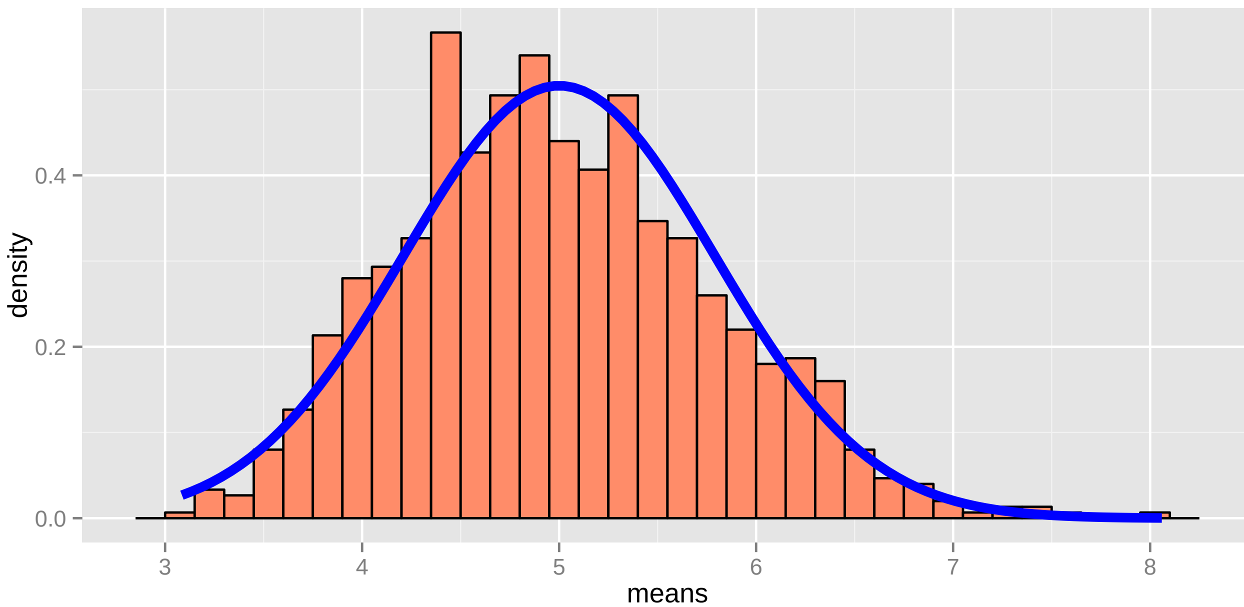
```
## [1] 0.7859435
```

```
varOfSimulation <- var(simulationMeans$means)
varOfSimulation
```

```
## [1] 0.6177072
```

Q3. Show the distribution is approximately normally distributed

```
g <- ggplot(data = simulationMeans,
            aes(x = means))
g <- g + geom_histogram(aes(y = ..density..),
                        fill = "salmon1",
                        binwidth = 0.15,
                        color = "black")
g <- g + stat_function(fun = dnorm,
                       arg = list(mean = mu , sd = sigma),
                       colour = "blue",
                       size = 2)
g
```



As the graph shows the distribution of the sample means is approximately normally distributed.