Statistical inference course project | Part 1

Mick van Hulst

30 Jan 2016

Synopsis

This document contains the first part of the project for the statistical inference class. In this first part of the project a few simulations were done to explore the inference. In the second part a inferential data analysis was done.

Task

Simulating an exponential distribution in R can be done using `rexp(n, lambda)`. In this case lambda is the rate parameter. The mean of exponential distribution is '1/lambda' and the standard deviation is also '1/lambda'. For all the simulations, lambda is set as 0.2. In the simulation there will be an investigation about the distribution of averages of 40 expentionals. Note that you will need to do around a thousand simulated averages of 40 exponentials.

```
# set seed for reproducability
set.seed(31)

# set lambda to 0.2
lambda <- 0.2

# 40 samples
n <- 40

# 1000 simulations
sim <- 1000

# simulate
simExp <- replicate(sim, rexp(n, lambda))

# calculate mean of exponentials
meanExp <- apply(simExp, 2, mean)
```

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

Question 1: Show the sample mean and compare it to the theoretical mean of the distribution.

First the analytical mean is calculated.

```
analyticalMean <- mean(meanExp)
analyticalMean
```

```
## [1] 4.993867
```

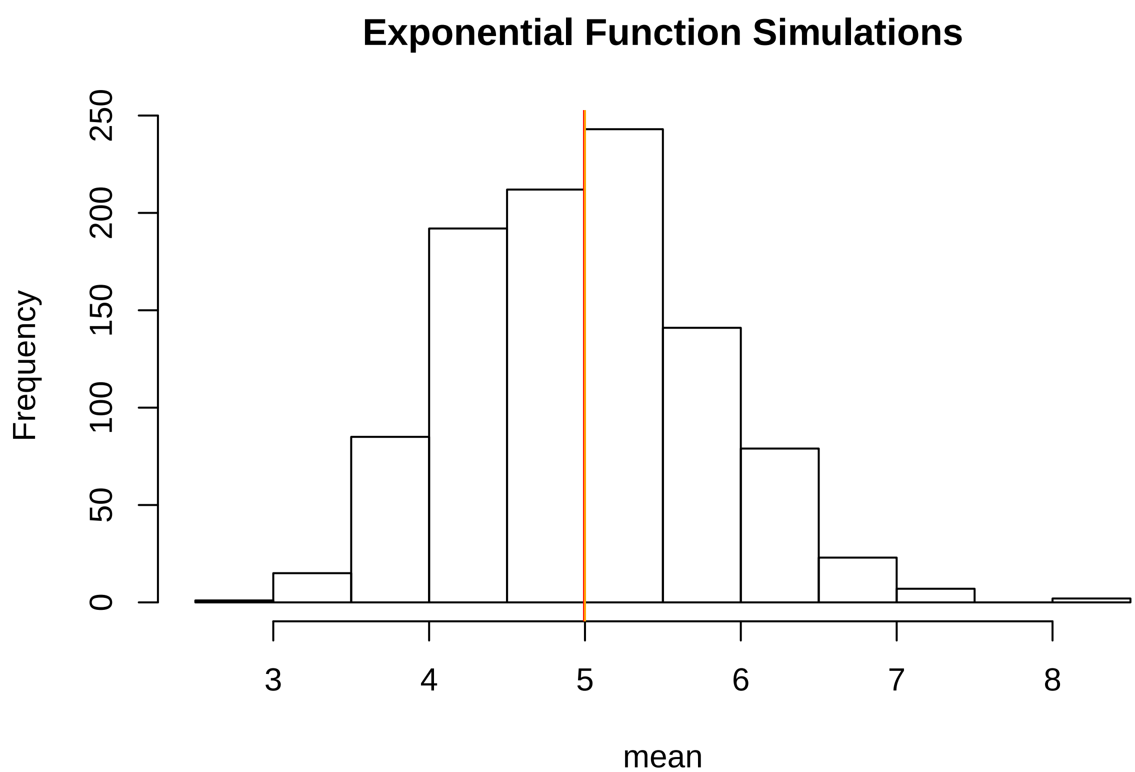
Then calculate the theoretical mean.

```
theoreticalMean <- 1/lambda
theoreticalMean
```

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

Create a histogram

```
# visualization
hist(meanExp, xlab = "mean", main = "Exponential Function Simulations")
abline(v = analyticalMean, col = "red")
abline(v = theoreticalMean, col = "orange")
```



Answer: The analytical mean is 4.993867 and the theoretical mean is 5. The center of distribution of averages of 40 exponentials is very close to the theoretical center of the distribution.

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

Calculate the standard deviation of the distribution.

```
standDevDis <- sd(meanExp)
standDevDis
```

```
## [1] 0.7931608
```

Calculate the standard deviation from analytical expression.

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

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

Calculate the variance of distribution

```
varDist <- standDevDis^2
varDist
```

```
## [1] 0.6291041
```

Calculate the variance from the analytical expression

```
varTheory <- ((1/lambda)*(1/sqrt(n)))^2
varTheory
```

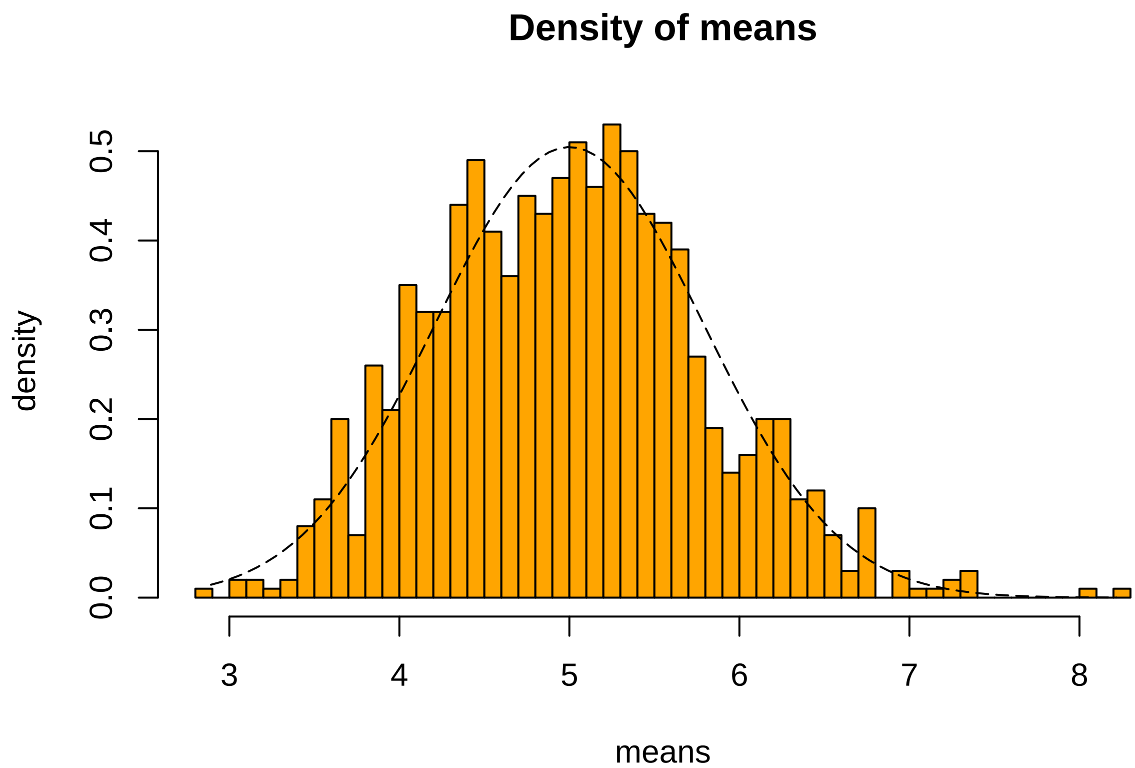
```
## [1] 0.625
```

Answer: The standard deviation of the distribution is 0.7931608 with the theoretical standard deviation of 0.7905694. The theoretical variance is calculated as (1?????1n???)2(1?????1n)2 = 0.625. The actual variance of the distribution is 0.6291041.

Question 3: Show that the distribution is approximately normal.

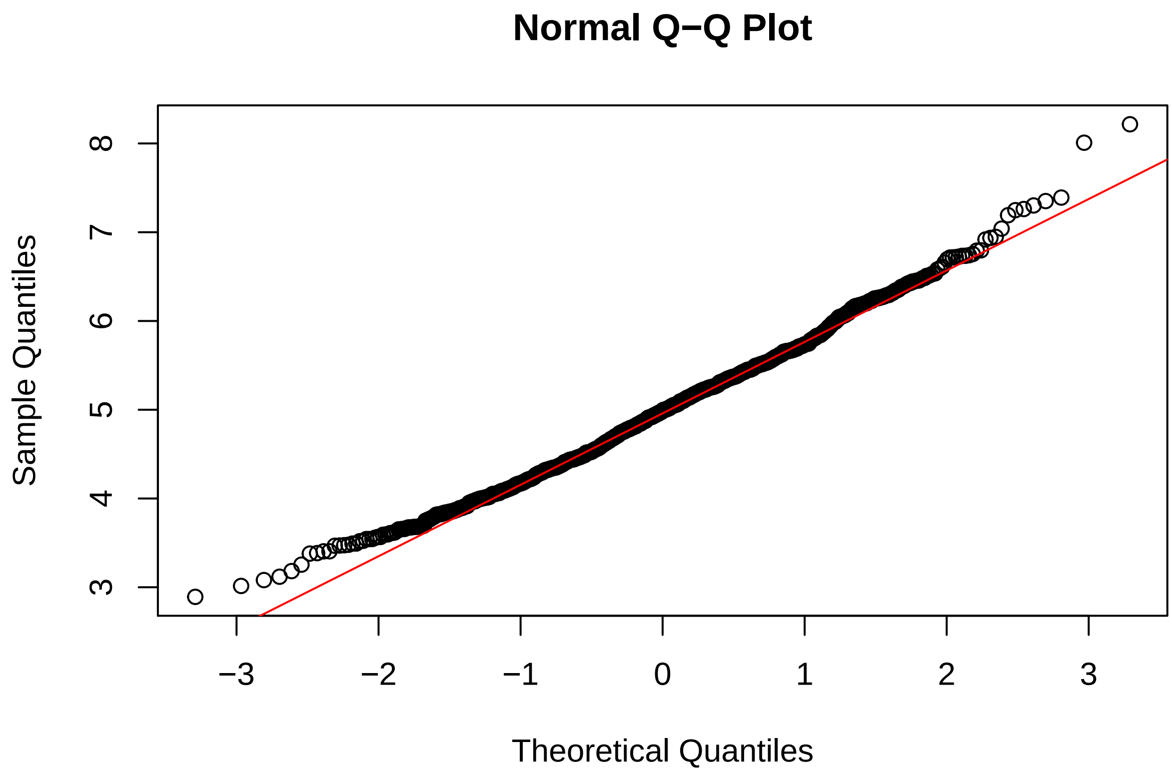
The calculations below are made to show that the distribution is approximately normal

```
xfit <- seq(min(meanExp), max(meanExp), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(n)))
hist(meanExp,breaks=n,prob=T,col="orange",xlab = "means",main="Density of means",ylab="density")
lines(xfit, yfit, pch=22, col="black", lty=5)
```



Compare the distribution of avarages of 40 exponentials to a normal distribution

```
qqnorm(meanExp)
qqline(meanExp, col = 2)
```



Answer: Due to the central limit theorem, the distribution of averages of 40 exponentials is close to a normal distribution.