Central Limit Theorem applied to Exponential Distribution

Nida Batt

6/29/2017

Overview

Simulation Exercise Instructions

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The project illustrates via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials.

```
# Set working directory
setwd("~/Desktop/coursera")
```

```
# Load the required libraries
install.packages("ggplot2", repos="REPOLINK")
library(ggplot2)
```

Sample simulation

The sample mean compared it to the theoretical mean of the distribution

In order to reproduce the data in the future, we use set.seed

```
set.seed(1234) # to reproduce results
```

The exponential distribution is simulated in R with rexp(n,lambda), where lambda=0.2, sample size n is 40, and the number of simulation is 1000 times.

Variables that control simulation

```
nSim <- 1000     # number of simulations
n <- 40          # sample size
lambda <- 0.2    # lambda for the exponential distribution
```

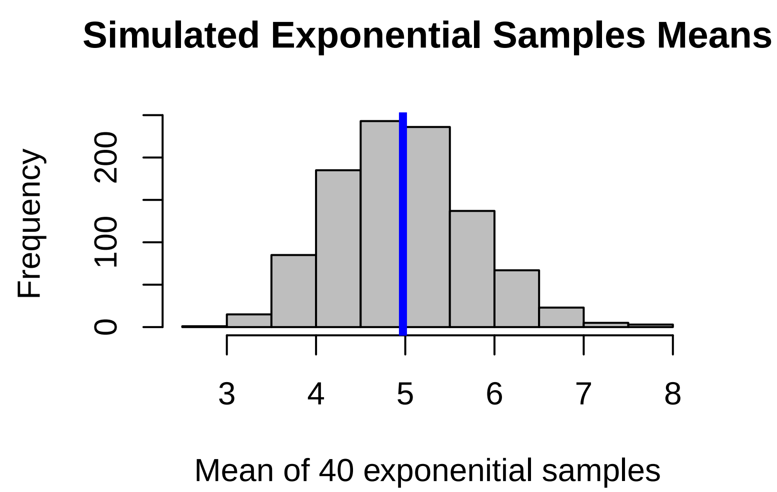
Generate 1000 simulations, each with 40 samples

```
simMatrix <- matrix(rexp(nSim * n, rate=lambda), nSim, n)
```

Calculate the mean of each row and show the distribution

```
simMean <- rowMeans(simMatrix)
```

```
hist(simMean, main="Simulated Exponential Samples Means", xlab = "Mean of 40 exponenitial samples", col
abline(v=mean(simMean), lwd = 4, col = "blue")
```



1. Show where the distribution is centered compared to the theoretical center of the distribution.

```
mean(simMean) #Center of simulated distribution
```

```
## [1] 4.974239
```

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

```
theoMeanExp <- 1/lambda
theoMeanExp
```

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

Actual mean of sample data is close to the theoretical mean of normal data.

2. Show how variable it is compared to the theoretical variance of the distribution.

As for the simulated actual Variance are:

```
actualVar <- var(simMean)
actualVar
```

```
## [1] 0.5949702
```

In comparison, the expected theoretical Variance are:

```
theoryVar <- (1/lambda)^2/n
theoryVar
```

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

The actual variance from sample data is close to the theoretical variance.

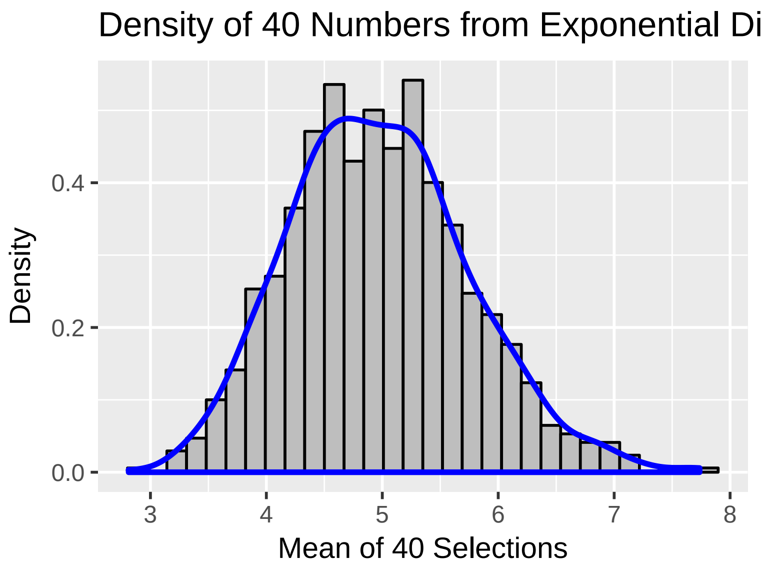
3. Approximately Normal Distribution

To prove the concept, we use the following three steps: Step 1: Create an approximate normal distribution and see how the sample data aligns with it. Step 2: Compare the confidence interval along with the mean and variance with normal distribution. Step 3: q-q plot for quantiles.

```
# Step 1
plotSimMean <- data.frame(simMean)
p <- ggplot(plotSimMean, aes(x =simMean))
```

```
p + geom_histogram(aes(y=..density..), color="black", fill = "grey")+
    geom_density(colour="blue", size=1) +
    labs(title="Density of 40 Numbers from Exponential Distribution", x="Mean of 40 Selections", y="Dens
```

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



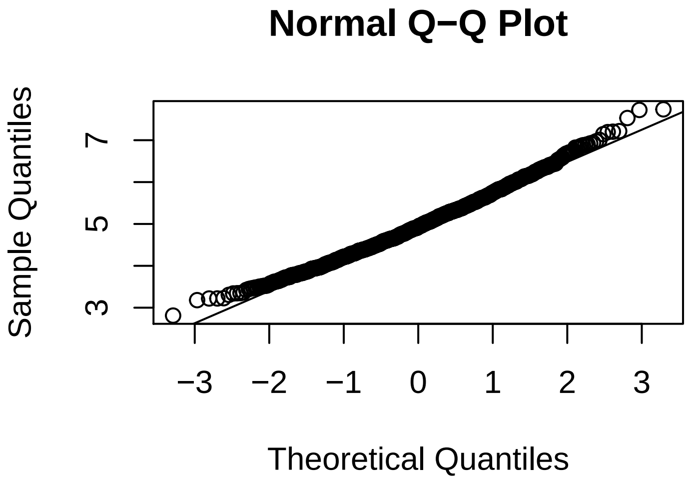

```
# Step 2
round(mean(simMean) + c(-1,1)*1.96*sd(simMean)/sqrt(n),3)
```

```
## [1] 4.735 5.213
```

```
theoMeanExp + c(-1,1)*1.96*sqrt(theoryVar)/sqrt(n);
```

```
## [1] 4.755 5.245
```

```
# Step 3
qqnorm(simMean); qqline(simMean)
```



As we can see, the mean and variance of the simluation are very close to that of a Normal Distribution, which confirms the Central Limit Theorem.