Statistical Inference Project 1

Igor Hut

October 25, 2015

Overview

The idea behind this project assignment 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. We will set lambda = 0.2 for all of the simulations. The distribution of averages of 40 exponentials will be investigated. We will perform 1000 simulations.

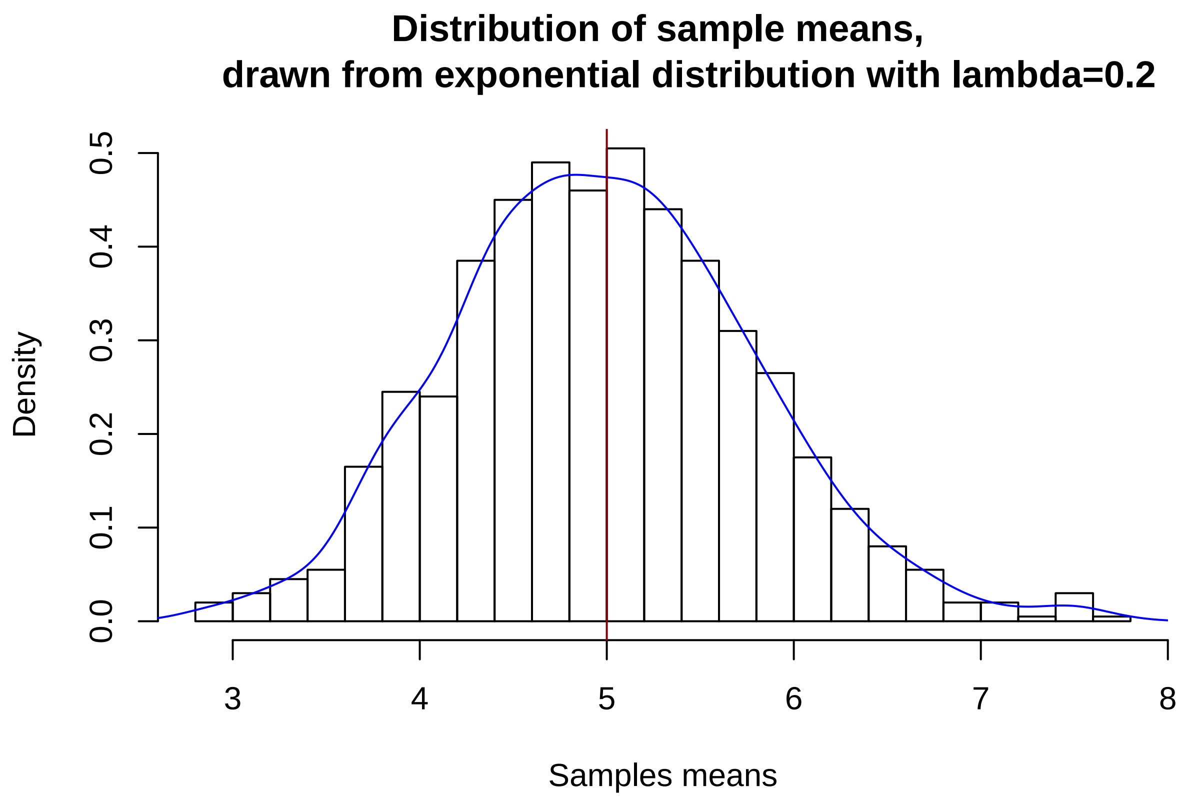
1. Show the sample mean and compare it to the theoretical mean of the distribution.

```
# setting constants
lambda <- 0.2
n <- 40 # number of exponetials
numSim<- 1000 # number of simulations

# provide reproducability by setting the seed
set.seed(1808)

# Simulation vs theory
sim <- matrix(data=rexp(n * numSim, lambda),numSim,n)
simMeans <- rowMeans(sim)
meanSim <- mean(simMeans)
sdSim <- sd(simMeans)
varSim<-var(simMeans)
meanTheory <- 1/lambda
```

```
# plot the means
hist(simMeans, breaks=20, freq=FALSE, main="Distribution of sample means,
     drawn from exponential distribution with lambda=0.2", xlab="Samples means")
lines(density(simMeans),col="blue")
abline(v=meanTheory, col="darkred")  # theory mean line
```



So, as can be seen, simulated mean 4.9850712 is rather close to theoretical mean $\frac{1}{\lambda} = 5$.

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

Standard deviation for simulated sample means is 0.801991 while the theoretical expected standard deviation σ of a exponential distribution of rate λ is $\sigma = \frac{1/\lambda}{\sqrt{n}}$, i.e.:

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

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

Variance for simulated sample means is 0.6431895 while the theoretical variance Var of standard deviation σ is $Var = \sigma^2$, i.e.:

```
varTheory <- sdTheory^2
varTheory
```

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

As can be observed standard deviations are very close and the same goes for the variance.

3. Show that the distribution is approximately normal.

Due to the central limit theorem, the sample means follow the normal distribution, which can be observed already with the first plot. Furthermore the normal q-q plot below inicates that the distribution is quite close to normal.

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