Statistical Inference - Course Project- Part 1

Boris Badmaev

August 18, 2015

Overview

This is the project for the statistical inference class. In it, we will use simulation to explore inference and do some simple inferential data analysis. The project consists of two parts:

1. A simulation exercise.
2. Basic inferential data analysis

The goal of this report is to illustrate the properties of the distribution of the mean of 40 exponentials. There are three tasks in this research:

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.

Simulations

First, let's simulates exponential distribution using `rexp(n, lambda)` function in R, where lambda is the rate parameter. Let's set lambda = 0.2 for all of the simulations and number of simulations nosim = 1000 with each simulation generating 40 exponentials.

```
# set the seed for reproducibility
set.seed(79)
# set lambda
lambda<-0.2
# set sample size
n<-40
# set number of simulations
nosim<-1000
# simulated exponentials
sim.exp<-replicate(nosim, rexp(n,lambda))
# let's calculate means for each simulation of 40 exponentials
means.exp<-apply(sim.exp, 2, mean)
```

Question 1

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

```
# let's calculate the distribution mean
sample_mean<-mean(means.exp)
sample_mean
```

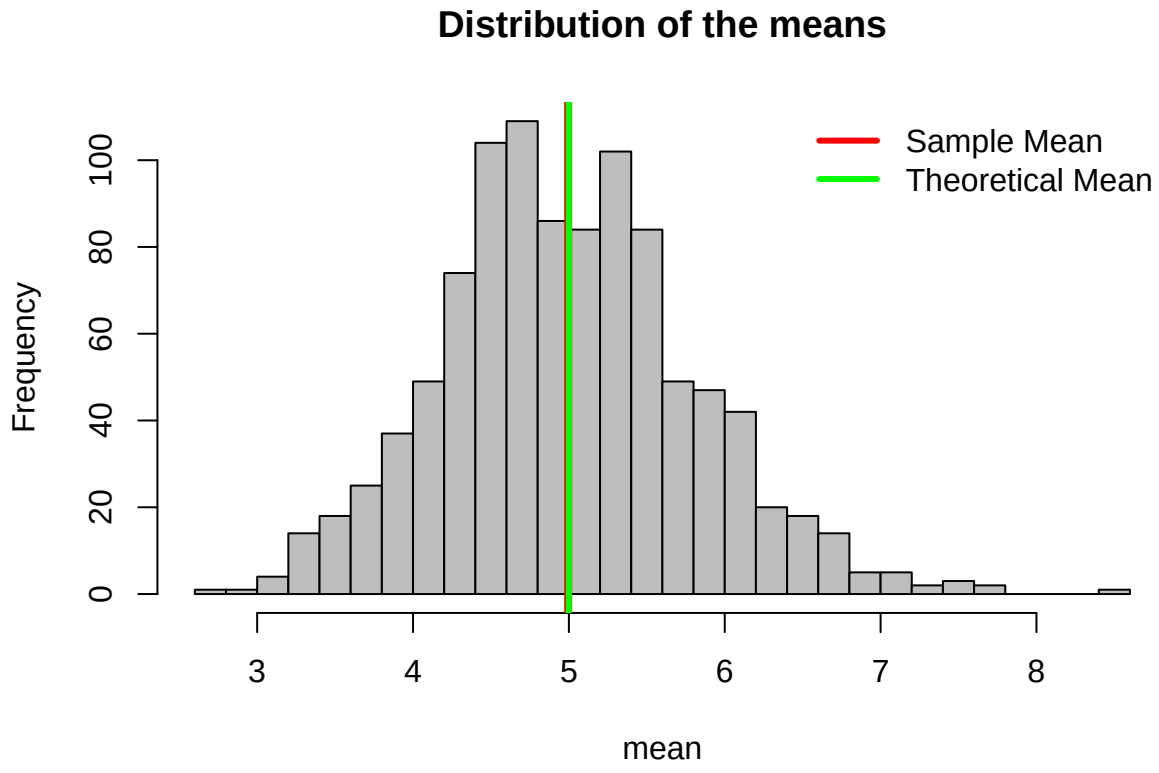
```
## [1] 4.995706
```

```
# let's compare it with theoretical mean
theory_mean<-1/lambda
theory_mean
```

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

Let's plot the distribution and show the sample and theoretical means

```
hist(means.exp, xlab = "mean", main = "Distribution of the means", col="gray", breaks=n)
abline(v = sample_mean, col = "red", lwd=4)
abline(v = theory_mean, col = "green", lwd=3)
legend("topright", lty=1, lwd=3, bty = "n", col = c("red","green"), legend = c("Sample Mean", "Theoreti
```



The sample mean = 4.995706 and the theoretical mean = 5. Thus 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.

```
# variance of distribution
var.dist<-var(means.exp)
var.dist
```

```
## [1] 0.6515704
```

```
# theoretical variance
var.theory<-((1/lambda)*(1/sqrt(n)))^2
var.theory
```

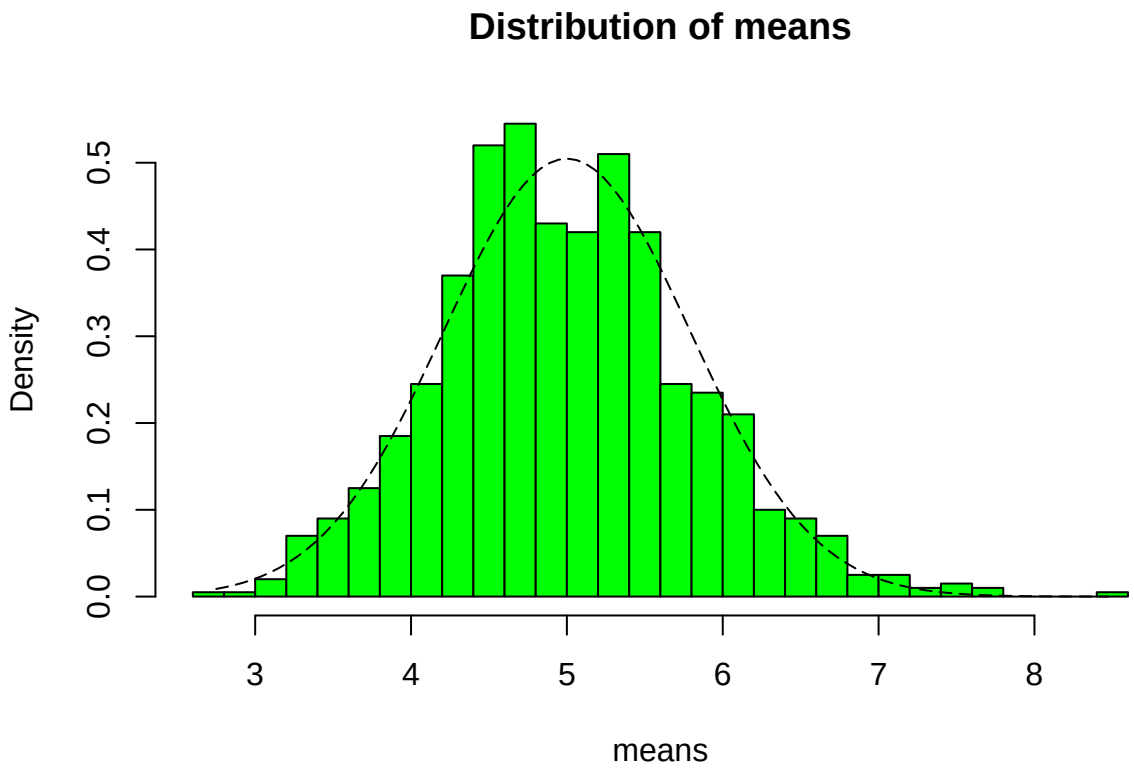
```
## [1] 0.625
```

Thus, as we can see, the sample variance 0.6515704 is very close to the theoretical one 0.625.

Question 3

Show that the distribution is approximately normal.

```
norm.x <- seq(min(means.exp), max(means.exp), length=100)
norm.y <- dnorm(norm.x, mean=1/lambda, sd=(1/lambda/sqrt(n)))
hist(means.exp,breaks=n,prob=T,col="green",xlab = "means",main="Distribution of means")
lines(norm.x, norm.y, pch=22, col="black", lty=5)
```



As we can see the distribution of the mean follow the normal distribution curve (black line).

We can compare the distribution of the means with the distribution of exponentials (see Appendix)

Appendix

```
hist(sim.exp, xlab = "mean", main = "Distribution of exponentials", col="lightblue", breaks=n)
```