Statistical Inference: Part 1

Paul Degtyariov

02 05 2020

Overview

This is a part 1 for project in Statistical Inference cource by JHU. It's dedicated to exploring properties of an exponential distribution

Questions

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.

Activating packages

```
library(ggplot2)
```

Simulation

```
#Setting seed for reproducibility purposes
set.seed(42)

#Simulation exponentials - 1000 simulations, 40 observations in each with a labmda of 0.2
sim_exponentials <- replicate(1000, rexp(40, 0.2))
```

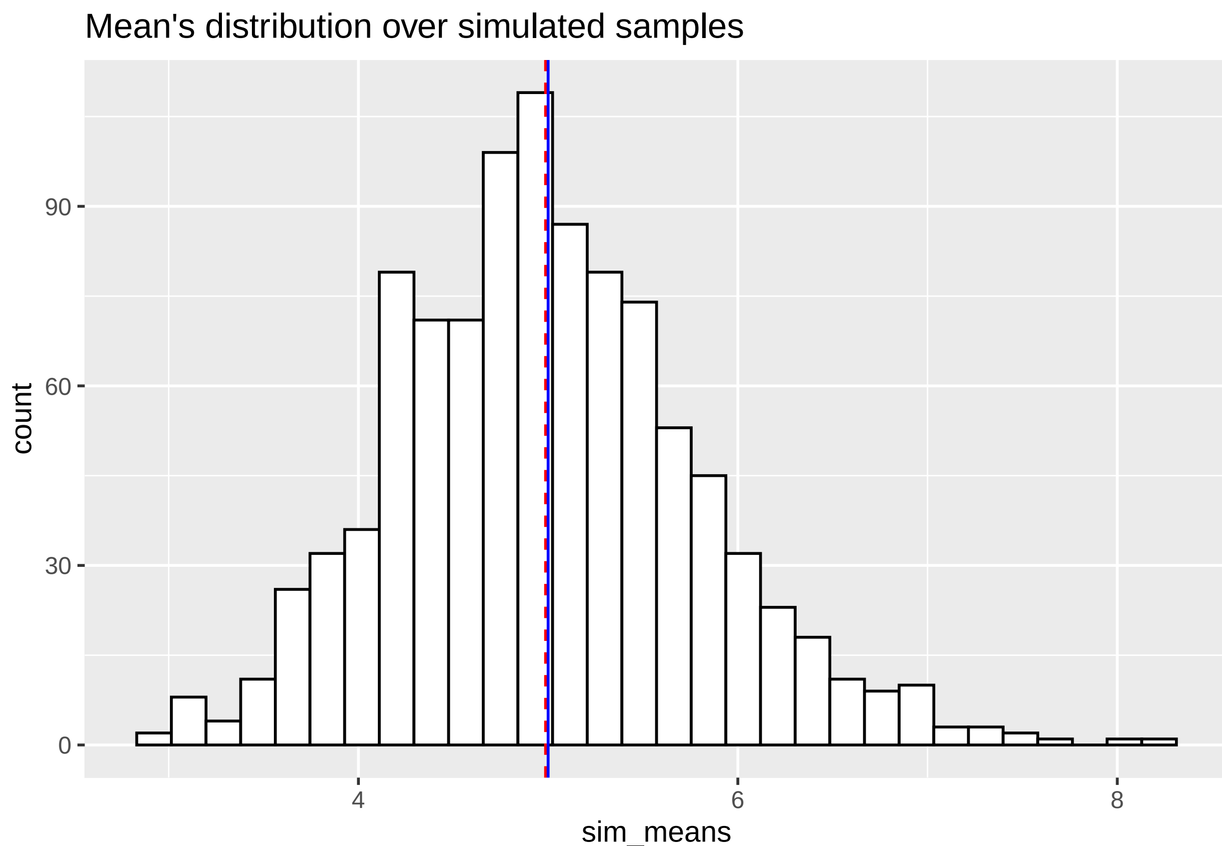
Question 1

```
sim_means <- apply(sim_exponentials,2,mean)

#Calculating means
emperical_mean <- mean(sim_means)
theoretical_mean <- 1/0.2
```

```
#Visualising means
ggplot(data.frame(sim_means=sim_means),aes(x=sim_means))+
    geom_histogram(color="black",fill="white")+
    geom_vline(xintercept=emperical_mean,color="red",linetype="dashed")+
    geom_vline(xintercept=theoretical_mean,color="blue")+
    ggtitle("Mean's distribution over simulated samples")
```

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



Emperical mean is 4.9865083, theoretical mean is 5, which are quite close.

Question 2

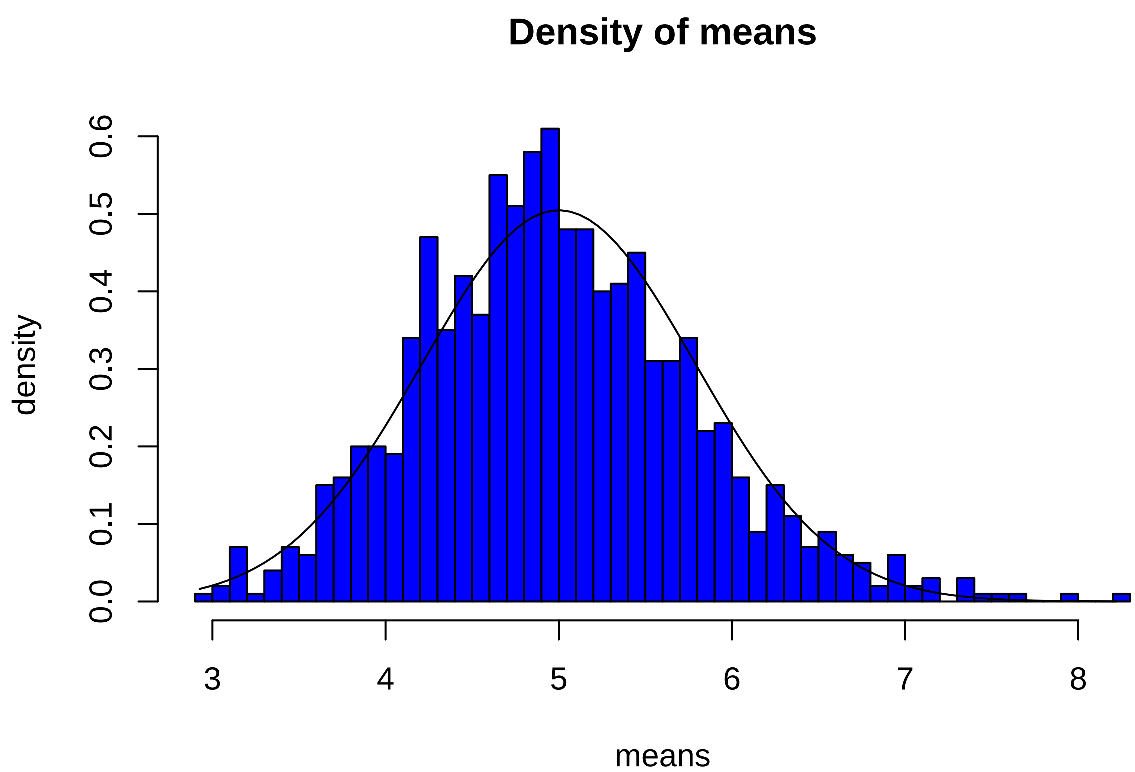
```
#Calculating standard deviations
empirical_sd <- sd(sim_means)
theoretical_sd <- (1/0.2)/sqrt(40)

#Calculating variance
empirical_variance <- empirical_sd^2
theoretical_variance <- theoretical_sd^2
```

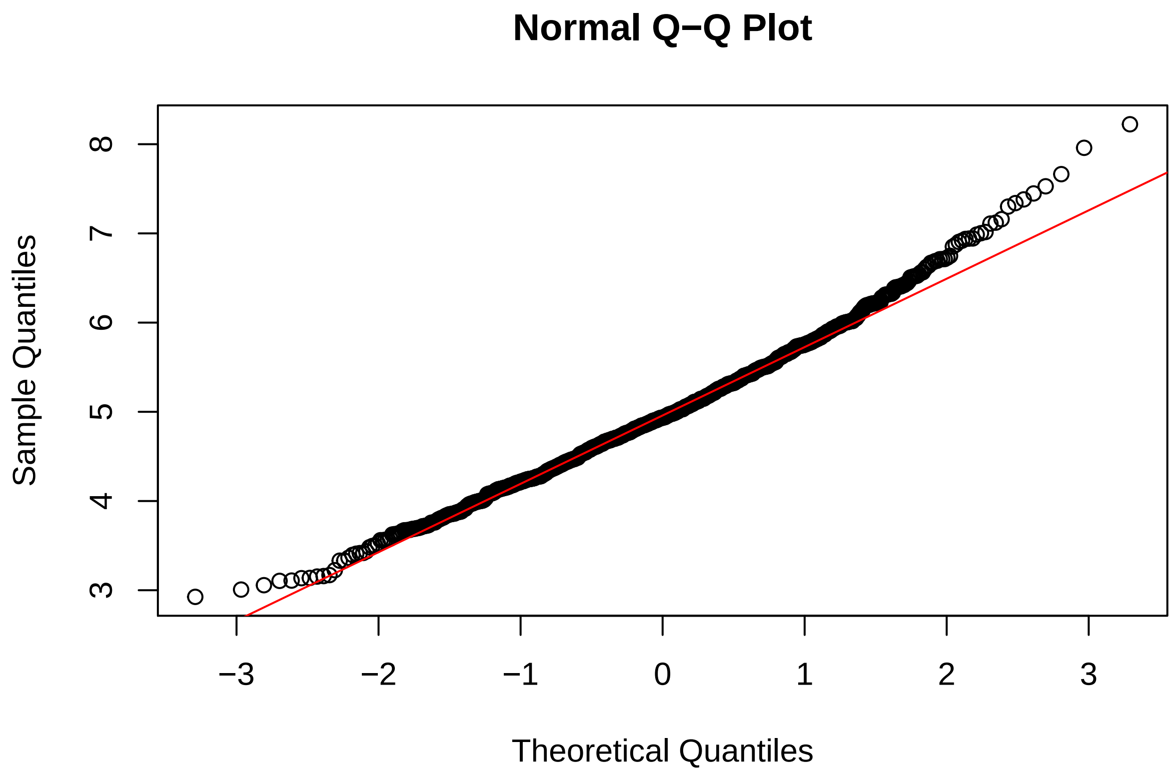
Emperical variance is 0.6344405, theoretical variance is 0.625, which are quite close.

Question 3

```
#Constructing a plot of means density distribution
xfit <- seq(min(sim_means), max(sim_means), length=100)
yfit <- dnorm(xfit, mean=1/0.2, sd=(1/0.2/sqrt(40)))
hist(sim_means,breaks=40,prob=T,col="blue",xlab = "means",main="Density of means",ylab="density")
lines(xfit, yfit, pch=22, col="black", lty=1)
```




```
#Comparing  distribution of averages of 40 exponentials to a normal distribution
qqnorm(sim_means)
qqline(sim_means, col = 2)
```



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