Statistical Inference Project

MiRo

22 Jul 2015

Overview

The exponential distribution can be simulated in R with `rexp(n, lambda)` where `lambda` λ is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. For this simulation, we set $\lambda = 0.2$. In this simulation, we investigate the distribution of averages of 40 numbers sampled from exponential distribution with $\lambda = 0.2$.

Main question

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

Loading requirements and data.

```
packages <- c("base","data.table", "ggplot2", "dplyr", "scales", "vwr", "gridExtra")
sapply(packages, require, character.only=TRUE, quietly=TRUE)

set.seed(10001)
```

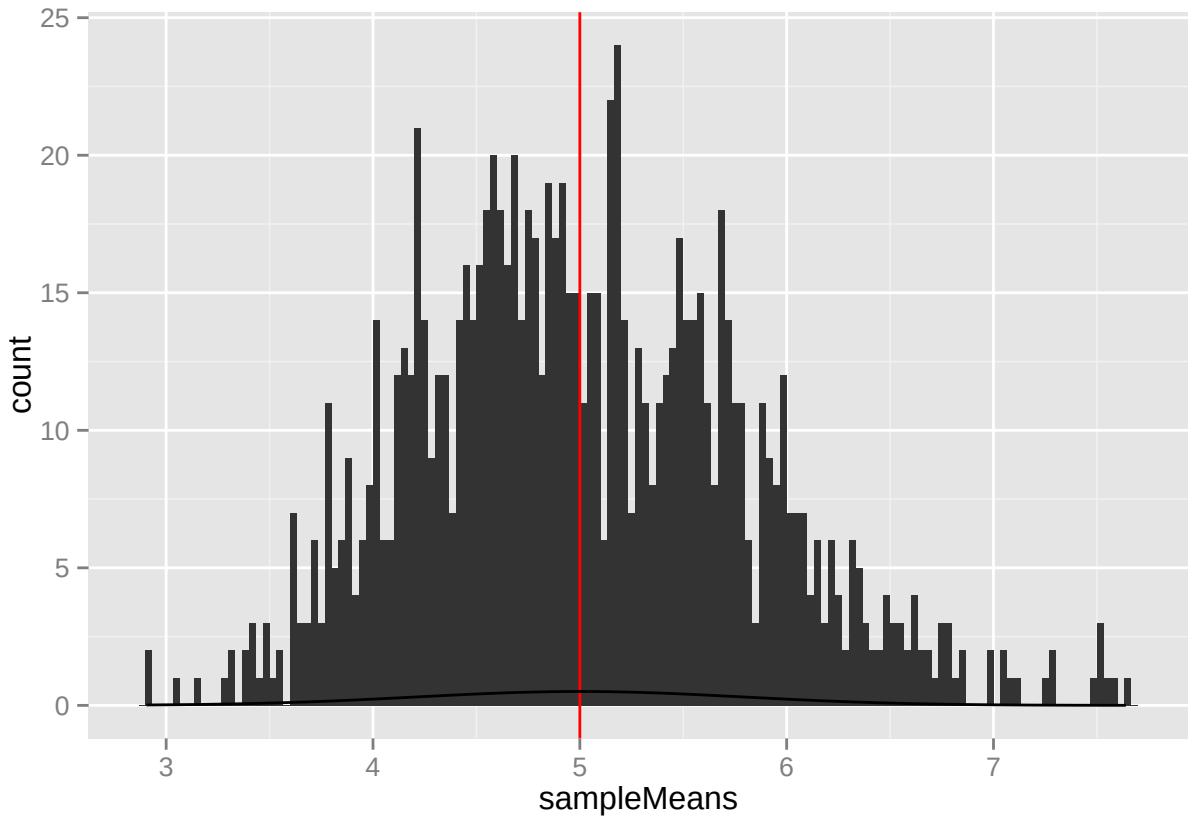
Simulations

We are about to simulate 1000 tests with n=40 of exponential distribution. The mean of exponential distribution is $1/\lambda$ and the standard deviation also equals to $1/\lambda$.

```
testsN <- 1000
n <- 40
lambda <- 0.2

meanTested <- 1/lambda
sdTested <- 1/lambda
alpha <- 0.05

sampleMeans <- c()
sampleSDs <- c()

for(i in 1:testsN){
  sample <- rexp(n, lambda)
  sampleMeans <- c(sampleMeans, mean(sample))
  sampleSDs <- c(sampleSDs, sd(sample))
}
```

Testing and plotting data

```
xfit <- seq(min(sampleMeans), max(sampleMeans), length=testsN)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(n)))

gplot <- ggplot(as.data.frame(x=sampleMeans),aes(sampleMeans)) +
  geom_histogram(binwidth = 1/(30)) +
  labs(main = "Distribution of averages of samples, lambda=0.2",
       ylab ="Count", xlab = "Observed means") +
  geom_vline(xintercept = 1/lambda, col="red") +
  geom_line(data=data.frame(xfit,yfit), aes_string(x="xfit", y="yfit"), width=0.1)

gplot
```

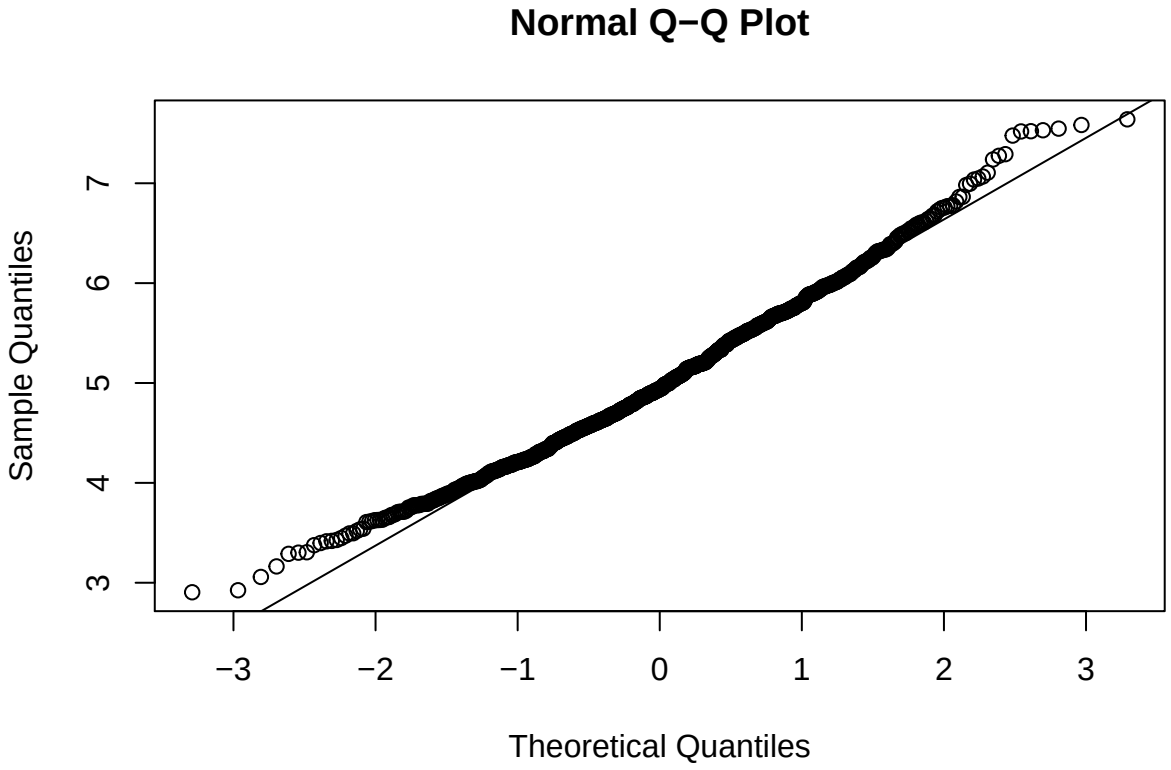



```
meanMeans <- mean(sampleMeans)
varMeans <- var(sampleMeans)
```

Theoretical mean is equal to 5. Let's test our example mean (5.0167018) with t test. Variation of sample means is equal to 0.6442362 where the theoretical variation of the distribution is $\sigma^2/n = 1/(\lambda^2 n) = 1/(0.04 \times 40) = 0.625$.

Due to the central limit theorem, the averages of samples follow normal distribution.

```
qqnorm(sampleMeans)
qqline(sampleMeans)
```


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
3
4
5
6
7
−3
−2
−1
0
1
2
3