Statistical Inference Course Project, Part 1

Dmitry Lishmanov

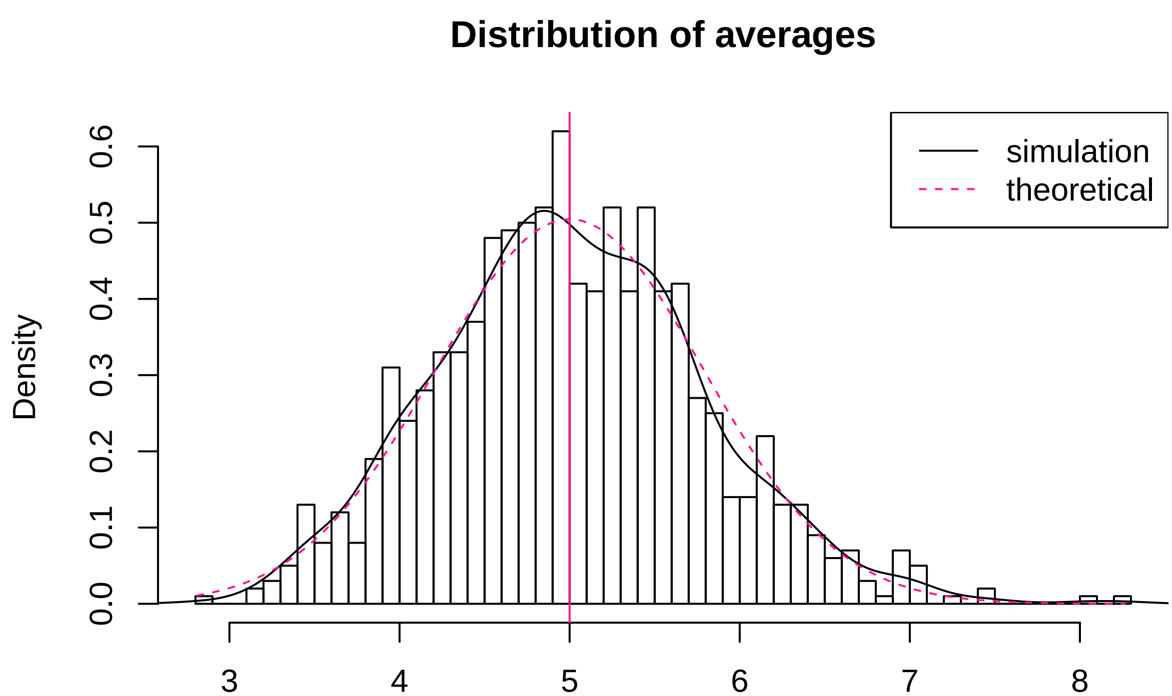
30-01-2016

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 $\lambda = 0.2$.

First we simulate 1000 averages of 40 exponentials.

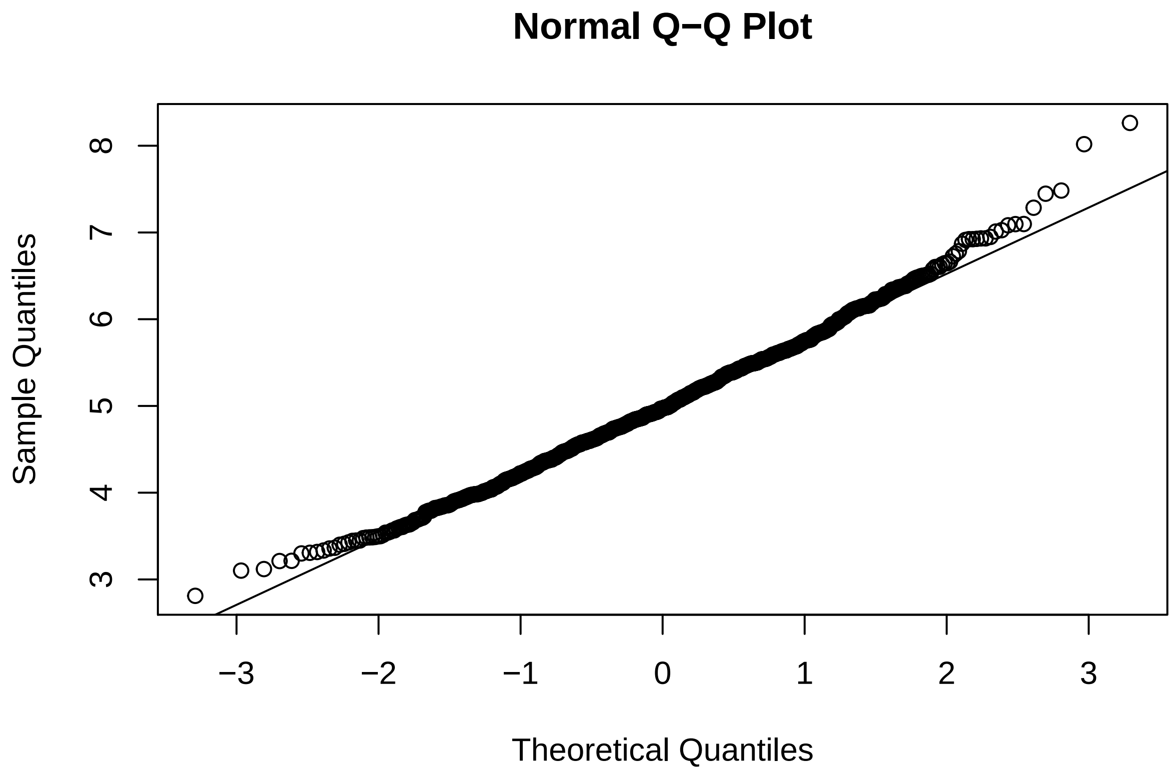
```
set.seed(12)
lambda <- 0.2
nSims <- 1000
sample_size <- 40
sim <- matrix(rexp(nSims*sample_size, rate=lambda), nSims, sample_size)
row_means <- rowMeans(sim)
```

The distribution of sample means is as follows.

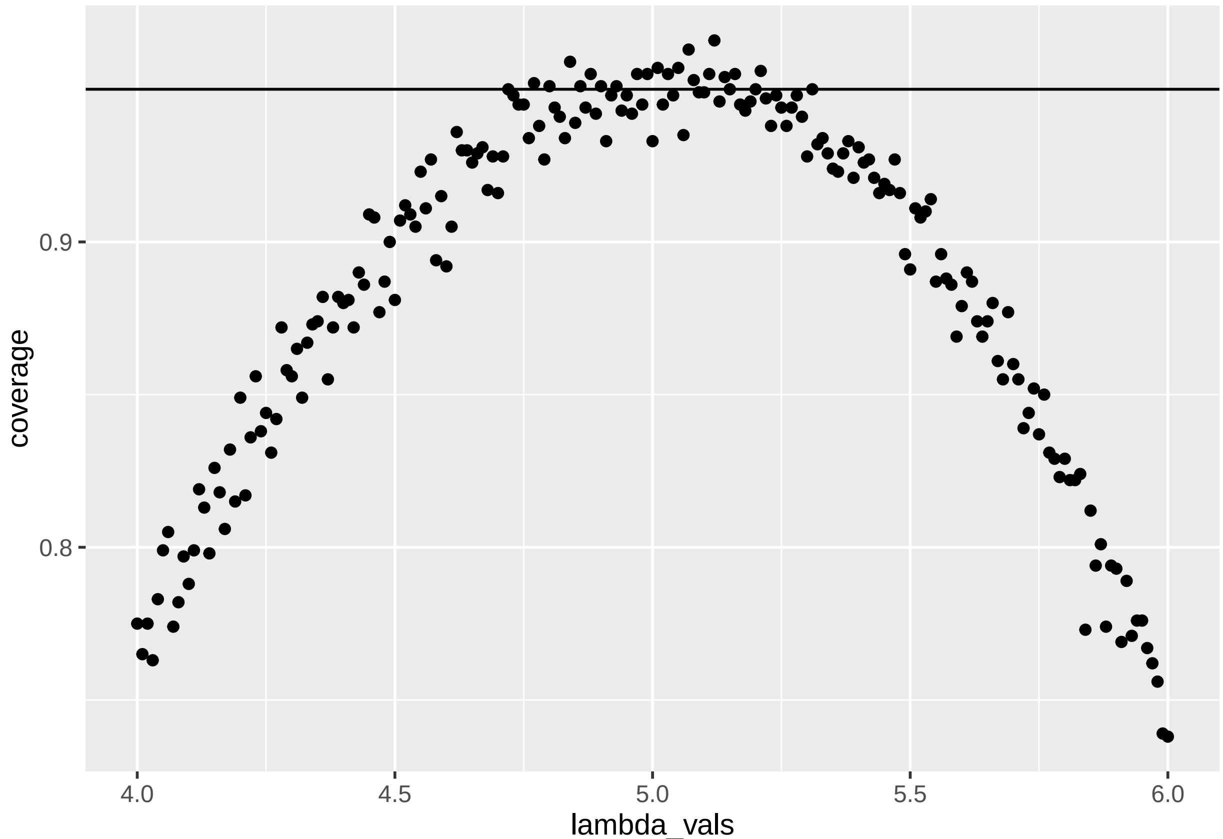


The distribution of sample means is centered at 5.0100152 and the theoretical center of the distribution is $\lambda^{-1} = 5$. The variance of sample means is 0.6145512 where the theoretical variance of the distribution is $\sigma^2/n = 1/(\lambda^2 n) = 1/(0.04 \times 40) = 0.625$.

According to the central limit theorem, the averages of samples follow normal distribution. The figures above illustrate the density computed using the histogram and the normal density plotted with theoretical mean and variance values. The q-q plot below suggests the normality.



Finally we evaluate the coverage of the confidence interval for $1/\lambda = \bar{X} \pm 1.96\frac{S}{\sqrt{n}}$



The 95% confidence intervals for the rate parameter (λ) to be estimated ($\hat{\lambda}$) are $\hat{\lambda}_{low} = \hat{\lambda}(1 - \frac{1.96}{\sqrt{n}})$ and $\hat{\lambda}_{upp} = \hat{\lambda}(1 + \frac{1.96}{\sqrt{n}})$. As can be seen from the plot, for selection of $\hat{\lambda}$ around 5, the average of the sample mean falls within the confidence interval at least 95% of the time. Note that the theoretical rate λ is 5.

The report including the code for plots is available at https://github.com/lidv/statistical-inference/