Statistical Inference Course Project, Part 1

Kelly Ng

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 $\lambda = 0.2$. Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

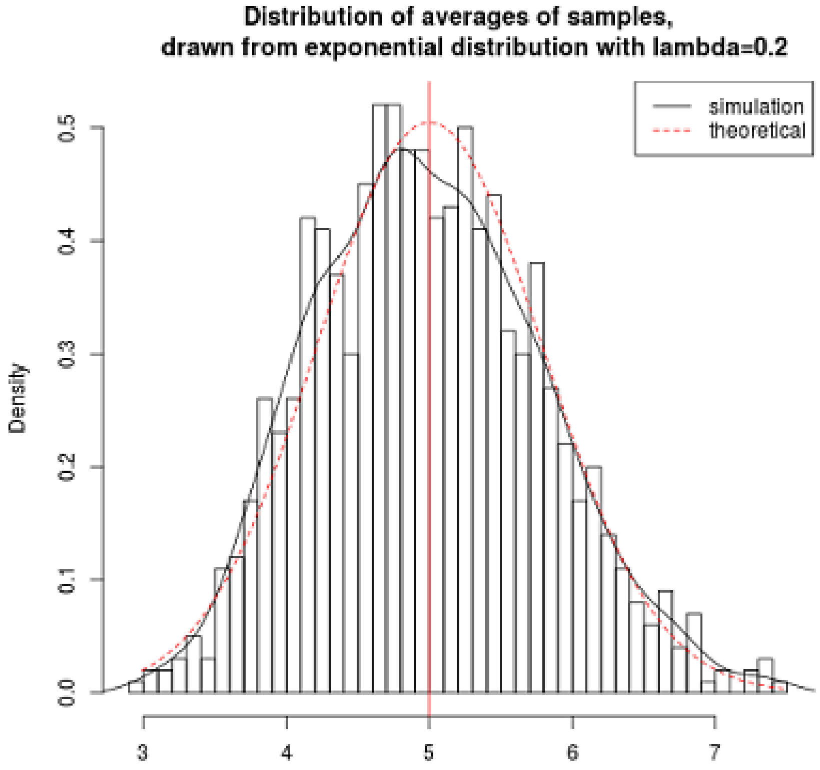
1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Let's do a thousand simulated averages of 40 exponentials.

Simulation

```
#settings
set.seed(3)
lambda <- 0.2
num_sim <- 1000
sample_size <- 40
sim <- matrix(rexp(num_sim*sample_size, rate=lambda), num_sim, sample_size)
row_means <- rowMeans(sim)
```

The distribution of sample means is as follows.

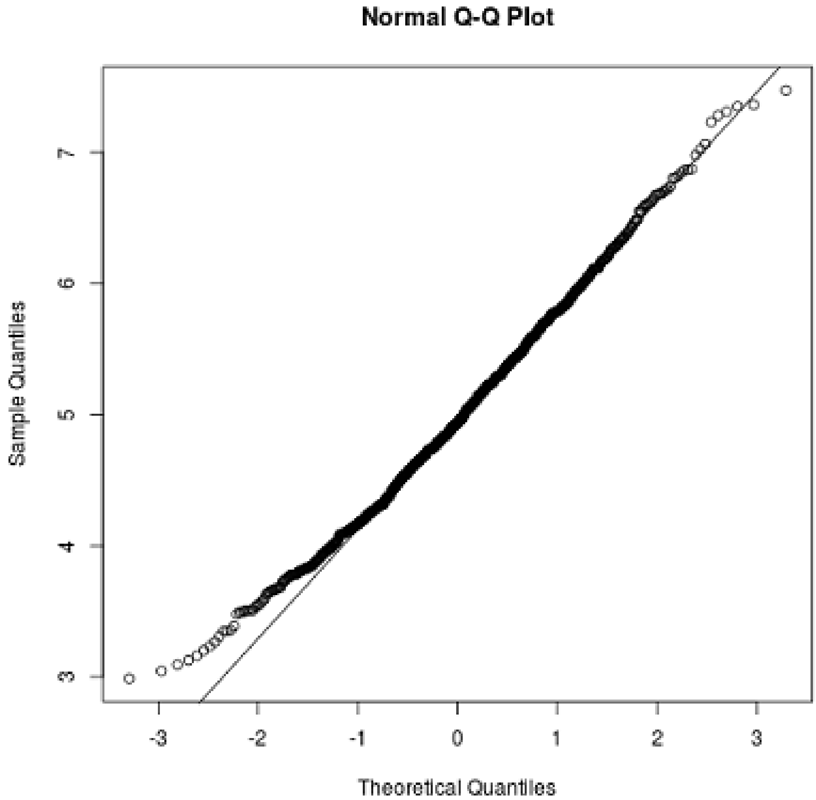


The distribution of sample means is centered at 4.9866 and the theoretical center of the distribution is λ^{-1} = 5. The variance of sample means is 0.6258 where the theoretical variance of the distribution is $\sigma^2/n = 1/(\lambda^2 n) = 1/(0.04 \times 40)$ = 0.625.

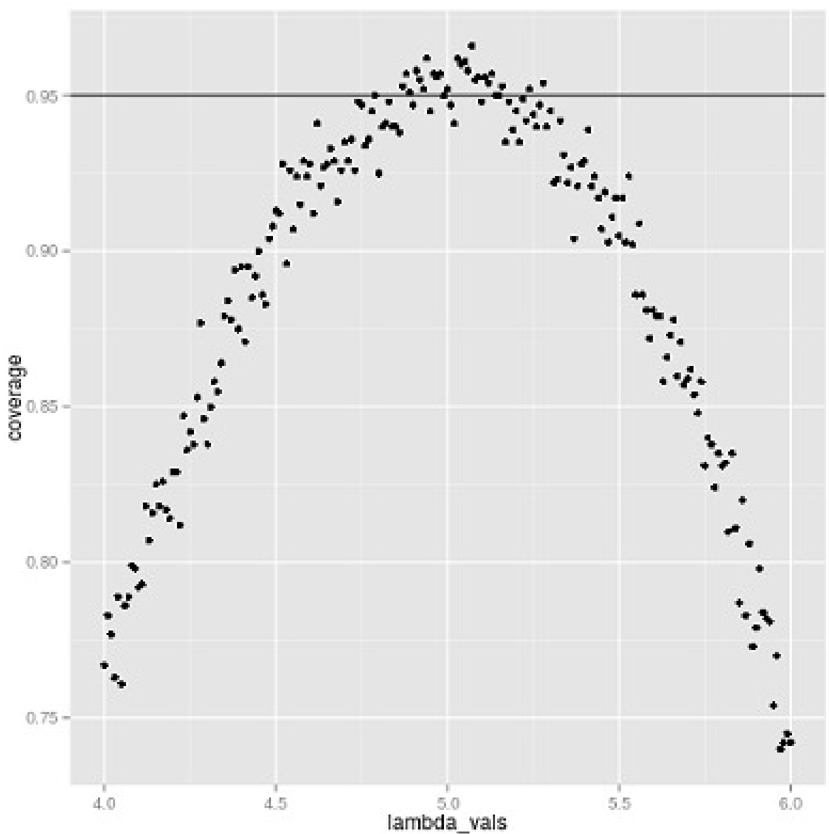
Distribution

Due to the central limit theorem, the averages of samples follow normal distribution. The figure above also shows the density computed using the histogram and the normal density plotted with theoretical mean and variance values.

Also, the q-q plot below suggests the normality.



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



From the plot above, the average of the sample mean falls within the confidence interval at least 95% of the time. 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 $\hat{\lambda}_{upp} = \hat{\lambda}(1 + \frac{1.96}{\sqrt{n}})$. Note that the true rate, λ is 5.