Statistical Inference Project

18th december 2014

Synopsis

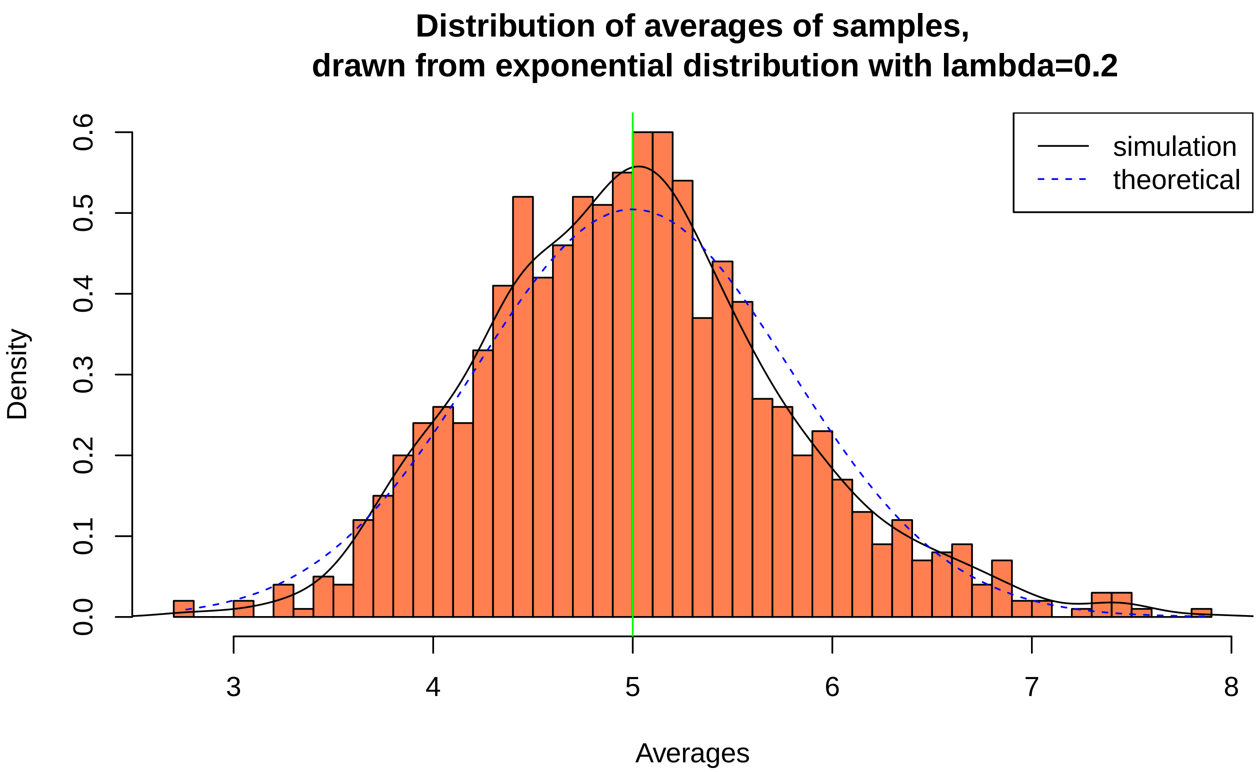
The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s. You should:

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.
2. Show how variable it is and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Data processing

The distribution of sample means is as follows:

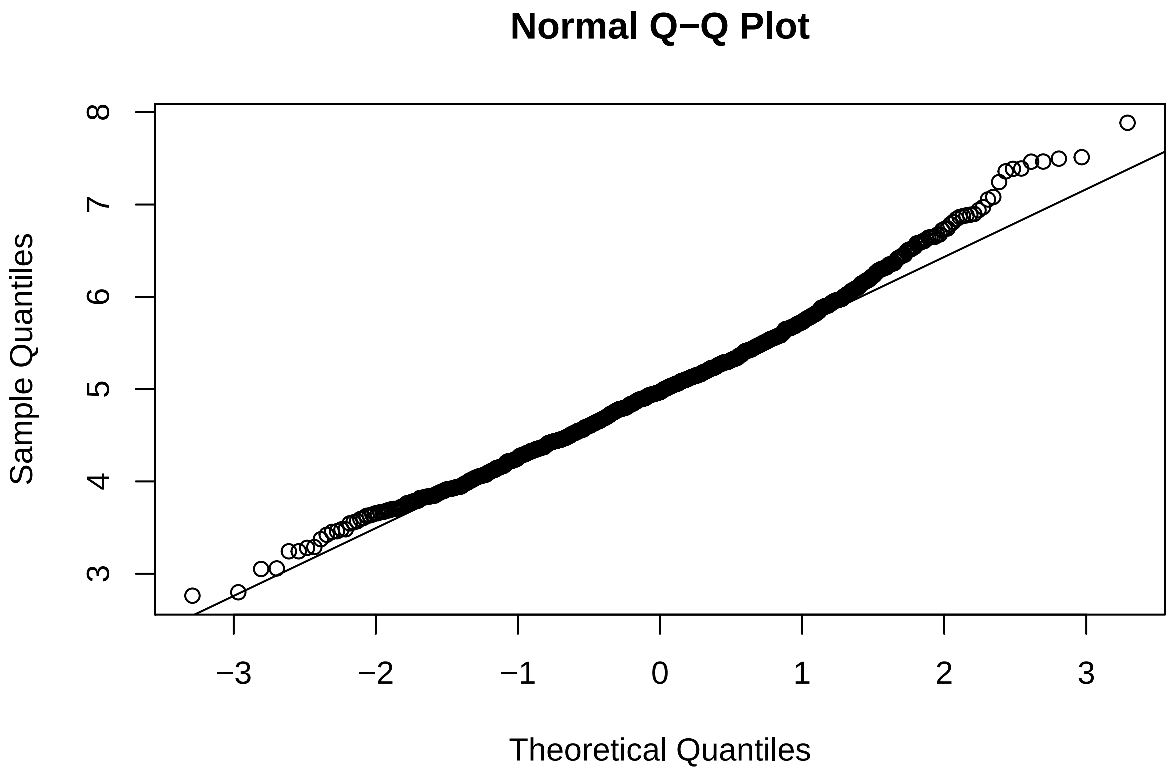



```
## [1] 5.006442
```

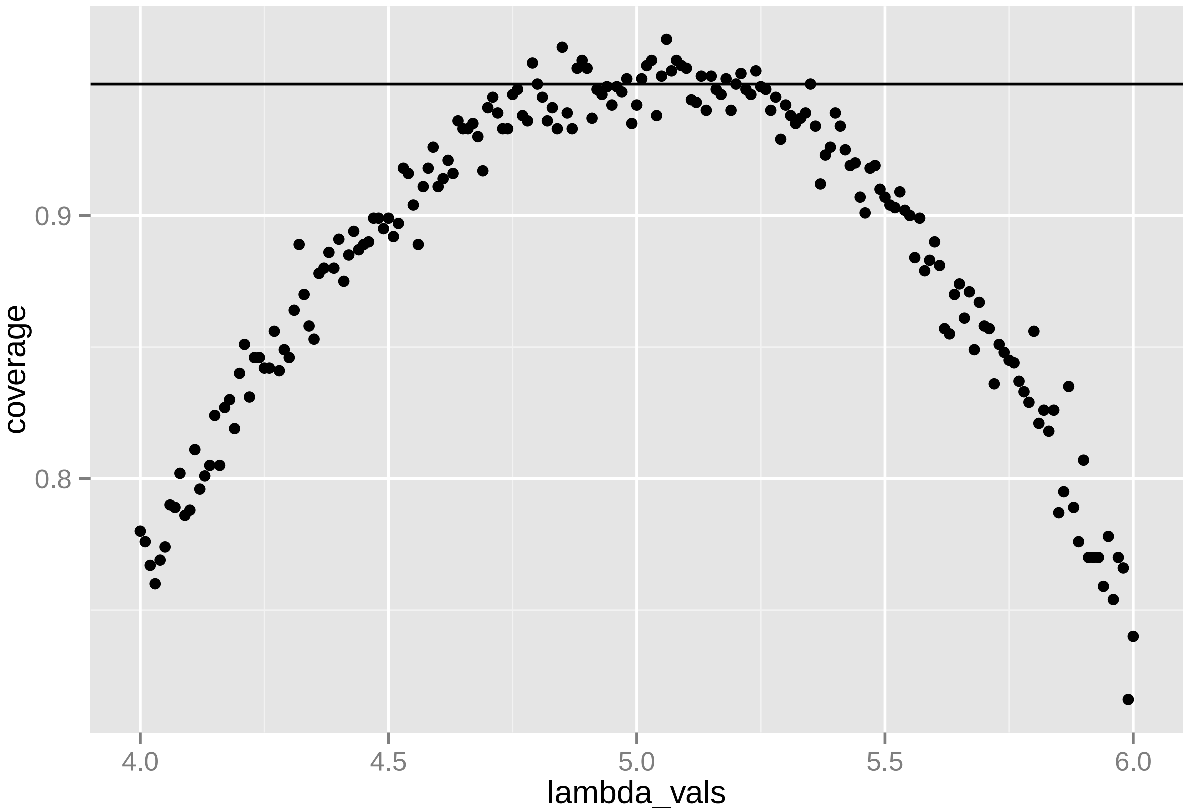
```
## [1] 0.5908158
```

The distribution of sample means is centered at 5.0064417 and the theoretical center of the distribution is λ^{-1} = r 1/lambda. The variance of sample means is 0.5908158 where the theoretical variance 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. 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 $1/\lambda = \bar{X} \pm 1.96\frac{S}{\sqrt{n}}$



Results

The center of the distribution is 5,006 and the expected center is 5,0. The mean of the means of the exponential of 1000 simulations of 40 exponential(0.2)s is 5.0064417, which is very close to the expected mean of 1/0.2 = 5.0.

Regarding the variability of the distibution. The standard deviation of 0.7686454 is also close to the expected standard deviation of 0.79056. Likewise, the variance and expected variance are 0.591 and 0.625, respectively.

The distribution of our simulations appears to be normal based on the plots shown in the Data processing section.

Appendix

R Code

simulation

set.seed(8)
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)
 # plot averages
hist(row_means, breaks=50, prob=TRUE, main="Distribution of averages of samples,

```
drawn from exponential distribution with lambda=0.2”,
xlab=“Averages”, col=“coral”)
# density of the averages of samples
lines(density(row_means))
# theoretical center of distribution
abline(v=1/lambda, col=“green”)
# theoretical density of the averages of samples
xfit <- seq(min(row_means), max(row_means), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(sample_size)))
lines(xfit, yfit, pch=22, col=“blue”, lty=2)
# add legend
legend(‘topleft’, c(“simulation”, “theoretical”), lty=c(1,2), col=c(“black”, “blue”))
qqnorm(row_means); qqline(row_means)
# confidence intervals
lambda_vals <- seq(4, 6, by=0.01)
coverage <- sapply(lambda_vals, function(lamb) {
mu_hats <- rowMeans(matrix(rexp(sample_sizenum_sim, rate=0.2),
num_sim, sample_size))
ll <- mu_hats - qnorm(0.975) * sqrt(1/lambda2/sample_size)
ul <- mu_hats + qnorm(0.975) * sqrt(1/lambda2/sample_size)
mean(ll < lamb & ul > lamb)
})
# loading dependencies and plotting
library(ggplot2) qplot(lambda_vals, coverage) + geom_hline(yintercept=0.95)
```