StatisticalInference

ArjunVenkat

October 25, 2015

Statistical inference course project

Synopsis

This is the project for the statistical inference class. In it, you will use simulation to explore inference and do some simple inferential data analysis.The project consists of two parts:

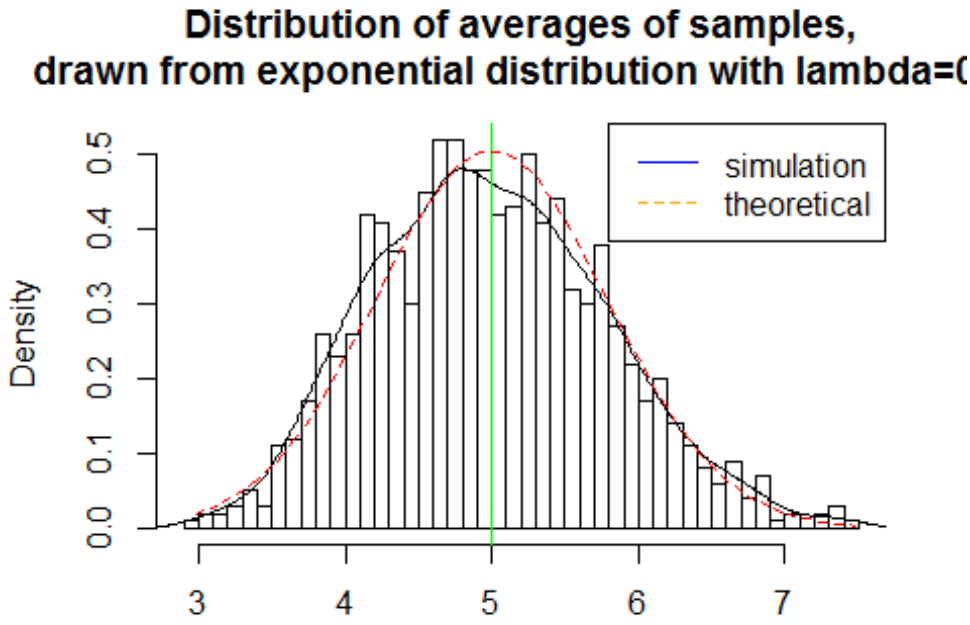
1.A simulation exercise. 2. Basic inferential data analysis. This project investigates the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution is simulated with rexp(n, lambda) where lambda is the rate parameter,theoretical mean of exponential distribution is 1/lambda and theoretical standard deviation is also 1/lambda. This project performs a thousand simulations to get the distribution of averages of 40 exponentials, where the lambda is set to 0.2 for all of the simulations. The simulated samples are used to illustrate and explain the properties of the distribution of the mean of 40 exponentials in the following ways: 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 Task We start by simulating a thousand sets of 40 exponentials using lambda 0.2 and calculate the mean for each set.

```
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)
```

Show the sample mean and compare it to the theoretical mean of the distribution.With thesimulated data, we calculate the theoretical mean and sample mean for the exponential distribution.Check the figure below,the sample mean is very close to the theoretical mean at 5: #plot the histogram of averages

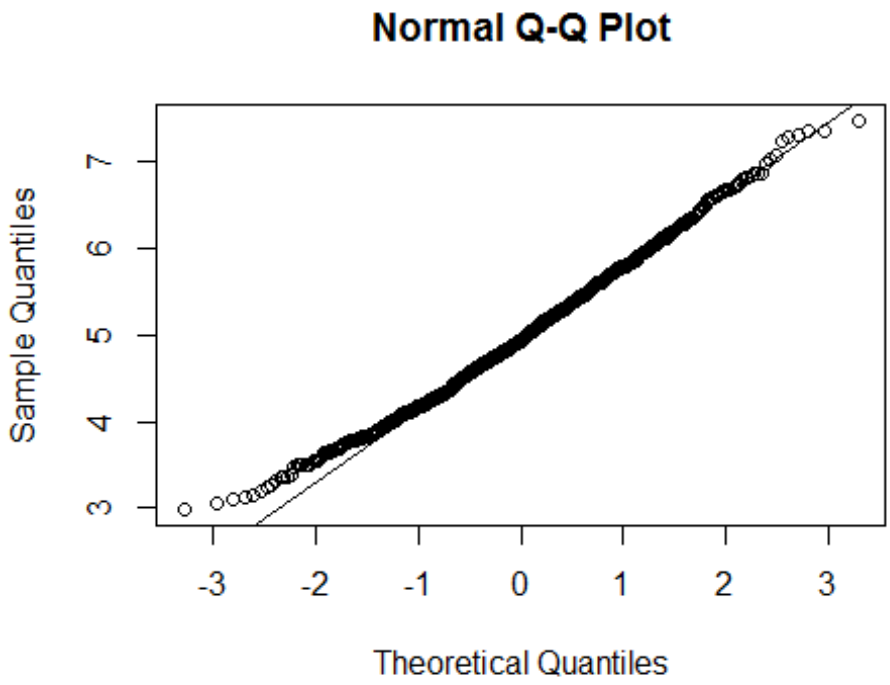
```
hist(row_means, breaks=50, prob=TRUE,
main="Distribution of averages of samples,
drawn from exponential distribution with lambda=0.2",
xlab="")
# density of the averages of samples
lines(density(row_means))
# theoretical center of distribution
abline(v=1/lambda, col="green")
# theoretical density 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="red", lty=2)
# add Legend
legend('topright', c("simulation", "theoretical"), lty=c(1,2), col=c("blue",
"orange"))
```



The distribution of sample means is centered at 4.9866 and the theoretical center of the distribution is ????????? = 5. The variance of sample means is 0.6258 where the theoretical variance of the distribution is ??????/n = ????/??????????n???? = ????/????????.????4 × 4???????? = 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.

```
qqnorm(row_means); qqline(row_means)
```

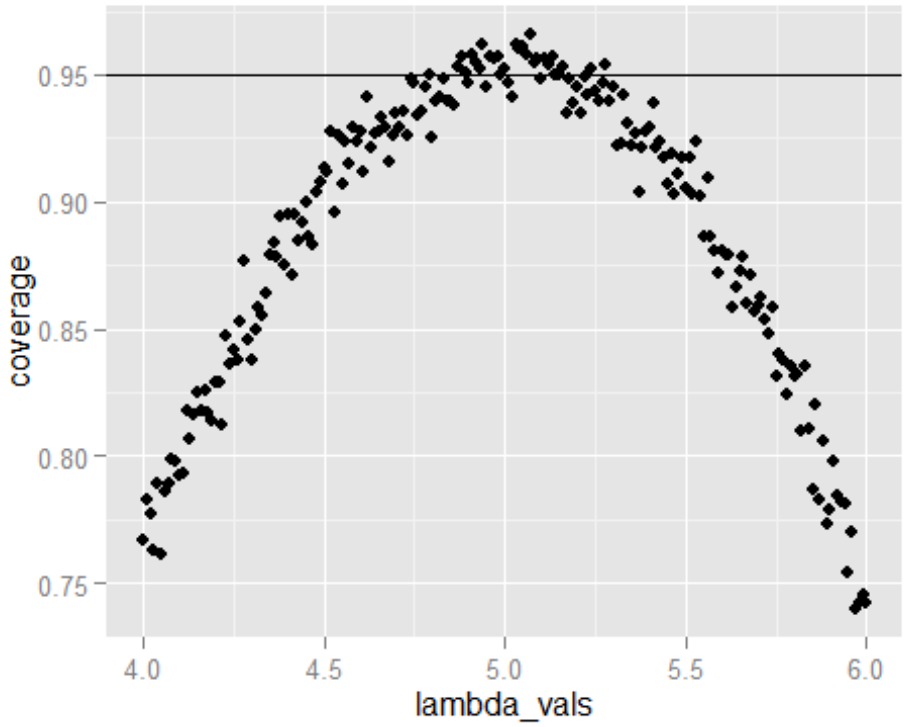


Finally, let's evaluate the coverage of the confidence interval for ????/?? = X ± ????.96 S ???n

```
lambda_vals <- seq(4, 6, by=0.01)
coverage <- sapply(lambda_vals, function(lamb) {
mu_hats <- rowMeans(matrix(rexp(sample_size*num_sim, rate=0.2),
num_sim, sample_size))
ll <- mu_hats - qnorm(0.975) * sqrt(1/lambda**2/sample_size)
ul <- mu_hats + qnorm(0.975) * sqrt(1/lambda**2/sample_size)
mean(ll < lamb & ul > lamb)
})
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.2.2

qplot(lambda_vals, coverage) + geom_hline(yintercept=0.95)
```



Answer: Due to the central limit theorem (CLT), the distribution of averages of 40 exponentials is very close to a normal distribution.