Statistical Inference - Exponential Distribution

Jeroen Zonneveld

May, 2015

Overview

In this project we will investigate the exponential distribution and compare it with the Central Limit Theorem. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$, where λ is the rate parameter. We will illustrate the properties of the distribution of the mean of 40 exponentials. We will:

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

Simulations

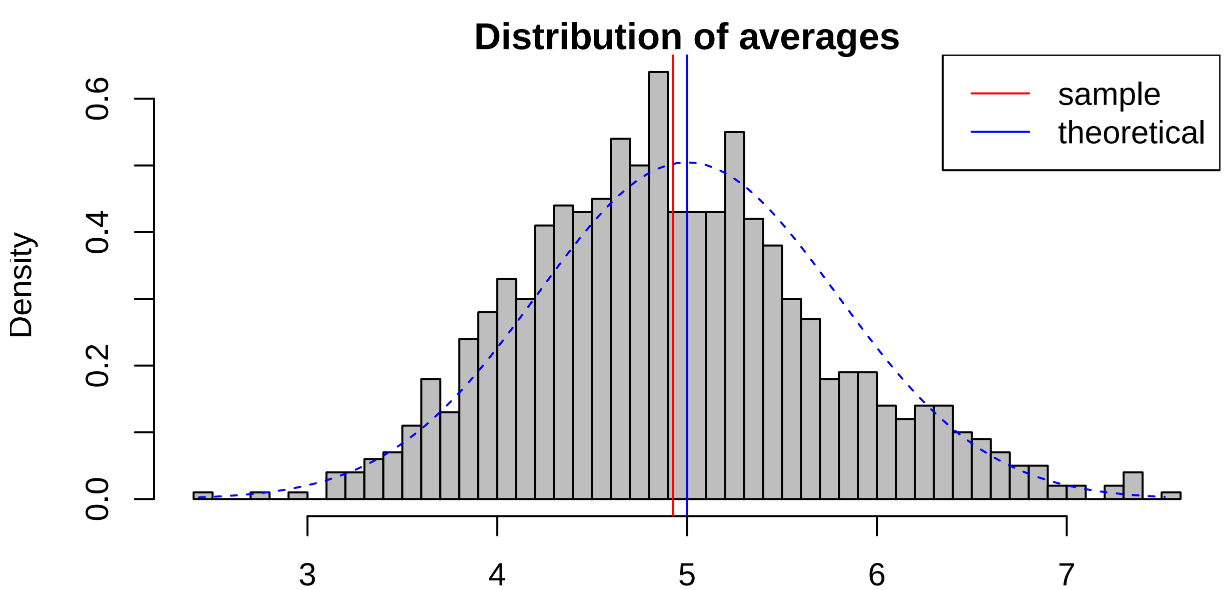
We will use the following parameters

```
set.seed(1910)
lambda <- 0.2
sample_size <- 40
simulations <- 1000
```

Now we do a thousand simulations of sample size 40. Then we calculate the means and variance for each simulation.

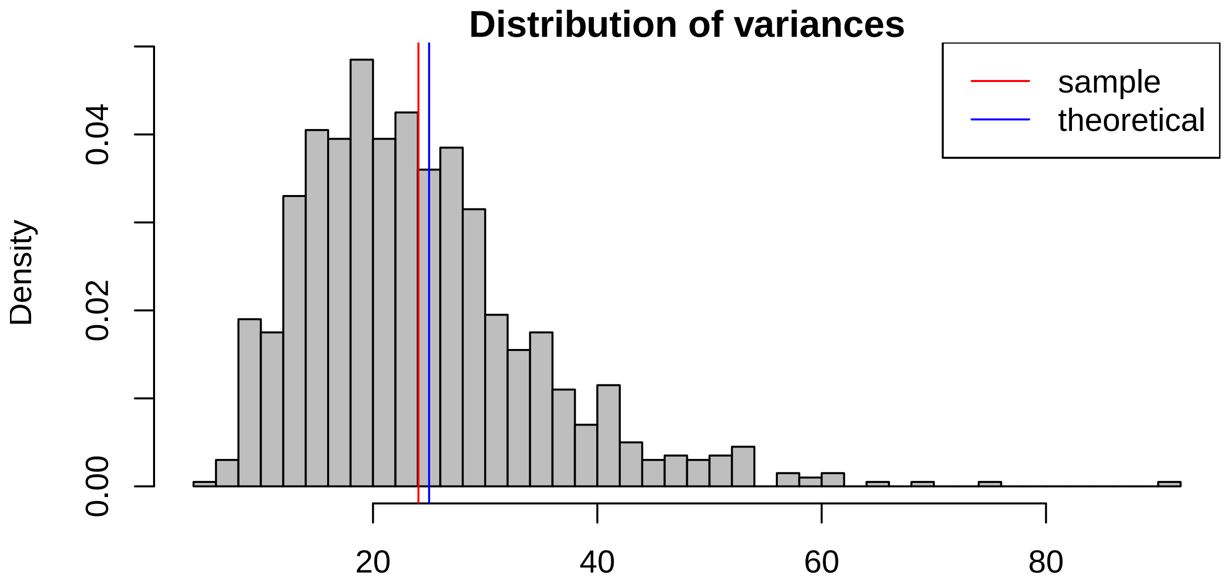
```
sim_exp <- matrix(rexp(simulations*sample_size, rate = lambda), simulations, sample_size)
row_means <- rowMeans(sim_exp)
row_variance <- apply(sim_exp, 1, var)
```

Sample Mean versus Theoretical Mean



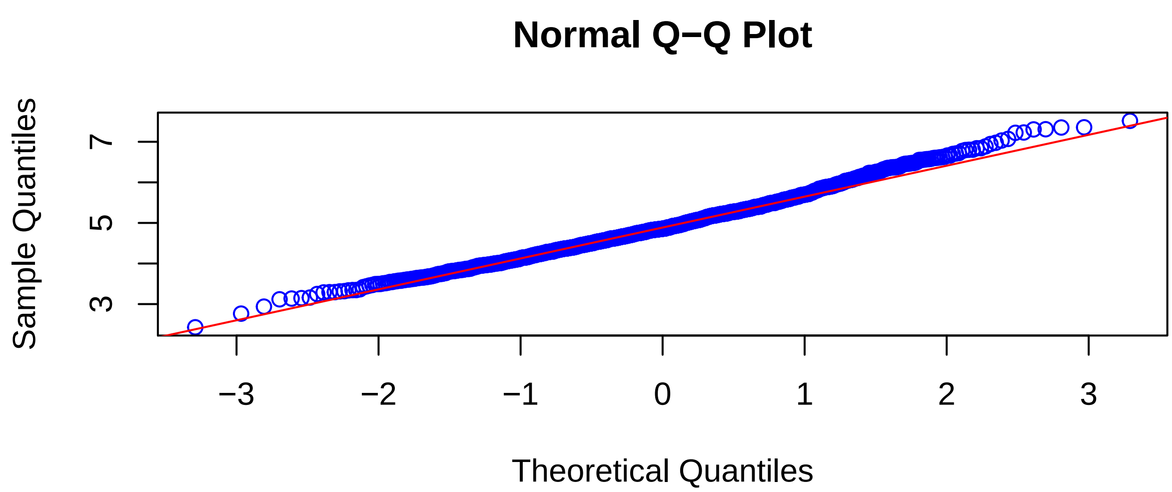
The sample mean is 4.93 while the theoretical mean is $\lambda^{-1} = 5$. The above histogram shows the distribution of averages and the sample mean in red, compared to the theoretical mean in blue.

Sample Variance versus Theoretical Variance



The mean sample variance is 24.05 while the theoretical variance is $\lambda^{-2} = 25$. The above histogram shows the distribution of variances and the mean sample variance in red, compared to the theoretical variance in blue.

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 suggests the distribution of averages of 40 exponentials is very close to a normal distribution (represented by the straight red line).

Appendix

The following code was used to make the plots above.

Averages

```
# plot the histogram of averages
hist(row_means, breaks=50, prob=TRUE,
     main="Distribution of averages",
     xlab="", col = "Grey")
# sample center of distribution
abline(v=mean(row_means), col="red")

# theoretical center of distribution
abline(v=1/lambda, col="blue")

# 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('topright', c("sample", "theoretical"), lty=c(1,1), col=c("red", "blue"))
```

Variance

```
# plot the histogram of variances
hist(row_variance, breaks=50, prob=TRUE,
     main="Distribution of variances",
     xlab="", col = "Grey")
# mean sample variance
abline(v=mean(row_variance), col="red")

# theoretical variance of distribution
abline(v=1/(lambda^2), col="blue")

# add legend
legend('topright', c("sample", "theoretical"), lty=c(1,1), col=c("red", "blue"))
```

Distribution

```
# use qqplot and qqline to compare the distribution of averages of 40 exponentials
# to a normal distribution
qqnorm(row_means, col="blue")
qqline(row_means, col = 2)
```