

# Investigation of Exponential Distribution

Author: Pieter Overdevest

## Overview

In this short report we intend to investigate the exponential distribution and compare it with the Central Limit Theorem. 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  $1/\lambda$ .

## Simulations of Means

In all simulations of the exponential distribution `lambda` has been set to 0.2. This means that we can expect an average of the means and of the standard deviation of 5. We investigate the distribution of averages of 40 exponentials. We repeat the simulation 1000 times.

First, we set a few parameters,

```
set.seed(16091970)
mean_vector <- NULL
sd_vector <- NULL
sim_vector <- NULL
lambda <- 0.2
```

Now, we are ready to get to the real deal! Run the 1000 simulations and store the means and standard deviations in a vector. All simulated data are stored in a matrix.

```
for (i in 1:1000) {
  sim <- rexp(40, lambda)
  mean_vector <- c(mean_vector, mean(sim))
  sd_vector <- c(sd_vector, sd(sim))
  sim_vector <- cbind(sim_vector, sim)
}
```

Let's see the distribution of the 1000 averages (each of 40 sampled data), see left figure below. The right figure shows the distribution of all 40,000 sampled data, just because we can. The red vertical lines indicate the average of the 1000 means. The blue dotted lines indicate the mean of the population (5).

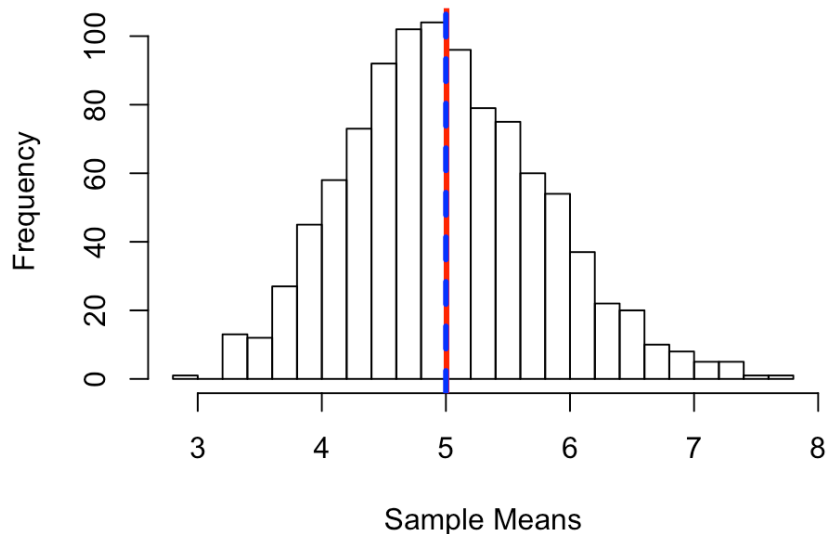
```

par(mfrow = c(1, 2))
hist(mean_vector, 25, xlab="Sample Means", main="Distribution 1,000 means means")
abline(v = mean(mean_vector), col = "red", lwd = 3)
abline(v = 1/lambda, col = "blue", lwd = 3, lty = 2)

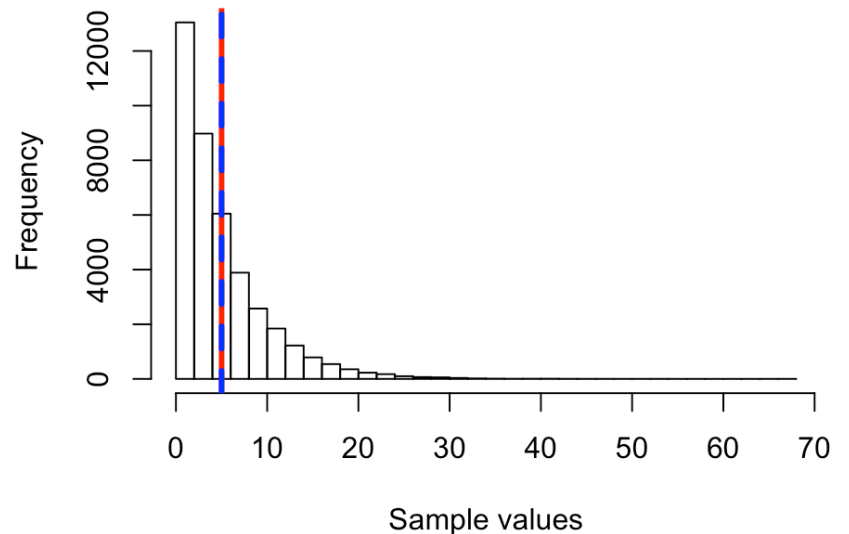
hist(sim_vector, 25, xlab="Sample values", main="Distribution 40,000 samples")
abline(v = mean(mean_vector), col = "red", lwd = 3)
abline(v = 1/lambda, col = "blue", lwd = 3, lty = 2)

```

**Distribution 1,000 means means**



**Distribution 40,000 samples**



The average of the 1000 means is very close to the mean of the population (5.01 and 5, resp). In the left figure, we see that the means more or less follow a normal distribution. The right figure shows how the original sampled data are distributed, clearly an exponential distribution (not normal)!

By plotting a so-called 'normal QQ plot', we can investigate how well the distribution of the means follows a normal distribution (left). And for the fun we also plotted the original sampled data in the QQ plot.

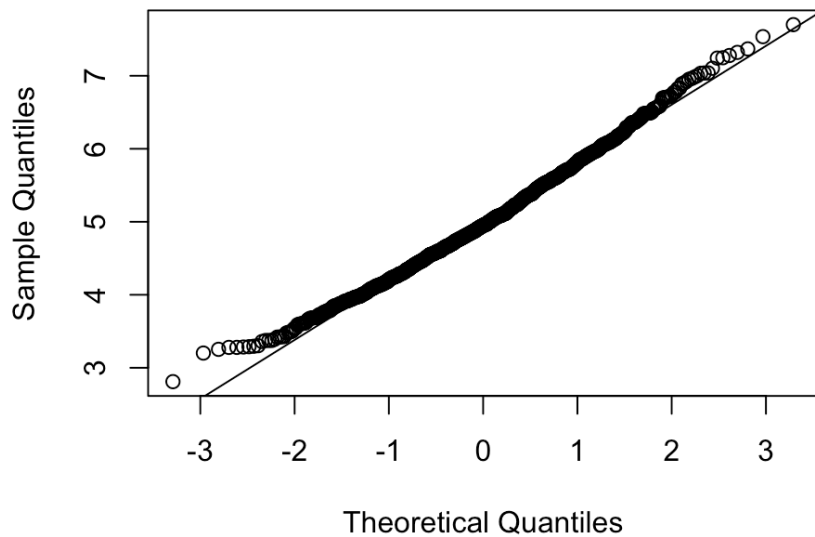
```

par(mfrow = c(1, 2))
qqnorm(mean_vector, main="Normal Q-Q Plot of 1,000 sample means")
qqline(mean_vector)

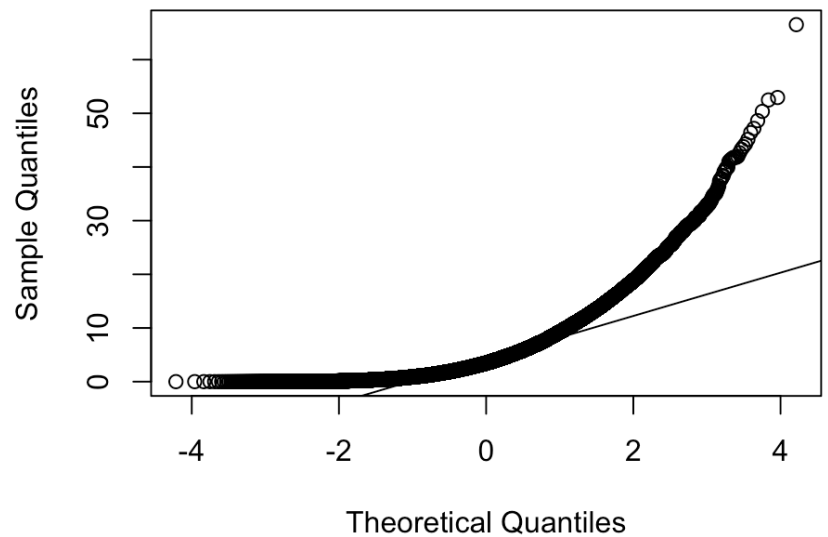
qqnorm(sim_vector, main="Normal Q-Q Plot of 40,000 samples")
qqline(sim_vector)

```

Normal Q-Q Plot of 1,000 sample means



Normal Q-Q Plot of 40,000 samples



This confirms that the 1000 means follow a normal distribution (left). As expected, the exponential distribution is not a normal distribution (right).

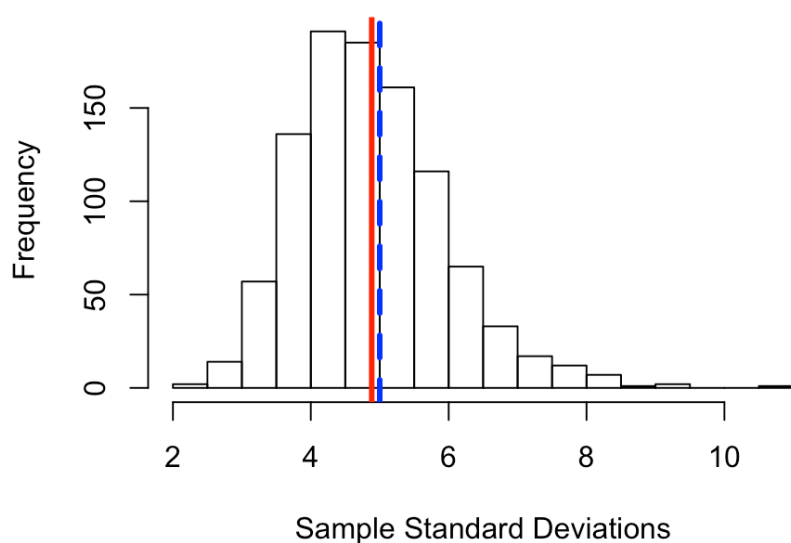
## Simulations of Standard Deviations

We do the same for the standard deviation. Distribution of the 1,000 standard deviations is shown in the left figure. In the right figure, the same data are plotted in a QQ plot. The red vertical lines indicate the average of the 1000 standard deviations. The blue dotted lines indicate the standard deviation of the population (5).

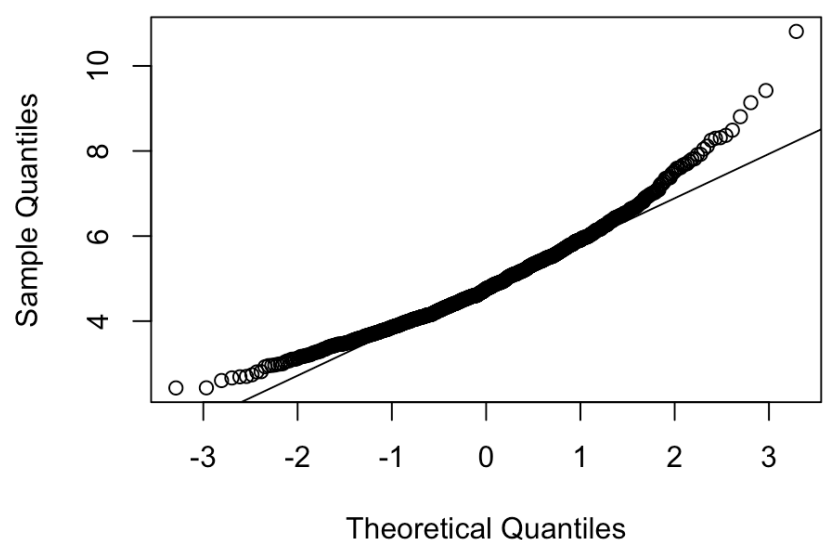
```
par(mfrow = c(1, 2))
hist(sd_vector, 25, xlab="Sample Standard Deviations", main="Distribution 1,000 sd's")
abline(v = mean(sd_vector), col = "red", lwd = 3)
abline(v = 1/lambda, col = "blue", lwd = 3, lty = 2)

qqnorm(sd_vector, main="Normal Q-Q Plot of 1,000 sample sd's")
qqline(sd_vector)
```

Distribution 1,000 sd's



Normal Q-Q Plot of 1,000 sample sd's



The average of the 1000 standard deviations is not as close to the standard deviation of the population (4.88 and 5, resp), as the average is, see above. As with the means, we see that the standard deviations approach a normal distribution. However, the distribution of means seem to be closer to a normal distribution, than the distribution of the standard deviations. You can see this by comparing the two QQ plots. Probably, for standard deviations a larger number of simulations is needed to approach a normal distribution.