Investigation of exponential distribution by simulation

Ilya Maltsev

March 10, 2017

Overview

The project is focused on simulation and properties of mean of several IID exponentials. Sample mean and sample variance of the mean are presented and compared to the theoretical ones.

Simulations

R libraries for the simulation and the presentation:

```
library(ggplot2)
```

Consider exponential distribution $f(x) = \lambda e^{-\lambda x}$. Its population mean is $\frac{1}{\lambda}$ and standard deviation is $\frac{1}{\lambda}$.

Following parameters are used for the simulation:

```
lambda <- 0.2 # Rate parameter (the distribution parameter)
sample_size <- 40 # Number of exponentials to build an average variable
sim_count <- 1000 # Number of averages for the simulation

# Random number generator is used. It need to set random seed explicitly
# for full replicability
seed = 234567890
set.seed(seed)
```

Let's generate random numbers using exponential distribution (*rexp* R function), fill 1000 samplings and evaluate mean in each sampling:

```
sim_data <- matrix(rexp(sim_count * sample_size, lambda),
                   nrow = sim_count, ncol = sample_size)
sampling_averages <- apply(sim_data, 1, mean)
```

Results

Sample Mean versus Theoretical Mean

Sample mean could be evaluated as a mean of sampling averages:

```
sample_mean <- mean(sampling_averages)
sample_mean
```

```
## [1] 5.029905
```

Theoretical (population) mean for exponential distribution is:

```
theor_mean <- 1/lambda
theor_mean
```

```
## [1] 5
```

The sample mean is pretty close to the theoretical one, the difference is 0.0299052 and just 0.5945477% of theoretical mean.

Sample Variance versus Theoretical Variance

Sample variance of a mean could be obtained from simulation data as:

```
sample_variance <- var(sampling_averages)
sample_variance
```

```
## [1] 0.6203462
```

Theoretical (population) variance is:

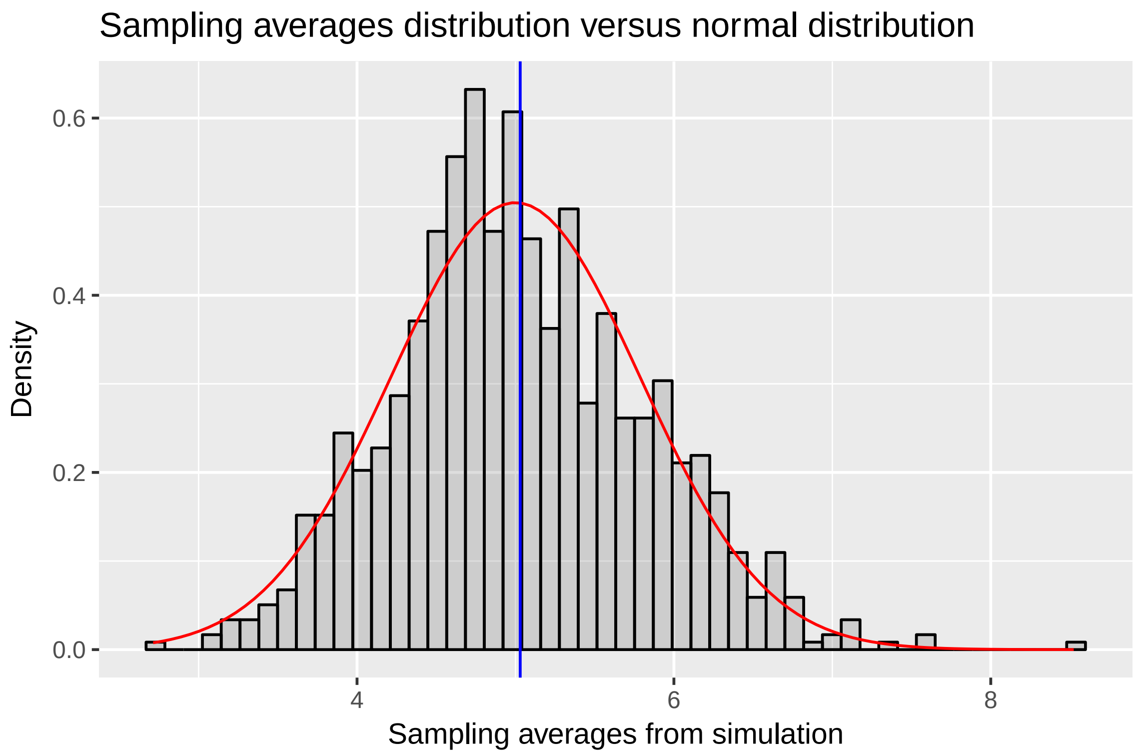
```
theor_variance <- (1/lambda)^2/sample_size
theor_variance
```

```
## [1] 0.625
```

The sample variance is pretty close to the theoretical one, the difference is 0.0046538 and just 0.7501999% of theoretical variance.

Distribution

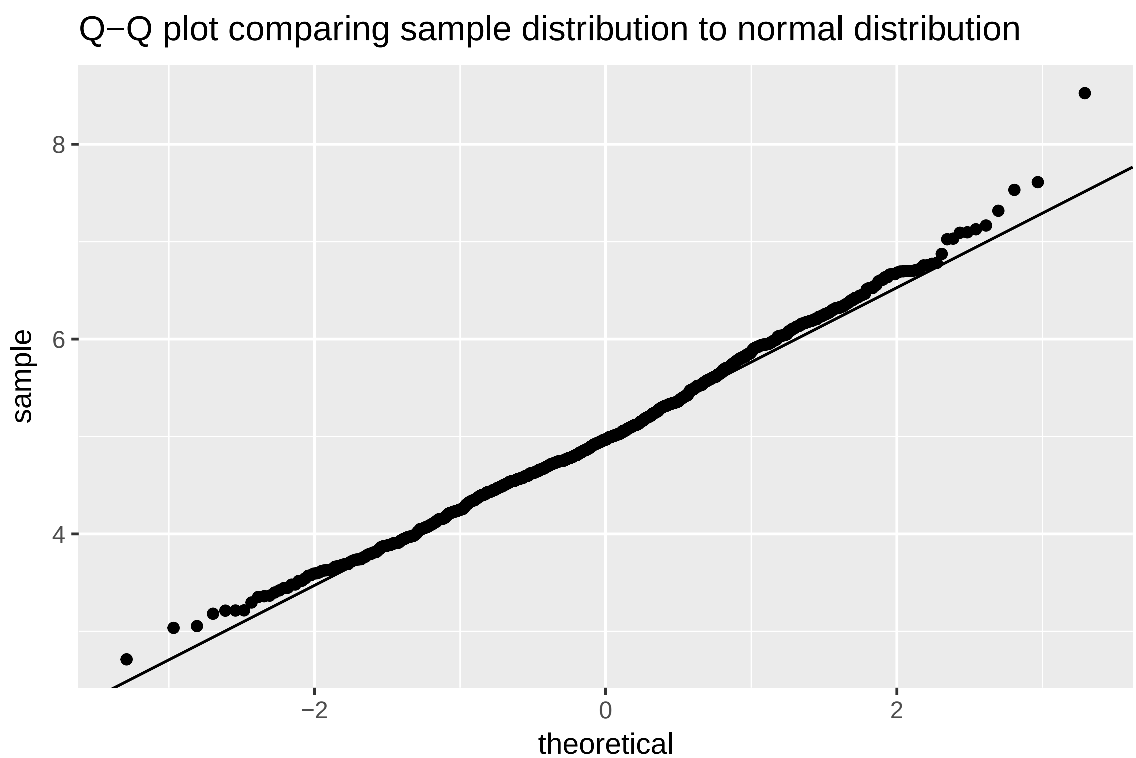
Number of samplings is large so a distibution of sampling averages must follow normal distribution according to Central Limit Theorem.



The sample distribution looks like a normal distribution.

```
sample_distribution_quartiles <- quantile(sampling_averages, c(0.25, 0.75))
normal_distribution_quartiles <- qnorm(c(0.25, 0.75))
```

```r
slope <- diff(sample_distribution_quartiles)/diff(normal_distribution_quartiles)
ggplot(data.frame(sampling_averages = sampling_averages)) +
    stat_qq(aes(sample = sampling_averages)) +
    geom_abline(slope = slope, intercept = theor_mean) +
    ggtitle("Q-Q plot comparing sample distribution to normal distribution")
```



The Quantile-Quantile plot shows that sample quantiles corresponds to normal distribution quantiles closely.

Conclusion

The simulation shows that considered population properties (mean and variance) are very close to theoretical ones. The simulated distribution of averages of exponentials is close to normal distribution. This is proved by quantile-quantile plot for the simulated distribution.