Statistical Inference Course Project part 1

Jure Bordon

Sunday, February 22, 2015

Overview

This project tests our understanding of distributions, standard deviation, variances, etc. and is a par to Statistical Inference course from Coursera. In the project, we will simulate 40 exponential distributions and calculate the distribution of their means, the variance of obtained distribution and compare it to the normal distribution.

```
set.seed(505) # in order results to be reproducible, I set the seed to a value 505
```

We set parameters for our exponential distribution:

```
lambda <- 0.2
nosim <- 1000
n <- 40
```

Run the simulations (1000x40 matrix):

```
exp_simulated <- matrix(rexp(nosim*n,lambda),nosim)
```

... and calculate the means of simulated exponentials:

```
exp_sim_means <- apply(exp_simulated,1,mean)
```

Question 1

Show the sample mean and compare it to the theoretical mean of the distribution! Our calculated distribution mean:

```
sim_mean <- mean(exp_sim_means)
```

```
## [1] 5.001057
```

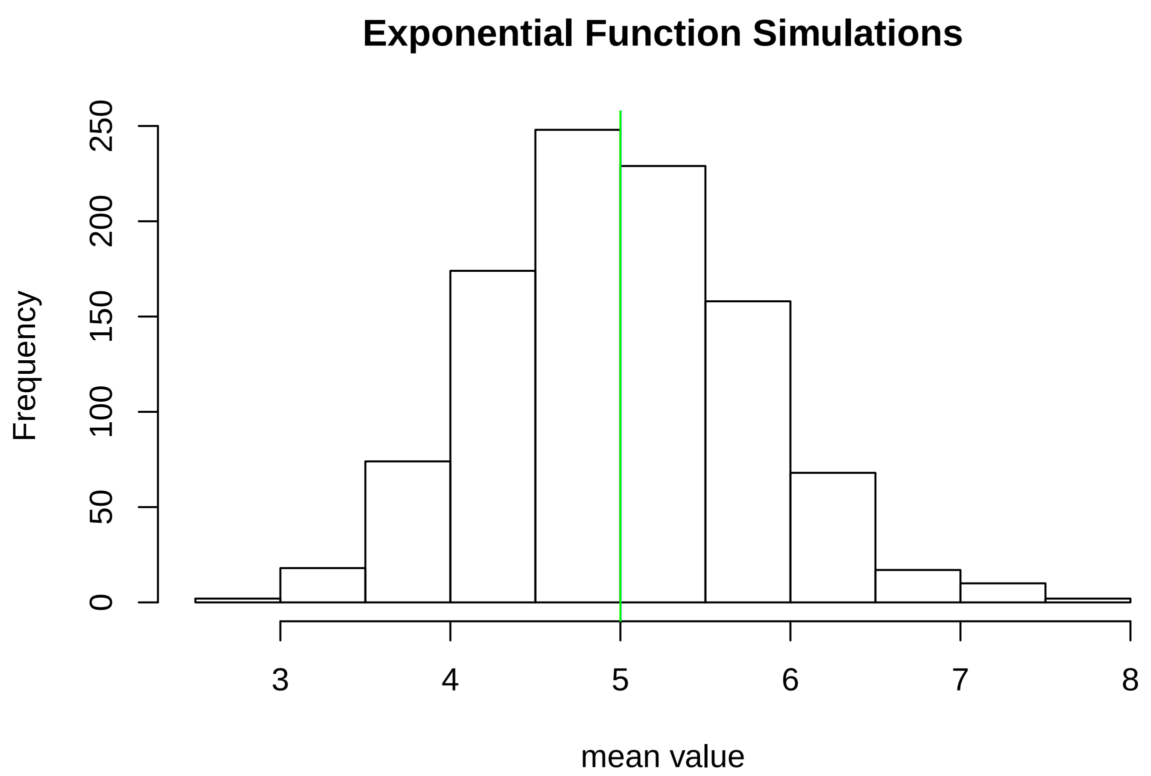
... and calculated theoretical mean:

```
theoretical_mean <- 1/lambda
```

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

In order to show the results, we plot a histogram with calculated (blue vertical line) and theoretical (green vertical line) means:

```r
hist(exp_sim_means, xlab = "mean value", main = "Exponential Function Simulations")
abline(v = sim_mean, col = "blue")
abline(v = theoretical_mean, col = "green")
```



The mean we obtained by simulating 40 exponential distributions is 5.001057, while the theoretical mean is 5. The center of distribution of simulated exponentials is therefore very close to the theoretical mean.

Question 2

Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

In order to obtain variance, we first calculate the standard deviation of our exponential means distribution from the 1000 simulations:

```r
exp_sim_sd <- sd(exp_sim_means)
```

```
## [1] 0.7749455
```

And now the variance:

```r
exp_sim_var <- exp_sim_sd^2
```

```
## [1] 0.6005405
```

… which the same as:

```
exp_sim_var <- var(exp_sim_means)
```

```
## [1] 0.6005405
```

Theoretical standard deviation is calculated by the expression

```
exp_sim_sd_theoretical <- (1/lambda)/sqrt(n)
```

```
## [1] 0.7905694
```

… and theoretical variance is standard deviation squared

```
exp_sim_var_theoretical <- ((1/lambda)/sqrt(n))^2
```

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

Standard deviation of our simulated exponentials is 0.7749455, with the theoretical standard deviation of 0.7905694. Variances obtained from simulations and theoretical calculations (sandard deviation squared) are 0.6005405 and 0.625, respectively. If we used more simulation runs and larger sample sizes, the calculated numbers would be even closer to the theoretical ones (CLT).

Question 3

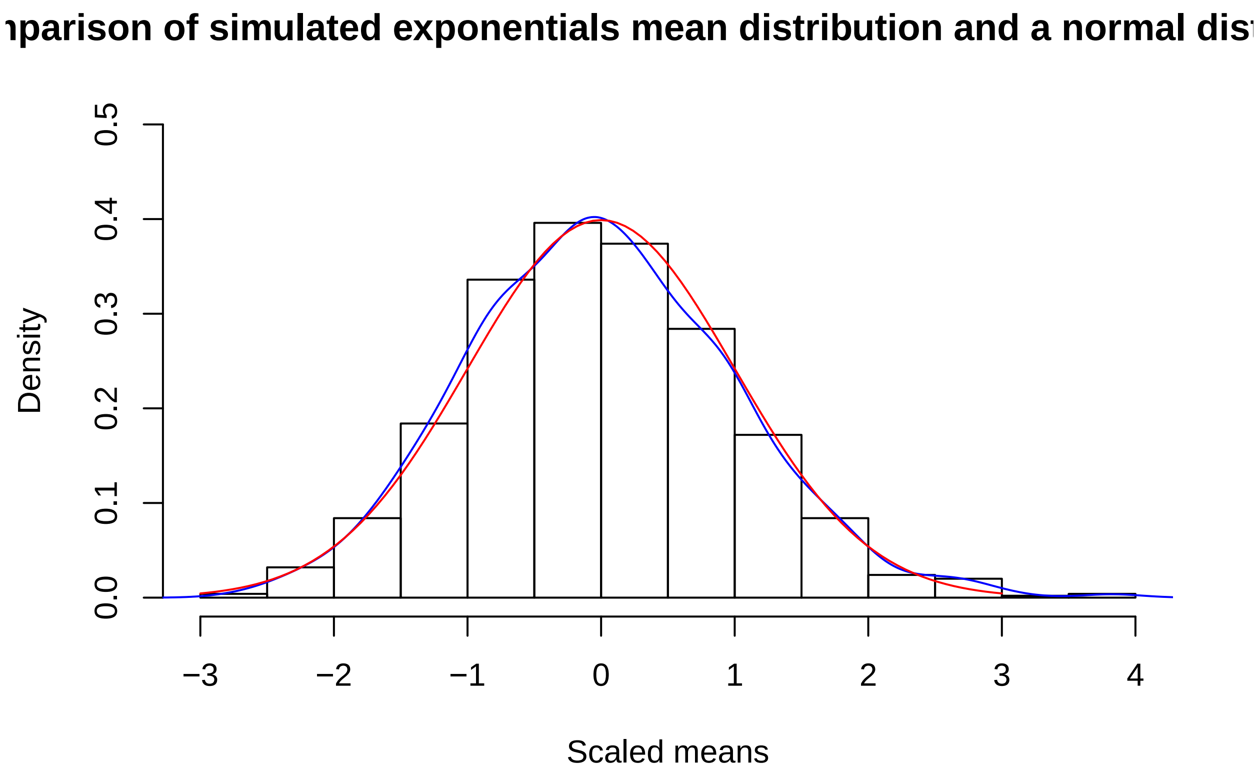
Show that the distribution is approximately normal.

scale() function comes in handy when plotting the distribution:

```
exp_sim_means_scaled <- scale(exp_sim_means)

hist(exp_sim_means_scaled,
     probability=T,
     main="Comparison of simulated exponentials mean distribution and a normal distribution",
     xlab="Scaled means",
     ylim=c(0, 0.5))
lines(density(exp_sim_means_scaled), col="blue") # distribution of our means

curve(dnorm(x,0,1), -3, 3, col="red", add=T) # normal distribution
```



The normal distribution (blue line) is fairly similar to the distribution of our means, so we can say our distribution is approximately normal.