Statistical Inference Course project

Vadzim Dabravolski

This report is generated as a result Inference Course Project. The main theme of the project is to discover how distribution parameters of samples are connected to theroretical parameters of the given distribution.The simulation was done using R-language and the source file of calcualtions is located in author's git repo.

For the set of experments Exponential Distribution with parameters **rate=0.2** and **n=40** was chosen. This distribution has following parameters:
- **theoretical mean** is equal to *1/rate*
- **theoretical variance** is equal to *mean^2*

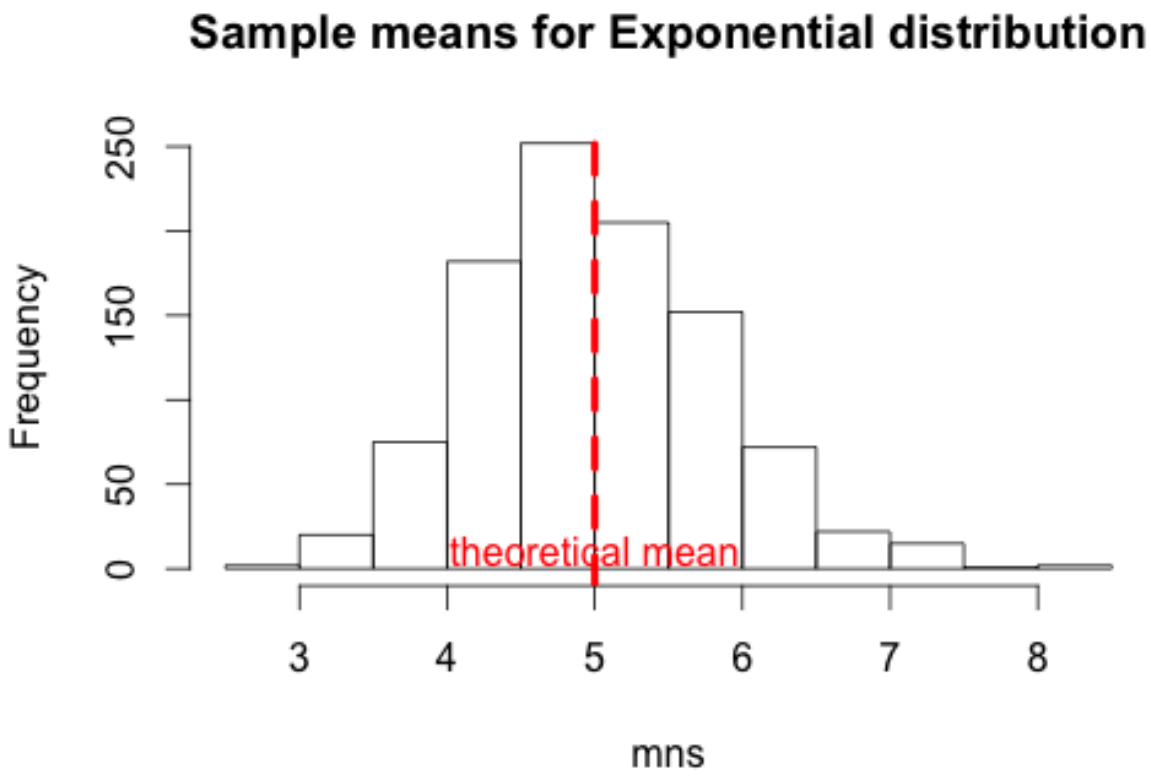
TASK 1: "Show the sample mean and compare it to the theoretical mean of the distribution."
First of all, let's simulate Exponential Distribution (with 1000 observation) and calcualte mean of samples:

```
for (i in 1 : 1000) mns = c(mns,mean(rexp(n, rate)))
```

Then, let's create the plot to visually compare means of the samples and the theoretical mean of the distribution . Following R-code generates the appropriate histogram:

```
hist(mns, main="Sample means for Exponential distribution")
abline(v=th_mean,col="red",lty="dashed", lwd=4)
text(th_mean,10, "theoretical mean", col = "red")
```

The following histogram is being generated:



Task summary
Based on the instagram we can make following conclusions which are aligned with statistical theory:
- Distribution of means is close to Normal distribution;
- Distirbution is centered around theoretical mean (red line).

TASK 2: "Show the sample variance and compare it to the theoretical variance of the distribution."

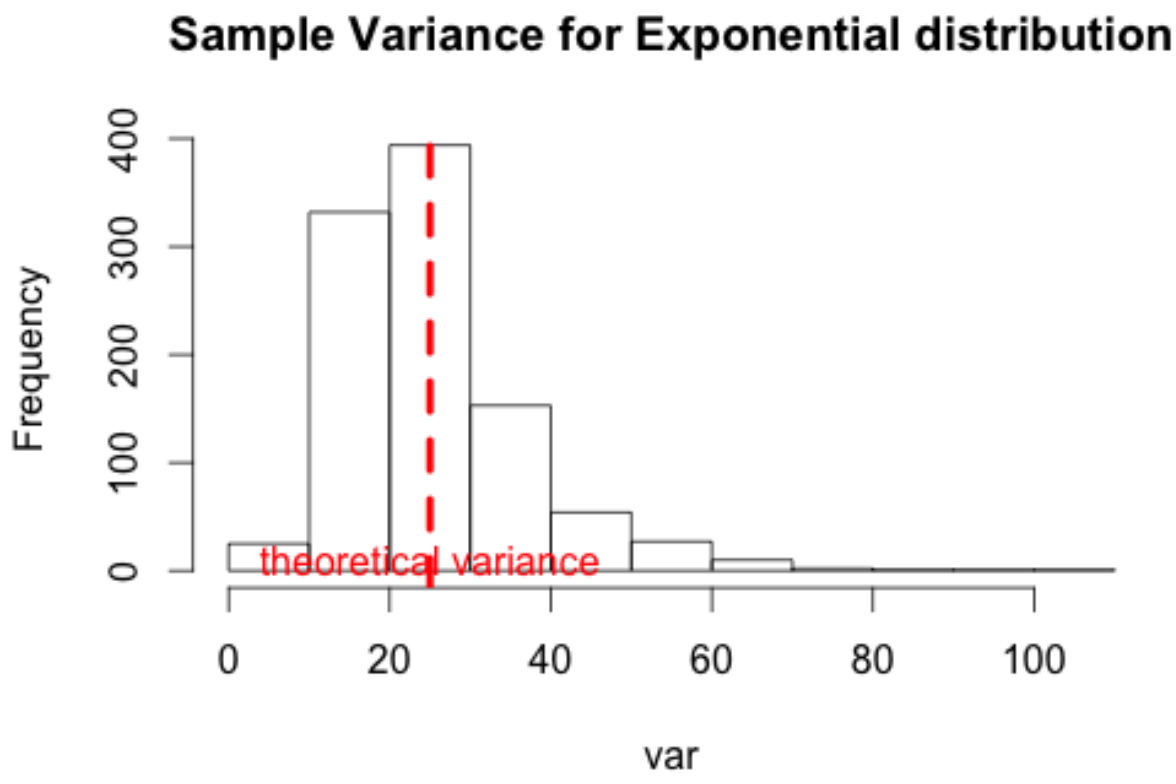
First of all, let's simulate Exponential Distribution (with 1000 observation) and calcualte variance of all samples:

```
for (i in 1 : 1000) {
  exp <- rexp(n, rate)
  sample_mean <- mean(exp)
  # calculating the variance of each sample.
  sample_variance <- sum((exp - sample_mean)^2)/(n-1)
  var = c(var,sample_variance)
}
```

Now we need to generate histogram using following R-code to visualize distribution of sample variances:

```
hist(var, main="Sample Variance for Exponential distribution")
abline(v=th_variance,col="red",lty="dashed", lwd=4)
text(th_variance,10, "theoretical variance", col = "red")
```

The following histogram is being generated:



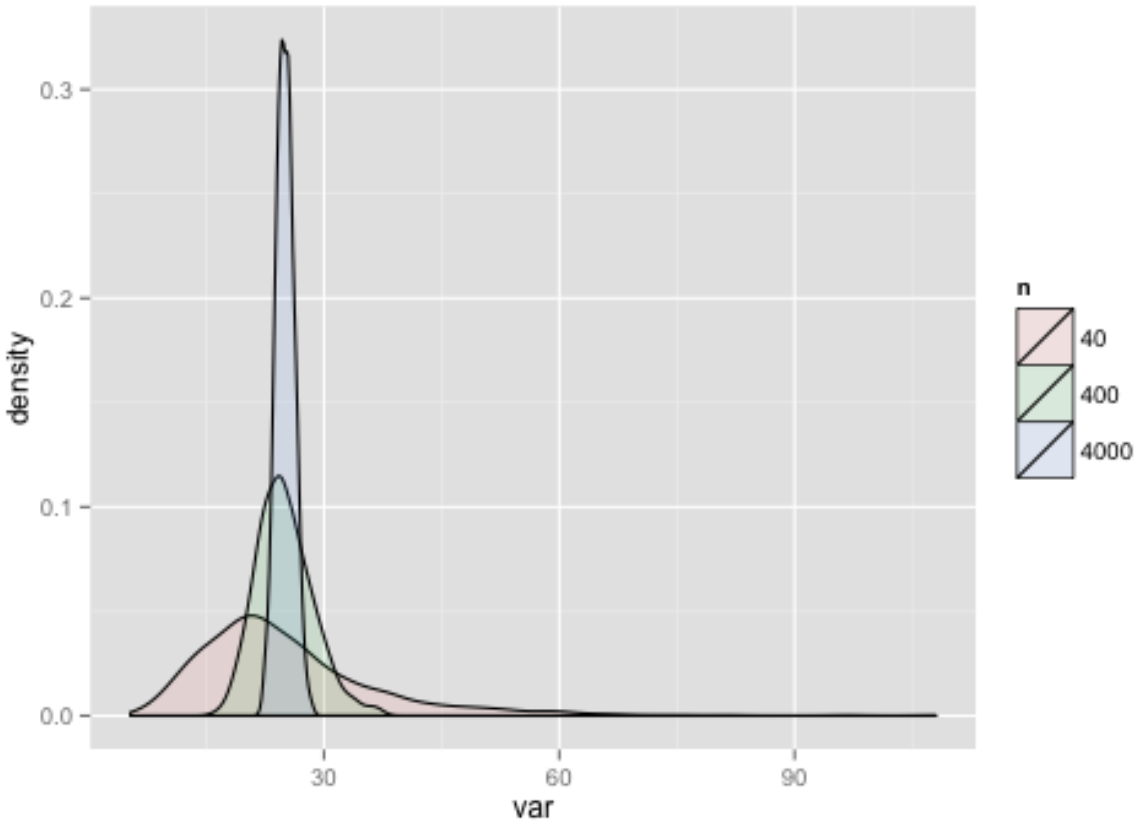
Task summary
Based on the instagram we can make following conclusions which are aligned with statistical theory:
- Distribution of sample variances is visually centered around theoretical variance (red line).

Task 3. Show that the distribution is approximately normal.
In order to confirm that dsitrubtion of variances is approximately normal let's see how the dsitribution function will behave if we increase amount of samples. For this we will run 3 simulations with n=40, n=400 and n=4000.

The following R-code creates the general plot to represent 3 simulations at the same time:

```
 # merging resutls of 3 simulations into 1 data frame
variance_mr <- rbind(n_s,n_m,n_l)
ggplot(variance_mr, aes(var, fill = n)) + geom_density(alpha = 0.1)
```

The following plot is being generated:



Task summary
Based on the histagram we can make following conclusions which are aligned with statistical theory:
- With increase of n the distribution of variance is leaning towards "Bell" shape (indicator of Normal distribution)
- For large n (in our case 4000) distribution functional is leaning to it theoretical meaning.