Checking CLT via exponential samples

Reza

April 14, 2016

Overview

In this report we will validate the law of large numbers and central limit theorem (CLT) via a huge random sample of exponential distribution.

Simulation

First let us simulate 1000 simulations of exponential samples each having 40 elements with λ equal .2. Recall that the mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$.

I simulated 40000 values and then arranged then in a 1000x40 matrix to have 1000 samples each of 40 values.

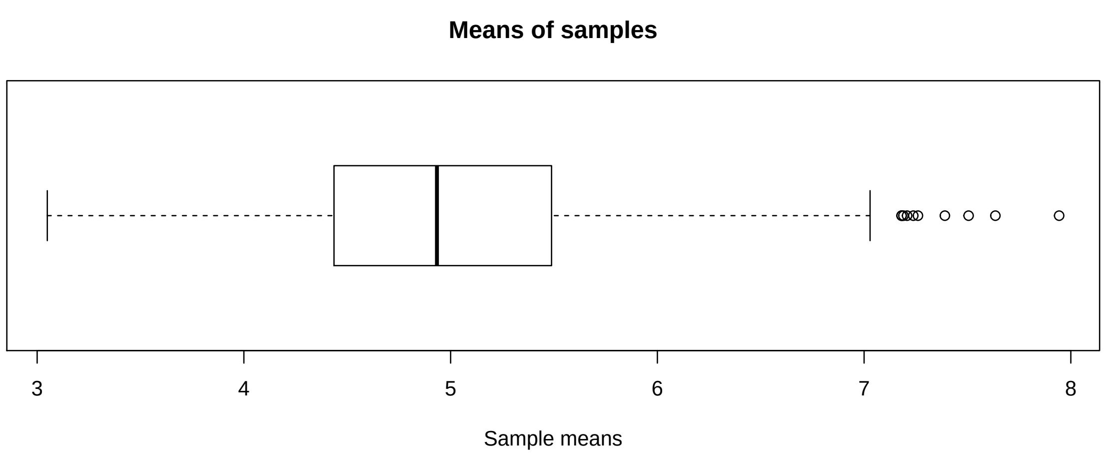
```
#simulating 40000 of exponential distributions
set.seed(200)
exp <- rexp(40000, .2)
#arraging them in 1000 samples each of quantity 40
expmat <- matrix(exp, 1000, 40)
```

Sample mean vs. the theoretical mean

In this section we will examine if the sample mean and compare it to the distribution mean which is 5. First compute the means of samples.

```
means <- apply(expmat, 1, mean)
```

The mean of means is equal 4.984, quite close to expected mean 5. Now let's look at boxplot of means. It is centered around 4.93 with half of the numbers between 4.44 and 5.49



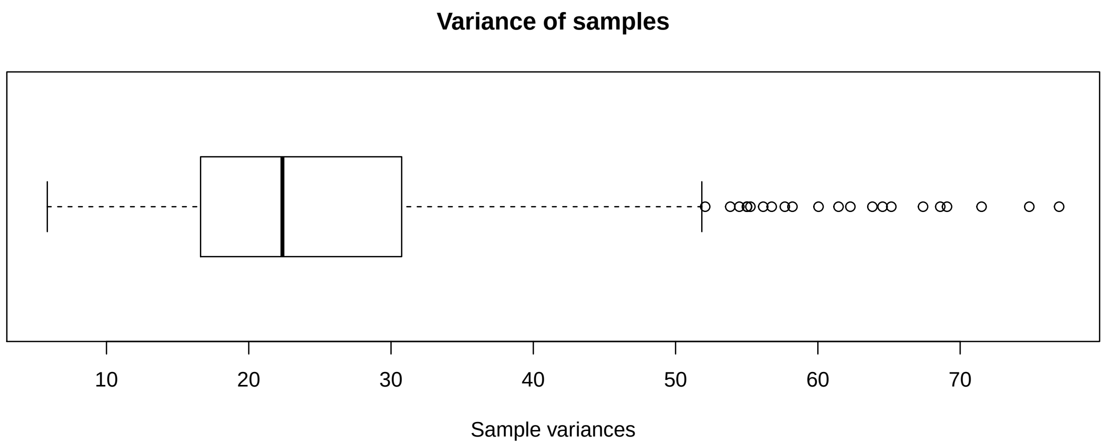
Finally running a t-test one can see that the 95% confidence interval is $(4.935, 5.033)$ apparently containing 5 the distribution's mean.

Sample variance vs. the theoretical variance

One can do similar test done in the last section on the variance of the samples. Note that the expected variance is 25. Fist we compute the samples variances.

```
vars <- apply(expmat, 1, var)
```

The average variance is 24.699 quite close to 25. Now drawing a boxplot of sample variances.



It can be seen that the plot is centered at 22.37 with half of the data between $16.63, 30.75$. Not that far from the expected value of 25.

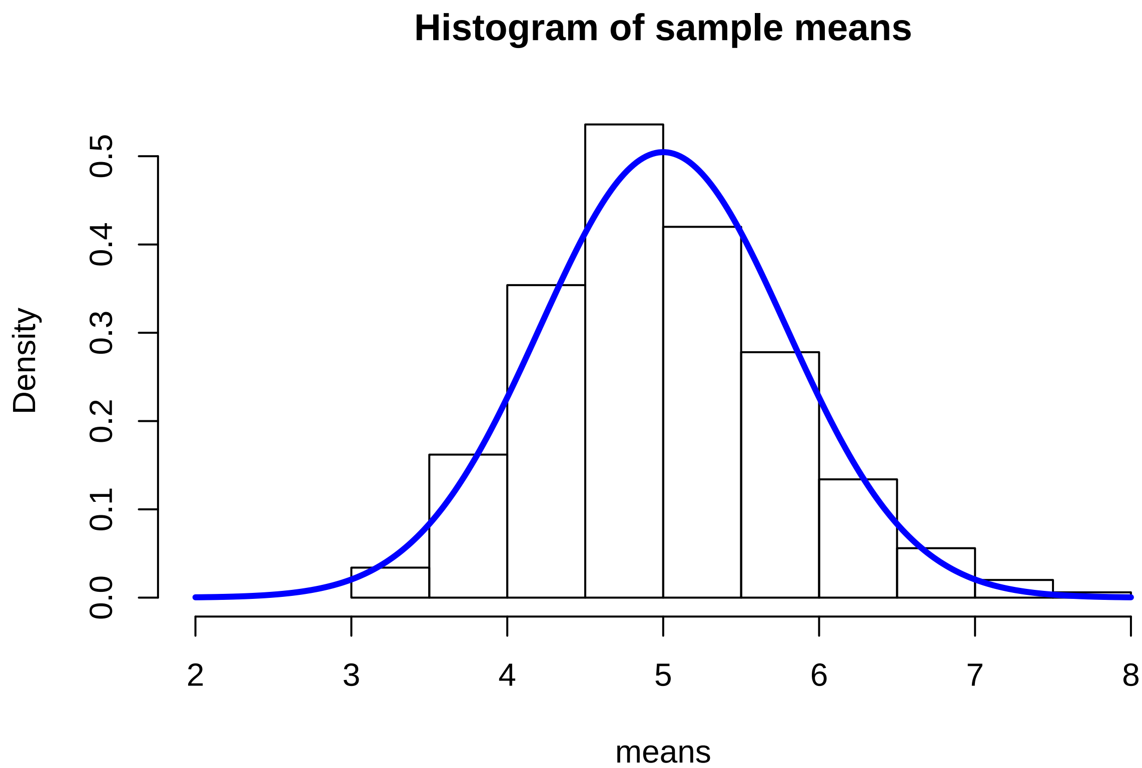
Finally let's see the 95% t-confidence interval. It is $(24.02, 25.38)$, 25 is in it so the sample's variance is a good approximation of the distribution variance.

Distribution

In this section we will test the CLT. i.e. we are going to see if the sample means actually obey a normal distribution. Recall that CLT states that as the number of samples tends to infinity the distribution of means of iid samples tends to be $N(\mu, \sigma^2/\sqrt(n))$ where μ and σ are sample mean and standard deviation and n the sample size. In our case $\mu = \sigma = 5$ and $n = 40$. So by the CLT we expect sample means to be almost normally distributed with mean 5 and standard deviation 0.79.

Let's compare the plot of means and normal distribution.

```
hist(means, freq = F, xlim = c(2,8), main = "")
x <- seq(2, 8, length=300)
hx <- dnorm(x, mean = 5, sd = 5/sqrt(40))
lines(x,hx, type = "l", col = "blue", lwd = 3)
title(main = "Histogram of sample means")
```



As you can see the plot of our samples' mean acceptingly matches the normal distribution which is drawn in blue.