Assignment_P1

Marcelladane

18 August 2017

Statistical Inference Course Project

In this project you will investigate the exponential distribution in R 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. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should - Show the sample mean and compare it to the theoretical mean of the distribution. - Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. - Show that the distribution is approximately normal. In point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

Simulation

Start by loading the necessary packages, here, just one for plot.

```
packages <- c("ggplot2")
sapply(packages, require, character.only = TRUE, quietly = TRUE)
```

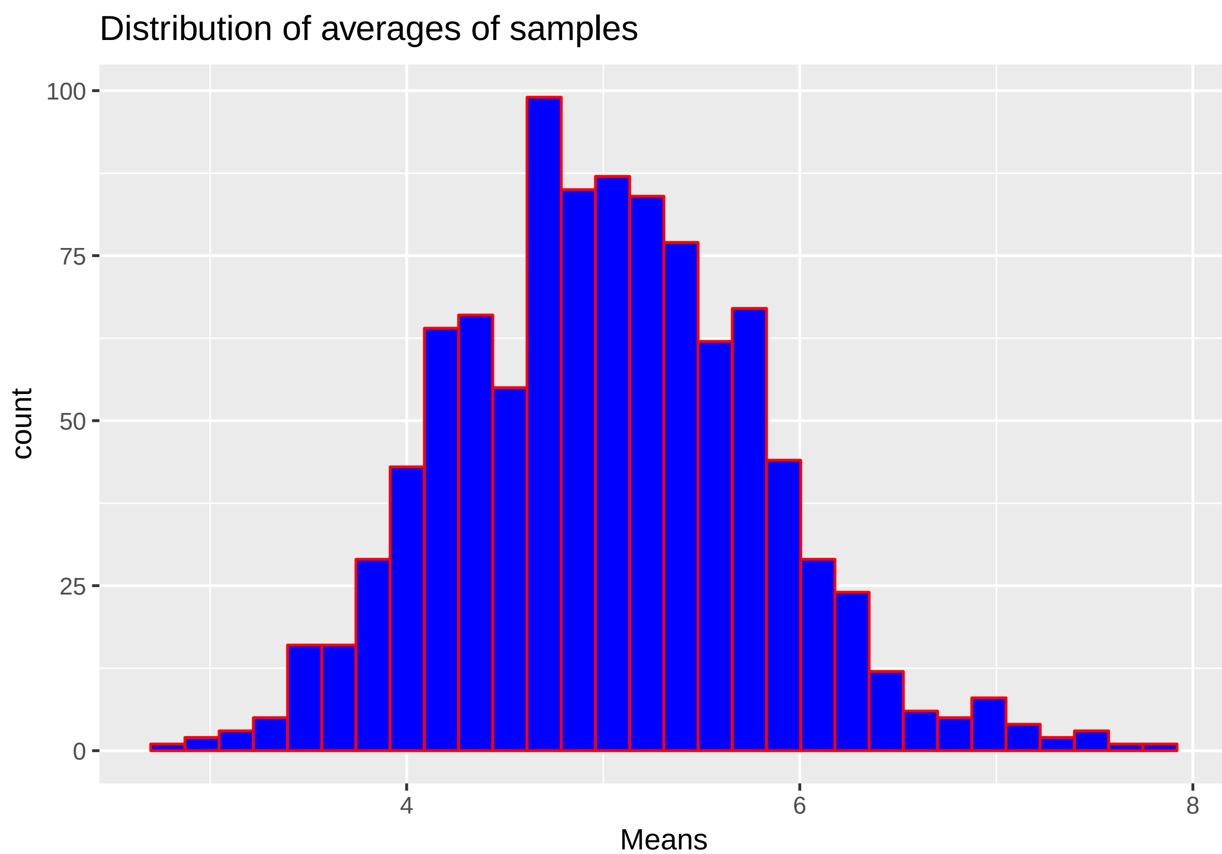
```
## ggplot2
##    TRUE
```

Now coming to the simulations, first let's set a seed (to create a set of values) and then set the needed constants (lambda, n and number of simulations, in this order)

```
set.seed(123)
lambda <- 0.2
n <- 40
NS <- 1000
```

Now run the exponential distribution and plot it. Calculate the means of distribution. Plot a histogram with the mean values.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Comparing Sample values versus Theoretical values

MEANS

The expected mean μ (mu) of a exponential distribution of rate λ (lambda) is:

mu = 1 / lambda

And

Call X (in the classes was X line, but here I call X) the average of your 1000 simulations (NS) of 40 samples (n), so the theoretical mean is the mean of the means.

```
mu = 1 / lambda
mu
```

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

```
MofMeans <- mean(ExpD_means$means)
MofMeans
```

```
## [1] 5.011911
```

They are not completely identical, but quite close means.

VARIANCE

The expected standard deviation σ (sigma) of a exponential distribution of rate lambda is:

sigma = (1 / lambda)/ sqrt(n)

And

the variance (Var) of sigma is: Var=sigma^2 (to calculate the SD for the second example I will call sigma, SigmaX)

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

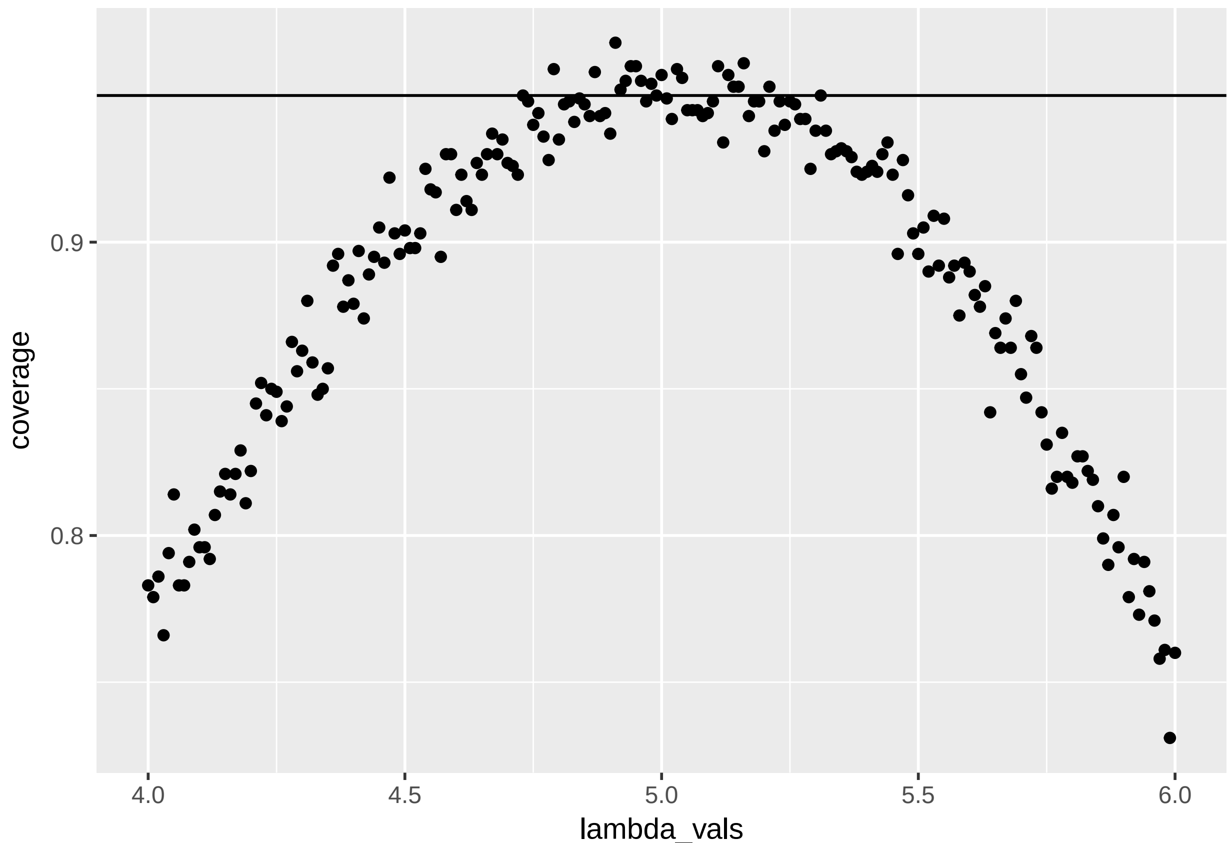
```
## [1] 0.7905694
```

```
SigmaX <- sd(ExpD_means$means)
Var <- SigmaX^2
Var
```

```
## [1] 0.6088292
```

Here the difference is a bit bigger then the mean one, but still fairly close.

Finally, lets plot the distribution of means and see if it looks normal distribution



Here you can see that for selection of $\hat{\lambda}$ around 5, the average of the sample mean falls within the confidence interval at least 95% of the time.

Full code: https://github.com/marcelladane/statistical_inference