StatisticalInferenceW4

Me

23/10/2021

```
#Load packages
library(dplyr, warn.conflicts = F)
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
library(ggplot2)

#Exponential function parameters
lambda <- 0.2
n <- 40
num.of.sim <- 1000

#set the seed
set.seed(119983)

#Create a 1000x40 matrix containing the results of the simulation
sim.distrib <- matrix(data=rexp(n * num.of.sim, lambda), nrow=num.of.sim)
```

###Sample mean vs theoretical mean

we compute the means and store the results in a dataframe which is what the dplyr and ggplot2 packages take as input and it's also the typical datastructure in R. So we create a dataframe, *sim_mns*.

Here we want to compare the theoretical mean for an exponetial distribution, given by mu = 1/lambda = 5, to the mean of our simulated distribution.

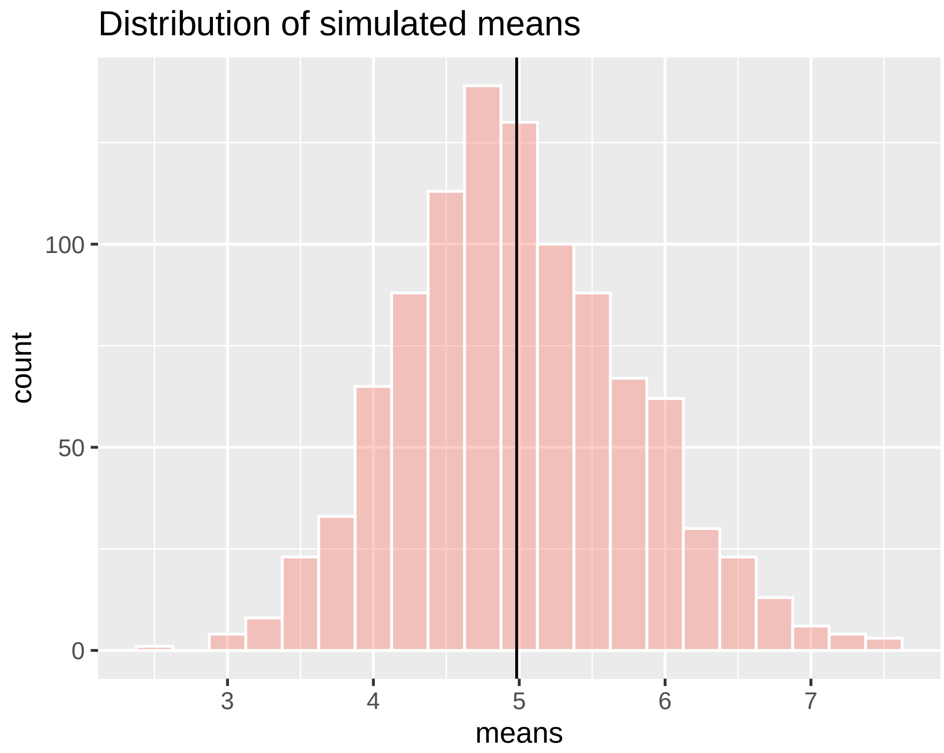
```
#compute the mean for each of the 1000 simulations(rows)
sim_mns <- data.frame(means=apply(sim.distrib, 1, mean))
sim<-sim_mns

#Convert dataframe to tbl_df object for more convenient printing
sim_mns <- tbl_df(sim_mns)
```

```
## Warning: `tbl_df()` was deprecated in dplyr 1.0.0.
## Please use `tibble::as_tibble()` instead.
```

```
#compute the mean of the simulated means
(mean_sim <- sim_mns %>% summarize(simulated.mean = mean(means)) %>% unlist())
```

```
## simulated.mean
##       4.982365
```

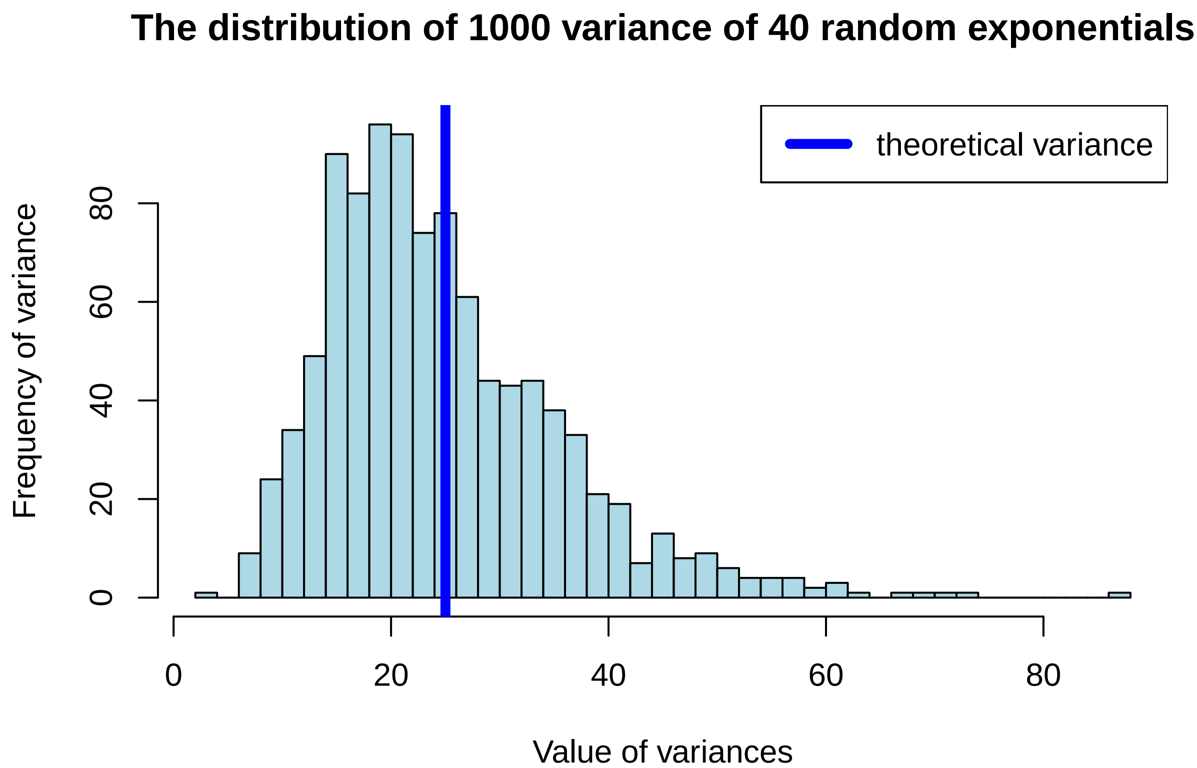
```r
#Sample Mean Distribution Plot
sim_mns  %>%
  ggplot(aes(x = means) ) + geom_histogram(alpha=0.4, binwidth= .25, fill = "salmon", col = "white") +
  geom_vline(xintercept = mean_sim, color="black", size = 0.5) +
  ggtitle("Distribution of simulated means")
```



From the plot we can see that the sample mean(the vertical line) is **4.982365**and is very close to the theoretical mean 1/lambda = **5**

###Sample Variance VS Theoretical Variance

```r
Var.distrib <- apply(sim.distrib, 1, var)
hist(Var.distrib, breaks = 50, main = "The distribution of 1000 variance of 40 random exponentials", xla
abline(v = (1/lambda)^2, lty = 1, lwd = 5, col = "blue")
legend("topright", lty = 1, lwd = 5, col = "blue", legend = "theoretical variance")
```




```
#Compute the variance of the sample means
sd.samp <- sim_mns  %>% select(means)  %>% unlist() %>%  sd()
(var.samp <- sd.samp ^ 2)
```

```
## [1] 0.628253
```

```
#Theoretical variance of the exponential distribution
(((1/lambda))/sqrt(40))^2
```

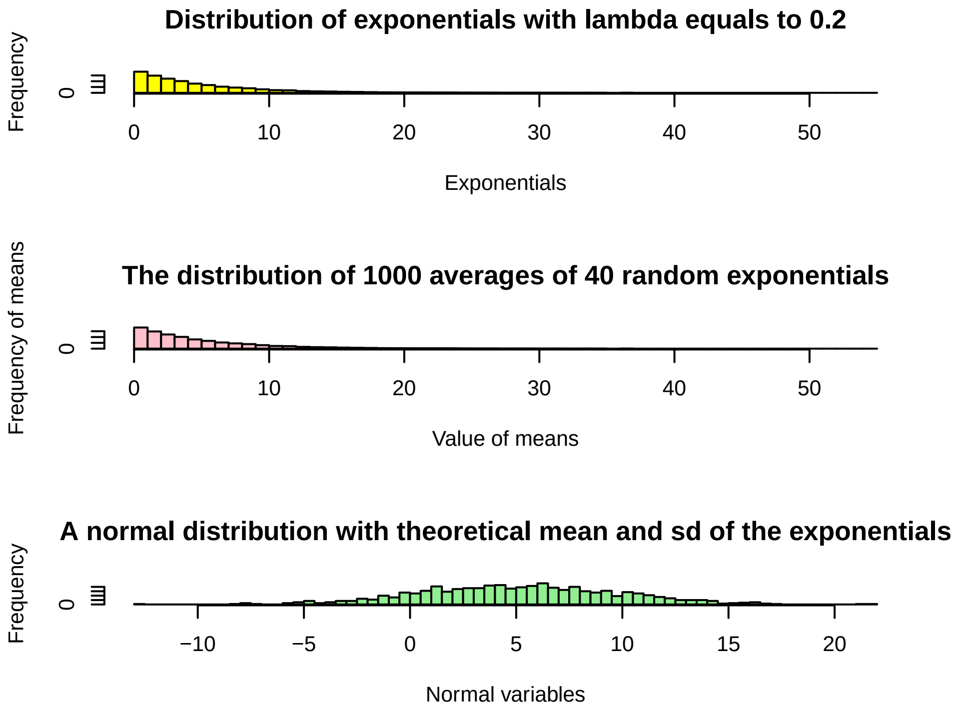
```
## [1] 0.625
```

The sample variance and the theoretical variance are very close, 0.628253 and 0.625, respectively.

###Normality of the Distribution

From the Central limit theorem we know that the distribution of averages of normalized variables becomes that of a standard normal distribution as the sample size increases.

```
par(mfrow = c(3, 1))
hist(sim.distrib, breaks = 50, main = "Distribution of exponentials with lambda equals to 0.2", xlab =
hist(sim.distrib, breaks = 50, main = "The distribution of 1000 averages of 40 random exponentials", xla
simN <- rnorm(1000, mean = mean(sim.distrib), sd = sd(sim.distrib))
hist(simN, breaks = 50, main = "A normal distribution with theoretical mean and sd of the exponentials"
```


Distribution of exponentials with lambda equals to 0.2
Frequency
0
0
10
20
30
40
50
Exponentials
The distribution of 1000 averages of 40 random exponentials
Frequency of means
0
0
10
20
30
40
50
Value of means
A normal distribution with theoretical mean and sd of the exponentials
Frequency
0
−10
−5
0
5
10
15
20
Normal variables