Statistical Inference Course Project

David Contreras Carretero

Part 1: Simulation Exercise Instructions

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

1. Show the sample mean and compare it to the theoretical mean of the distribution.

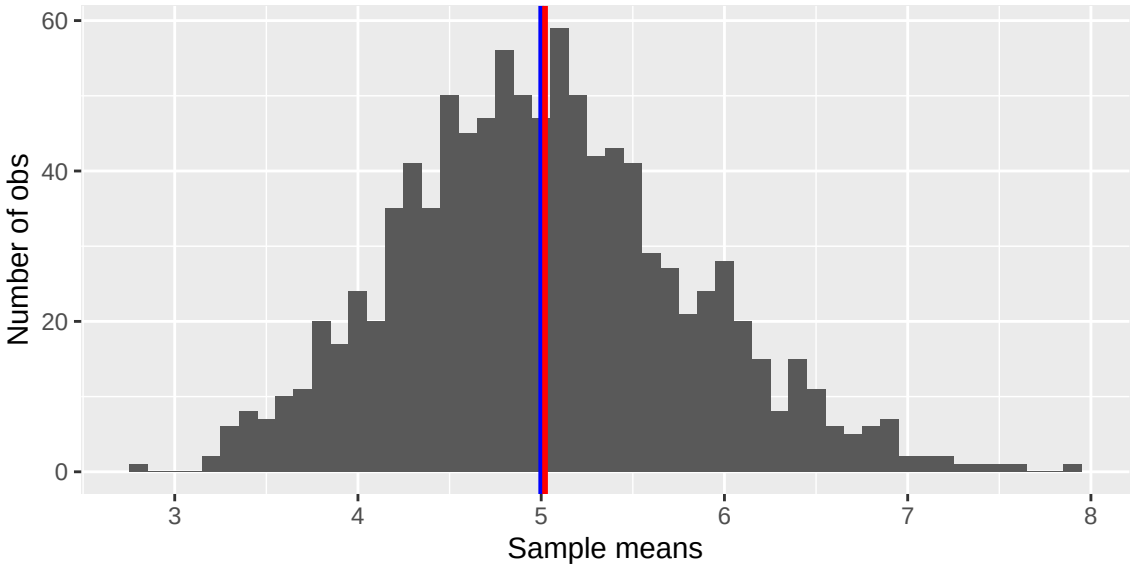
First we create a thousand simulations of the 40 exponentials and calculate and store their means.

```
set.seed(1928)
lambda = 0.2
n = 40
nsim=1000

sims <- data.frame(mean=numeric(nsim))

for (i in 1 : nsim) {
  sample <-  rexp(n,lambda)
  sims[i,1] <- mean(sample)
}
```

```
tmean = 1/lambda
smean <-mean(sims$mean)
```

Sample mean : 5.0209585 Theoretical mean: 5

Figure 1 − Distribution of sample means of 40 exponentials



The blue line indicates theoretical mean, red line sample mean

The histrogram show how close theoretical and sample mean are.

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

We know that theoretical standard deviation is 1/lambda, so theoretical variance is 1/lambda^2

```
tvar <-  (1/lambda^2)/n
tvar
```

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

```
Then we calculate de sample variance as the variance of the means.
```

```
svar <- var(sims$mean)
svar
```

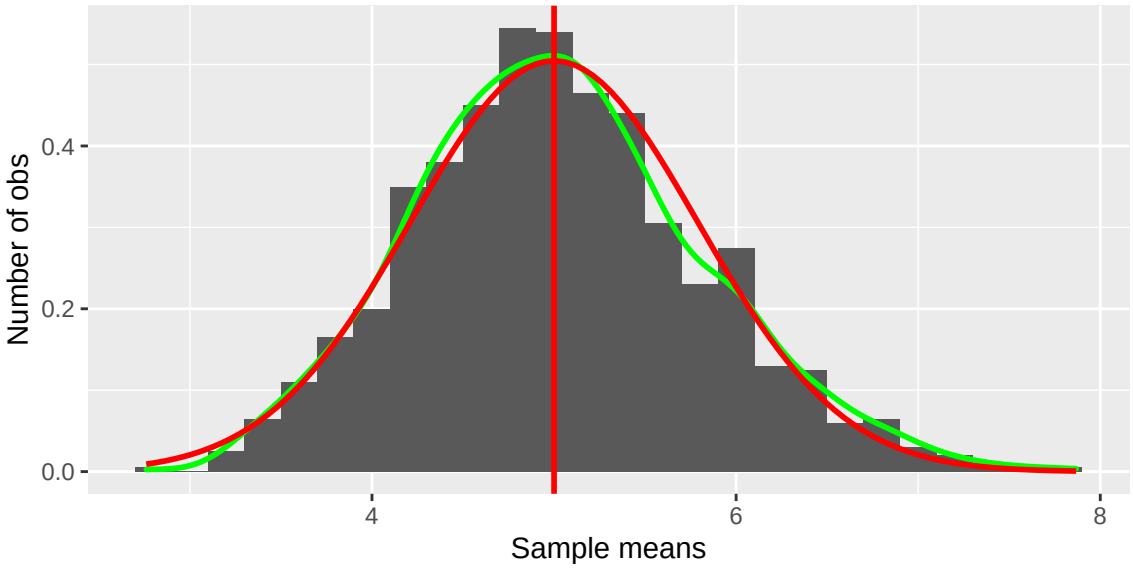
```
## [1] 0.6276333
```

As we can see theoretical variance and sample variance are also very close.

3. Show that the distribution is approximately normal.

In order to compare the distribution of averages of samples with a normal distribution with theoretical mean and standard deviation.

Figure 2 − Distribution of sample means of 40 exponentials



Green line show distribution density, red lines show normal distribution and theoretical mean

Figure 2 shows how sample distribution almost match normal distribution line. We would say it is aproximatly normal.

Appendix - R Code

```
# Figure 1 code
ggplot(sims, aes(x=mean)) +
           geom_histogram(binwidth=0.1) +
           labs(y = "Number of obs",
                x = "Sample means",
                title ="Figure 1 - Distribution of sample means of 40 exponentials",
                caption ="The blue line indicates theoretical mean, red line sample mean") +
           geom_vline(xintercept=tmean,  color = "blue", size=1) +
           geom_vline(xintercept=smean,  color = "red", size=1)
```

```
# Figure 2 code

ggplot(sims, aes(x=mean)) +
       labs(y = "Number of obs",
            x = "Sample means",
            title =" Figure 2 - Distribution of sample means of 40 exponentials",
            caption ="Green line show distribution density,
                      red lines show normal distribution and theoretical mean") +
      geom_histogram(aes(y=..density..), size=1, binwidth=0.2) +
      geom_density(color="green", size=1) +
      stat_function(fun=dnorm, args=list(mean=tmean, sd=sqrt(tvar)), color = "red", size=1) +
      geom_vline(xintercept=tmean, color="red", size=1)
```