SI Course Project Part 1

Sasa Pakvovic

Course project task 1 description

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 also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s. You should

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.
2. Show how variable it is and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.
4. Evaluate the coverage of the confidence interval for 1/lambda: X¯±1.96Snv.

Introduction

According to Central Limit Theorem the distribution of averages of iid variables properly normalized becomes that of a standard normal distribution as the sample increases. It is proper to say that this theorem does not say how many samples we need to average to see the effect, but as we near infinity the effect should be more and more obvious.

Also, the Law of Large Numbers says that the average limits to what it is estimating. for example averages of the sample mean converge to the population mean. We can also say that the sample mean of iid samples is consistent for the population mean.

Understaning these concepts can tell us what we can expect for some of the qustions asked in the project part 1.

Simulations

We will define some constants for the simulations.

```r
library(ggplot2)
##No. of simulations, recommended to set to 1000
nosim <- 1000
##No. of exponential iids, defined by task, set to 40
n <- 40
##Lambda, defined by task to be 0.2
lambda <- 0.2
##Population mean is 1/lambda
popmean <- 1/lambda
##Population sd is 1/lambda
popsd <- 1/lambda
##Population Var is sqrt(1/lambda)
popvar <- sqrt(popsd)

set.seed(777)
```

1. Show where the distribution is centered at and compare it to the

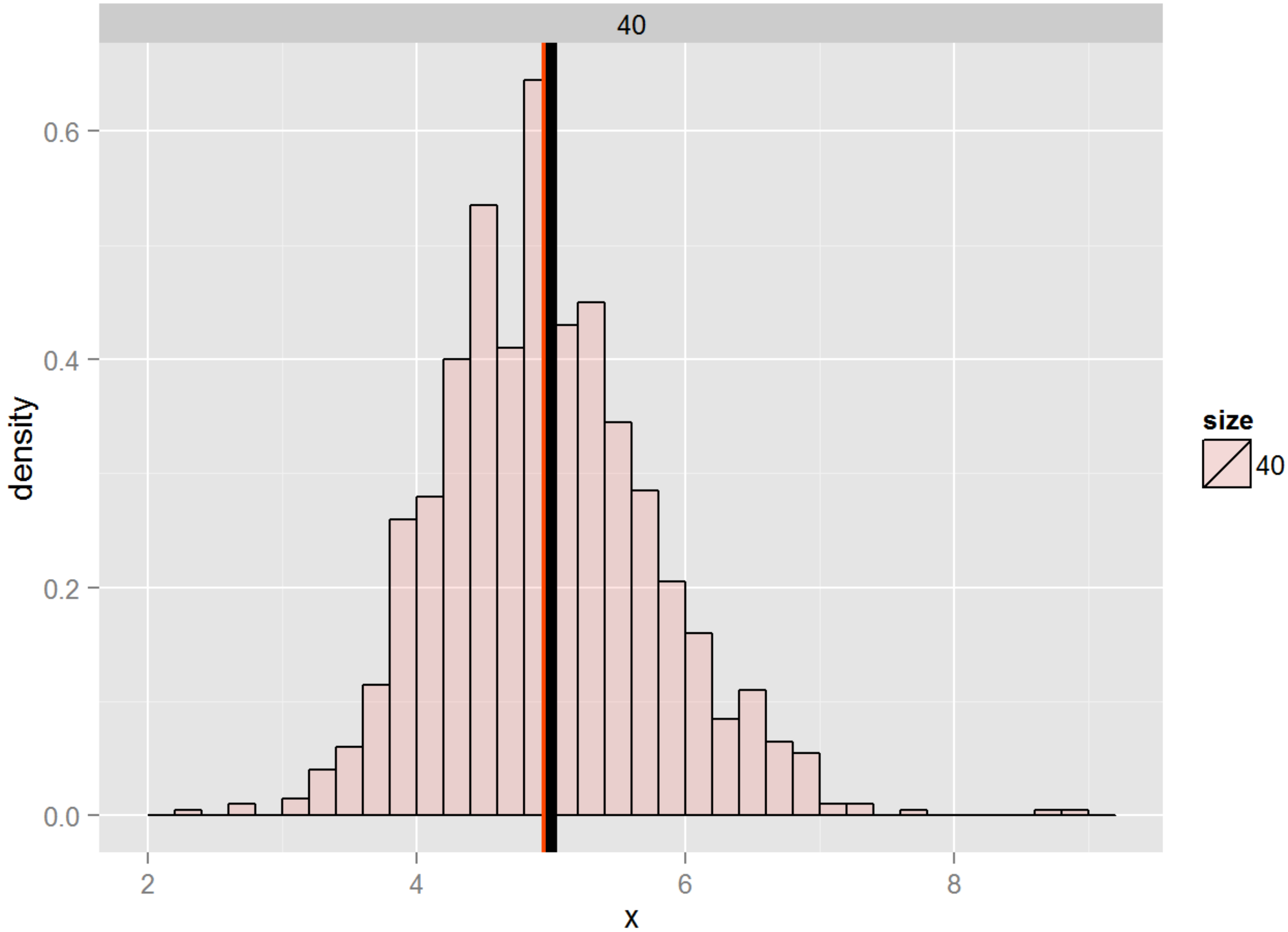
theoretical center of the distribution.

To show this we can create a dataframe in which we will store averages of 40 iid exponential variables, and then take the mean of all of the 1000 obeservations. The sample mean should be close to what it is estimating. It should be close to 5.

```r
cfunc <- function(x,n) mean(x)
dat <- data.frame(
        x = apply(
                replicate(nosim,rexp(n,lambda))
                ,2,cfunc,nosim
                ),
                size = factor(rep(c(40),rep(nosim,1))))
samplemean <- mean(dat[,1])
samplemean
```

```
## [1] 4.97
```

We can also show this with a graph showing the distribution of averages of 40 exponential(0.2). Mean of averages is shown with red, while theoretical mean is shown with the black line. Also, we can note that the distribution os already taking bell shape that is associated with the normal distribution.



2. Show how variable it is and compare it to the theoretical variance of the distribution.

Theoretical variance for exponential(0.2) is square root of population standard deviation, so we can get it by:

```
popvar <- sqrt(popsd)
popvar
```

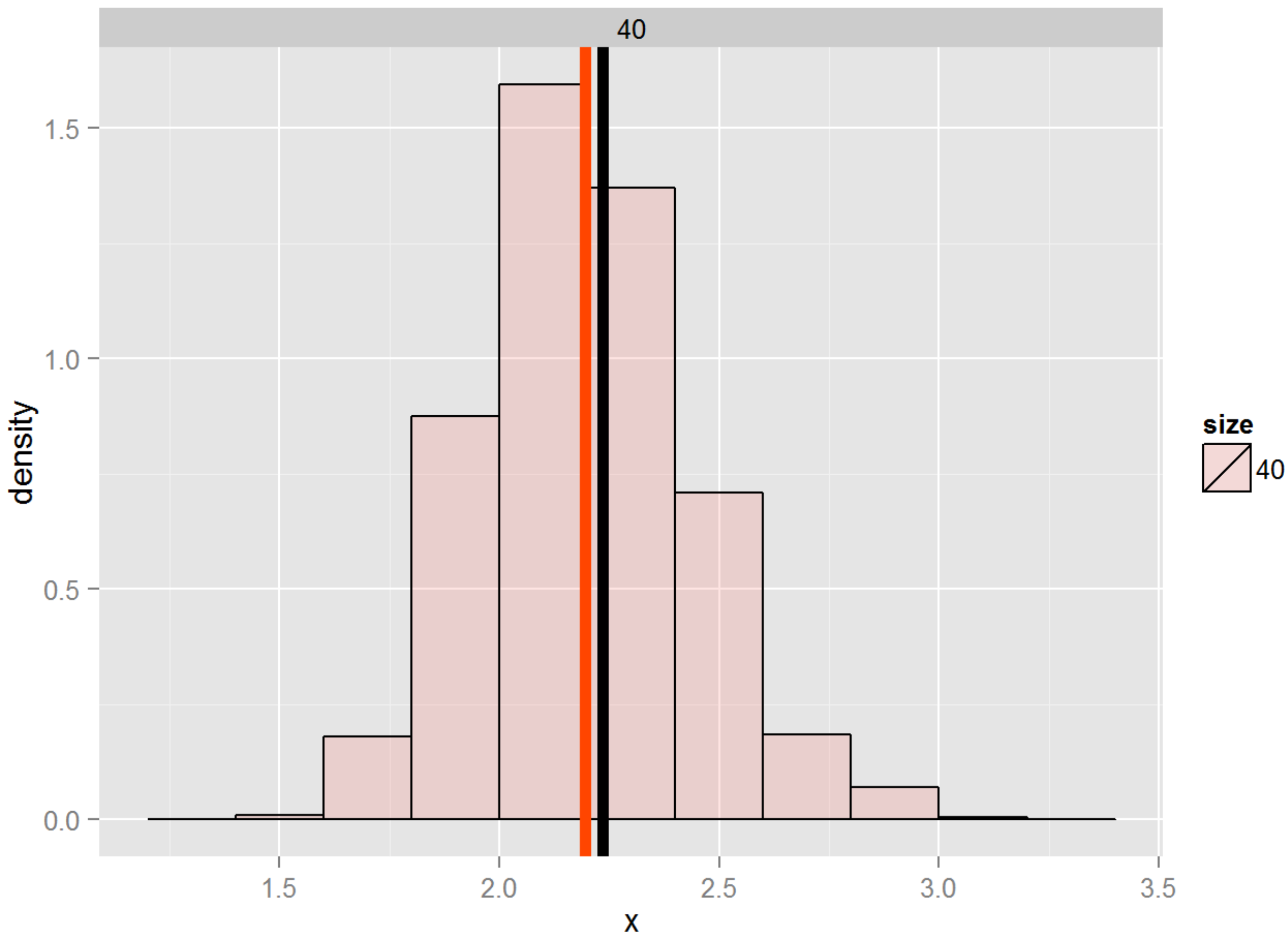
```
## [1] 2.236
```

To show the variability of our 40 iid exponential(0.2) variables we can create a dataframe in which we will store average variance of 40 iid exponential variables, and then take the average of all of the 1000 obeservations. The sample variance should be close to what it is estimating. It should be close to population variance.

```
cfunc <- function(x,n) sqrt(sd(x))
datvar <- data.frame(
        x = apply(
                replicate(nosim,rexp(n,lambda))
                ,2,cfunc,nosim
                ),
                size = factor(rep(c(40),rep(nosim,1))))
samplevar <- mean(datvar[,1])
samplevar
```

```
## [1] 2.196
```

We can also show this with a graph showing the distribution of averages of sd of 40 exponential(0.2). Sd of averages is shown with red, while theoretical sd is shown with the black line. Also, we can note that the distribution is already taking bell shape that is associated with the normal distribution.



3. Show that the distribution is approximately normal.

We have already seen in both cases of sample mean and sample varaiance that the distribution is normal.

4. Evaluate the coverage of the confidence interval for 1/lambda: X¯±1.96Snv.

First i want to show the confindence interval taking into account population mean, population sd and number of measurements taken.

We can calculate the lower limit by:

```
samplemeanll <- samplemean - qnorm(.975) * popsd / sqrt(n)
samplemeanll
```

```
## [1] 3.42
```

We can caluculate the upper limit by:

```
samplemeanul <- samplemean + qnorm(.975) * popsd / sqrt(n)
samplemeanul
```

```
## [1] 6.519
```

We can say that the interval from 3.4203 to 6.5193 contains the population mean with probability of 95% if we were to take that sample from the same exponential distribution exponential(0.2).

As we already have the data about the sample mean as a vector we can use this to check this assumption and calculate the coverage for our 1000 means taken from 40 exponential(0.2).

```
samplemeansll <- dat[,1] - qnorm(.975) * popsd/sqrt(n)
samplemeansul <- dat[,1] + qnorm(.975) * popsd/sqrt(n)

coverageall <- sapply(popmean,
                   function(param){
                           ll <- dat[,1] - qnorm(.975) * popsd/sqrt(n)
                           ul <- dat[,1] + qnorm(.975) * popsd/sqrt(n)
                           mean(samplemeansll < popmean &
                                samplemeansul > popmean)
                   }
             )
coverageall
```

```
## [1] 0.949
```

We can see that the coverage is 94.9 percent.