# Statistical Inference Course Project-1

Shivangi

8/8/2020

## PART-1: SIMULATION

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.

### Observing the population:

In R, we can obtain n samples from an exponential distribution with the following code:

```
rexp(10,0.2)
```

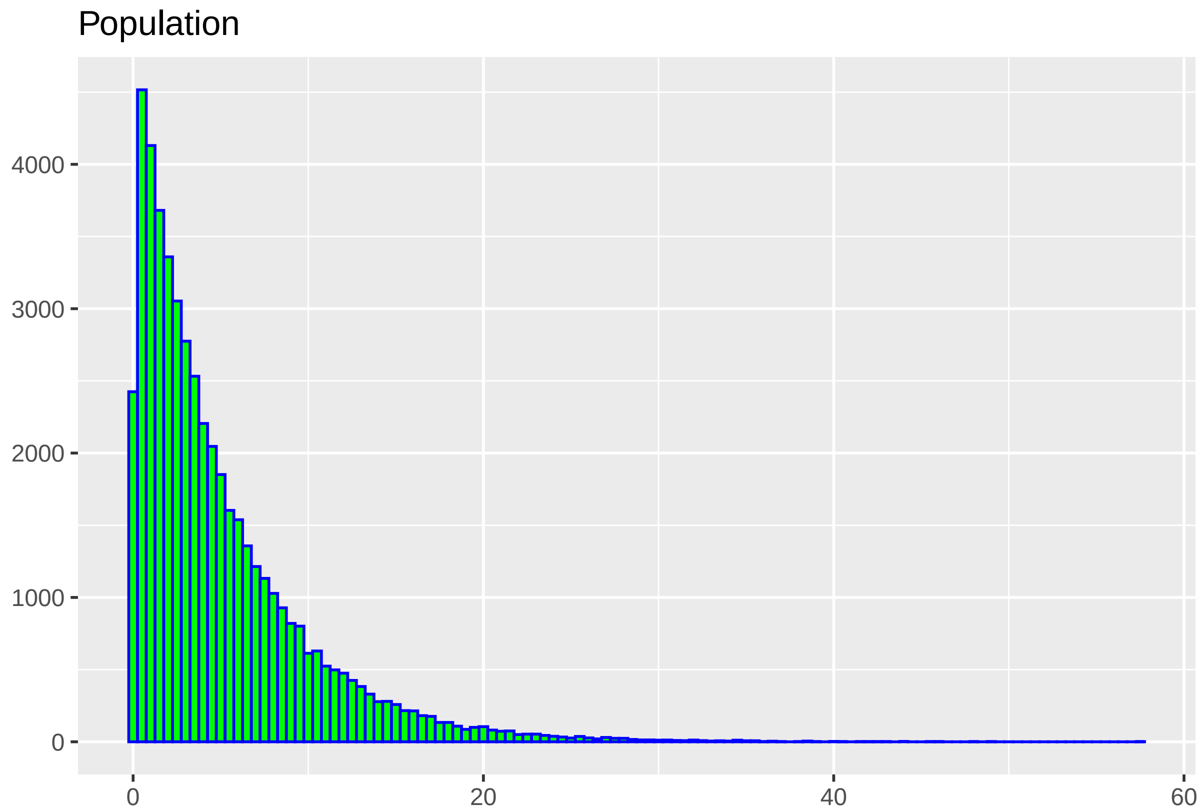
```
##  [1] 4.1459726 8.9412334 2.1036556 2.4309440 2.2686856 2.0802944 7.3294407
##  [8] 0.4649387 5.7457492 1.7137721
```

Now we take the exponential population with 50,000 subjects.

```
pop <- rexp(50000,0.2)
mu <- mean(pop)
sigma <- sd(pop)
cat("mu= ",mu)
cat("\nsigma= ",sigma)
```

Histogram of the population having exponential ditribution:

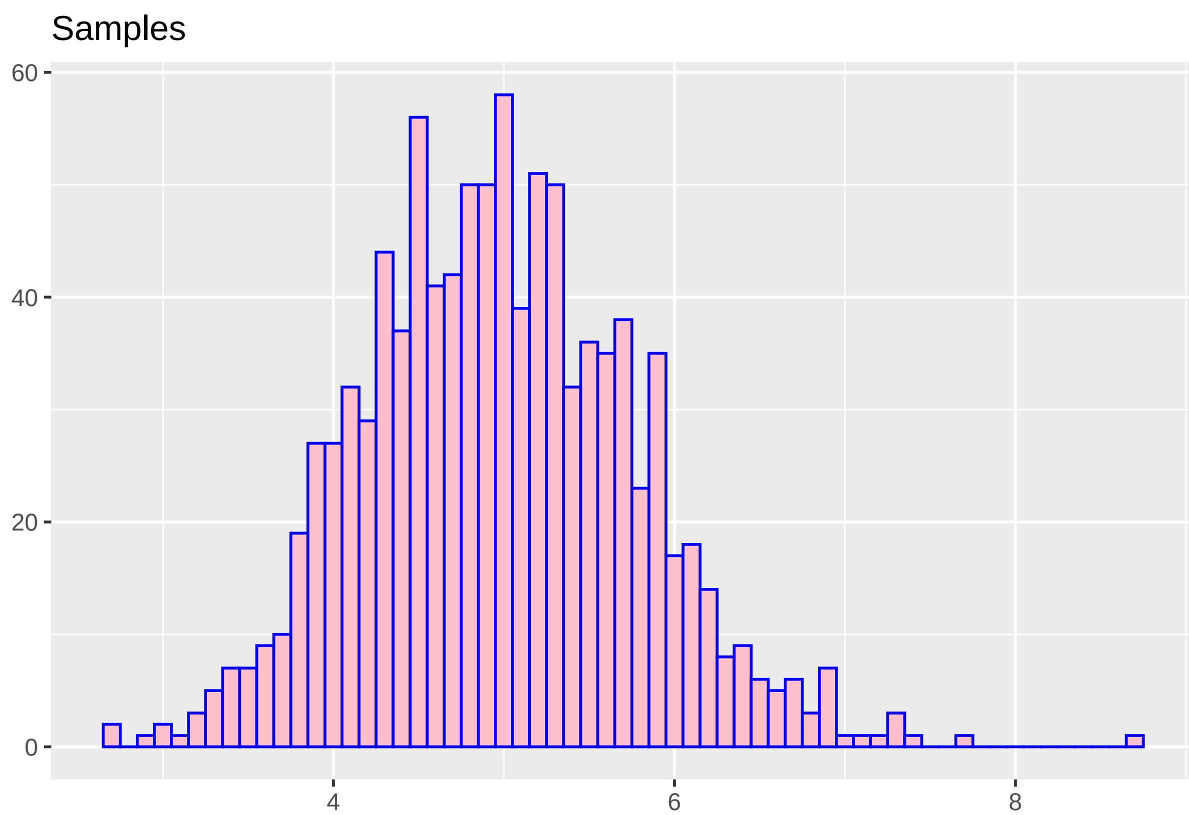
```
library(ggplot2)
qplot(pop,col=I("blue"),fill=I("green"),main="Population",binwidth=0.5,xlab = "")
```




**Sampling 40 samples from exponential population:**

Distribution of 1000 averages of 40 random exponentials are:

```
mean = NULL
for (i in 1:1000){
mean <- c(mean, mean(rexp(40,0.2)))
}
qplot(mean,col=I("blue"),fill=I("pink"),main="Samples",binwidth=0.1,xlab = "")
```




Although it's slightly skewed to the right, the bell shape of the Normal distribution is evident. Strikingly, the mean appears to be around 5, which is what we expected.

**Calculating sample mean and standard deviation from distribution function:**

Now we calculate the expected mean and variances for the exponential distribution of samples(n) equal to 40.

```
n <-40
lambda <- 0.2
exp_mean <- 1/lambda
exp_sd <- 1/lambda/sqrt(40)
sprintf("expected sample mean = %s",exp_mean)
```

```
## [1] "expected sample mean = 5"
```

```
sprintf("expected sample std deviation = %s",exp_sd)
```

```
## [1] "expected sample std deviation = 0.790569415042095"
```

**Comparing the theoretical or expected sample mean of the population from the sample mean:**

We know the expected value of sample mean is the estimate of the population mean.
Mean of sample means is the expected value of mean.

```r
cat("The expected mean =",exp_mean)
```

```
## The expected mean = 5
```

```r
cat("\t  Actual mean = ",mean(mean))
```

```
##    Actual mean =  4.972216
```

**Comparing the theoretical or estimated sample standard deviated of the population from the sample mean:**

It is known that standard deviation of sample mean

```r
cat("Expected SD =",exp_sd)
```

```
## Expected SD = 0.7905694
```

```r
cat("\t Actual SD =",sd(mean))
```

```
##   Actual SD = 0.7948669
```

Unsurprisingly, the CLT works.