Statistical Inference Course Project Part 1

TK

8 septembre 2016

Synopsis

In this project I 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.

For this project:

- Set lambda = 0.2 for all of the simulations.
- Investigate the distribution of averages of 40 exponentials.
- We need to do a thousand simulations.
- Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials.

I 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.

Simulation of an Exponential Distribution

I calculated the average of 40 samples drawn from the exponential distribution a thousand times

```r
set.seed(1000)

# set lambda
lambda <- 0.2

# samples
n <- 40

# simulations
NSim <- 1000

# simulate
simExp <- replicate(NSim, rexp(n, lambda))
```

Calculate mean of exponentials

```
meanExp <- apply(simExp, 2, mean)
```

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

```
# Simulated mean
smean <- mean(meanExp)
print(smean)
```

```
## [1] 4.986963
```

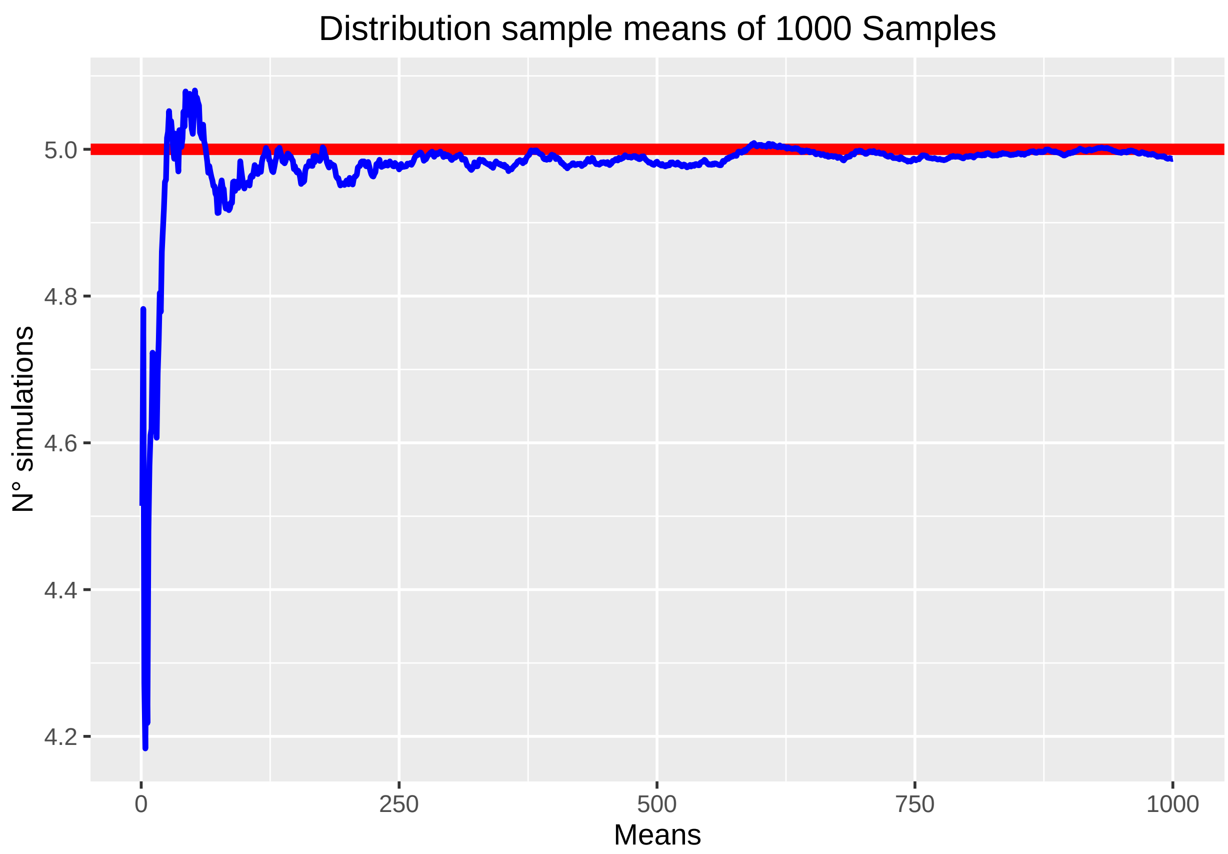
```
# Theoretical mean
tmean <- 1/lambda
print(tmean)
```

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

```
means <- cumsum(meanExp)/1:NSim

# Construction plot

library(ggplot2)
g <- ggplot(data.frame(y=means, x=1:NSim), aes(x=x, y=y ))+
     geom_hline(yintercept = tmean, color = "red", size = 2)+
     geom_line(size=1, color = "blue")+
     labs(title="Distribution sample means of 1000 Samples",x = "Means", y="N° simulations")
print(g)
```



We can see that simulated mean is very close to theoretical mean.

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

```r
# standard deviation and variance of distribution of averages of 40 exponentials

sdSim <-sd(meanExp)
print(sdSim)
```

```
## [1] 0.8089147
```

```r
varSim <-  sdSim^2
print(varSim)
```

```
## [1] 0.654343
```

```r
# theoretical standard deviation and variance

Thsd <-(1/lambda)/sqrt(n)
print(Thsd)
```

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

```
varTh <- Thsd^2
print(varTh)
```

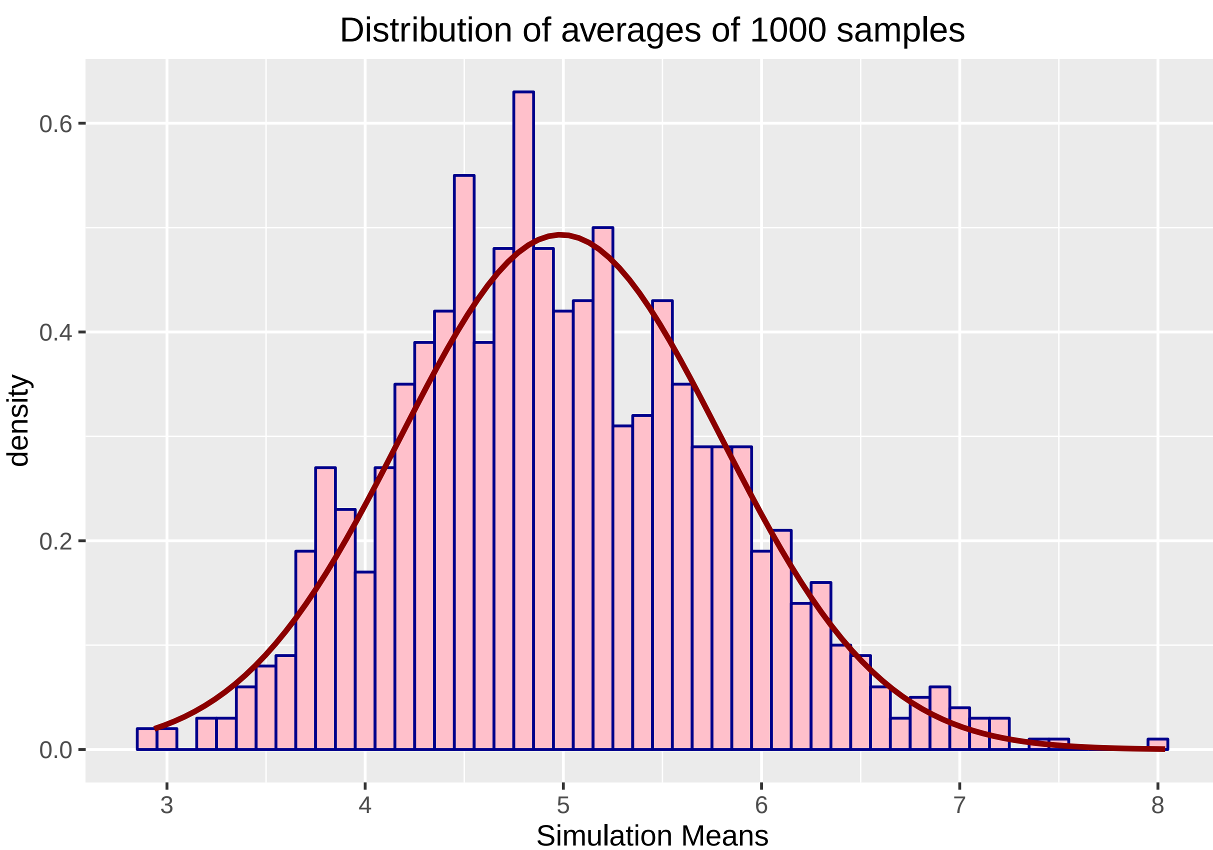
```
## [1] 0.625
```

The simulated standard deviation and variance are closed to theoretical standard deviation and variance.

3. Show that the distribution is approximately normal.

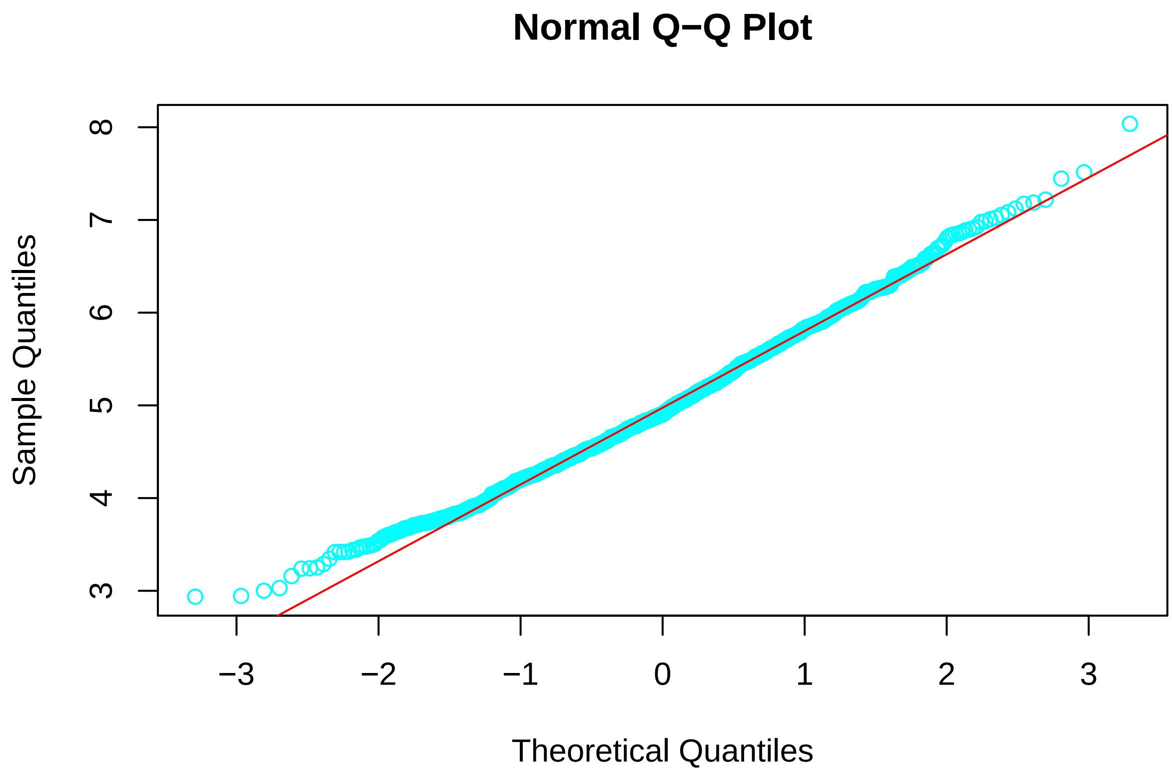
1. Plot of distrubution simulated means.

```
data <- data.frame(meanExp, 1:NSim)
ggplot(data.frame(y=meanExp), aes(x=y)) +
   geom_histogram(aes(y=..density..), binwidth=0.1, fill="pink",
                  color="darkblue") +
   stat_function(fun=dnorm, args = list(mean=smean,
                                    sd=sdSim), color = "darkred",
                 size=1) +
   labs(title="Distribution of averages of 1000 samples", x="Simulation Means")
```



2. Compare our distribution with normal distribution.

```
qqnorm(meanExp, col="5"); qqline(meanExp, col="2")
```



This plots show us that distribution of simulated mean is approximately normal.