Statistical Inference Project Part 1

Lulu

6/30/2020

Project Descriptions

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.

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

```
## Set up parameters
lambda <- 0.2
n <- 40
simulations <- 1000

# Set seed for reproduciblity
set.seed(123)

# Create a vector containing 1000 means of 40 random exponentials.
mns <- NULL
for(i in 1:simulations){mns <- c(mns, mean(rexp(n,lambda)))}
length(mns)
```

```
## [1] 1000
```

Point 1

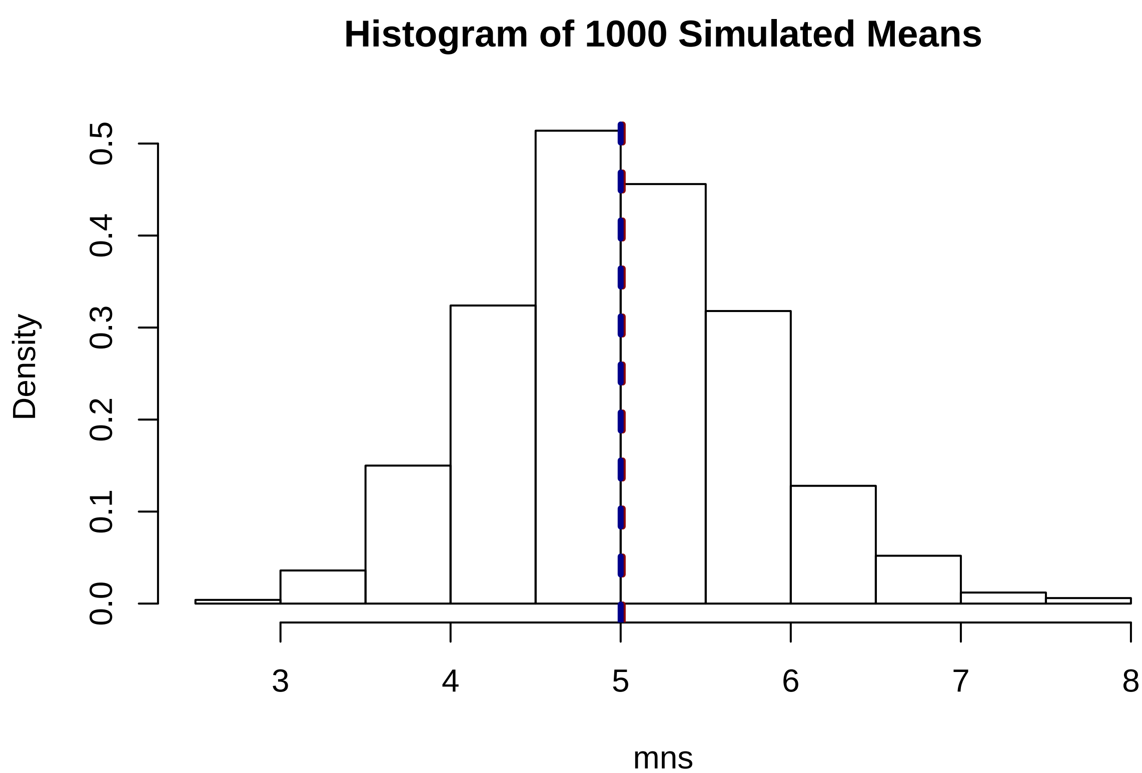
Show the sample mean and compare it to the theoretical mean of the distribution.

```
simulated_mean <- mean(mns)
simulated_mean
```

```
## [1] 5.011911
```

```r
theoretical_mean <- 1/lambda
theoretical_mean
```

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



The theoretical mean is 5. The simulated mean is 5.011911. The center of the simulated distribution (1000 means of 40 random exponentials) is very close to the theoretical mean.

Point 2

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

```r
simulated_sd <- sd(mns)
simulated_sd
```

```
## [1] 0.7749147
```

```r
theoretical_sd <- (1/lambda)/sqrt(n)
theoretical_sd
```

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

```
simulated_variance <- var(mns)
simulated_variance
```

```
## [1] 0.6004928
```

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

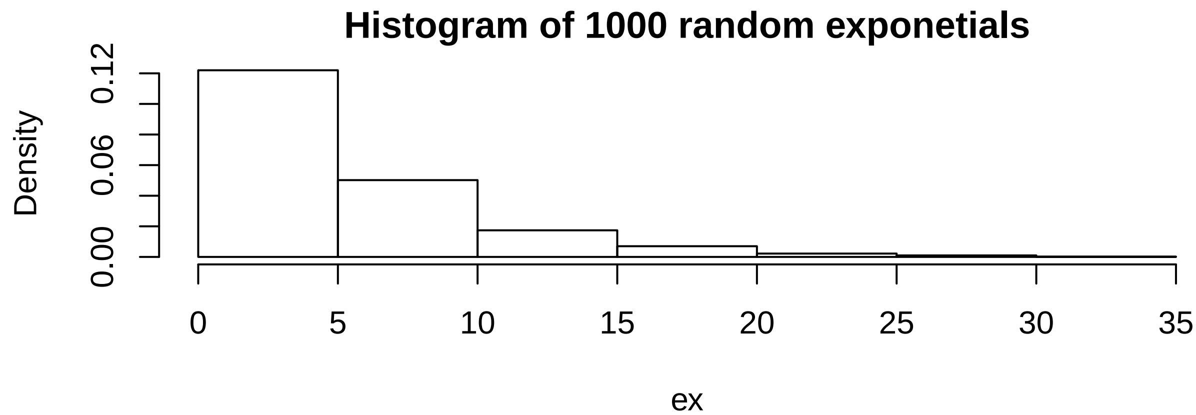
```
## [1] 0.625
```

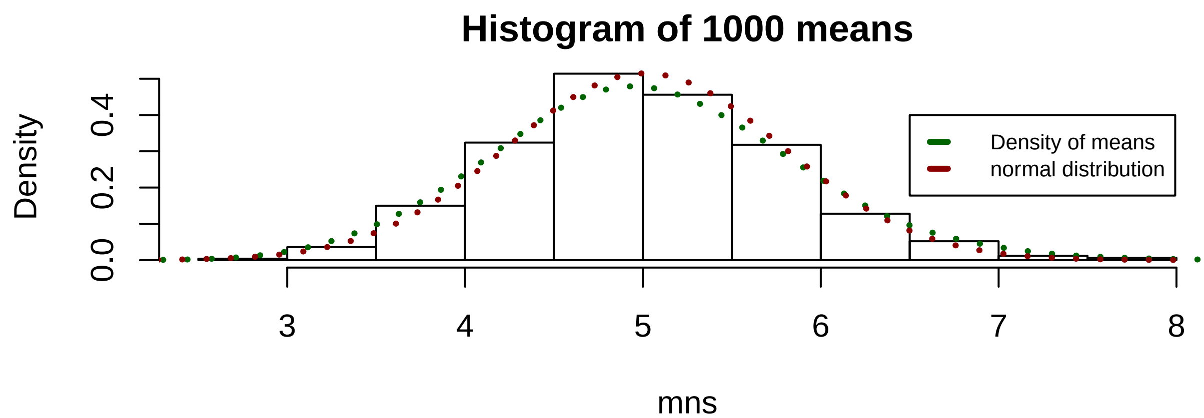
The theoretical_variance is (1/lambda)^2/n = (1/0.2)^2/40 = 0.625. The simulated_variance (1000 means of 40 random expoentials) is 0.6004928 and it is very close to the theoretical variance of the distribution.

Point 3

Show that the distribution is approximately normal.

```
# Make a plot of two rows and 1 column
par(mfrow=c(2,1), mar=c(4,4,2,1))

# Create a vector containing 1000 random exponetials
ex=rexp(1000,lambda)

# Generate a histgram plot for 1000 random exponentials
hist(ex, prob=TRUE, main="Histogram of 1000 random exponetials")

# Generate a histgram plot for the 1000 means with prob=TRUE for probabilities not counts
hist(mns, prob=TRUE, main="Histogram of 1000 means")

# Add a density plot line
lines(density(mns, adjust=2), lty="dotted", col="darkgreen", lwd=3)

# Add a normal distribution line
x <- seq(2,8, length=1000)
y <- dnorm(x, mean=mean(mns),sd=sd(mns))
lines(x,y, lty="dotted", col="darkred", lwd=3)

# Add a legend
legend(6.5,0.4, legend=c("Density of means","normal distribution"),
       col=c("darkgreen", "darkred"),
       lty=2, lwd=3,cex=0.7)
```





Compared to the distribution of 1000 random exponentials, the distribution of 1000 means of 40 random exponentials is approximately normal.