Statistical Inference Course Project

K.Sedes.

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.
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.

Task I

Show where the distribution is centered at and compare it to the theoretical center of the distribution.

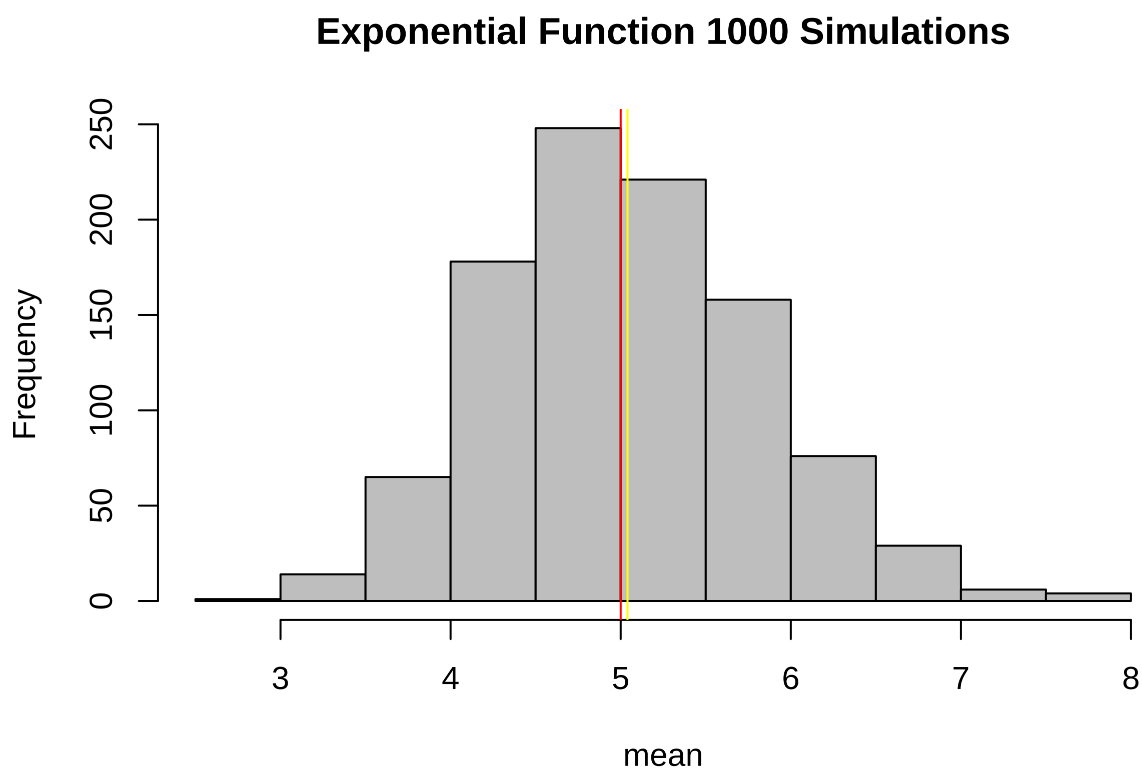
```
# 40 samples
n <- 40

# 1000 simulations
simulations <- 1000

# set lambda to 0.2
lambda <- 0.2

# simulate exp distributions rexp
sim_expDistributions <- replicate(simulations, rexp(n, lambda))

# asociated means
means_sim_expDistributions <- apply(sim_expDistributions, 2, mean)

# final mean of created exp. distributions
simulated_mean <- mean(means_sim_expDistributions)
simulated_mean
```

```
## [1] 5.040342
```

```
# theoretical mean
theoretical_mean <- 1/lambda
theoretical_mean
```

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

```r
# plotting theoretical mean vs simulated
hist(means_sim_expDistributions, xlab = "mean", main = "Exponential Function 1000 Simulations" ,col = "g
abline(v = simulated_mean, col = "yellow")
abline(v = theoretical_mean, col = "red")
```



If we compare the values obtained above,the theoretical mean "5" versus the center of distribution of averages "5.00475" are very closed.

Task II

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

```r
# standard deviation of distribution
sd_simulated <- sd(means_sim_expDistributions)
sd_simulated
```

```
## [1] 0.7862016
```

```r
# standard deviation from theoretical expression
sd_theoretical <- (1/lambda)/sqrt(n)
sd_theoretical
```

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

```r
# variance of distribution
variance_simulated <- sd_simulated^2
variance_simulated
```

```
## [1] 0.618113
```

```r
# variance from analytical expression
variance_theoretical <- ((1/lambda)*(1/sqrt(n)))^2
variance_theoretical
```

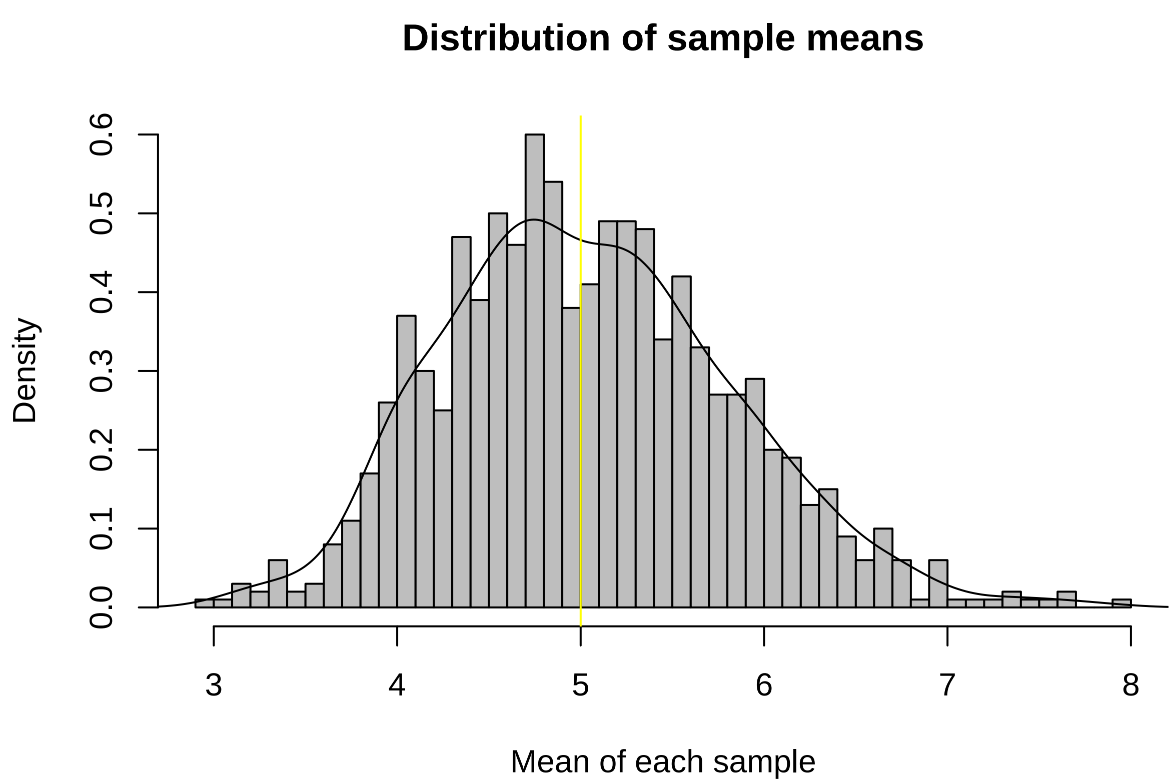
```
## [1] 0.625
```

The variance of our distribution is 0.619 with the theoretical variances calculated as 0.625.

Task III

Show that the distribution is approximately normal.

```r
hist(means_sim_expDistributions,  xlab="Mean of each sample", ylab = "Density",
     main="Distribution of sample means", breaks=n, prob=TRUE ,col="grey")
lines(density(means_sim_expDistributions))
abline(v = 1/lambda, col = "yellow")
```



The plot shows that the distribution is approximately normal. Let's check CI for both theoretical and simulated distributions

#Compute CI for theoretical mean

```
theoretical_ci<-  theoretical_mean +c(-1,1)*qnorm(0.95) * sd_theoretical/n
theoretical_ci
```

```
## [1] 4.967491 5.032509
```

```
#Compute CI in actual
simulated_ci<- simulated_mean +c(-1,1)*qnorm(0.95) * sd_simulated/sqrt(n)
simulated_ci
```

```
## [1] 4.835871 5.244813
```

Another time, as we have seen with mean and variance, CI are similar. In conclusion, as we may expect due to the CLT, the distribution of averages of 40 exp. tends to follow a normal distribution