Simulation exercise

I. Goikoetxea

Overview

In this project the exponential distribution in R will be investigated and compared with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The results will:

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.

Simulations

1. Sample mean vs. theoretical mean

1000 simulations will be run for comparing the sample results with theory. Each simulation will contain 40 exponentials with lambda = 0.2, following an exponential distribution function 'rexp(40, 0.2)'.

```
nsim=1000; n=40; lambda = 0.2

means = NULL
for (i in 1:nsim){
    means <- append(means,mean(rexp(n,lambda)))
}

sample_mean <- mean(means)
print(sprintf("The sample mean is : %f", sample_mean))
```

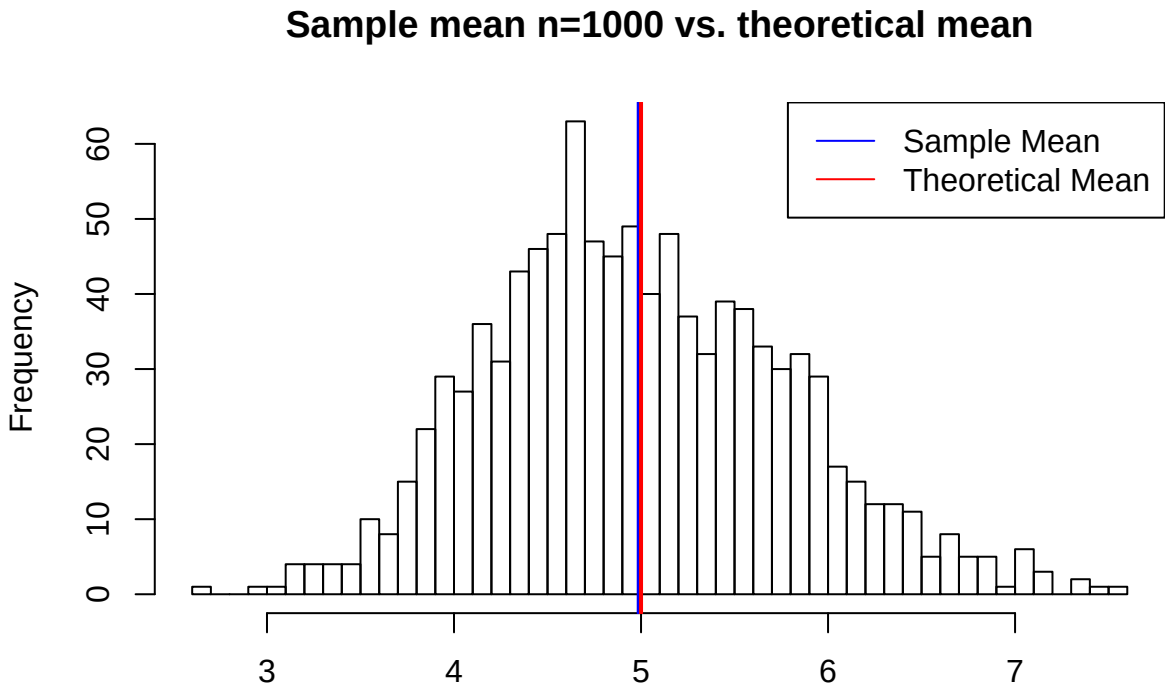
```
## [1] "The sample mean is : 4.987673"
```

```
theoretical_mean <- 1/lambda
print(sprintf("The theoretical mean is : %f", theoretical_mean))
```

```
## [1] "The theoretical mean is : 5.000000"
```

In the follwoing histogram for the sample distribution with a mean of 5.015 is compared to the theoretical mean, 5.0. As shown, the distribution of the sample means has converged to the theoretical mean.

```
hist(means, breaks=50, main = "Sample mean n=1000 vs. theoretical mean", xlab="")
abline(v=sample_mean,col="blue",lwd=2)
abline(v=theoretical_mean,col="red", lwd=2)
legend('topright', c("Sample Mean", "Theoretical Mean"), lty = c(1,1),col=c("blue","red"))
```



2. Sample variance vs. Theoretical variance of the distribition

The variance of the sample means is compared to the theoretical varience of the population. For this, the variance of the sample means is multiplied by the sample size. As the number of simulations increases the variance of the sample means converges to the theoretial variance of the distribution.

```
sample_variance <- var(means)*n
print(sprintf("The sample variance is : %f", sample_variance))
```

```
## [1] "The sample variance is : 25.712754"
```

```
theoretical_variance <- (1/lambda)^2
print(sprintf("The theoretical variance of the distribution is : %f", theoretical_variance))
```

```
## [1] "The theoretical variance of the distribution is : 25.000000"
```

3. Distribution of Sample Means vs Normal Distribution

Due to the central limit theorem, the averages of samples follow normal distribution. The next plot shows the density of the sample means and the normal density using the theoretical values. It can be observed that the distribution of the sample means can be compared to a normal distribution.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```r
means_data <- data.frame(means)
plot1 <- ggplot(means_data, aes(x=means)) +
geom_histogram(aes(y = ..density..), color="black",fill="blue",binwidth=0.25) +
labs(title="Distribution of Sample Means vs Normal Distribution", x="sample mean", y="density") +
stat_function(fun=dnorm,args=list(mean=1/lambda, sd=1/lambda/sqrt(40)), col="red", size=1.5)
plot1
```