Coursera Statistical Inference Course Project Part_1

ZiadAbdelnabi

7/4/2020

Overview

In part 1 of this project we are investigating the exponential distribution in R and comparing it with the Central Limit Theorem using a simulation exercise.

Part 1: Simulation Exercise

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

As per the instructions, the exponential distribution can be simulated in R with the function rexp(n, lambda) where lambda is the rate parameter. The simulation can be repeated multiple times using the repetition function.

The theoretical mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda.

For the 1000 simulations, lambda is assumed to be 0.2 and the sample size n is 40.

```r
n <- 40
Simulations <- 1000
Lambda <- 0.2

SampleMean <- NULL
for(i in 1:Simulations) {
  SampleMean <- c(SampleMean, mean(rexp(n, Lambda)))
}
mean(SampleMean)

## [1] 4.950109
```

So, as we can see, compared to the theoretical mean distribution of 5 , our mean 5 is close.

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

The theoretical standard deviation of the distribution is also 1/lambda , which, for a lambda of 0.2 , equates to 5 . The variance is the square of the standard deviation, which is 25 .

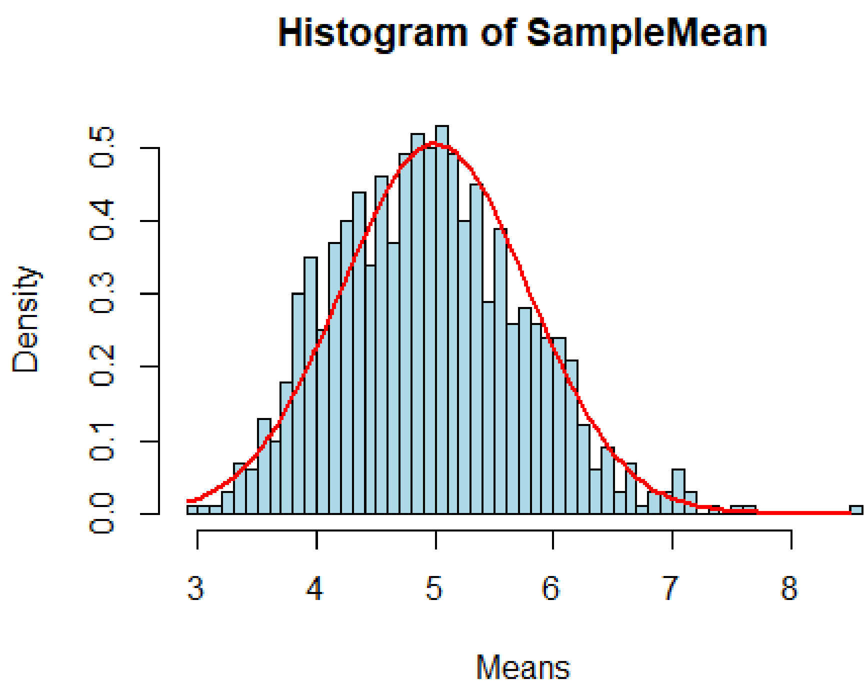
```
Variance <- var(SampleMean)
Variance

## [1] 0.6403031
```

0.6 is close to the theoretical distribution.

1.3 Show that the distribution is approximately normal.

```
hist(SampleMean, breaks = n, prob = T, col = "light blue", xlab = "Means")
x <- seq(min(SampleMean), max(SampleMean), length = 100)
lines(x, dnorm(x, mean = 1/Lambda, sd = (1/Lambda/sqrt(n))), pch = 25, col =
"red", lty=1, lwd=2.5)
```



A normal distribution follows a bell shaped curve, where the black line represents the calculated normal distribution, which we can compare shape of the histogram.

The central limit theorom states that the sample means would become that of a standard normal distribution as the sample size increases whilst meeting the two conditions of independence (n < 10%) and normal, or if skewed distribution, that n > 30.