Stat_inf_pr1

Harsha

Sunday, December 27, 2015

Project 1

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.

Overview

1000 simulations of 40 samples of exponential distribution would be investigated. The sample means, standard deviation would be compared with the theoritical mean and the standard deviation.

The exponential distribution is defined by lambda - Mean is 1/lambda and the standard deviation is also 1/lamdba

Samples simulation

Samples are generated with rexp function. Exploration is done with head and summary on the the data. Data has 40 rows and 1000 columns. Each column represents a sample. Using column means, each sample mean is found out.

```
n <- 40
sim <- 1000
lambda <- 0.2

data <- replicate(sim, rexp(n, lambda))

nrow(data)
```

```
## [1] 40
```

```
ncol(data)
```

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

```
data_with_mn <- rbind(data, colMeans(data))

nrow(data_with_mn)
```

```
## [1] 41
```

```
sample_means <- data_with_mn[41,]
```

```
head(data)
summary(data)
```

Q1 and 2

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.

```
thmean <- 1/lambda
thsd <- (1/lambda)/sqrt(n)

calcmean <- mean(sample_means)
calcsd <- sd(sample_means)

thmean
```

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

```
calcmean
```

```
## [1] 5.033402
```

```
thsd
```

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

```
calcsd
```

```
## [1] 0.7765546
```

It is to be noted that the theoritical mean and the calculated means are close. The same can be said about the standard deviation (a measure of variance)

Q3