Comparing Exponential Distribution in R with CLT

Srividya Bobji

Saturday, January 24, 2015

Overview: 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 also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials. Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s.

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

```
lambda = 0.2
n = 40
sim = 1:1000
set.seed(1000)
sampleMeans <- data.frame(x = sapply(sim, function(x) {mean(rexp(n,
lambda))}))

sampleMean <- mean(sampleMeans$x)
theoreticalMean <- 1/lambda
sampleMean
```

```
## [1] 4.987
```

```
theoreticalMean
```

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

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

```
variance <- sd(sampleMeans$x)
theoreticalVariance <- (1/lambda)/sqrt(40)
variance
```

```
## [1] 0.8089
```

```
theoreticalVariance
```

```
## [1] 0.7906
```

Sample mean: 4.987 Theoretical mean: 5

Variance of the distibution: 0.8089 Theoretical Variance : 0.7906

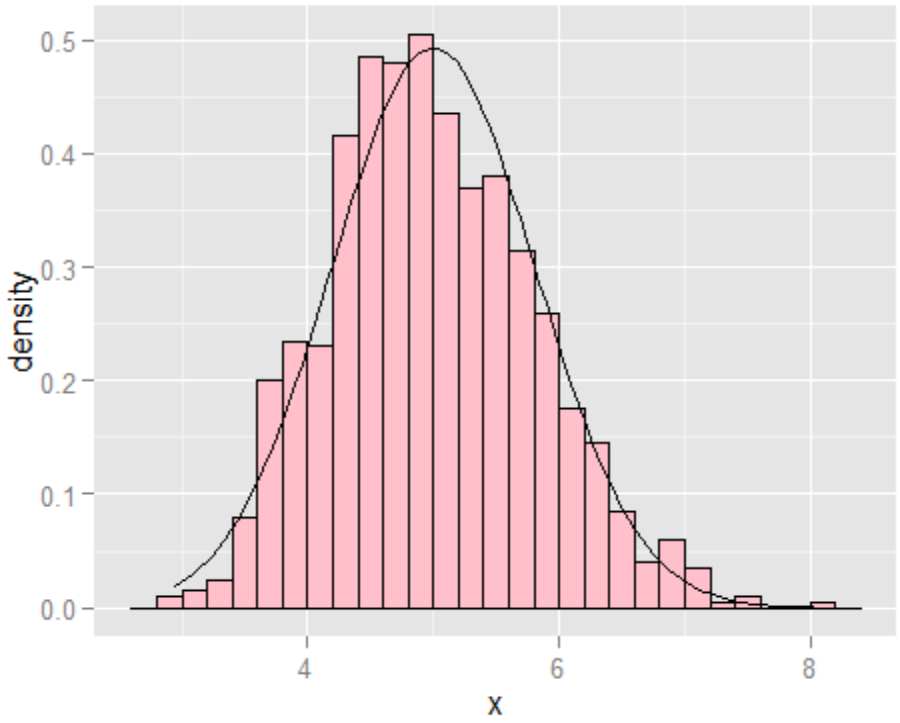
3. Show that the distribution is approximately normal.

Below is a histogram plot of the means of the 1000 simulations of rexp(n, lambda). It is overlaid with a normal distribution with mean 5 and variance 0.8089.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.0.3
```

```
ggplot(data = sampleMeans, aes(x = x)) +
    geom_histogram(aes(y=..density..), fill = I('pink'),
                   binwidth = 0.20, color = I('black')) +
    stat_function(fun = dnorm, arg = list(mean = theoreticalMean, sd =
variance))
```



The distribution of our simulations appears normal.

Focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

```
sampleMean + c(-1,1)*1.96*variance/sqrt(nrow(sampleMeans))
```

```
## [1] 4.937 5.037
```