Statistical Inference Project - Part 1 : Simulation Exercise

S C Jayaprakash

Thursday, January 22, 2015

Overview

Investigate the exponential distribution and compare it with the Central Limit Theorem.

The exponential distribution can be simulated in R with rexp(n, lambda) where n is the number of observations and lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda.

In these simulation exercises, we investigate the distribution of averages of 40 exponentials over a thousand observations (n=1000), assuming lambda = 0.2

Simulation Results

```
set.seed(123)
Lambda = 0.2
n = 40
nSims = 1:1000
Means <- data.frame(x = sapply(nSims, function(x) {
    mean(rexp(n, Lambda))
}))
head(Means)
```

```
##       x
## 1 4.811
## 2 5.360
## 3 4.593
## 4 4.900
## 5 5.517
## 6 5.613
```

1. Show the sample mean and compare it to the theoretical mean of the distribution 1/lambda = 5

Mean of simulations

```
mean(Means$x)
```

```
## [1] 5.012
```

From above, we can see that the calculated mean is 5.012 and the theoretical mean is 5; therefore the variation is negligible.

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

Expected (theoretical) Variance

```
((1/Lambda)/sqrt(40))^2
```

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

Variance of simulations

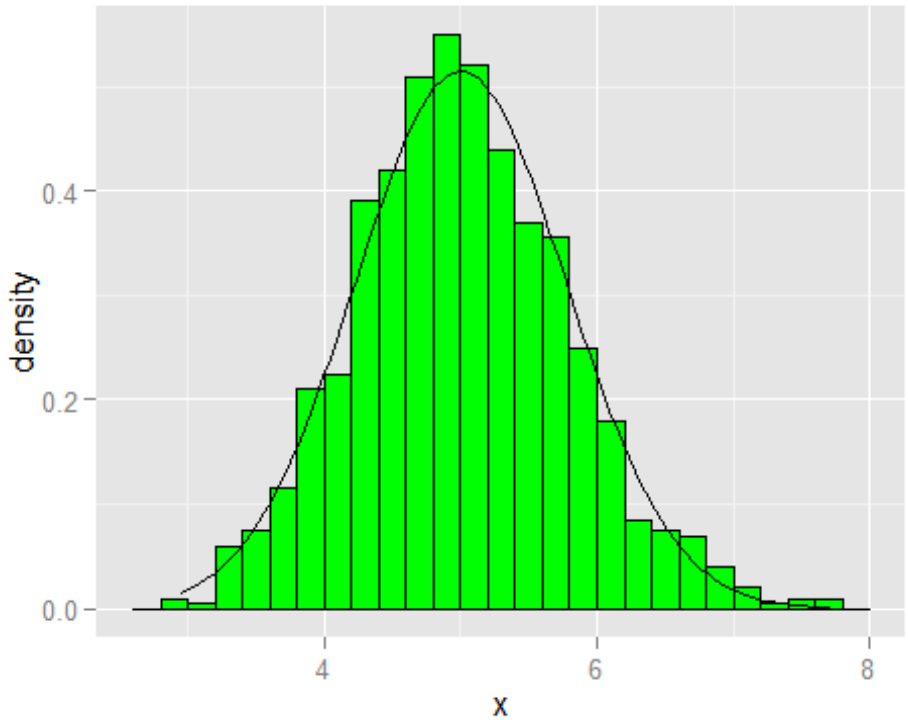
```
var(Means$x)
```

```
## [1] 0.6005
```

From above, we can see that the calculated variance is 0.6005 and the theoretical variance is 0.625; therefore both distributions have similar variability.

3. Show that the distribution is approximately normal.

```
library(ggplot2)
ggplot(data = Means, aes(x = x)) +
    geom_histogram(aes(y=..density..), fill = I('green'),
                   binwidth = 0.20, color = I('black')) +
    stat_function(fun = dnorm, arg = list(mean = 5, sd = sd(Means$x)))
```



The histogram plot depicts a distribution that is approximately normal with a standard deviation of 0.7749 and the 95% confidence interval for the sample mean is (4.9639, 5.0599).