Statistics Inference Assignment - 1

Liu Lee

24/10/2014

Contents

Details : This PDF contains all mathematical calculation and graph required to answer questions for assignment 1

Problem Statement - Statistics Inference 1

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. You should

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.
2. Show how variable it is and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.
4. Evaluate the coverage of the confidence interval

Simulation to Sample 1000 Means of 40 Exponentials

```r
lambda = 0.2
n = 40
nsims = 1:1000
set.seed(12345)
means <- data.frame(x = sapply(nsims, function(x) {mean(rexp(n, lambda))}))
head(means)
```

```
##          x
## 1 5.553693
## 2 4.670847
## 3 4.256227
## 4 4.614148
## 5 5.304342
## 6 3.627482
```

Solution for Q1 & Q2

```
# Sample Mean
mean(means$x)
```

```
## [1] 4.971972
```

```
# Population Mean
1/lambda
```

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

```
# Sample Standard Deviation
sd(means$x)
```

```
## [1] 0.7716456
```

```
# Population Standard Deviation
(1/lambda)/sqrt(40)
```

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

```
# Sample Variance
var(means$x)
```

```
## [1] 0.5954369
```

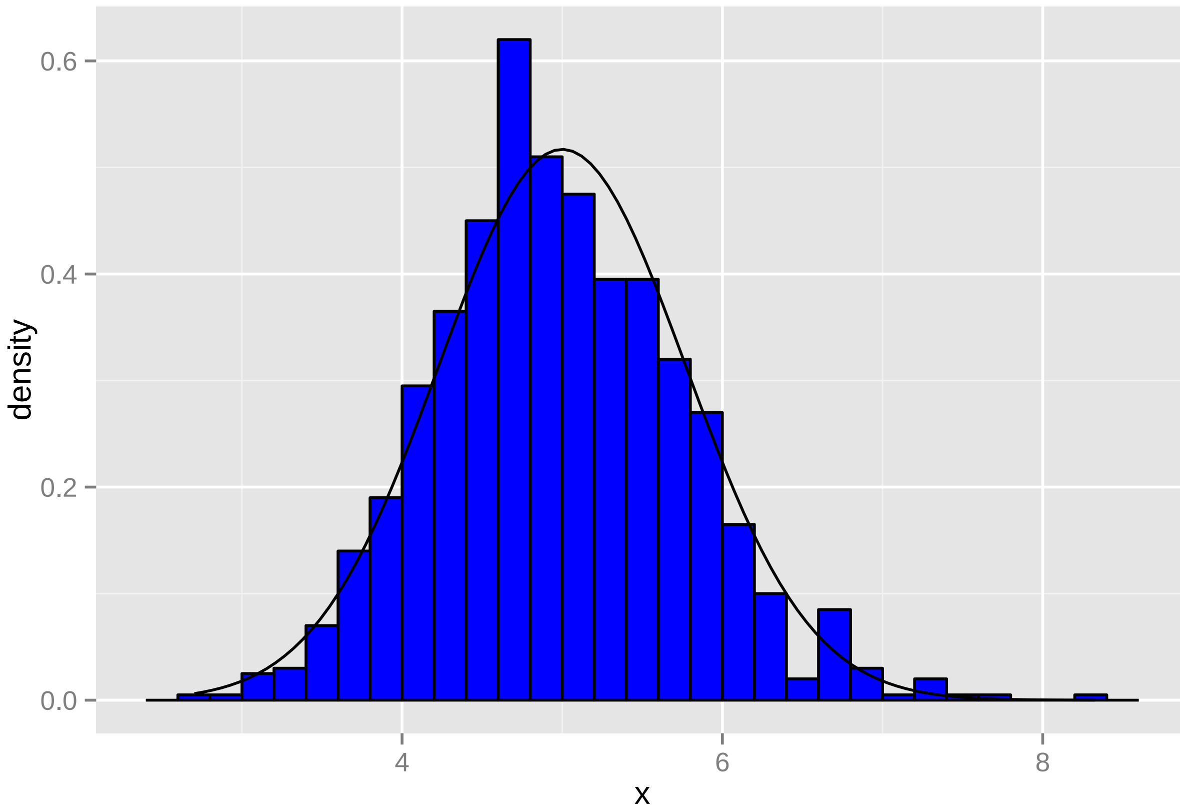
```
# Population Variance
((1/lambda)/sqrt(40))^2
```

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

> Explanation : The mean of the means of the exponential of 1000 sim. of 40 exponential(0.2)s is 4.971972, which is very close to the population mean of $1/0.2 = 5.0$.
> The standard deviation of sample, 0.7716456, is also close to the population standard deviation, 0.7905694. Population standard deviation using Central Limit Theorem . Population variance and sample variance are 0.625 and 0.5954369.

Solution for Q3

```
require(ggplot2, quietly = TRUE )
ggplot(data = means, aes(x = x)) +
    geom_histogram(aes(y=..density..), fill = I('blue'),
                   binwidth = 0.20, color = I('black')) +
    stat_function(fun = dnorm, arg = list(mean = 5, sd = sd(means$x)))
```



Above histogram shows that the distribution of our simulations appears normal.

Solution for Q4

```
mean(means$x) + c(-1,1)*1.96*sd(means$x)/sqrt(nrow(means))
```

```
## [1] 4.924145 5.019799
```

The 95% confidence interval for the mean of the sampling means is (4.924145, 5.019799).

Appendix

Sample Mean - mean(means$x)

Sample Std Deviation - mean(means$x)

Population standard deviation - (1/lambda)/sqrt(40)

Sample Variance - var(means$x)

Population Variance - ((1/lambda)/sqrt(40))^2

Graph Plot - Installed ggplot2 package and used ggplot to plot histogram chart