Class06-Part.1

Maria Koh

Saturday, September 20, 2014

Contents

This is the project for the statistical inference class. I used simulation to explore inference and do some simple inferential data analysis using random variables.

The exponential distribution can be simulated in R with rexp(nosim, 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.

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 3.1.1
```

```
library(ggplot2)
```

You can also embed plots, for example:

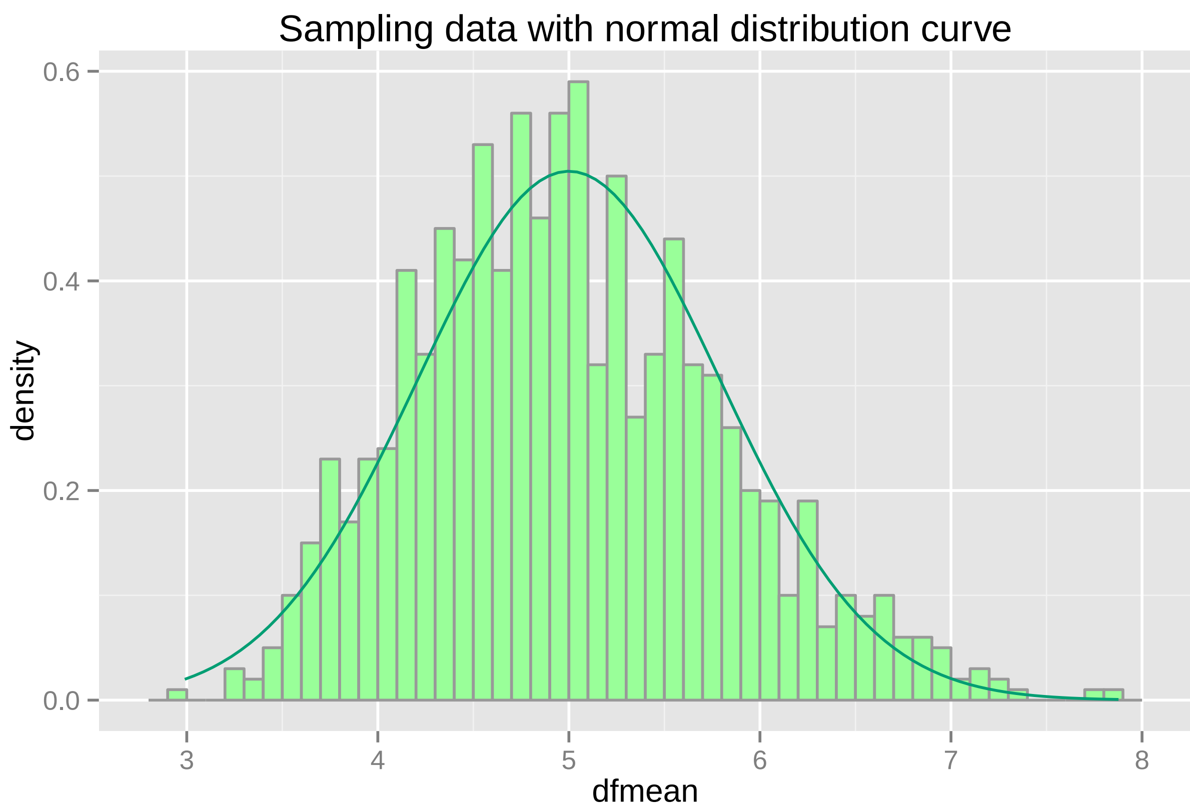
```
nn <- 40
  lambda <- 0.2
  mean_data <- 1/lambda
  sd_data <- 1/lambda*1/sqrt(nn)
```

I repeated the simulation for 1000 times. I make the plot using ggplot for Sample and normal distribution using line.

```
nosim <- 1000
  reap_matrix <- matrix( rexp(nosim*nn , rate=lambda), nosim)
  rexp_mean <- apply(reap_matrix, 1, mean)
  rexp_sd <- apply(reap_matrix, 1, sd)
rexp_df <- data.frame(dfmean=rexp_mean, dfsd=rexp_sd)
```

Show the plot

```
ggplot(rexp_df, aes(x = dfmean)) +
geom_bar(binwidth = 0.1, aes(y = ..density.. ), fill = "#99FF99",  colour ="#999999" , position = "ide
stat_function( fun = dnorm, colour = "#009E73", arg = list (mean = mean_data, sd = sd_data))+
ggtitle("Sampling data with normal distribution curve")
```



that is the mean for sample and theoretical mean.

```
sample_mean <- mean(rexp_mean)
  theoretical_mean <- 1/ lambda
  print(paste("sample_mean =", sample_mean))
```

```
## [1] "sample_mean = 4.9951234255298"
```

```
  print(paste("theoretical_mean =", mean_data))
```

```
## [1] "theoretical_mean = 5"
```

2. Show how variable it is and compare it to the theoretical variance of the distribution.

```
sample_sd <- sd(rexp_mean)
  print(sample_sd)
```

```
## [1] 0.7976
```

```
  theoretical_sd <- 1/lambda * 1/sqrt(nn)
  print(theoretical_sd)
```

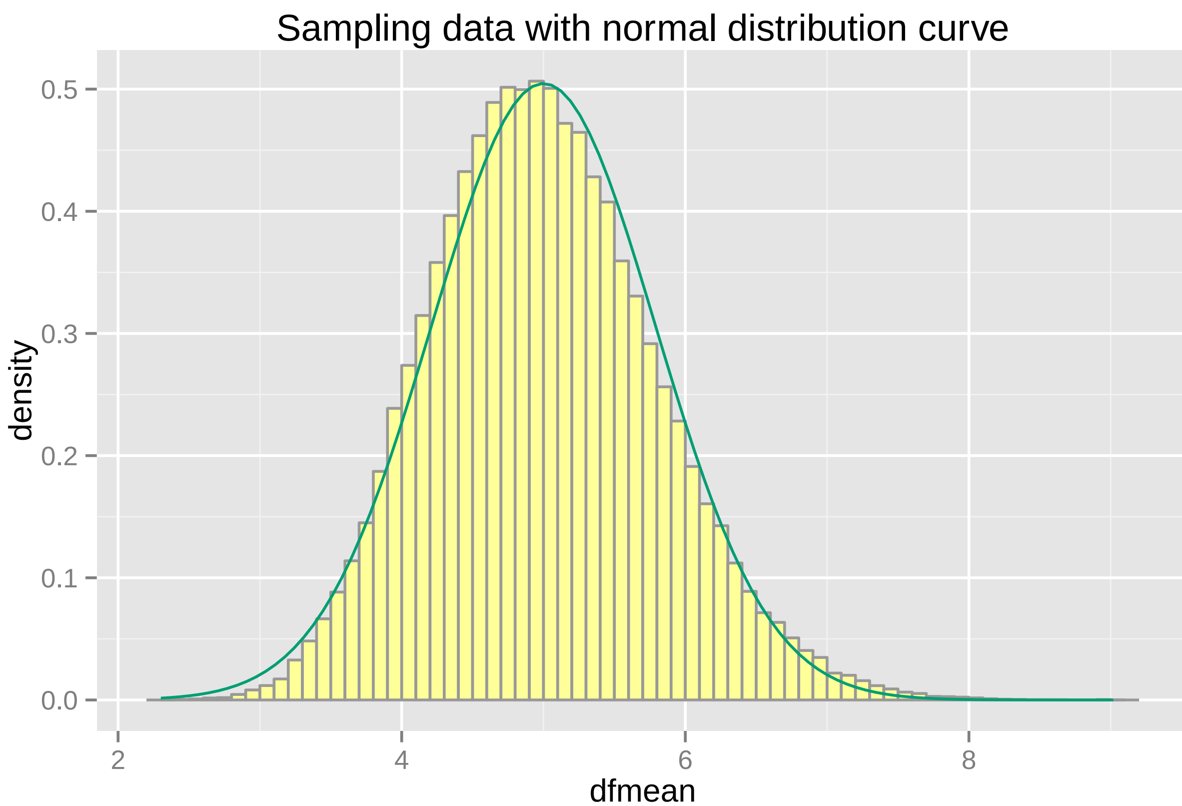
```
## [1] 0.7906
```

3. Show that the distribution is approximately normal.

```
nosim2 <- 100000
   reap_matrix2 <- matrix( rexp(nosim2*nn , rate=lambda), nosim2)
     rexp_mean2 <- apply(reap_matrix2, 1, mean)
     rexp_sd2 <- apply(reap_matrix2, 1, sd)
 rexp_df2 <- data.frame(dfmean=rexp_mean2, dfsd=rexp_sd2)
```

I repeated the simulation for 100000 times.

```
ggplot(rexp_df2, aes(x = dfmean)) +
 geom_bar(binwidth = 0.1, aes(y = ..density..), fill = "#FFFF99", colour ="#999999" , position = "identi
 stat_function( fun = dnorm, colour = "#009E73", arg = list (mean = mean_data, sd = sd_data))+
 ggtitle("Sampling data with normal distribution curve")
```



4. Evaluate the coverage of the confidence interval for 1/lambda : X ±1. 96 * S * 1/sqrt(n).

```
large_sample_mean <- mean(rexp_mean2)
large_sample_sd <- sd(rexp_mean2)

sample_interval <- sample_mean+ c(-1, 1) * 1.96 * sample_sd * 1/sqrt(nn)
large_sample_interval <- large_sample_mean+ c(-1, 1) * 1.96 * large_sample_sd * 1/sqrt(nn)

theory_interval <- mean_data + c(-1, 1) * 1.96 * 1/lambda * 1/sqrt(nn)

print(paste("theoretical_mean =", mean_data))
```

```
## [1] "theoretical_mean = 5"
```

```r
print(paste("Sample's Interval",sample_interval))
```

```
## [1] "Sample's Interval 4.74793303755872"
## [2] "Sample's Interval 5.24231381350089"
```

```r
print(paste("Large Sample's Interval",large_sample_interval))
```

```
## [1] "Large Sample's Interval 4.75392206547575"
## [2] "Large Sample's Interval 5.24356187519734"
```

```r
print(paste("Theory's Interval",theory_interval))
```

```
## [1] "Theory's Interval 3.45048394651749"
## [2] "Theory's Interval 6.54951605348251"
```

The end.