Coursera, Statistical Inference, part 1

Heather Dye

3/16/2018

Overview: A simulation exercise

This excercise investigates the exponential distribution as it compares to the Central Limit Theorem. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. For lambda = 0.2 and using 1000 simulations we investigate the distribution of averages of 40 exponentials.

1. Sample mean is compared to the theoretical mean of the distribution.
2. Variablility of sample (via variance) is compared it to the theoretical variance of the distribution.
3. It is illustrated that the distribution is approximately normal.

Data

```
library(ggplot2)
```

```
# simulation data
lambda <- 0.2
n <- 40
simulation <- 1000
data <- replicate(simulation, mean(rexp(n, lambda)))

# simulations
smean <- mean(data)
svar <- var(data)
ssd <- sd(data)


# Theoretical values
tmean <- 1/lambda
tvar <- 1/(lambda^2*n)
tsd <- (1/lambda)/sqrt(n)
```

Sample Mean versus Theorecital Mean

```
# Theoretical value
tmean
```

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

```
# Simulated value
smean
```

```
## [1] 4.982217
```

Sample Variance versus Theoretical Variance

```
# Theoretical value
tvar
```

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

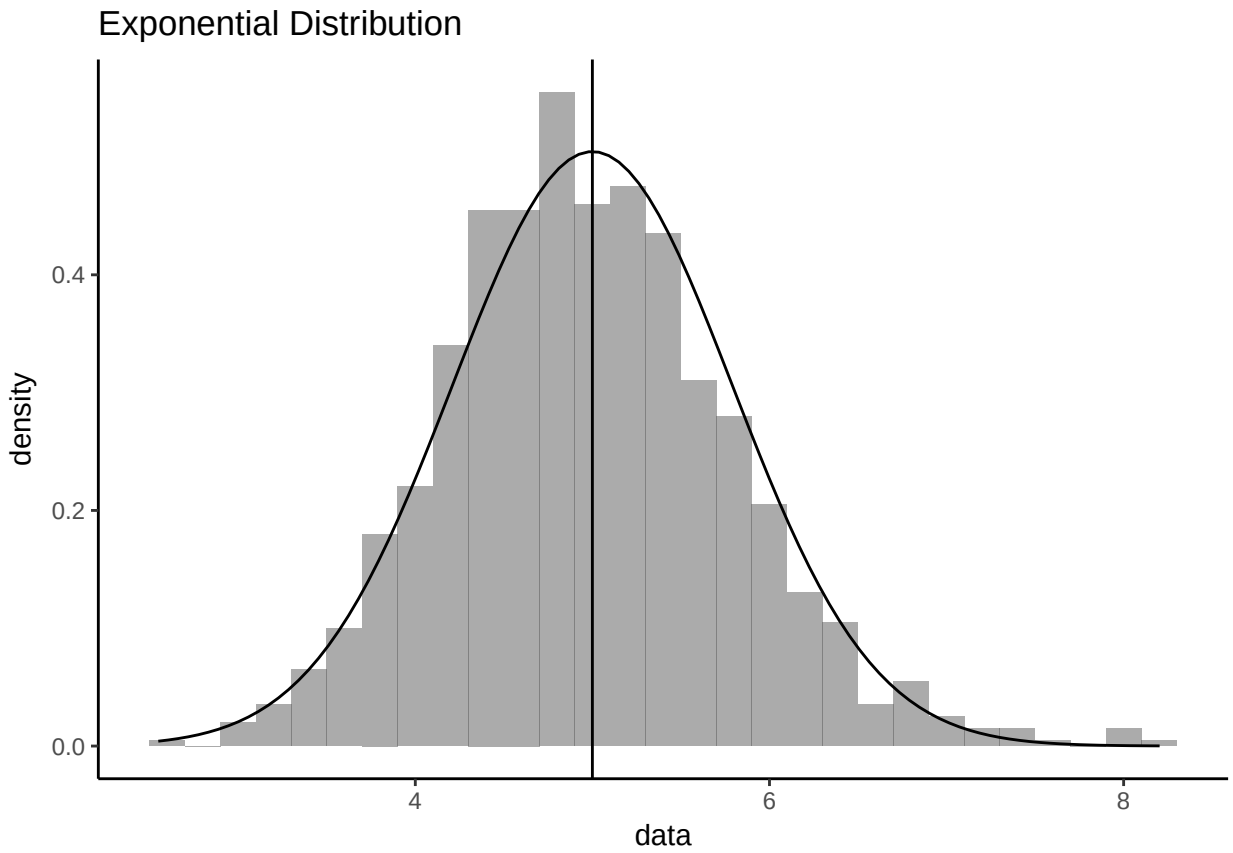
```
# Simulated value
svar
```

```
## [1] 0.6449959
```

Is Distribution Normal?

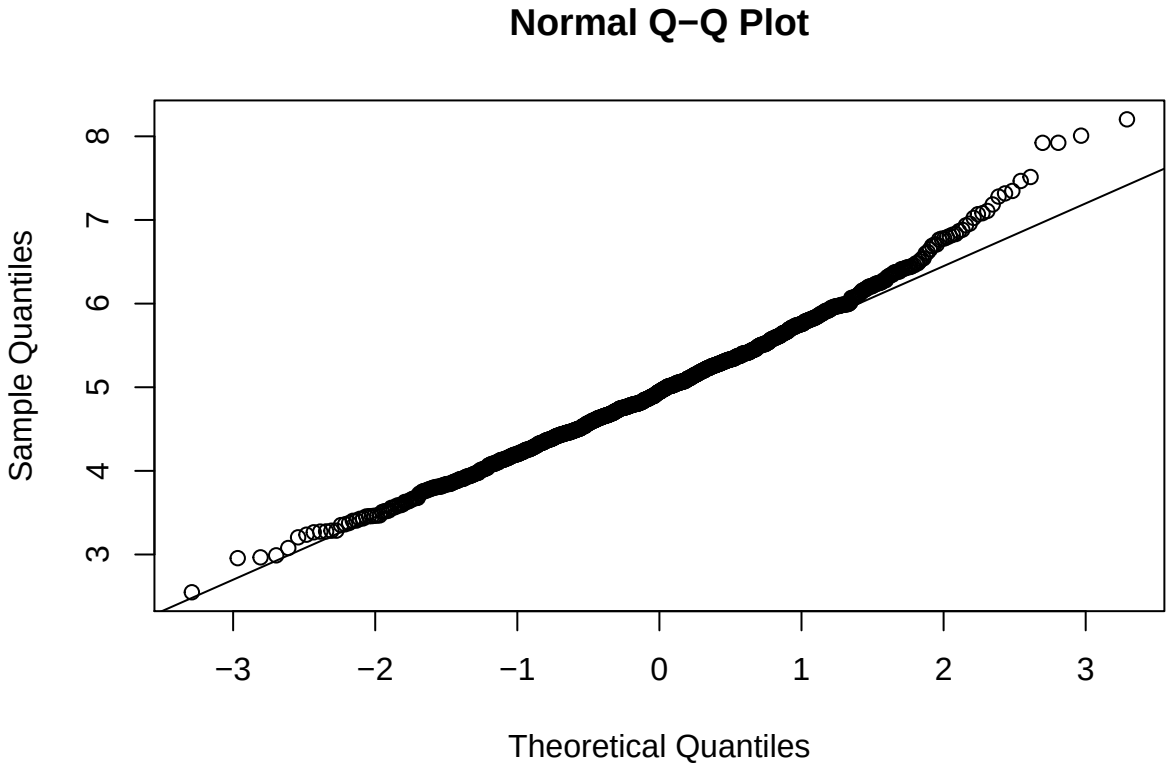
Plotting histogram of averages of 40 samples of exponentials against normal distribution.

```
g <- ggplot() + aes(data) + theme_classic()
g + geom_histogram(aes(y = ..density..),
                   binwidth = .2, alpha = 0.5) +
  stat_function(fun = "dnorm", args = list(mean = tmean, sd = tsd)) +
  geom_vline(xintercept=tmean, size=0.5) +
  ggtitle("Exponential Distribution")
```



Plotting qqplot to illustrate closness of the simulated distribution to normal distribution.

```
qqnorm(data); qqline(data)
```


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
3
4
5
6
7
8
−3
−2
−1
0
1
2
3