Statistical Inference Course Project Part1

Atsuko Yamamoto

January 24, 2015

This is the project for the statistical inference class. In it, you will use simulation to explore inference and do some simple inferential data analysis. The project consists of two parts:

1. A simulation exercise.
2. Basic inferential data analysis.

Part1 A simulation exercise

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. 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 1/lambda. Set lambda = 0.2 for all of the simulations.

Simulations

```
Sys.setlocale("LC_TIME","C")
```

```
## [1] "C"
```

```
setwd("~/Rdata/Inference")
num_sample <- 40
num_test <- 1000
lambda <- 0.2

mexp = NULL
for (i in 1 : num_test) mexp = c(mexp, mean(rexp(num_sample, rate=lambda)))

# mean of distribution of averages
mean_sample <- round(mean(mexp),3)
mean_theoretical <- round(1 / lambda,3)

# Standard deviation & Variance
sd_sample <- round(sd(mexp),3)
var_sample <- round(var(mexp),3)
sd_theoretical = round((1 / lambda) / sqrt(num_sample),3)
var_theoretical = round(1 / lambda^2 /num_sample,3)
```

Results

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

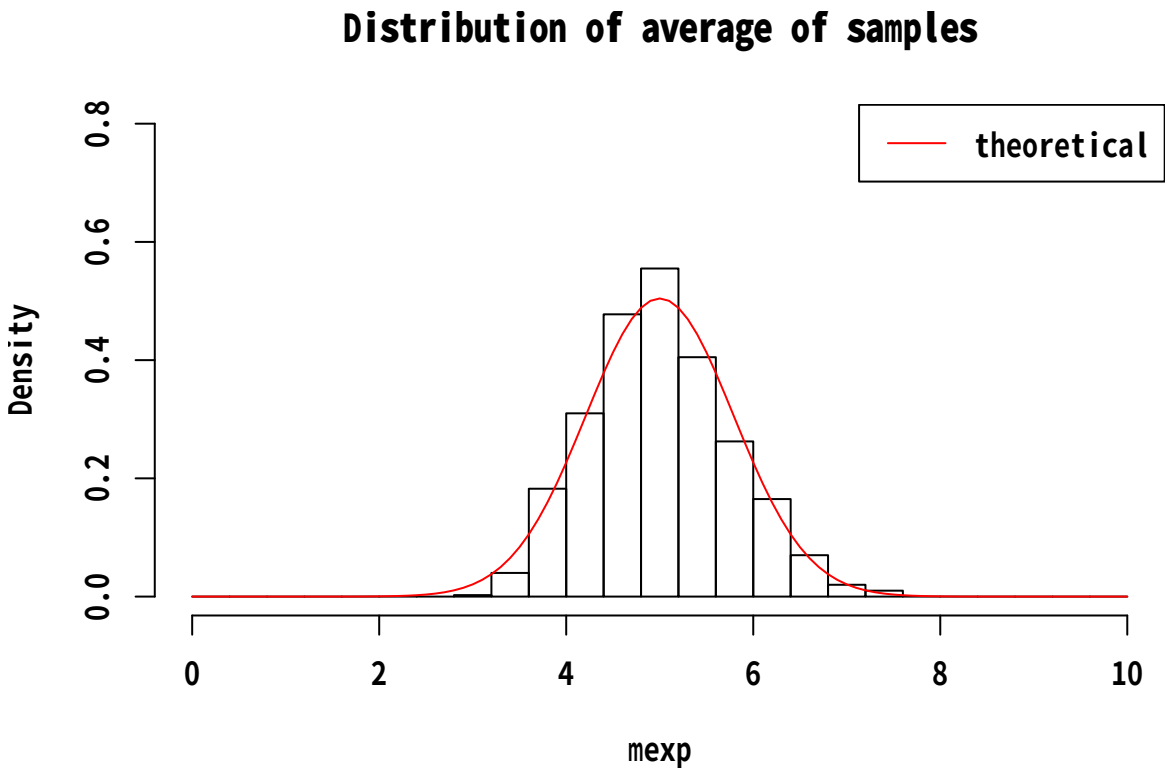
mean of distribution of averages

```
mean_sample
```

```
## [1] 5.011
```

```
mean_theoretical
```

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



Therefore, the center of distribution of averages of 40 exponentials is close to the theoretical center of the distribution.

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

```
# Standard deviation and Variance
sd_sample; sd_theoretical
```

```
## [1] 0.75
```

```
## [1] 0.791
```

```
var_sample; var_theoretical
```

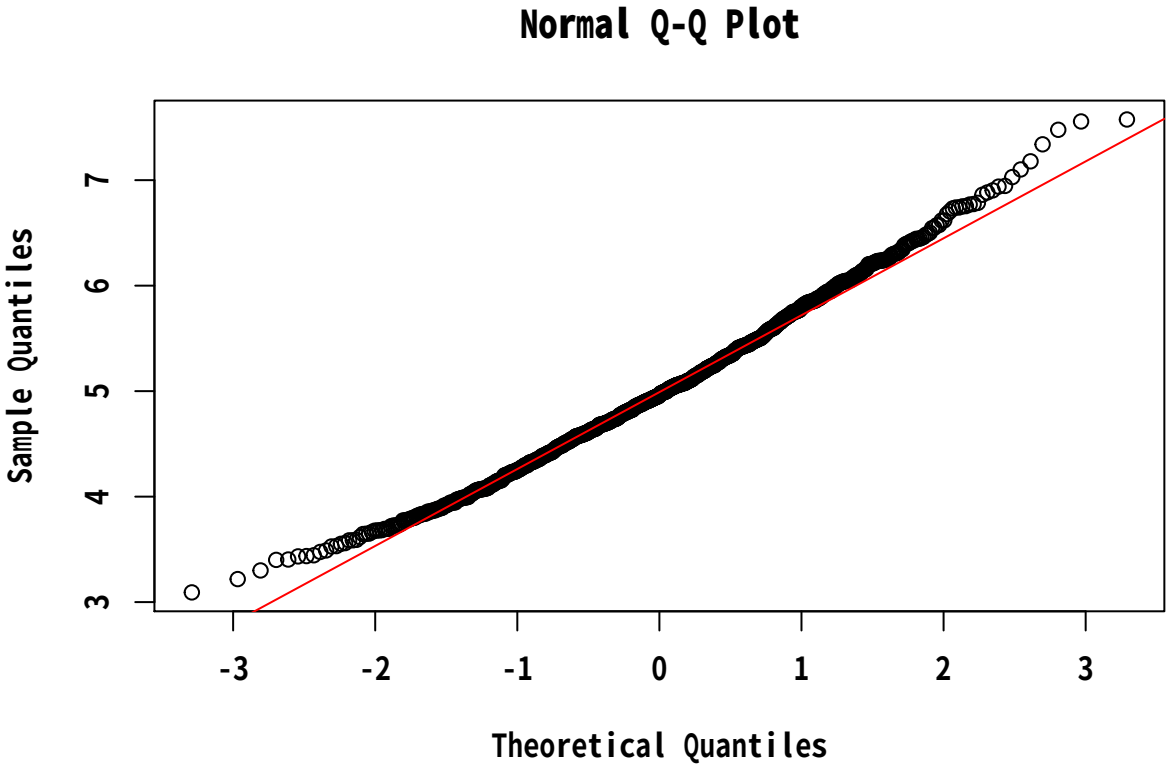
```
## [1] 0.562
```

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

Therefore, the variability in distribution of averages of 40 exponentials is close to the theoretical variance of the distribution.

3. Show that the distribution is approximately normal.

```
qqnorm(mexp)
qqline(mexp, col = 2)
```



From the central limit theorem, the averages of samples follow normal distribution. As evident from the Q-Q plot, the distribution of averages of 40 exponentials is close to a normal distribution.