Statistical Interference Project-Part 1

Albion Dervishi

March 21, 2015

This is the Statistical Inference course part one project. In this project I will use simulation to explore simple inferential data analysis. First step I will generate the mean, standard deviation, variance and standard error for a standard exponential distribution.

```
set.seed(30)
lambda <- 0.2
simulation<- 1000
# 40 samples
n<- 40
sim <- matrix(rexp(simulation*n, rate=lambda), simulation, n)
# mean of exponentials
row_means <- rowMeans(sim)
```

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

```
# distrribution mean
distribution_mean<-mean(row_means)
distribution_mean
```

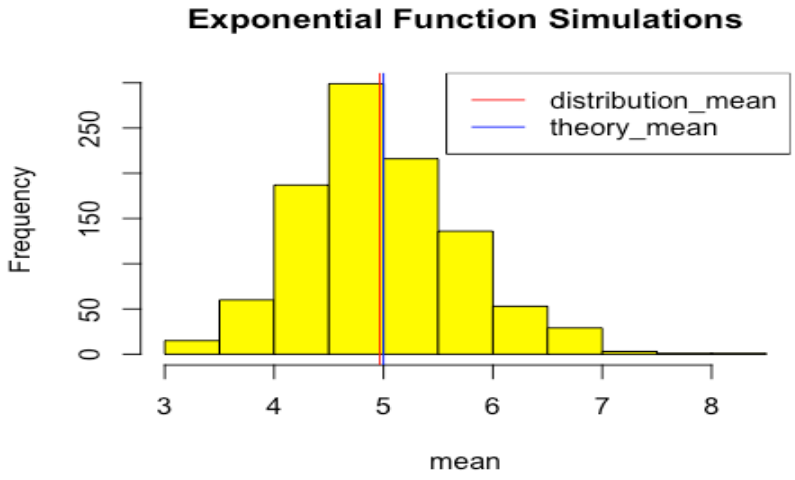
```
## [1] 4.967156
```

```
# theory mean
theory_mean <- 1/lambda
theory_mean
```

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

Visualization

```
hist(row_means, col="yellow", xlab = "mean", main = "Exponential Function Simulations")
abline(v = distribution_mean, col = "red")
abline(v = theory_mean, col = "blue")
legend('topright', c("distribution_mean", "theory_mean"), lty=c(1,1), col=c("red", "blue"
))
```



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

```
# standard deviation of distribution
sd_distribution<- sd(row_means)
sd_distribution
```

```
## [1] 0.7362476
```

```
# standard deviation from analytical expression
sd_theory <- (1/lambda)/sqrt(n)
sd_theory
```

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

```
# variance of distribution
var_distribution <- sd_distribution^2
var_distribution
```

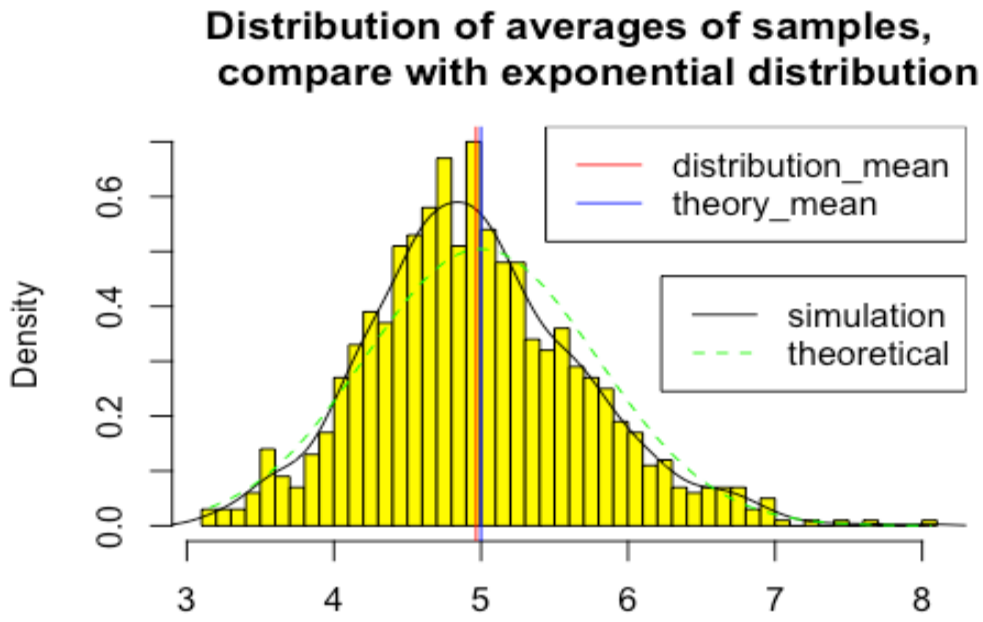
```
## [1] 0.5420605
```

```
# variance from analytical expression
var_theory <- ((1/lambda)*(1/sqrt(n)))^2
var_theory
```

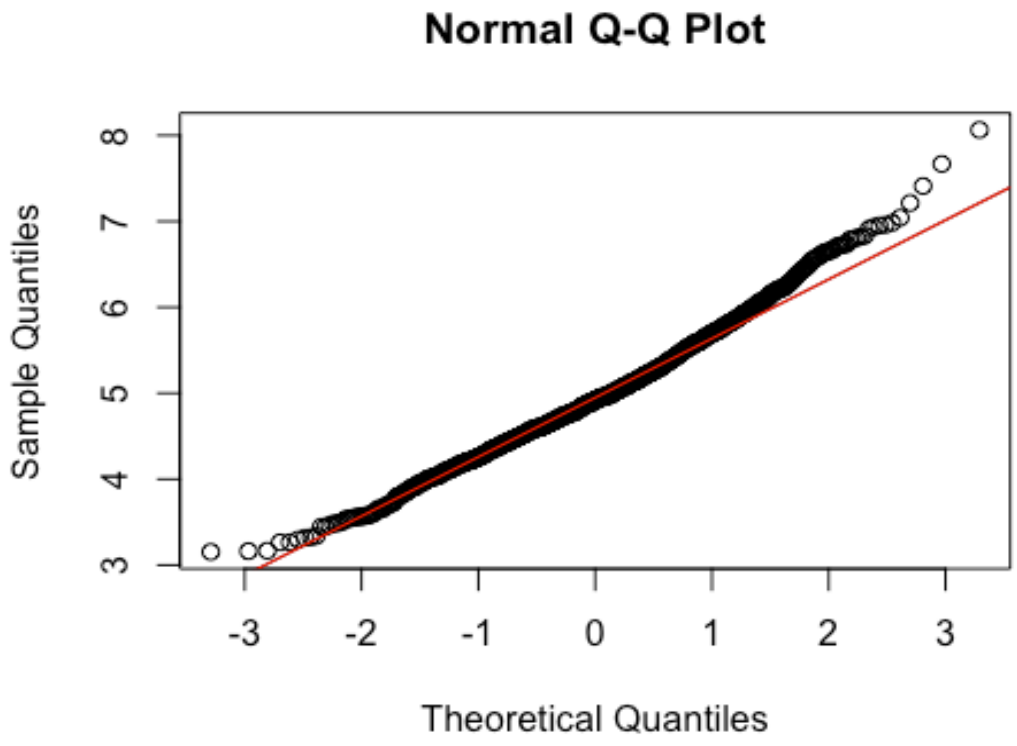
```
## [1] 0.625
```

Question 3-Show that the distribution is approximately normal.

```
hist(row_means, col="yellow", breaks=50, prob=TRUE,
     main="Distribution of averages of samples,
     compare with exponential distribution",
     xlab="")
lines(density(row_means))
abline(v = distribution_mean, col = "red")
abline(v = theory_mean, col = "blue")
xfit <- seq(min(row_means), max(row_means), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(n)))
lines(xfit, yfit, pch=22, col="green", lty=2)
legend('topright', c("distribution_mean", "theory_mean"), lty=c(1,1), col=c("red", "blue"
))
legend('right', c("simulation", "theoretical"), lty=c(1,2), col=c("black", "green"))
```




```r
qqnorm(row_means); qqline(row_means)
qqline(row_means, col="red", lty=1)
```



As per evidence from the Q-Q plot, the distribution of averages of a sample is pretty close to a normal distribution. This comprehended that, increasing the sample size, the data would eventually fit more closely to a normal distribution.