Part 1

Edmond Ho-Yin Lau

August 31, 2020

##Overview The purpose of this data analysis is to investigate the exponential distribution and compare it with the properties of a normal distribution.

##Exponential distribution simulation The simulation generates a distribution of averages of 40 exponentials (lambda set to 0.2) over 1000 simulations.

```
set.seed(1)
lambda <- 0.2; n <- 40; s <- 1000
sim <- replicate(s, rexp(n, lambda))
meanE <- apply(sim, 2, mean)
```

###Comparison of sample mean and theoretical mean of the distribution

```
meanS <- mean(meanE); meanS
```

```
## [1] 4.990025
```

```
meanT <- 1/lambda; meanT
```

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

The analytics mean is 4.99, the theoretical mean is 5. The center of distribution of averages of 40 exponentials is very close to the theoretical center of the distribution. The **Plot 1** shows the distribution with both the means.

###Comparison of sample variance and theoretical variance of the distribution

```
varS <- (sd(meanE))^2; varS
```

```
## [1] 0.6111165
```

```
varT <- ((1/lambda)*(1/sqrt(n)))^2; varT
```

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

Standard Deviation of the distribution is 0.782 with the theoretical SD of 0.791.

###Check if distribution is approximately normal

From **Plot 2**, the distribution of averages of 40 exponentials appears to be very close to a normal distribution. Also the C.I. for 1/lamda accomodates the simulated mean. H_0 that difference between both means is 0 is failed to be rejected.

```r
mean(meanS) + c(-1, 1) * 1.96 * sqrt(varS)/sqrt(n)
```

```
## [1] 4.747762 5.232289
```

From **Plot 3**, the Q-Q plot shows that normality is probably a reasonably good approximation. But the Shapiro-Wilk normality test rejects the H_0 that "the samples come from a normal distribution" against the H_1 that "the samples do not come from a normal distribution"

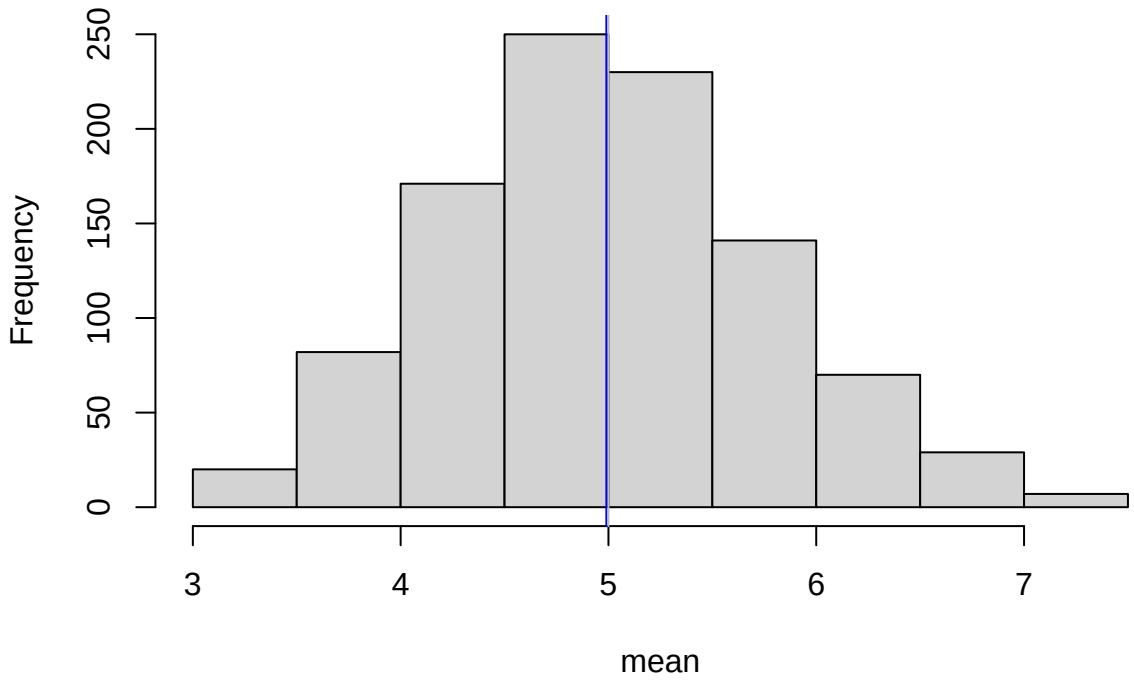
```r
shapiro.test(meanE)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  meanE
## W = 0.99352, p-value = 0.0002466
```

##Appendix

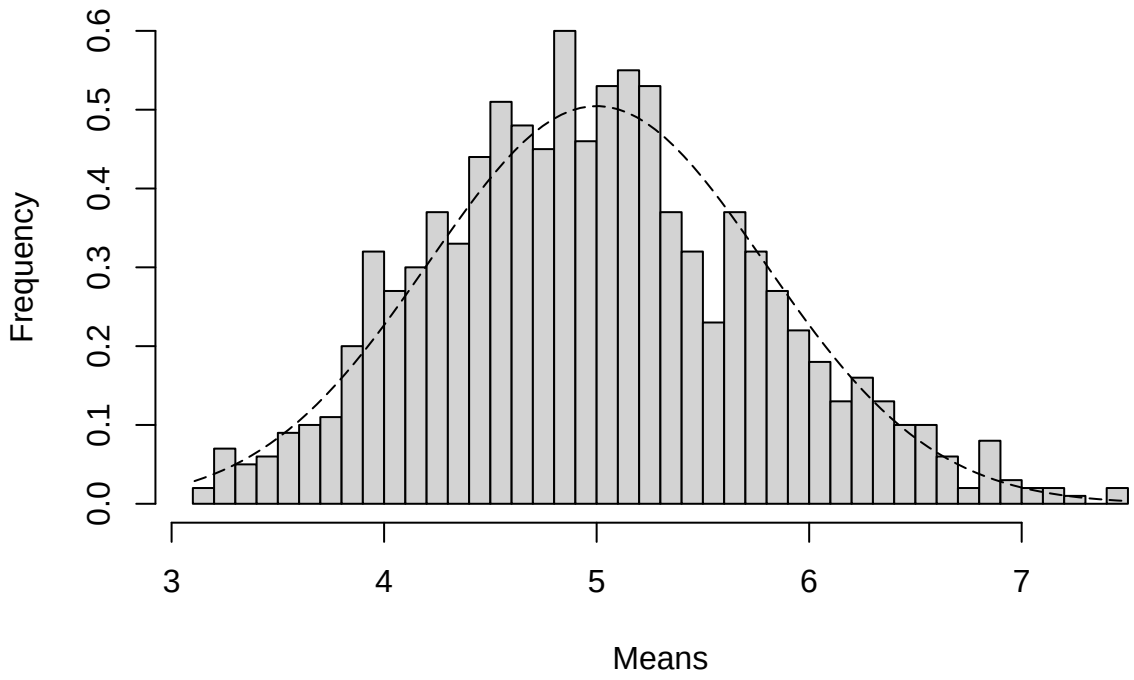
Plot 1: Histogram of means of exponential distribution

```r
hist(meanE, xlab = "mean", main = "")
abline(v = meanS, col = "blue")
abline(v = meanT, col = "grey")
```



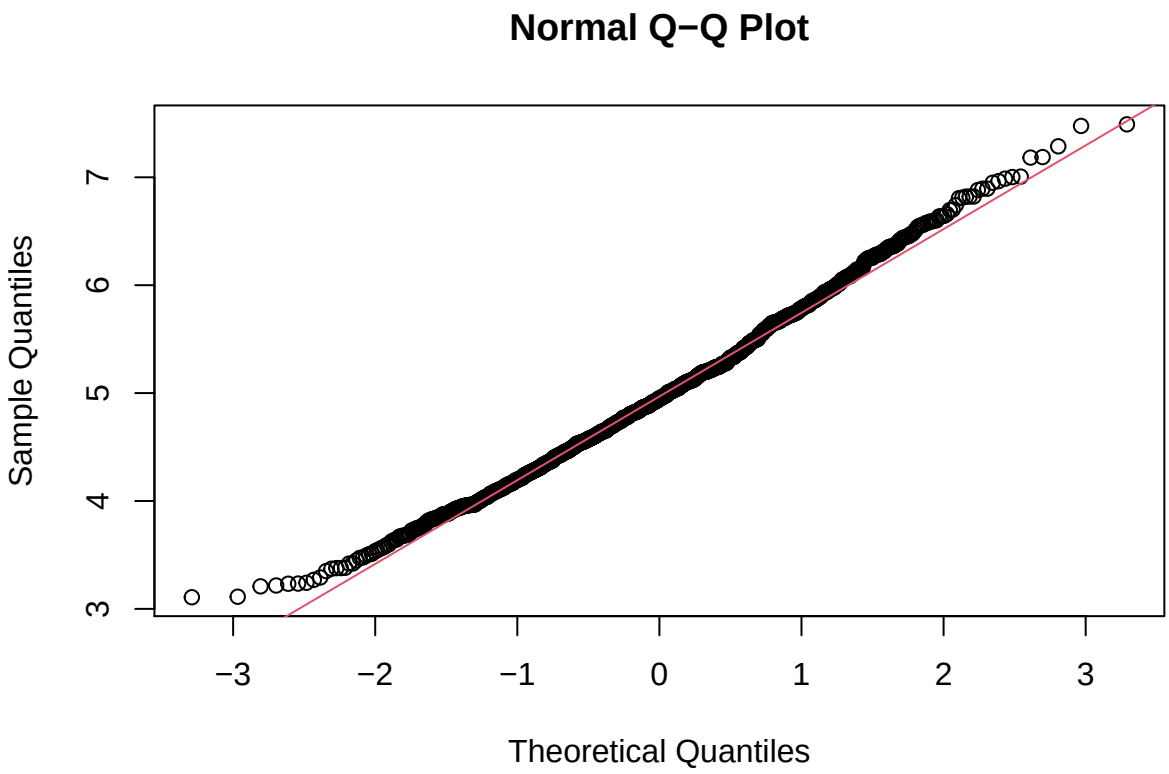
Plot 2: Exponential vs Normal Distribution

```
xfit <- seq(min(meanE), max(meanE), length=100)
yfit <- dnorm(xfit, mean=meanT, sd=sqrt(varT))
hist(meanE,breaks=n,prob=T,xlab = "Means",main="",ylab="Frequency"); lines(xfit, yfit, pch=25, lty=5)
```



Plot 3:

```
qqnorm(meanE); qqline(meanE, col = 2)
```


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