Statistical_Inference

This document examines the relationships between population mean, variance vs sample mean, variance via simulation in the light of the Central Limit Theorem.

Simulations

In our analysis we'll have sample dataset of 40 points from an exponential distribution. Exponential distribution has a lambda value of 0.2 Exponential distribution is a continuous function with mean and std deviation of 1/lambda. Let's do a thousand simulated averages of 40 exponentials.

Parameters

```
set.seed(3)
lambda <- 0.2
nosim <- 1000
sample_size <- 40
```

Matrix if 1000 Simulations, Exponential Distribution

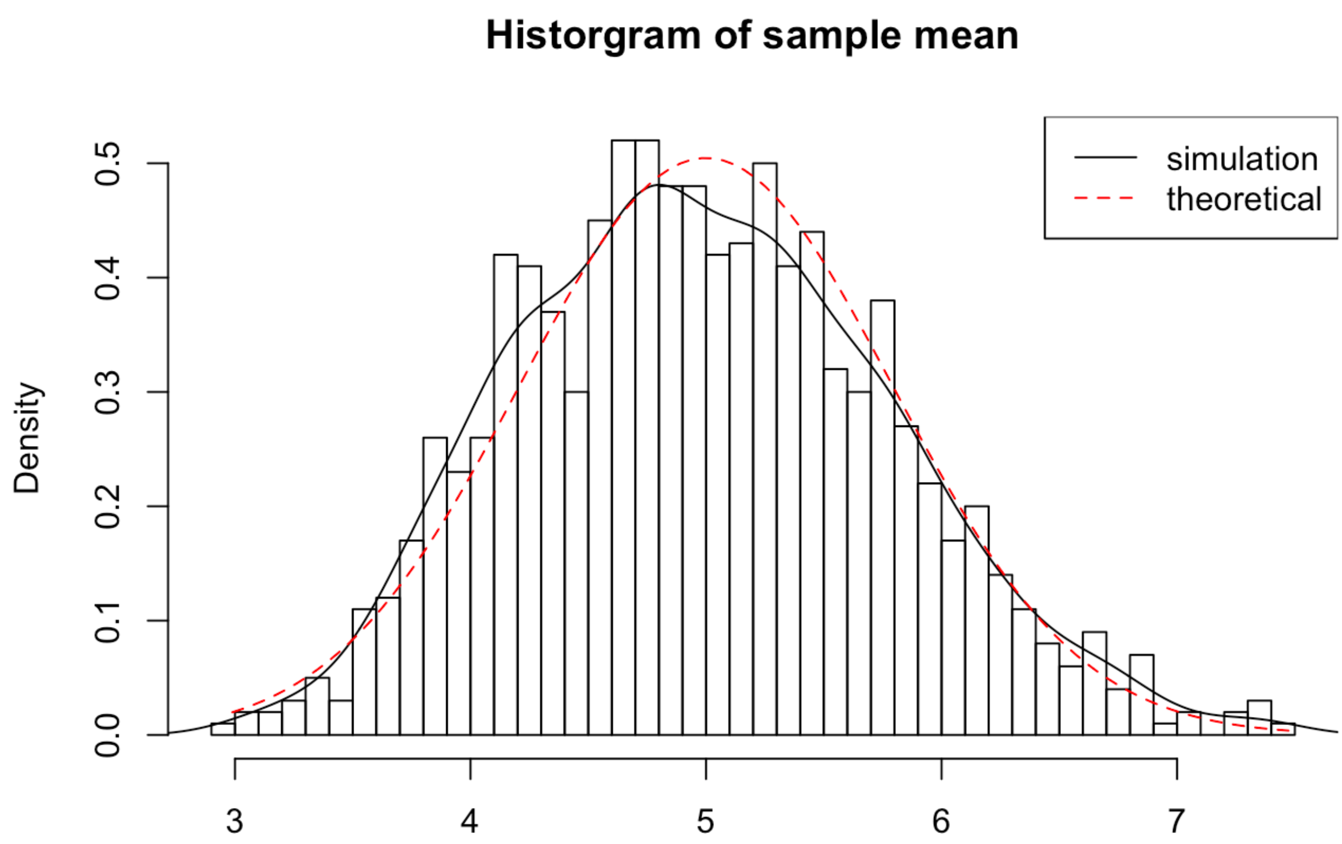
```
sim <- matrix(rexp(nosim*sample_size, rate=lambda), nosim, sample_size)
row_means <- rowMeans(sim)
```

The distribution of sample means is as follows.

```
# histogram
hist(row_means, breaks=40, prob=TRUE,
     main="Historgram of sample mean",
     xlab="")

# averages of samples
lines(density(row_means))

# theoretical average density
xfit <- seq(min(row_means), max(row_means), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(sample_size)))
lines(xfit, yfit, pch=22, col="red", lty=2)
legend('topright', c("simulation", "theoretical"), lty=c(1,2), col=c("black", "re
d"))
```



Theoretical Center

```
μ <- 1/lambda
μ
```

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

```
center <- mean(row_means)
```

The distribution of sample means is centered at 4.9866197 and the theoretical center of the distribution is 5.

Sample Variance vs Theoretical Variance

```
σ <- 1/lambda/sqrt(sample_size)
σ
```

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

```
Var <- σ^2
Var
```

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

The expected standard deviation is 0.7905694, the variance Var of standard deviation σ is 0.625

```
σ_x <- sd(row_means)
σ_x
```

```
## [1] 0.7910484
```

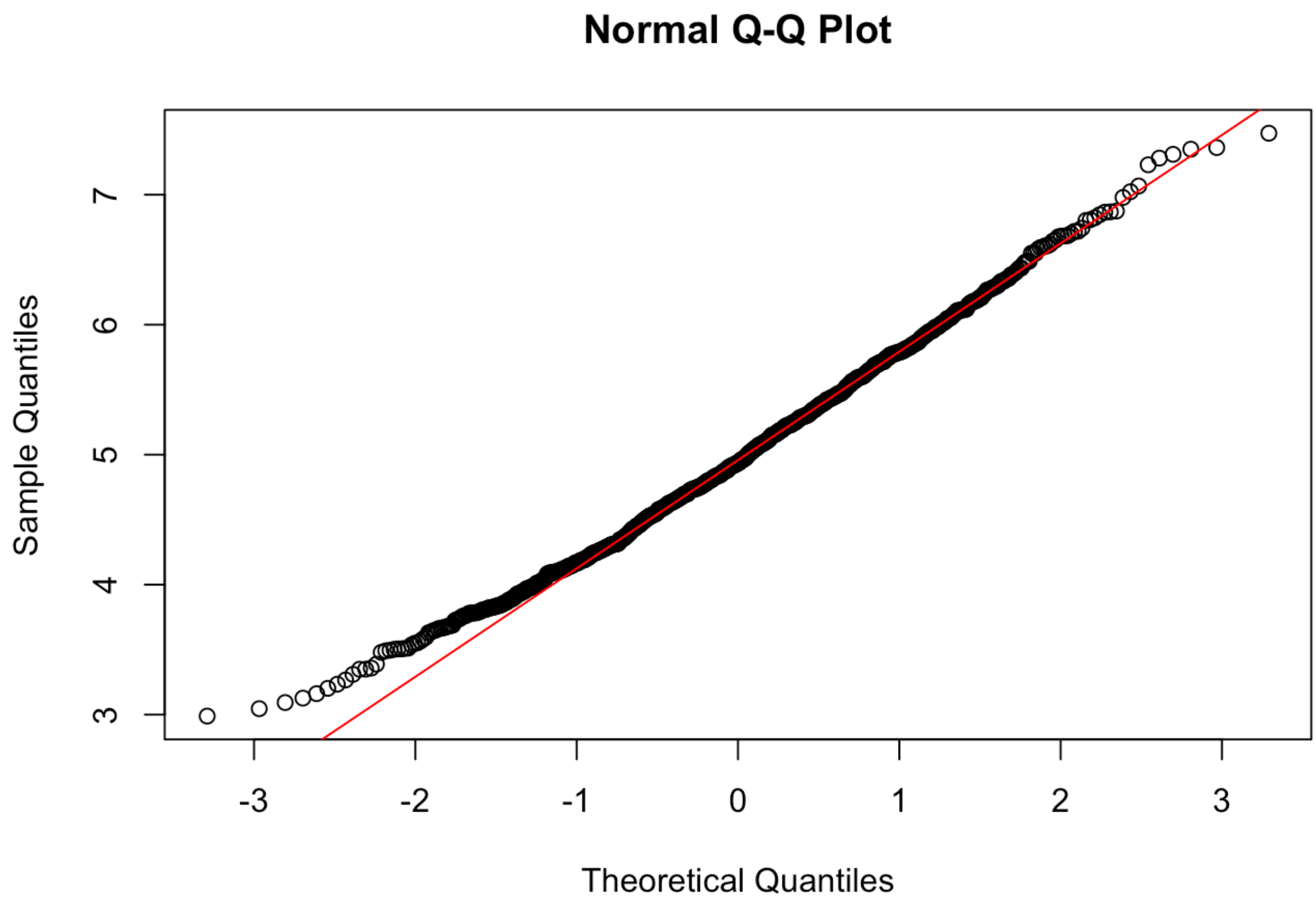
```
Var_x <- var(row_means)
Var_x
```

```
## [1] 0.6257575
```

Sample standard deviation is 0.7910484, sample variance is 0.6257575. Result: Expected and sample values are very close.

Distribution

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



The distribution of averages is close to a normal distribution.