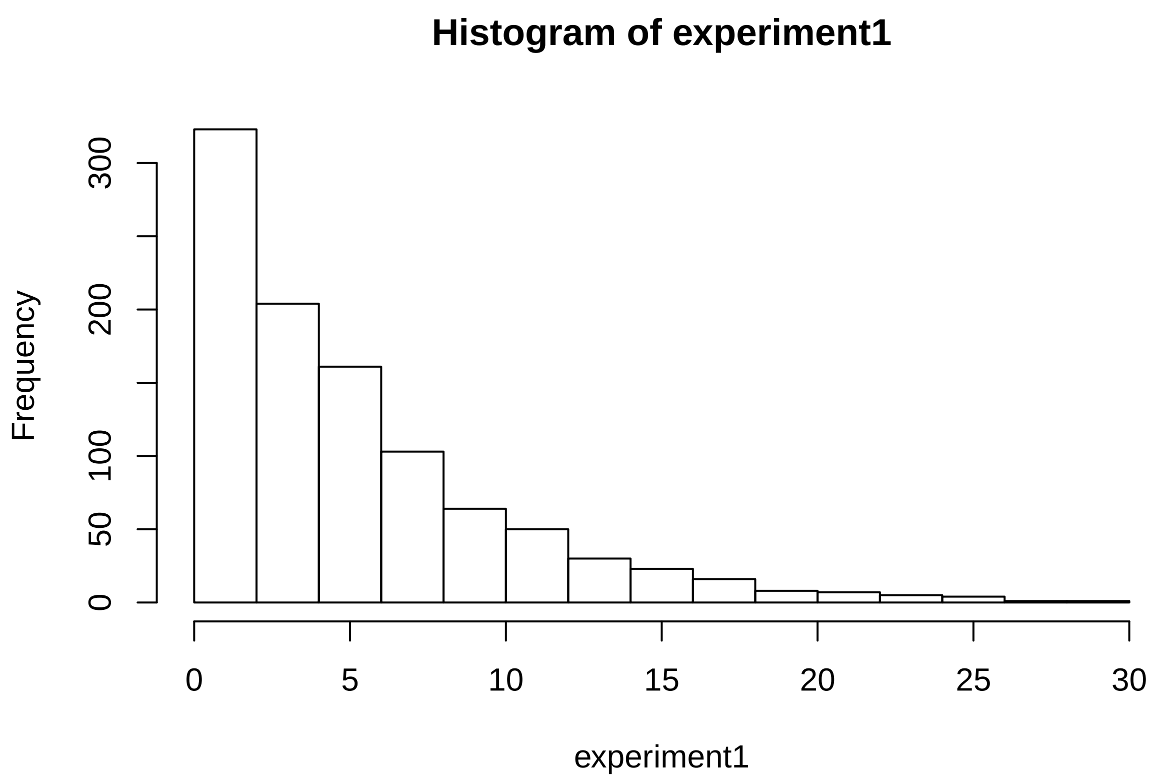
Statistical Inference: Course Project

rtaph

Part I: The Exponential Distribution

We can visualize the distribution with the histogram of an experiment in which we take 1000 random draws from the exponential distribution. Throughout this paper, we will use $\lambda = 0.2$:

```
set.seed(5678)
n = 1000; lambda = 0.2;
experiment1 = rexp(n, lambda)
hist(experiment1)
```



In this particular experiment, the mean is calculated to be 5.1165041. To make use of the properties of asymptotics, we run a monte carlo experiement where we repeat the original experiment 40 times. The means of each experiment is saved in a vector:

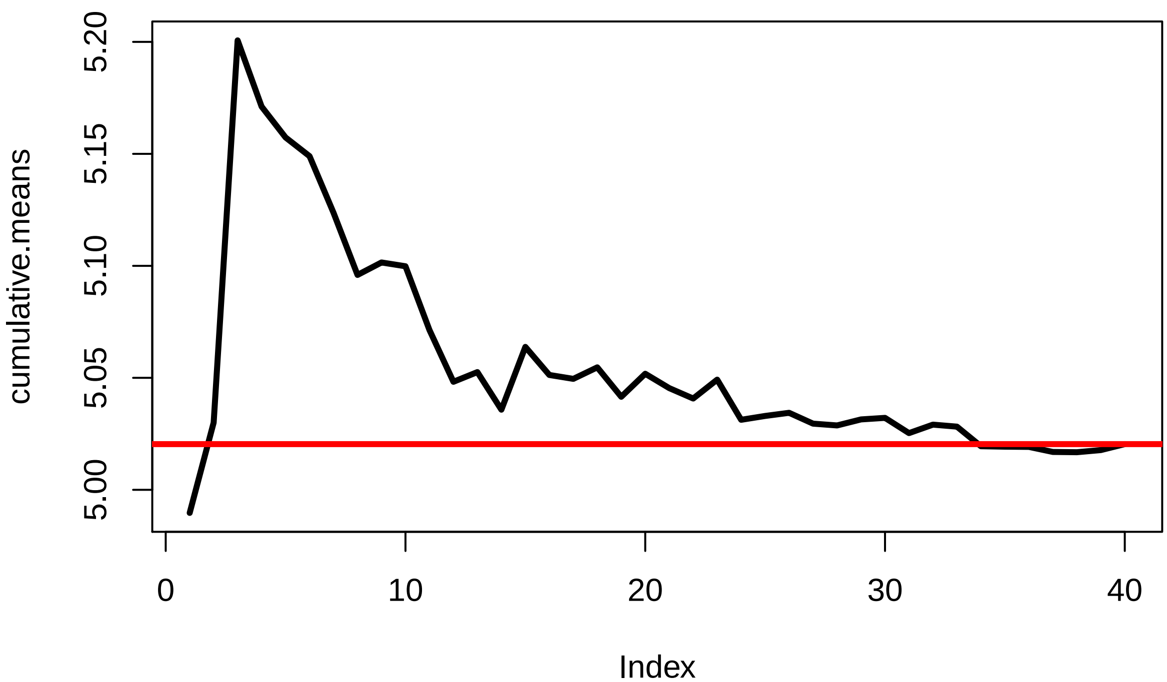
```
monte.carlo.means = NULL; m = 40;
for (i in 1 : m) monte.carlo.means = c(monte.carlo.means, mean(rexp(n, lambda)))
```

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

The expected value of an exponentially distributed random variable X with rate parameter λ is given by $E[X] = \frac{1}{\lambda}$. For the distribution chosen in this paper, the theoretical mean evaluates to $1/0.2 = 5$.

By plotting the cumulative mean for each additional experiment, we can see that the it asymptotes to the theoretical mean of 5 (shown in red).

```
theoretical.mean = (1/lambda)
cumulative.means <- cumsum(monte.carlo.means) / (1 : m)
plot(cumulative.means, type = "l", lwd = 3)
abline(h=cumulative.means[m], col = "red", lwd = 3)
```



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

The theretical variance of the distribution is given by the expression $Var[X] = \frac{1}{\lambda^2}$. We compare both the theoretical and empirical results below:

```
(theoretical.var = 1/(lambda^2) )
```

```
## [1] 25
```

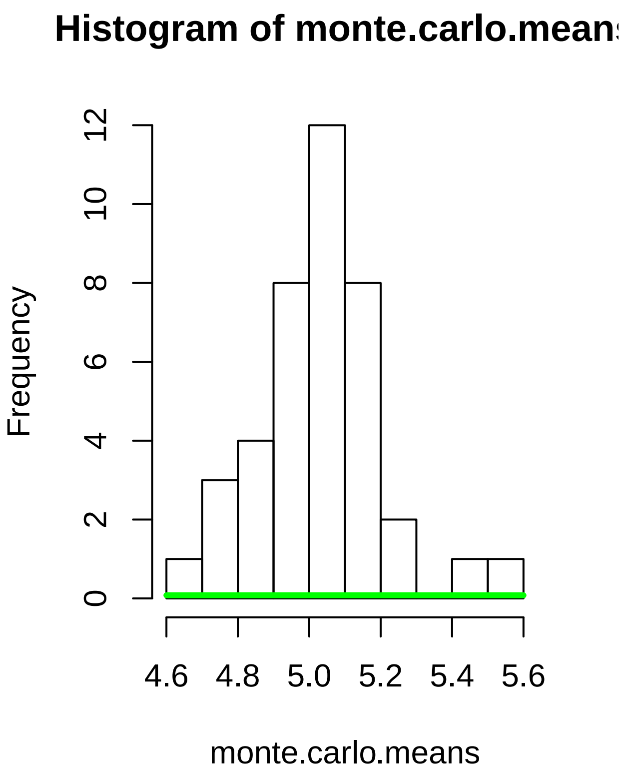
```
(empirical.var = var(experiment1))
```

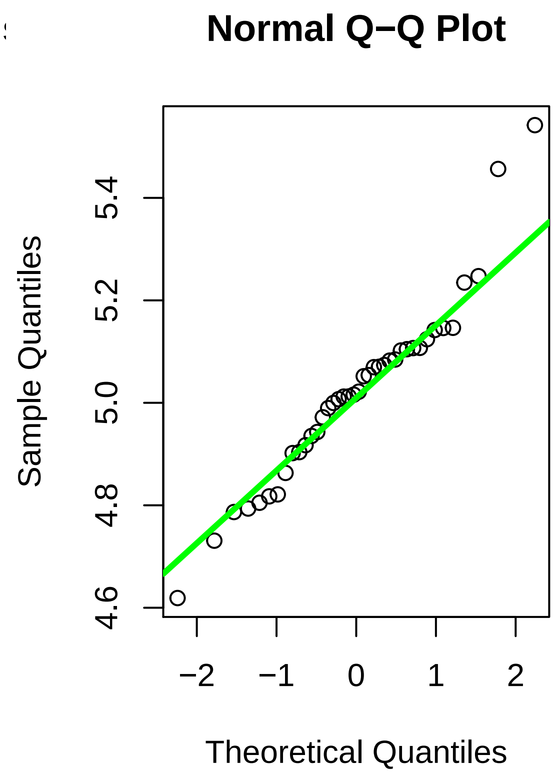
```
## [1] 23.67207
```

3. Show that the distribution is approximately normal.

The Central Limit Theorem (CLT) tells us that the resulting distribution of means from our 40 experiments should tend towards normality. We can visually test this on the basis of our empirical data by plotting a histogram of the means and producing a QQ plot.

```r
par(mfrow=c(1,2))
hist(monte.carlo.means);
x<-seq(min(monte.carlo.means),max(monte.carlo.means),0.1)
curve(dnorm(x, theoretical.mean, sqrt(theoretical.var)), col="green", lwd=3, add=TRUE, yaxt="n")
qqnorm(monte.carlo.means); qqline(monte.carlo.means, col="green", lwd=3)
```





The green line represents the normal distribution for comparison. Although the distribution is not perfectly normal, visual inspection reveals that the fit is pretty good. The deviations from normality fall within the expected range of noise for a monte carlo simulation with 40 draws.