Statistical Inference Course Project: Simulation Exercises

This is the project for the statistical inference class. The project consists of two parts:

1. Simulation exercises.
2. Basic inferential data analysis.

Simulation Exercises

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 also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

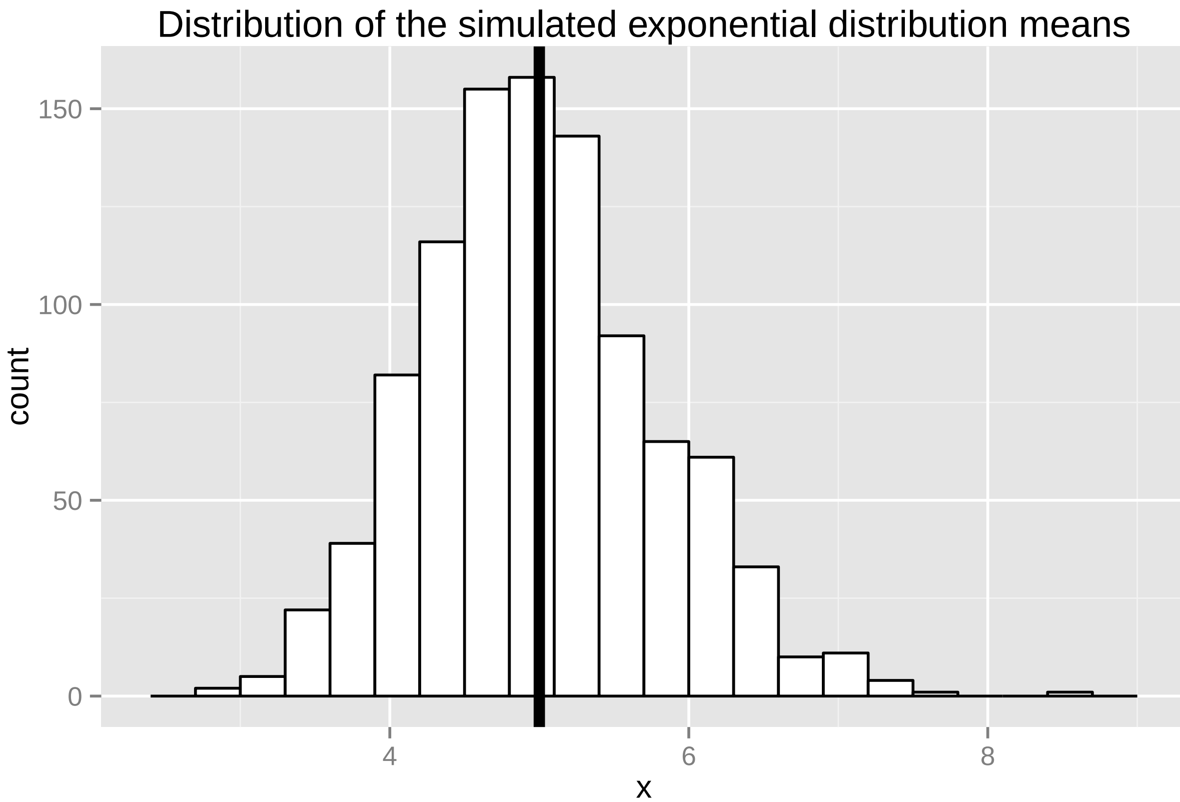
Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s. You should:

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.
2. Show how variable it is and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.
4. Evaluate the coverage of the confidence interval for 1/lambda.

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution. The theoretical center of the exponential distribution is 1/lambda (5 in this example). If we simulate 1000 samples of n = 40 exponentials and take the mean of each sample we get the following distribution, as seen in the figure below.

```
#1. Show where the distribution is centered at and compare it to the theoretical center of the distribu
n<-40
nosim<-1000
lambda<- 0.2
set.seed(100)
#create matrix of simulations and calculate means

test<-data.frame(x=c(apply(matrix(rexp(n*nosim,lambda), nosim), 1, mean)))
if (!require(ggplot2)) install.packages('ggplot2'); require(ggplot2)
ggplot(test, aes(x=x)) +
  geom_histogram(binwidth=.3,colour="black", fill="white")+
  geom_vline(xintercept = 5, size = 2)+ggtitle("Distribution of the simulated exponential distribution m
```




```
mean.dist<-mean(as.matrix(test))
```

The mean of this distribution is 5, which is the same as the theoretical center of 5. The theorictical center is marked by the black vertical line in the histogram above. As you can see, the distribution of the sample means peaks around the theoretical center.

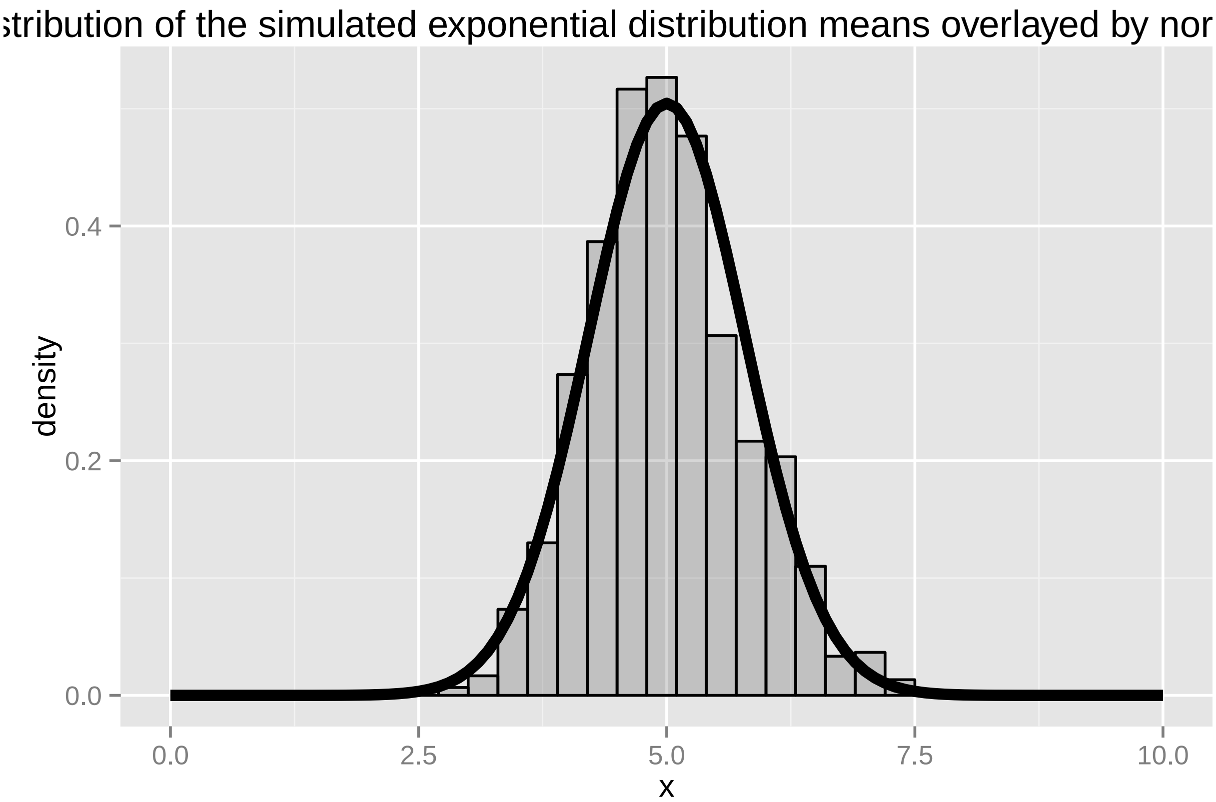
```
std.test<-sd(as.matrix(test))
se.test<-(1/lambda)/sqrt(n)
```

2. Show how variable it is and compare it to the theoretical variance of the distribution. The distribution of the mean of 40 exponentials has a standard deviation of 0.796. If this value is compared to the standard error (not the standard deviation of 5) 0.7906, we see that the sample distribution's variance is the same as the theorectical variance of the distribution.

3. Show that the distribution is approximately normal. If a normal distribution of mean = 1/lambda and sd = (1/lambda)/sqrt(n) is overlayed over a histogram of the sample mean distribution (refer to figure below), we see that the histogram follows the general shape of the normal distribution(black line in the figure).

```
if (!require(ggplot2)) install.packages('ggplot2'); require(ggplot2)
```

```
g <- ggplot(test, aes(x = x)) + geom_histogram(alpha = .20, binwidth=.3, colour = "black", aes(y = ..den
g + stat_function(fun = dnorm, size = 2, args = with(test, c(mean = 1/lambda, sd = (1/lambda)/sqrt(n)))
```



4. Evaluate the coverage of the confidence interval (95% of distribution) for 1/lambda.

```
## [1] 957
```

To find the coverage of 95% of the distribution, I found the mean values of the sample distribution at the 2.5% and 97.5% confidence intervals, 3.4502 and 6.5492 respectively. I then found the number of means in the 1000 sample distribution that lie within these intervals. A total of 957 are found with in the 2.5% and 97.5% confidence intervals of the distribution. This is 95.7% of the mean values in the distribution, which is a close approximation of of the 95% coverage.