Statistical Inference Course Project Part 1

Problem 1 description: 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: $\bar{x}$ś$1.96s/\sqrt{n}$.

```
lambda = 0.2
n = 40
nsims = 1:1000
set.seed(820)
means <- data.frame(x = sapply(nsims, function(x) {mean(rexp(n, lambda))}))
head(means)
```

Run Simulations

```
##       x
## 1 5.750
## 2 3.808
## 3 4.058
## 4 3.999
## 5 4.313
## 6 4.418
```

```
mean(means$x)
```

```
## [1] 4.999
```

```
sd(means$x)
```

```
## [1] 0.7909
```

```
# Expected standard deviation
(1/lambda)/sqrt(40)
```

```
## [1] 0.7906
```

```r
# Variance of our simulations:
var(means$x)
```

```
## [1] 0.6256
```

```r
# Expected variance
((1/lambda)/sqrt(40))^2
```

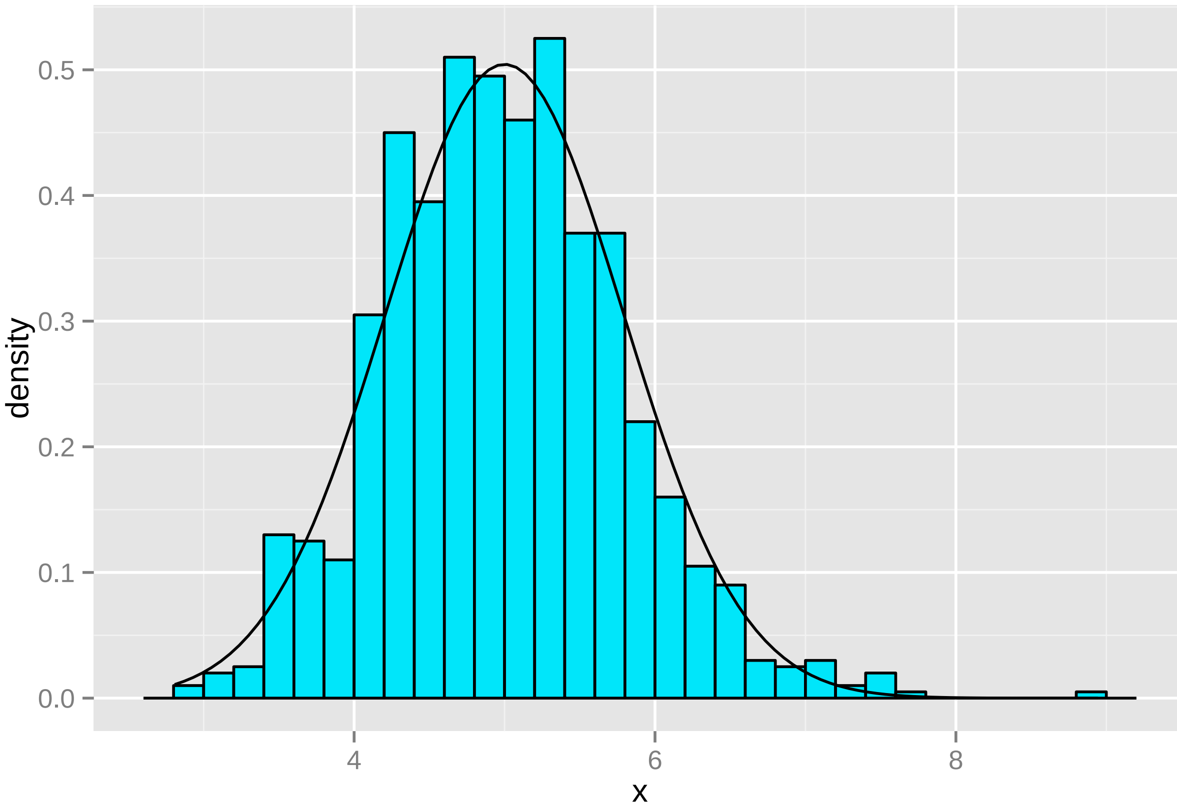
```
## [1] 0.625
```

Questions 1 1.Center of the distribution: 4.9988. Expected center: 5.0. The mean of the means of the exponential of 1000 simulations of 40 exponential(0.2)s is 4.9988, which is very close to the expected mean of $1/0.2 = 5.0$.

Questions 2 2.Variability of the distibution. The standard deviation of 0.7909 is also close to the expected standard deviation of 0.79056. (Expected standard deviation using Central Limit Theorem: $\sigma/\sqrt{n}$, or (1/lambda) $/\sqrt{n}$). Likewise, the variance and expected variance are 0.6256 and 0.625, respectively.

Question 3 Below is a histogram plot of the means of the 1000 simulations of rexp(n, lambda). It is overlaid with a normal distribution with mean 5 and standard deviation 0.7909. Yes, the distribution of our simulations appears normal.

```r
library(ggplot2)
ggplot(data = means, aes(x = x)) +
    geom_histogram(aes(y=..density..), fill = I('#00e6fa'),
                   binwidth = 0.20, color = I('black')) +
    stat_function(fun = dnorm, arg = list(mean = 5, sd = sd(means$x)))
```



Question 4 Evaluate the coverage of the confidence interval for 1/lambda: $\bar{x}$ś$1.96s/\sqrt{n}$.

```
mean(means$x) + c(-1,1)*1.96*sd(means$x)/sqrt(nrow(means))
```

```
## [1] 4.950 5.048
```

The 95% confidence interval for the mean of the means is 4.950-5.047.

More Simulations If we use 100,000 simulations instead of 1,000, it definitely converged toward the theoretical distribution.

```
lambda = 0.2
n = 40
nsims = 1:100000
set.seed(821)
means <- data.frame(x = sapply(nsims, function(x) {mean(rexp(n, lambda))}))
mean(means$x)
```

```
## [1] 4.998
```

```
sd(means$x)
```

```
## [1] 0.789
```

```
# Expected standard deviation
(1/lambda)/sqrt(40)
```

```
## [1] 0.7906
```

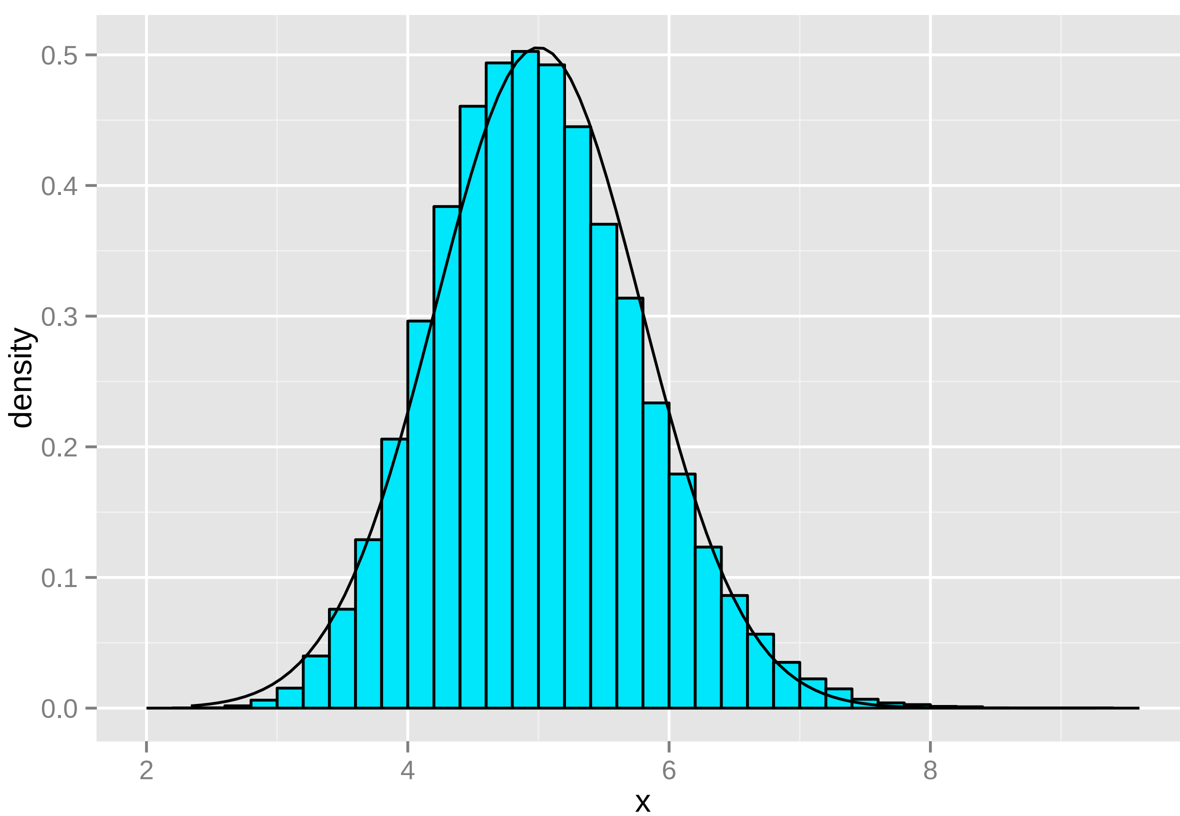
```
# Variance of our simulations:
var(means$x)
```

```
## [1] 0.6225
```

```
# Expected variance
((1/lambda)/sqrt(40))^2
```

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

```
ggplot(data = means, aes(x = x)) +
    geom_histogram(aes(y=..density..), fill = I('#00e6fa'),
                   binwidth = 0.20, color = I('black')) +
    stat_function(fun = dnorm, arg = list(mean = 5, sd = sd(means$x)))
```




```
mean(means$x) + c(-1,1)*1.96*sd(means$x)/sqrt(nrow(means))
```

```
## [1] 4.993 5.003
```