Statistical Inference Course Project Part 1

Author: Norman Benbrahim

Overview 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 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Question 1

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

Let's start by simulating some data

```
set.seed(0)
lambda <- 0.2
n <- 40
B <- 1000

# simulate and get means
sim_exp <- replicate(B, rexp(n, lambda))
means_exp <- apply(sim_exp, 2, mean)
```

The analytical mean is given by:

```
mean(means_exp)
```

```
## [1] 4.989678
```

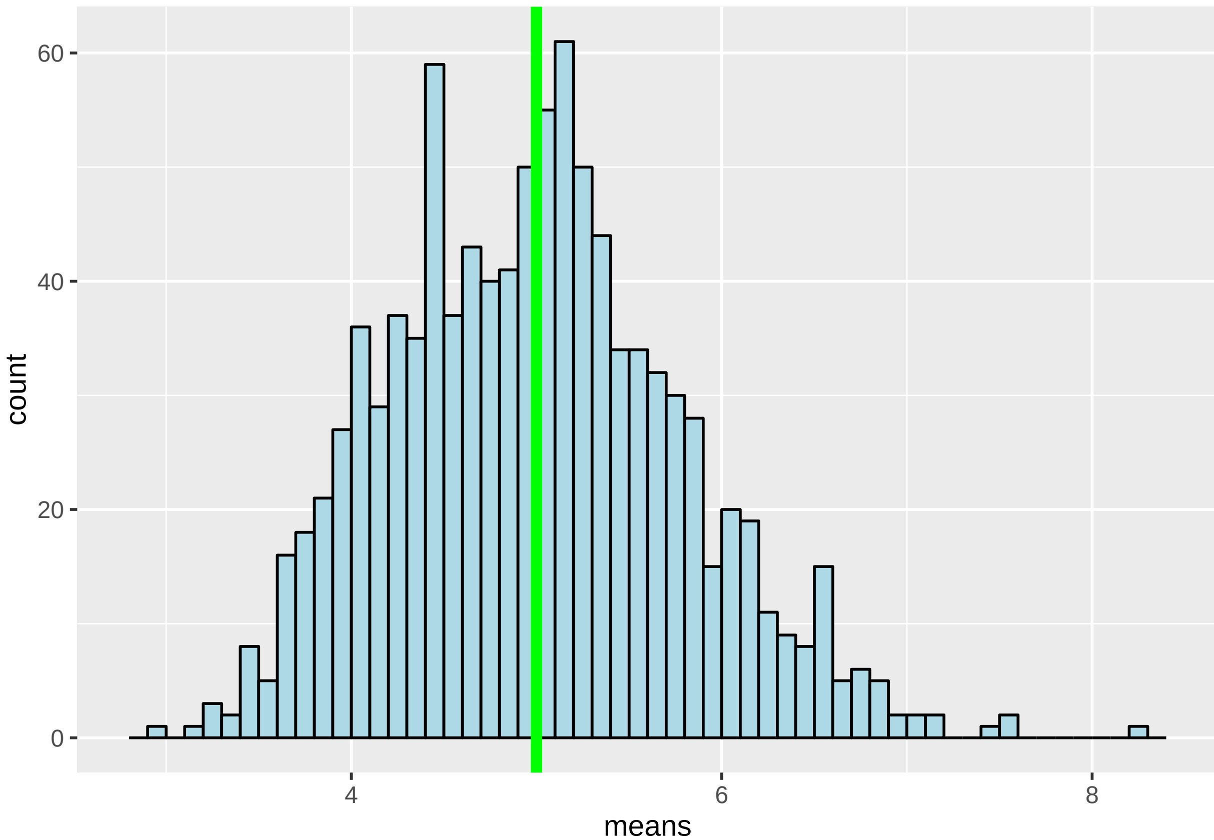
while the theoretical mean is given by

```
1/lambda
```

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

They're pretty close. Let's visualise this in a histogram though

```
library(ggplot2)
g <- ggplot(data.frame(means = means_exp), aes(x = means))
g <- g + geom_histogram(color="black", fill="lightblue", binwidth = 0.1)
g <- g + geom_vline(xintercept = 1/lambda, colour="green", size = 2)
g
```



So we see indeed that the center of distribution of averages of 40 exponentials is very close to the theoretical center of the distribution.

Question 2

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

The analytical expression gives standard deviation and variances as:

```
# standard deviation
(1/lambda)/sqrt(n)
```

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

```
# variance
((1/lambda)/sqrt(n))^2
```

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

and in the simulated case,

```
# standard deviation
sd(means_exp)
```

```
## [1] 0.7862304
```

```r
# variance
sd(means_exp)^2
```

```
## [1] 0.6181582
```

So in comparison, the "error" in our measurement can be evaluated by the absolute value in the difference. The "error"" in standard deviation is:

```r
abs(((1/lambda)/sqrt(n))^2 - sd(means_exp)^2)
```

```
## [1] 0.006841764
```

which is fairly small.

Question 3

Show that the distribution is approximately normal

We know that the central limit theorem guarantees this distribution is approximately normal. Let's demonstrate it with the Q-Q plot. If the quantiles fall along a line, this guarantees that the distribution estimated is approximately normal. In `R` this is easy:

```r
qqnorm(means_exp)
```