Part 1: Simulation

Lee Bunce

1 October 2017

This report investigates the exponential distribution. Using R's *rexp* function we ake 1000 sambles of size 40 fron an exponentiaal distribution with lamda equal to 0.2. We then consider the mean and standard deviations of these samples, comparing them to the values predicted by statistical theory. Finally we also consider whether our sample distribution is approximately normal.

1. Generating the sample

First we generate 1000 samples of size 40 fron an exponential distribution with lamda equal to 0.2. We do this by arranging 40 x 1000 samples into a matrix with 40 rows and 1000 columns.

```
set.seed(20171001)

sim <- matrix(rexp(40*1000, 0.2), 1000, 40)
```

2. Sample mean

First we investigate the means of our samples. Theory tells us that the mean of an exponential distribution is 1 divided by lamda, so in our case we should expect means of our samples to cluster around $1/0.2 = 5$.

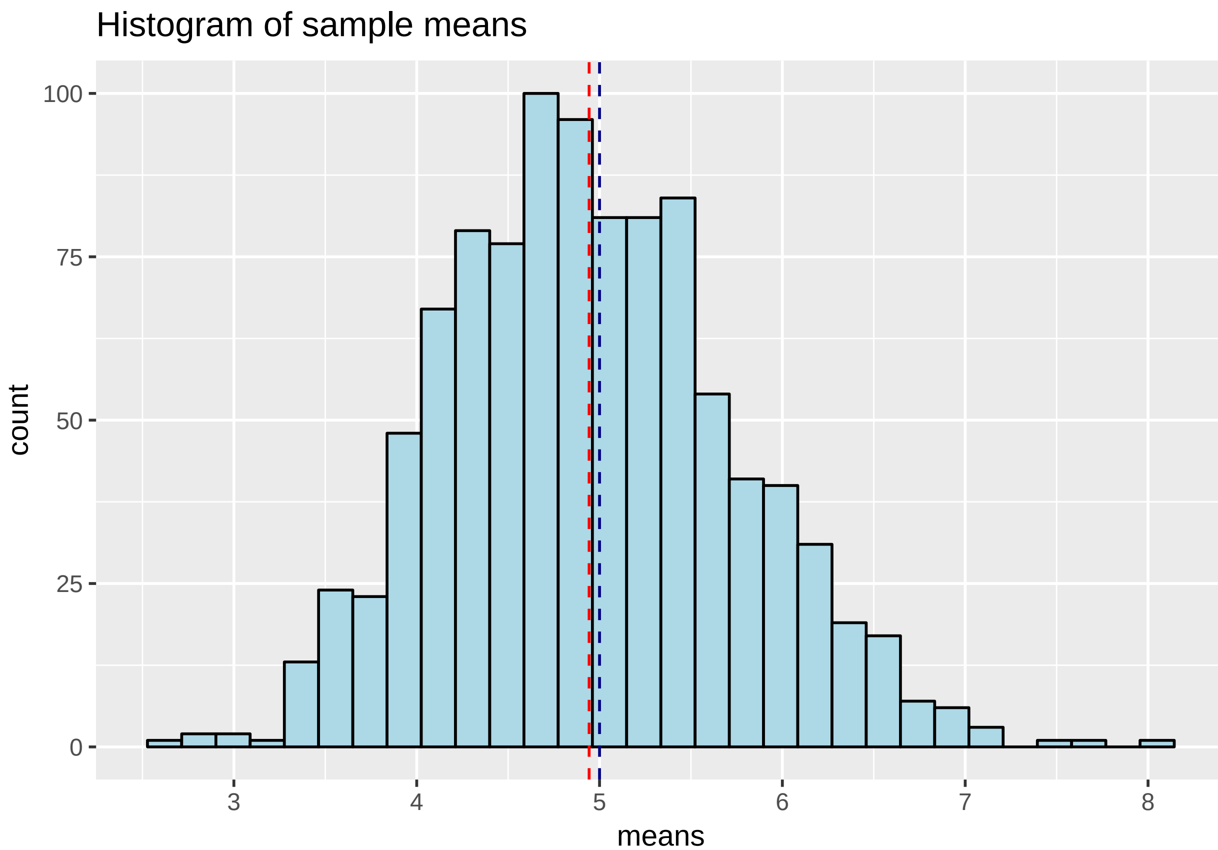
First we calculate the means of our samples using R's *apply* function, saving the results as a data frame. We then plot a histogram of these means.

The plot clearly shows the means cluster around 5, as theory predicts. The blue line indicates the theoretical mean and the read line gives the *mean of the means*. We can see the two values are very close.

```
means <- data.frame(means = apply(sim, 1, mean))

ggplot(means, aes(means)) +
  geom_histogram(colour = 'black', fill = 'lightblue') +
  scale_x_continuous(breaks = seq(1,10)) +
  geom_vline(xintercept = 5, linetype = 2, colour = 'navy') +
  geom_vline(xintercept = mean(means$means), linetype = 2, colour = 'red') +
  labs(title = 'Histogram of sample means')
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



3. Variance

Next we consider the variance of our samples. The standard deviation of the exponential distribution is also 1 divided by lamda, and the Central Limit Theorem tells us the standard deviation of our sample distribution will be 1 divided by lamda times the square root of the sample size. In our example then we should expect to find a sample standard distribution of 1 divided by 0.2 times the square root of 40, which equals 0.7905694.

Our observed standard deviation of the means is given by the calculation below:

```
sd <- sd(means$means)
print(sd)
```

```
## [1] 0.7888396
```

We see that the difference between the expected value and our observed value is small.

4. Normality

To check whether our distribution is approximately normal we compare it with a distribution of 1000 numbers from a normal distribution with the same mean and standard deviation as we would expect from our sample.

We plot our observed distribution below in blue, alongside a sample from a normal distribution in red. We see our exponential saample means closely follows the normal sample.

```
norm <- data.frame(norm = rnorm(1000, mean = 5, sd = 1/(0.2*sqrt(40))))

ggplot(means, aes(means)) +
```

```r
  geom_density(colour = 'blue') +
  geom_density(data = norm, aes(norm), colour = 'red') +
  labs(title = 'Comparison of our sample with a normal distribution')
```