Exponential distribution

2023-11-30

Synopsis

In this report I will explore the exponential distribution and compare it to the Central Limit Theorem by simulating 1000 runs of 40 exponents and comparing the means and variances of the simulations to the theoretical values.

Setup

For the simulation we are using a lambda of 0.2, we are doing 1000 simulations each consisting of 40 exponentials. To achieve this we populate a matrix where each row is a simulation of 40 columns.

```
lambda <- 0.2
simulations <- 1000
n <- 40

set.seed(1)
data <- matrix(rexp(n*simulations, lambda), nrow = simulations, ncol = n)
```

Means

The mean of an exponential distribution is given by $1/\lambda$, in this case $1/0.2 = 5$.

```
1/lambda
```

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

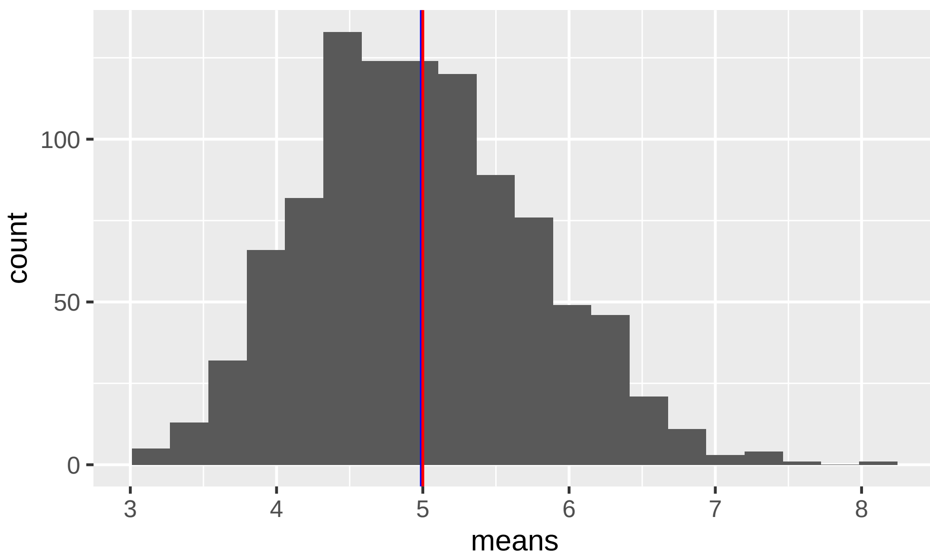
The sample mean of each simulation can be found using apply, we can then find the average of these to compare with the theoretical mean.

```
means <- apply(data, 1, mean)
mean(means)
```

```
## [1] 4.990025
```

Below is a histogram of all of the means of each simulation, the theoretical mean is shown by the red line, and the sample mean is shown in blue. We can see that despite some simulations having a mean far from the theoretical, overall the sample mean averages out to being almost identical to the theoretical mean.

```
ggplot(mapping = aes(means)) +
    geom_histogram(bins=20) +
    geom_vline(xintercept=mean(data), color="blue") +
    geom_vline(xintercept=1/lambda, color="red")
```



Variances

The variance of an exponential distribution is given by $1/\lambda^2$, in this case $1/0.2^2 = 25$.

```
1/lambda^2
```

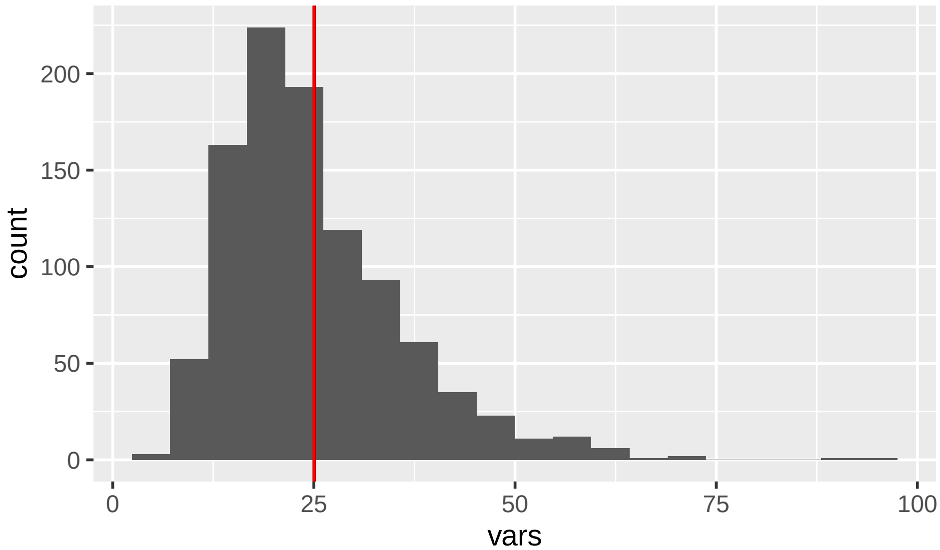
```
## [1] 25
```

We can find the variance of each simulation using apply, then we average this to find the average sample variance across simulations.

```
vars <- apply(data, 1, var)
mean(vars)
```

```
## [1] 25.05783
```

```
ggplot(mapping = aes(vars)) +
    geom_histogram(bins=20) +
    geom_vline(xintercept=mean(vars), color="blue") +
    geom_vline(xintercept=1/lambda^2, color="red")
```



Is the distribution normal

The Central Limit Theorem (CLT) states that $\bar{X}_n$ is approximately $N\left(\mu, \sigma^2/n\right)$ for large n. This suggests that the means of the simulations is a normal distribution. By plotting the distribution of all the data we can see that it is not normal, however by plotting the distribution of the means we can see that this appears to be approximately normal

```r
plot_grid(
    ggplot(mapping = aes(data)) +
        geom_histogram(),
    ggplot(mapping = aes(means)) +
        geom_histogram()
)
```