Simulaiton exercise

Josh McNamara

7/9/2019

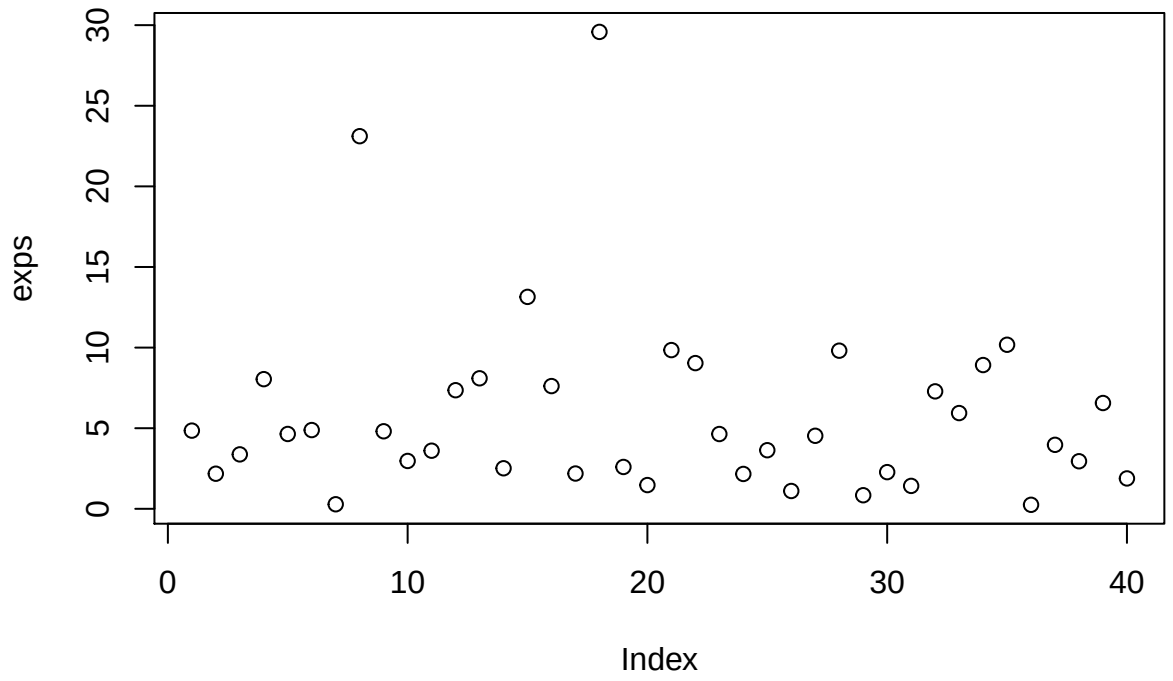
Overview

The purpose of this exercise is to explore some statistical theory and develop intuition via simulation. I will generate a list of simulations of samples from the exponential distribution and explore the mean and variance of their means, and how these compare to theory.

Simulations

One simulation result:

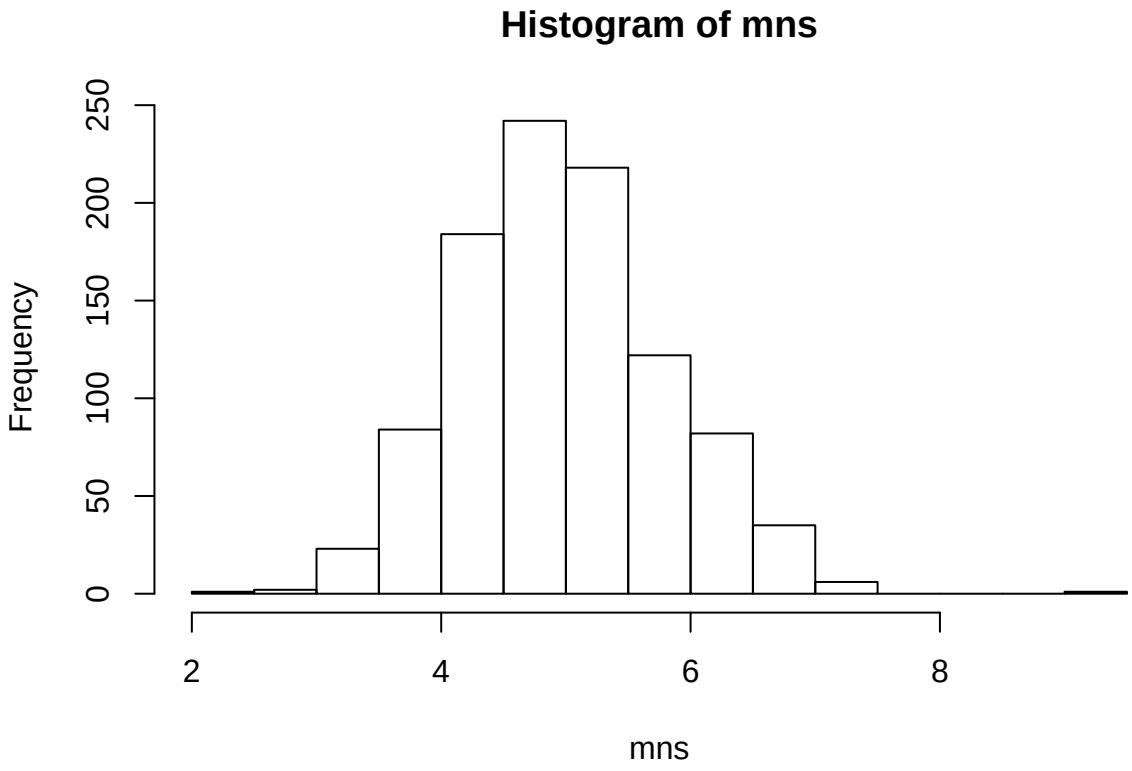
```
library(ggplot2)
library(lattice)
lambda <- 0.2
exps <- rexp(40, lambda)
plot(exps)
```



Sample Mean versus Theoretical Mean

How does the mean of the distribution means compare to the theoretical mean? The analytic value = 1/lambda. We will determine how the simulations approximate this.

```
mns = NULL
for (i in 1 : 1000) mns = c(mns, mean(rexp(40, lambda)))
hist(mns)
```




```
sim <- mean(mns)
theor <- 1/lambda
pct_err <- (sim-theor)/theor *100
```

The simulation means are very close to the theoretical mean.

Given the large sample size, it is not surprising hat the percent error between the analytic value and the simulation value is small.

Sample Variance versus Theoretical Variance

How does the mean of the variances vary between the simulation and the theoretical value? The expected value is lambda^(-2).

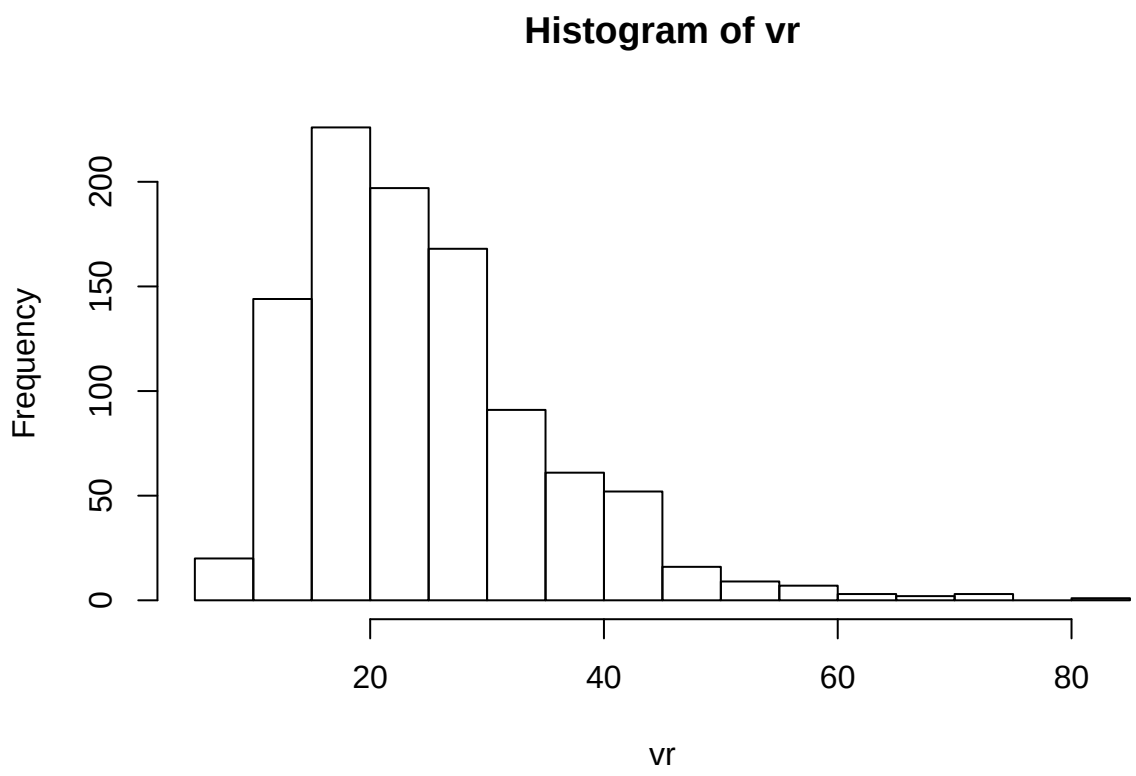
```r
# Variance
var(mns)
```

```
## [1] 0.6805593
```

```r
lambda ^(-2)
```

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

```r
vr = NULL
for (i in 1 : 1000) vr = c(vr, var(rexp(40, lambda)))
hist(vr)
```




```r
mean(vr)
```

```
## [1] 24.65526
```

We see that the simulation is close to the expected value.

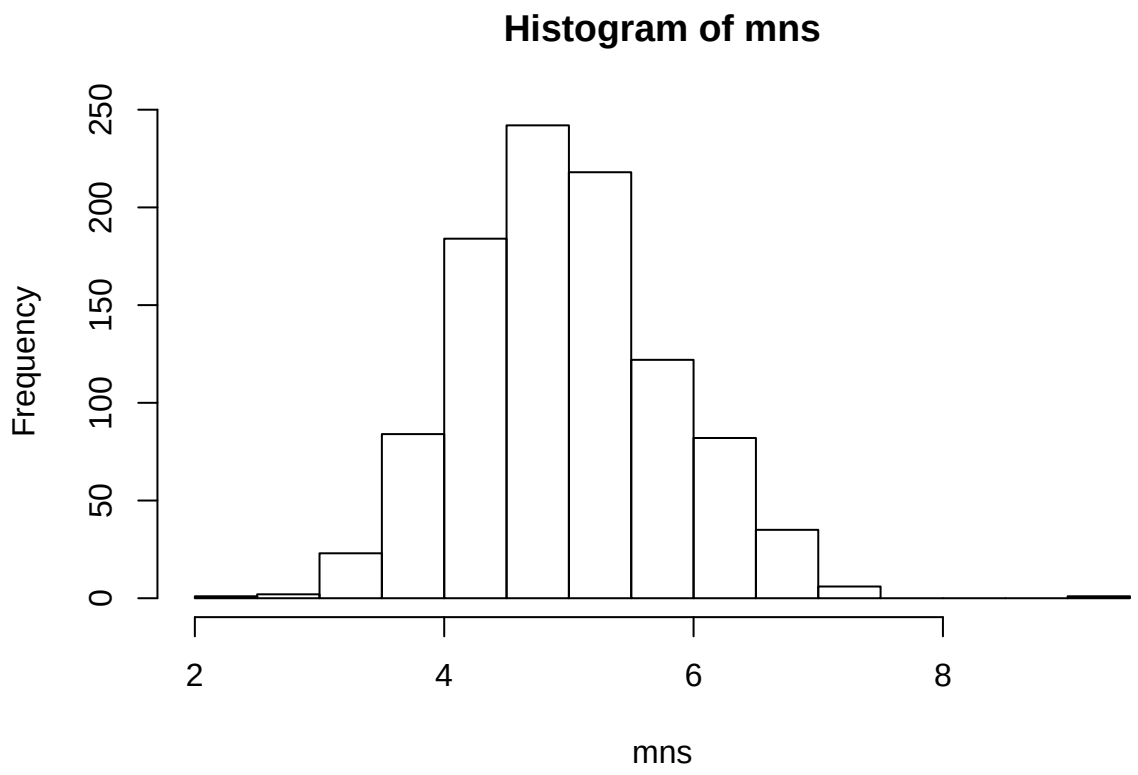
Is the distribution of means normal?

I will perform a Shapiro-Wilk test of normality.

```r
# Normality test
shapiro.test(mns)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  mns
## W = 0.99176, p-value = 2.25e-05
```

```r
hist(mns)
```




```r
mean(mns > sim) # distribution is roughly symmetrical
```

```
## [1] 0.476
```

It seems that the normality test conclusively rejected the null hypothesis that the distribution is drawn from a normally distributed sample.

Despite this, the distribution is at least roughly symmetrical, as evidenced by the histogram and the fact that roughly half of the samlpe is greater than the mean.