statistical inference project coursera

Abhishek Dubey

4/11/2018

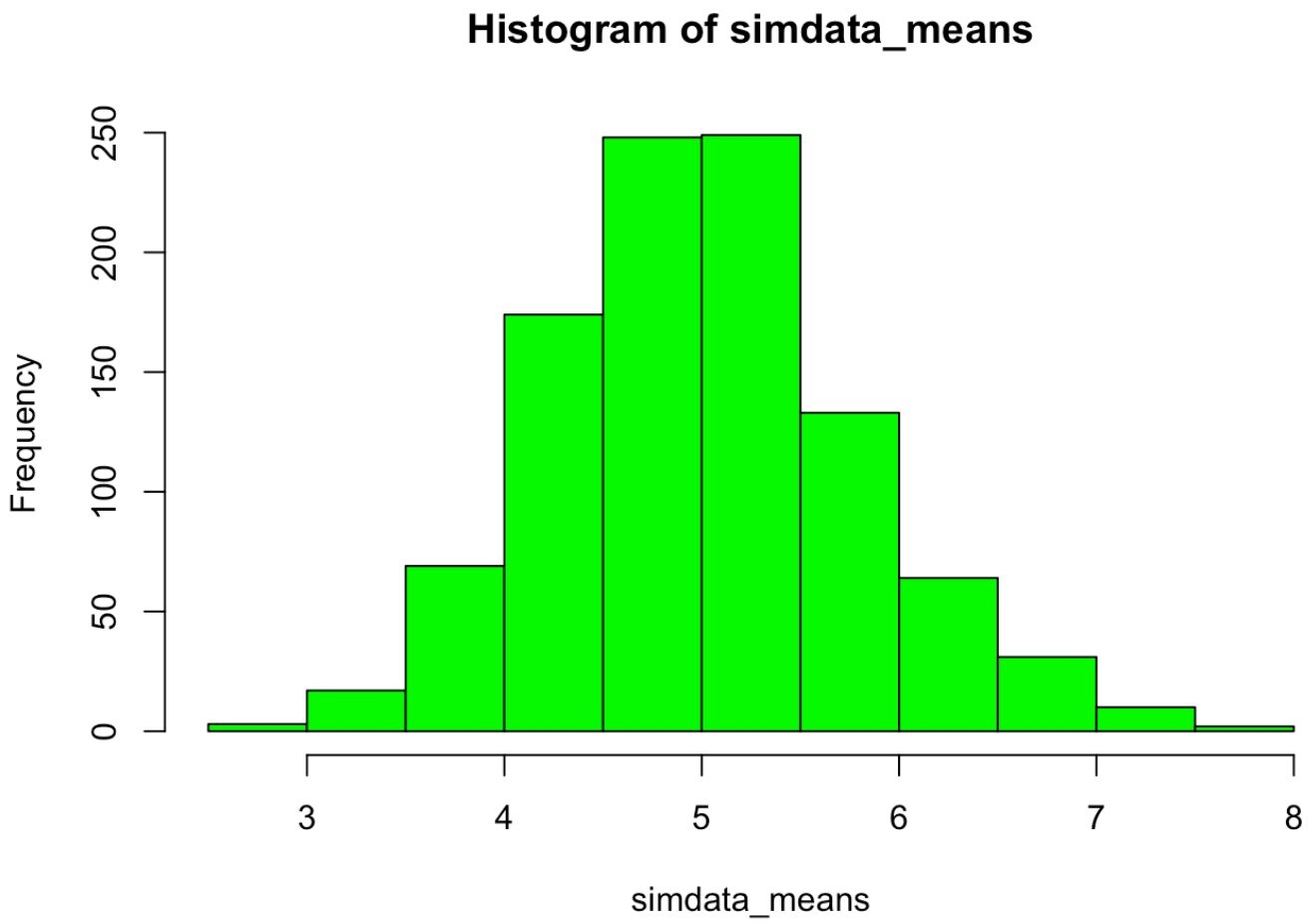
Objective

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. Investigate the distribution of averages of 40 exponentials over a thousand simulations.

```
#
set.seed(12323)
lambda <- 0.2
n <- 40
simulations <- 1000

# simulate
simdata <- matrix(rexp(simulations * n, rate=lambda), simulations)

# calculate mean of exponentials
simdata_means <- apply(simdata, 1, mean)
hist(simdata_means, col="green")
```



1. Show the sample mean and compare it to the theoretical mean of the distribution.

```
#Mean of exponential distribution is 1/lambda

mean_exp_dist <- 1/lambda
print (paste("Theoretical center of the distribution = ", mean_exp_dist))
```

```
## [1] "Theoretical center of the distribution =  5"
```

```
print (paste("Actual center of the distribution based on the simulations = ", round
(mean(simdata_means), 2)))
```

```
## [1] "Actual center of the distribution based on the simulations =  5.01"
```

2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
theoretical_variance <- (1/lambda)^2/n
print (paste("Theoretical variance = ", theoretical_variance))
```

```
## [1] "Theoretical variance =  0.625"
```

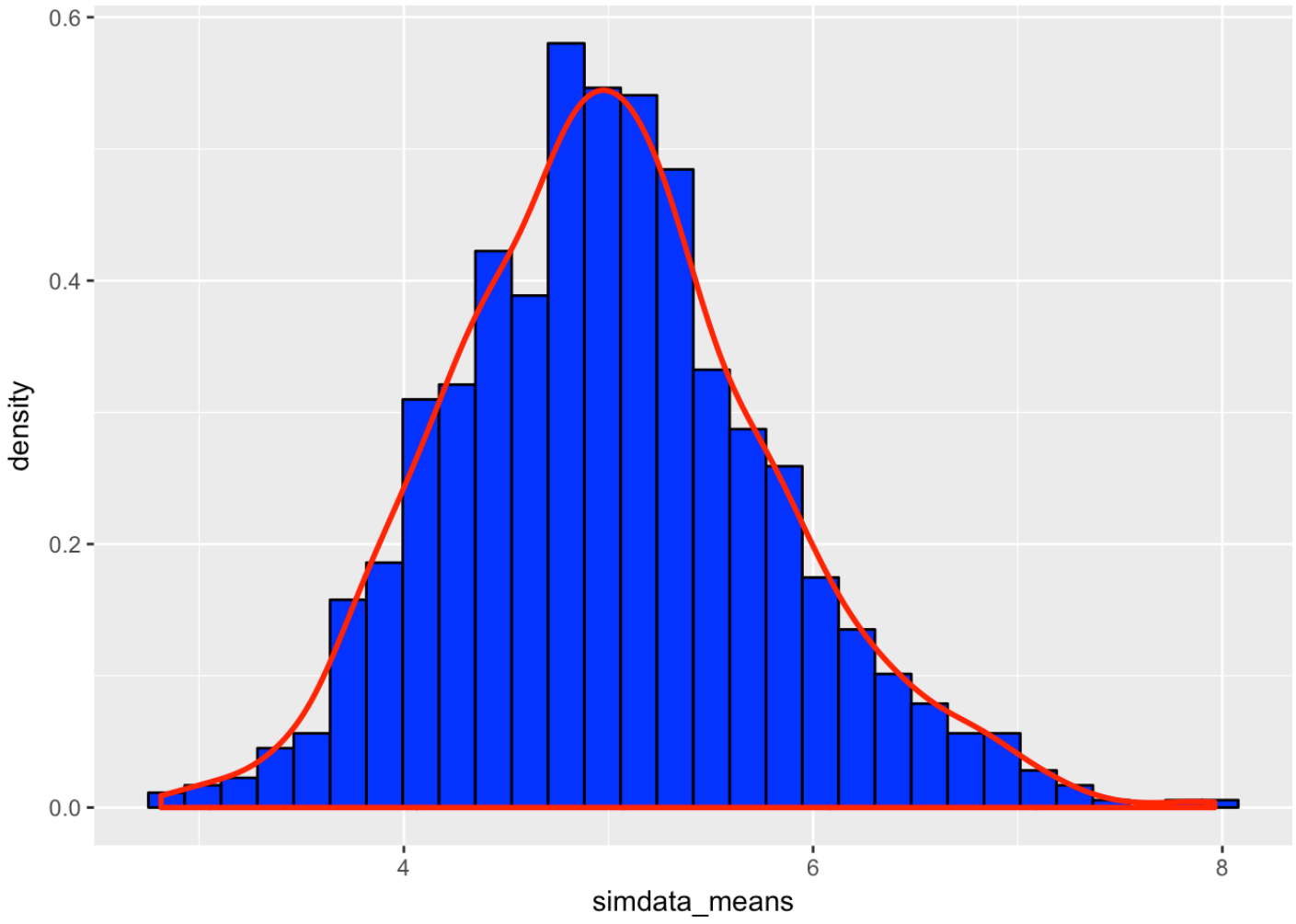
```
print (paste("Actual variance = ", round(var(simdata_means), 3)))
```

```
## [1] "Actual variance =  0.621"
```

3. Show that the distribution is approximately normal.

```
plotdata <- data.frame(simdata_means)
m <- ggplot(plotdata, aes(x = simdata_means))
m <- m + geom_histogram(aes(y=..density..), colour="black", fill = "blue")
m + geom_density(colour="red", size=1)
```

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



```
# compare the distribution of averages of 40 exponentials to a normal distribution
qqnorm(simdata_means)
qqline(simdata_means, col = 2)
```