Statistical Inference Project

Erik Rehnberg Steeb

7/23/2020

```
knitr::opts_chunk$set(echo = TRUE)
library(ggplot2)
```

Overview

Setup

The first step is to set our initial variables and set the seed so this is fully reproducible.

```
ECHO=TRUE
set.seed(1775)
lambda = 0.2
exp = 40
```

Next we run our simulation with the parameters set above.

```
sim_means = NULL
for (i in 1 : 1000) {
  sim_means = c(sim_means, mean(rexp(exp, lambda)))
}
```

Compare Sample and Theoretical Means

```


mean_sim_means <- mean(sim_means)
mean_sim_means
```

Sample Mean

```
## [1] 5.013338
```

Theoretical mean The theoretical mean of the distribution is lambda ˆ -1

```
theo_mean <- lambda^-1
theo_mean
```

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

Comparison The difference is fairly small.

```
abs(mean_sim_means - theo_mean)
```

```
## [1] 0.013338
```

Sample Variance and Theoretical Variance

Sample Variance Again, this is very simple

```
var_sim <- var(sim_means)
```

Theoretical Variance The theoretical variance follows

```
theo_var  <- (lambda * sqrt(exp))^-2
```

Compared variances Again, a very slight difference between the two

```
abs(var_sim - theo_var)
```

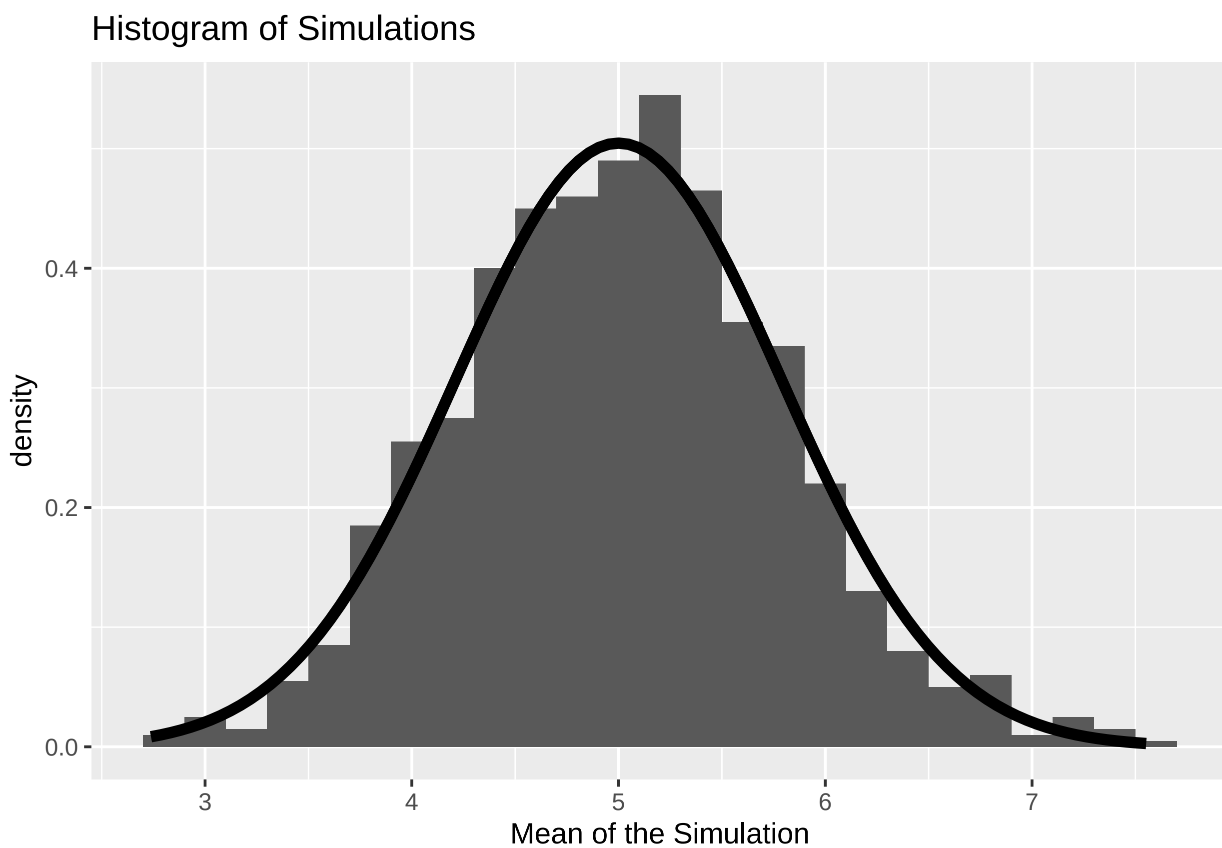
```
## [1] 0.02075072
```

Distribution

Next, I'll directly compare a density histogram of the one thousand simulations with the normal distribution, which has a mean of lambda^-1 and std deviation of (lambda*sqrt(n))^-1, which is the theoretical distribution for the simulations. They are fairly similar, with the exception of the peak of the simulation occuring just after the peak of the theoretical mean.

```
df <- data.frame(sim_means)

ggplot(df, aes(x = (unlist(df)))) +
  geom_histogram(aes(y= ..density..), binwidth = 0.2)+
  stat_function(fun = dnorm,
                args = list(mean = lambda^-1,
                            sd=(lambda*sqrt(exp))^-1),
                size = 2) +
  labs(title = "Histogram of Simulations", x = "Mean of the Simulation")
```


Histogram of Simulations
density
0.0
0.2
0.4
3
4
5
6
7
Mean of the Simulation