Statistical_Inference_Course_Project_Part_1

Jason B.

Overview

The purpose of this exercise is to demonstrate that the distribution of means of sets drawn from an exponential distribution is approximately normal with a mean of means that estimates the population mean.

```r
lambda <- 0.2
n <- 40
set.seed(20180720)

#calculate theoretical mean
e_mn <- 1/lambda

#calculate theoretical variance
e_var <- (e_mn/sqrt(n))^2
```

Simulations

This code creates a loop that calculates the mean of a set of 40 random variables drawn from an exponential distribution with lambda = 0.2. For each iteration it appends the new mean to the vector "mns".

```r
#create simulation means
mns = NULL
for(i in 1:1000) mns = c(mns, mean(rexp(n, lambda)))

#calculate sample mean = mean of means
sample_mn <- round(mean(mns), 3)

#calculate variance of sample = square of sd
sample_var <- round(sd(mns)^2, 3)
```

The sample mean is **5.017**, which is very close to the theoretical mean **5**.

The sample variance is **0.601**, which is similar to the theoretical variance **0.625**.

. . .

Distribution

This next bit of code samples 1,000 random variables from a exponential distribution with lambda = 0.2 and plots them as a histogram showing the typical decreasing exponential curved shape. I compare this to the histogram of the sample means and the theoretical mean of the distribution.

```
#plot hist of 1000 random variable of exp with lambda = 0.2
exps <- rexp(1000, lambda)
hist(exps, col = "cyan", breaks = 20, main = "Simulation of random variables and sample means \n for an

#plot means of 1000 simulations
hist(mns, add = T, col = "magenta")

#plot population mean
abline(v = e_mn, lwd = 3, lty =2)
legend("topright", legend = c("RVs", "means"), col = c("cyan", "magenta"), pch =15)
```

Simulation of random variables and sample means
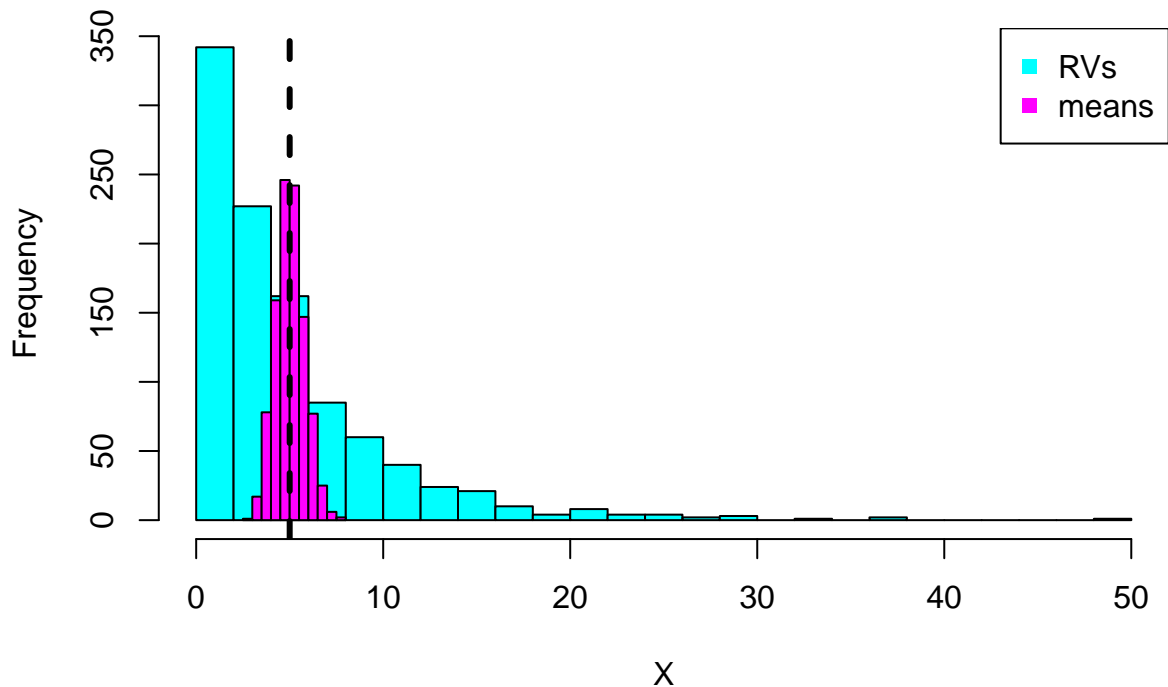## for an exponential distribution with lambda = 0.2

Figure 1. Histogram showing the simulation of 1,000 random variables (cyan) and 1,000 means (magenta) of simulated samples with n = 40 from a exponential distribution with lambda = 0.2. Note that the distribution of sample means is symmetric around the theoretical population mean (dashed line) and is approximately Gaussian.