

Pt1_SimulationExercise

Alex Fleming

1/29/2018

Exploration of the Exponential Distribution in R - Final Course Project

by Alex C. Fleming, student in the JHU Data Science Specialization

Overview

This document will walk through the analysis of an exponential distribution, and the sample means from it, and serve as a re-inforcement of the central limit theorem. We will compare statistics of the theoretical data and the sample data after 1000 simulations of draws from 40 exponential distributions (in a dataset called "mns"). There will be no variation in the lambda of the exponential simulations.

The exponential distribution is a function of x and lambda where

$$f(x) = \lambda e^{-\lambda x} \quad x \geq 0$$

The mean of this distribution is $1/\lambda$ and the variance is $1/\lambda^2$

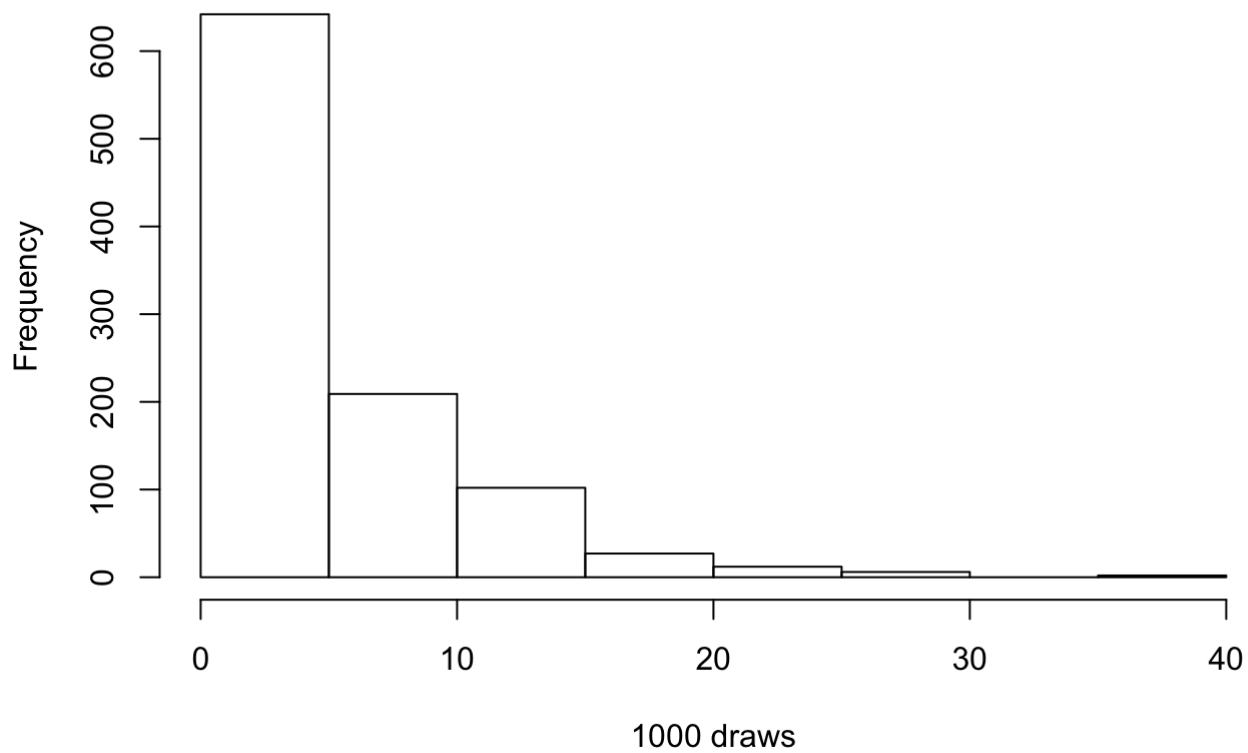
Simulations

```
knitr::opts_chunk$set(echo = TRUE)
# Basic setup including lambda, data, and the means
lambda <- .2
# Exponential data distribution for comparison
data <- rexp(1000, lambda)
# Create Means simulation
mns <- NULL
for (i in 1 : 1000){
  mns <- c(mns, mean(rexp(40, lambda)))
}
```

The graphs below illustrate the raw distribution vs. the means of repeated distributions

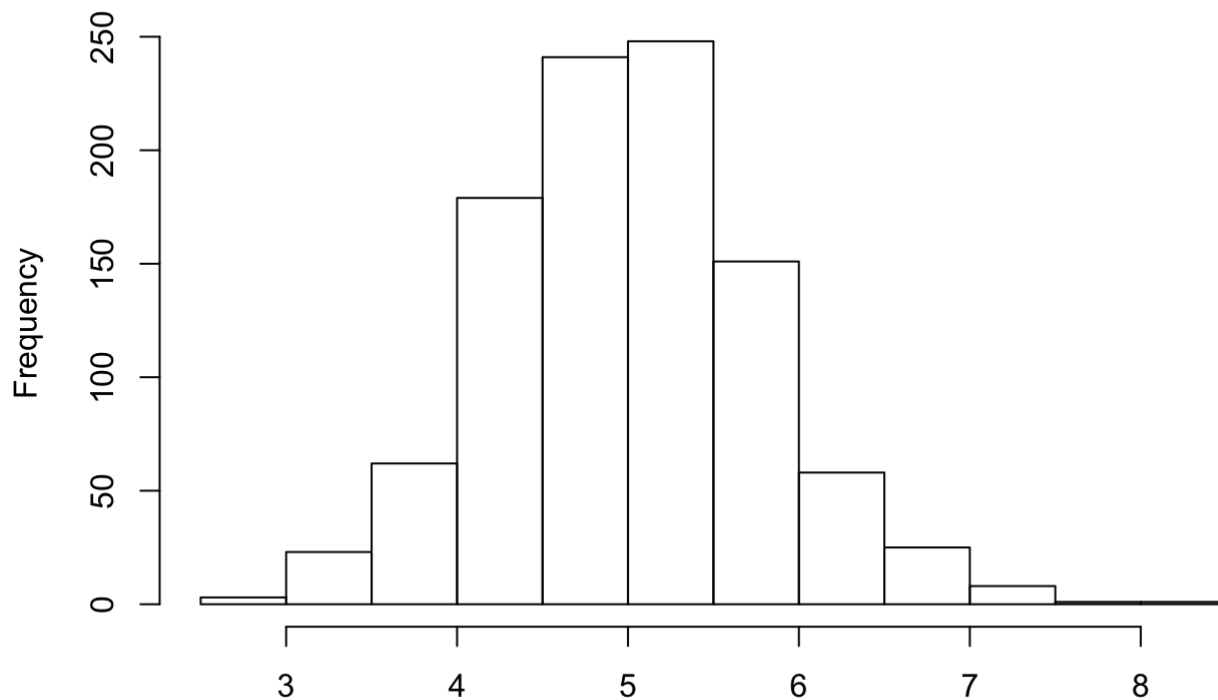
```
hist(data, main="Exponential Distribution", xlab="1000 draws")
```

Exponential Distribution



```
hist(mns, main="Means of Exponential Distributions", xlab="1000 repeats of Mean of 40 Ex  
p Dist")
```

Means of Exponential Distributions



1000 repeats of Mean of 40 Exp Dist

Sample Mean vs Theoretical Mean

The differences between the sample and theory are small, which does not surprise us with only 1000 simulations. They seem to be close enough to satisfy basic curiosity.

```
tmu <- 1/lambda
smu <- mean(mns)
diff <- smu-tmu
paste("The difference between the means is", diff, ", the theoretical mean is", tmu, ",
and the sample mean is", smu, ".")
```

```
## [1] "The difference between the means is -0.00344245030362433 , the theoretical mean
is 5 , and the sample mean is 4.99655754969638 ."
```

This is to be expected since the sample is coming from an exponential distribution.

Sample Variance vs. Theoretical Variance

The variance of the theory of the exponential distribution

```
tvar <- 1/(lambda^2)
svar <- var(mns)
diffv <- tvar-svar
paste("The difference between the variances is", diffv, ", the theoretical variance of a
n exponential distribution is", tvar, ", and the drawn sample variance is", svar)
```

```
## [1] "The difference between the variances is 24.3914276997347 , the theoretical varia
nce of an exponential distribution is 25 , and the drawn sample variance is 0.6085723002
6532"
```

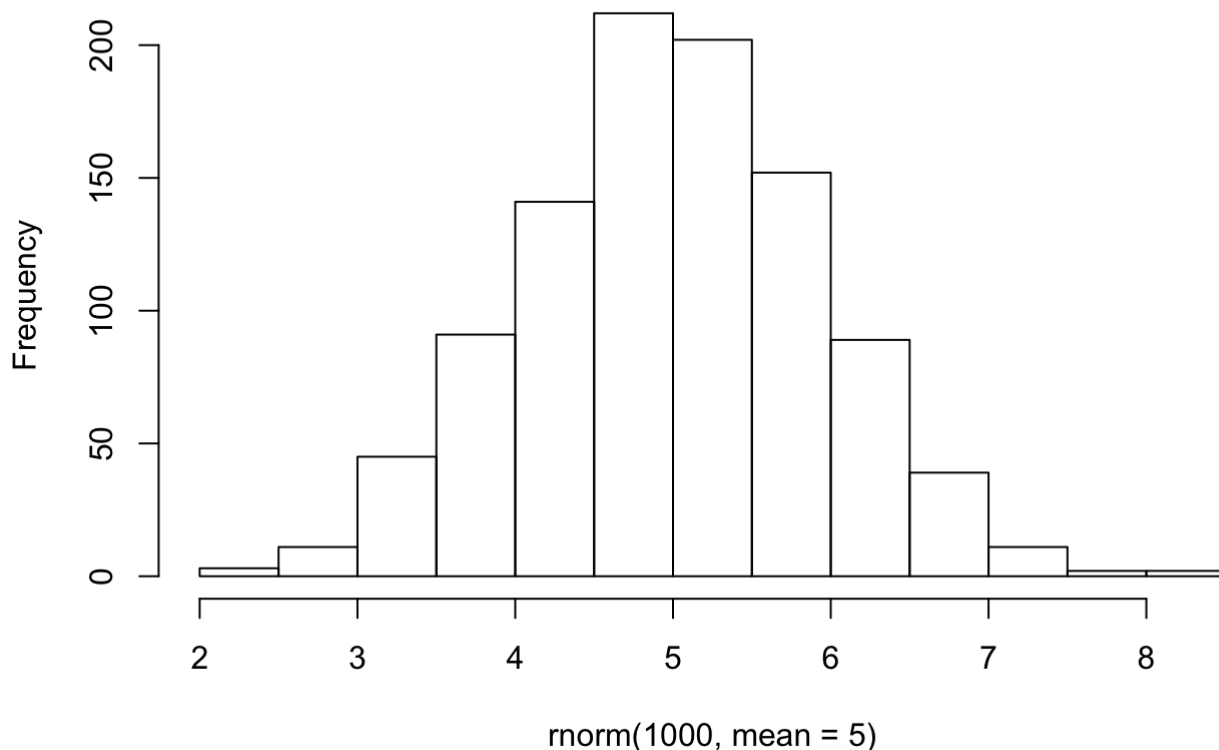
This is also not a surprising result since the central limit theorem tells us that as you have more and more independent samples from any distribution, the resulting sample is normal, not exponential.

Distribution

We would expect the distributon of the means to look more and more normal as the number of iterations increases, which we can see when we look at the plot of a normal distribution histogram with 1000 draws with mean=5 ($1/\lambda$). The variance of this distribution is not exactly our mns sample, but it is close.

```
hist(rnorm(1000, mean=5), main="Normal Distribution Draw")
```

Normal Distribution Draw



```
nvar <- var(rnorm(1000, mean=5))  
paste("Our sample variance from the mns dataset is", svar, "versus the variance of 1000  
normal draws of", nvar)
```

```
## [1] "Our sample variance from the mns dataset is 0.60857230026532 versus the variance  
of 1000 normal draws of 0.935319263062128"
```

We would expect this sample variance to approach as the number of iterations increases, but this shows that our sample is beginning to approximate a normal distribution. This reinforces what we would expect from the central limit theorem.

Contact: alex.c.fleming@gmail.com (mailto:alex.c.fleming@gmail.com)