Statistical Inference: Analyzing exponential distributions

Beth Skierski

April 2021

Overview

This report contains analysis from a simulation exercise, which investigates a sample exponential distribution by comparing it to the central limit theorem and highlighting the difference between the two.

Simulations

An exponential distribution can be modeled in R, which shows a distribution of 1000 random exponentials.

```
#set lambda to 0.2
lambda = 0.2
#simulate 1000 random exponentials
exp_sample <- rexp(1000,lambda)
```

The distribution created can be compared to the Central Limit Theorem by running 1000 simulations of a distribution of 40 exponentials.

```
#set lambda to 0.2
lambda = 0.2
#simulate 1000 averages of 40 exponentials
exps=NULL
for(i in 1:1000)exps=c(exps,mean(rexp(40,lambda)))
```

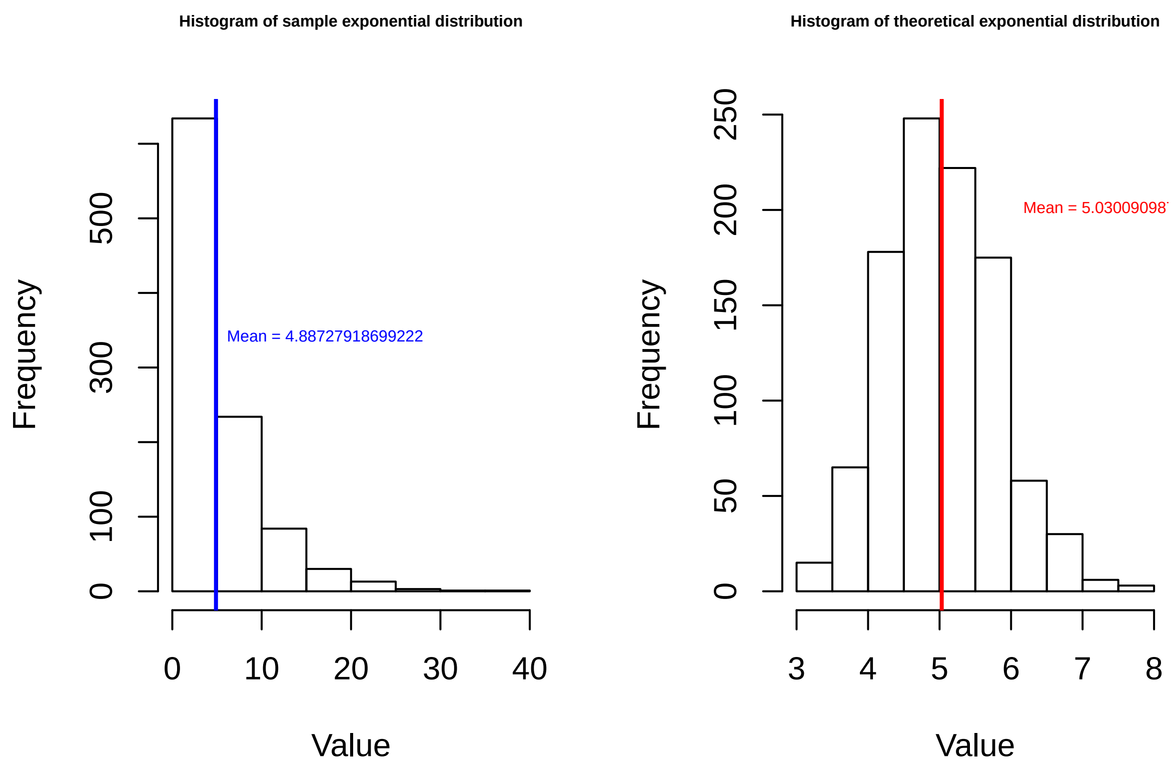
The code accomplishes the creation of two distributions. The first looks at 1000 random exponentials, where the second looks at 1000 averages of 40 exponentials. The two distributions are compared in the following section.

Sample Mean vs. Theoretical Mean

Next, we plot both distributions and highlight their means.

```
par(mfrow=c(1,2))
#plot histogram of sample distribution and highlight mean
hist(exp_sample, main = "Histogram of sample exponential distribution",
     xlab = "Value", cex.main = 0.5)
abline(v=mean(exp_sample), col = "blue", lwd = 2)
text(x = mean(exp_sample)*3.5, y = mean(exp_sample)*70,paste("Mean =",
```

```
    mean(exp_sample)),col = "blue",cex = 0.5)
#plot histogram of theoretical distribution and highlight mean
hist(exps, main = "Histogram of theoretical exponential distribution",
     xlab = "Value", cex.main = 0.5)
abline(v=mean(exps), col = "red", lwd = 2)
text(x = mean(exps)*1.5, y = mean(exps)*40,paste("Mean =",
    mean(exps)),col = "red",cex = 0.5)
```



As shown on the plots, the sample mean and theoretical mean are very close in value.

Sample Variance vs. Theoretical Variance

Next, we can look at the variances of both the sample population and the theoretical population.

```
#show variance of sample
var_sample <- var(exp_sample)
var_sample
```

```
## [1] 23.79303
```

```
#show theoretical variance
var_theo <- var(exps)
var_theo
```

```
## [1] 0.6032297
```

As calculated in R, the theoretical variance is much smaller than the sample variance. This makes sense, because the theoretical variance calculation divides the sample variance by the sample size, which in this case is 40.

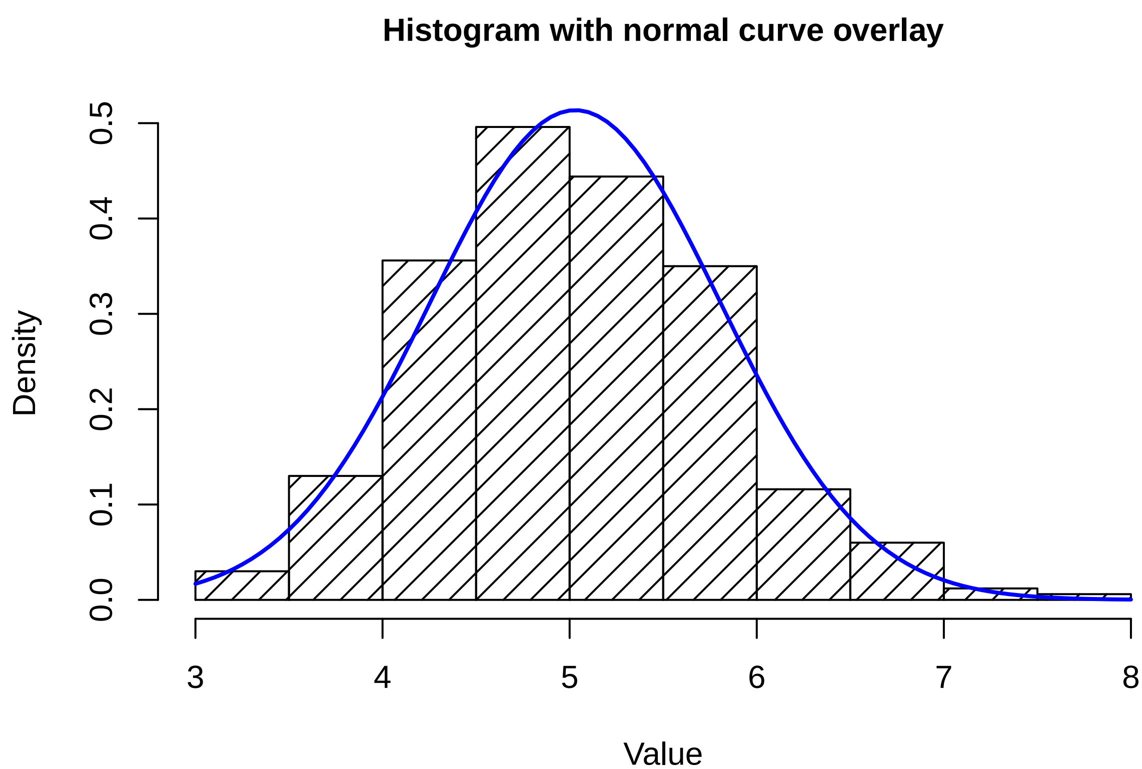
Sample variance calculation: 1/(lambda^2) = 1/(.2^2) = 25

Theoretical variance calculation: 1/(lambda^2)/n = 1/25/40 = 0.625

Distribution Conclusions

As shown in the plots above, one can see that the large collection of averages of 40 exponentials is approximately normal, while the large collection of exponentials is not. To prove this, we can look at the density of the large collection of averages with a normal curve overlay.

```r
#plot histogram of theoretical distribution with normal curve
mn <- mean(exps)
std<-sqrt(var(exps))
hist(exps, density = 10, breaks = 10, prob = TRUE,
     main = "Histogram with normal curve overlay",
     xlab = "Value", cex.main = 1)
curve(dnorm(x, mean=mn, sd=std), col="blue", lwd=2, add=TRUE, yaxt="n")
```



This confirms that the large collection of averages of 40 exponentials is indeed a normal distribution.