Statistics of the Exponential Distribution

Overview

This document briefly explores the statistical properties of the exponential distribution. Particularly, I will conduct 1000 simulations of 40 random variables drawn from an exponential distribution, calculate the sample means and sample variances, and compare the theoretically expected values. Finally, I will highlight properties of the distributions of the sample statistics. All code for this document is collected in the appendix at the end to allow clarity of discussion in the document.

Simulation and Results

First I generate the simulation data of $B = 1000$ samples of size $N = 40$ from an exponential distrubtion and store it in a 1000x40 matrix. Below I visualize the exponential distribution and one sample realization shown in Fig. 1.

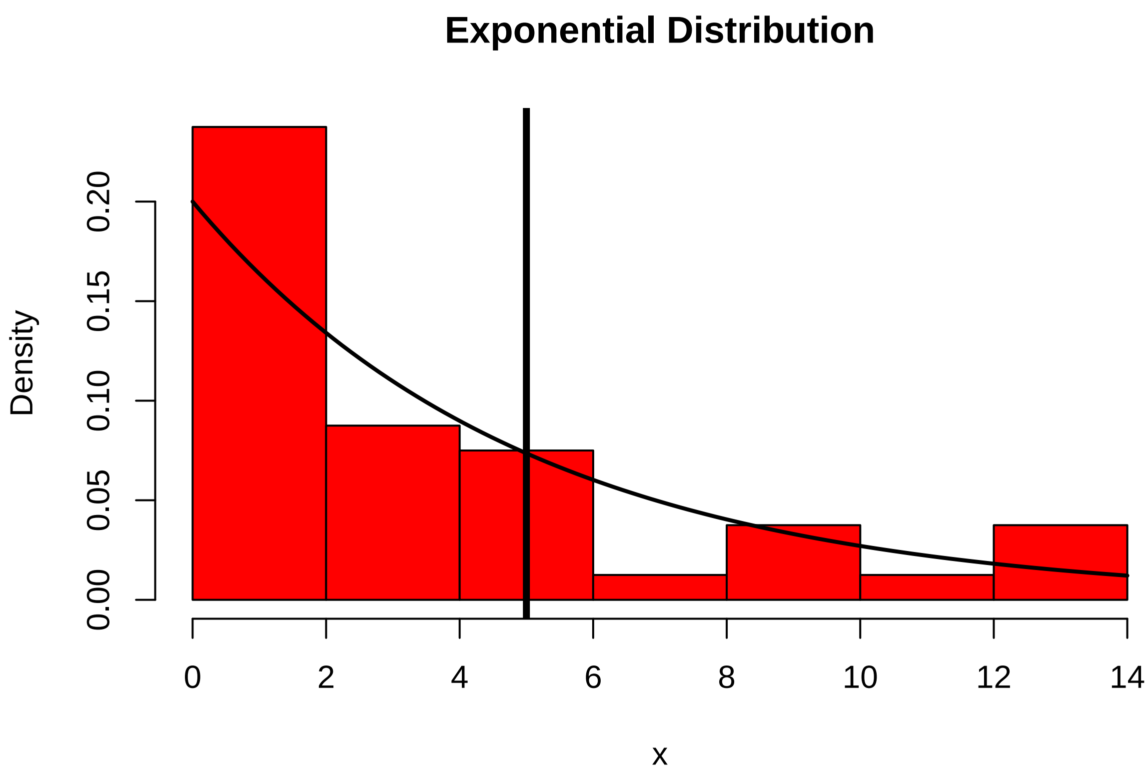

Figure 1: Probability density function of exponetial distribution (black curve). The theoretical mean $\mu = 1/\lambda$ (here $\mu = 5$) is shown as black vertical line. Single realization of a sample of size $N = 40$.

Sample Mean versus Theoretical Mean

The sample mean should be an unbiased random variable. By the central limit theorem, we expect that if the sample length N is large, and we take many samples (here $B = 1000$ samples of length $N = 40$), the distribution of the means obtained by taking the mean of each sample should be distributed normally. This distribution should be centered at the population mean $\mu = 1/\lambda$ (here $\mu = 5$) with a standard deviation of the distribution of the means of $\sigma = 1/(\lambda\sqrt{N})$ (here $\sigma = 0.62$).

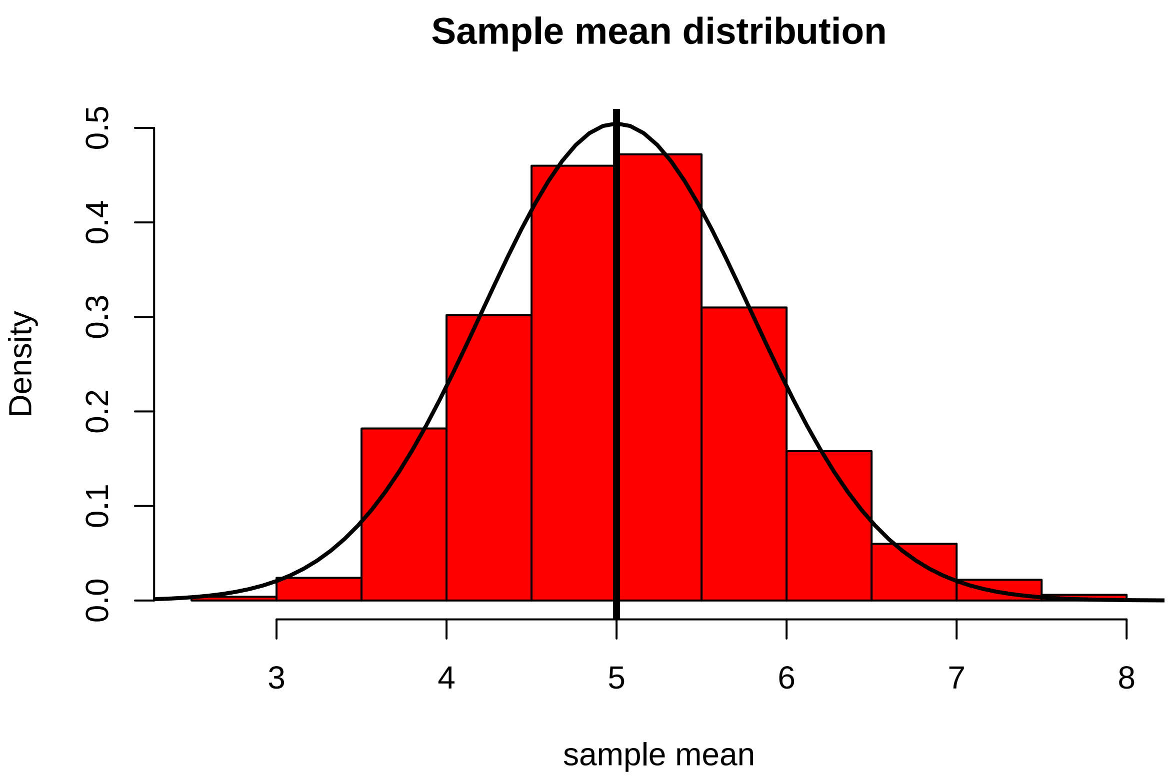

Figure 2: Histagram showing probability distribution of sample means for samples of length N=40 from the exponential distribution with $\lambda = 0.2$. The theoretical mean= $1/\lambda = 5$ is shown as the thick vertical black line. The theoretical normal distribution for the mean population distribution is shown in black (see text for details).

Note that there is a 4.67% difference in the theoretical variance of the sample means given by $1/(\lambda^2 N)$ and the realized variance of the sample means

Sample Variance versus Theoretical Variance

The sample variance is another random variable. However, the central limit theorem cannot be applied to general population distributions to extract the theoretical distribtuion of the sample variance (see e.g. https://ocw.mit.edu/courses/sloan-school-of-management/15-075j-statistical-thinking-and-data-analysis-fall-2011/lecture-notes/MIT15_075JF11_chpt05.pdf). Yet, one may hope that the distribution of the sample variance is centered the population variance which here is $1/\lambda^2$. Fig. 3 demonstrates that this is indeed apparently the case.

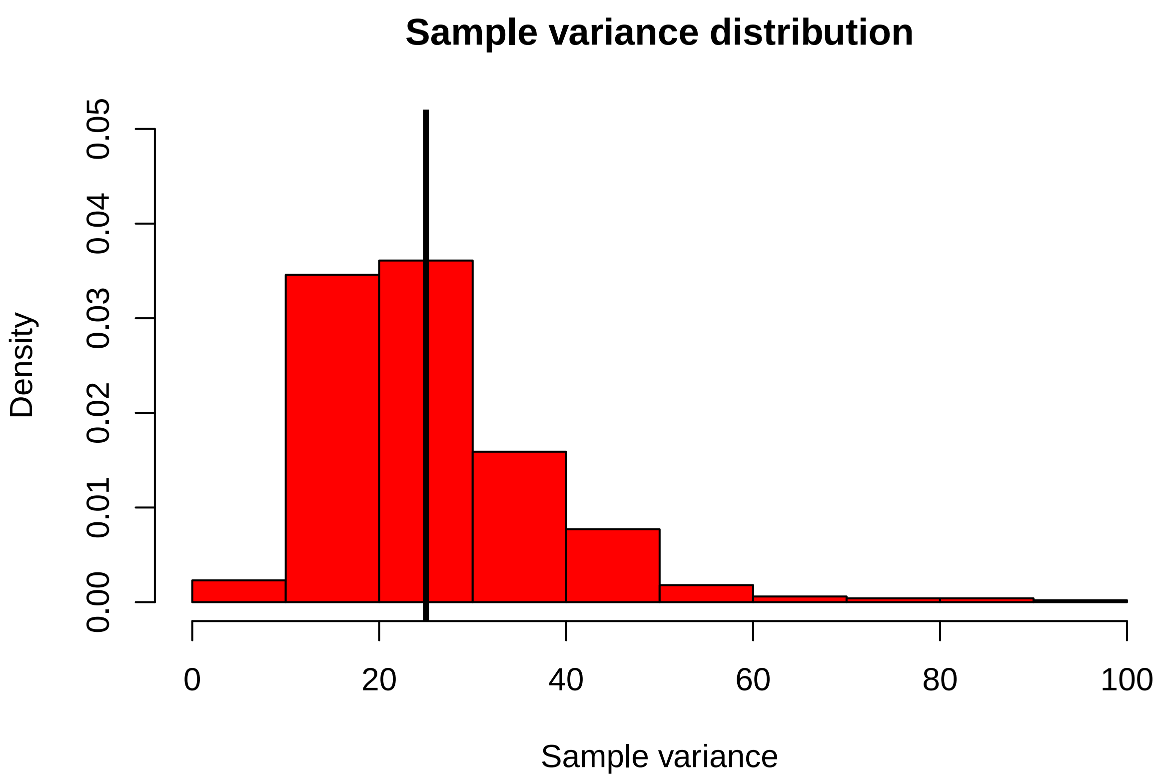

Figure 3: Histagram showing probability distribution of sample variance for samples of length N=40 from the exponential distribution with $\lambda = 0.2$. The theoretical variance= $1/\lambda^2 = 25$ is shown as the thick vertical black line.

Distribution Features

In summary, the distribution of sample means is approximately normal (see above for detailed discussion) and centered on the population mean as shown in Fig. 2. This distribution is very different from the exponential distribution shown in Fig. 1. Finally, while the sample variance distribution cannot be extracted from a central limit theorem, the distribution appears to be centered on the theoretical value of the sample variance given by the population distribution.

Appendix

```
knitr::opts_chunk$set(echo = FALSE)

##Perform Simulation and plot results and theoretical distribution
#Set seed for reproducibility
set.seed(10)
lambda<-0.2
samplen<-40
numsamp<-1000
samp<-matrix(data=rexp(samplen*numsamp,rate=lambda),numsamp,samplen)
#calculate theoretical distribution
thymean<-1/lambda
thymeanvar<-(1/lambda)^2*1/samplen
hist(samp[1,],prob=TRUE,col=2,xlab="x",main="Exponential Distribution")
curve(dexp(x,rate=lambda),from=0,to=14,lwd=2,add=TRUE)
abline(v=thymean,lwd=3.5)



##Calculate means and compare to theory.
means<-apply(samp,1,mean)

thymean<-1/lambda
sampmeanvar<-var(means)
thymeanvar<-(1/lambda)^2*1/samplen
meanvarerr<-abs(sampmeanvar-thymeanvar)/thymeanvar*100

#calculate theoretical distribution
hist(means,prob=TRUE,ylim=c(0,0.5),col=2,xlab="sample mean", main="Sample mean distribution")
curve(dnorm(x,mean=thymean,sd=sqrt(thymeanvar)),from=1,to=9,lwd=2,add=TRUE)
abline(v=thymean,lwd=3.5)



##Calculate sample variance and compare to theory.
vars<-apply(samp,1,var)
thyvar=1/lambda^2

hist(vars,col=2,prob=TRUE,xlab="Sample variance",ylim=c(0,.05),main=
             "Sample variance distribution")
abline(v=(1/lambda)^2,lwd=3)
```