

Statistical Inference Course Project Part 1

Sarah Fogel

Thursday, December 18, 2014

Instructions: The exponential distribution can be simulated in R with `rexp(n, lambda)` where `lambda` is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. Set $\lambda = 0.2$ for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

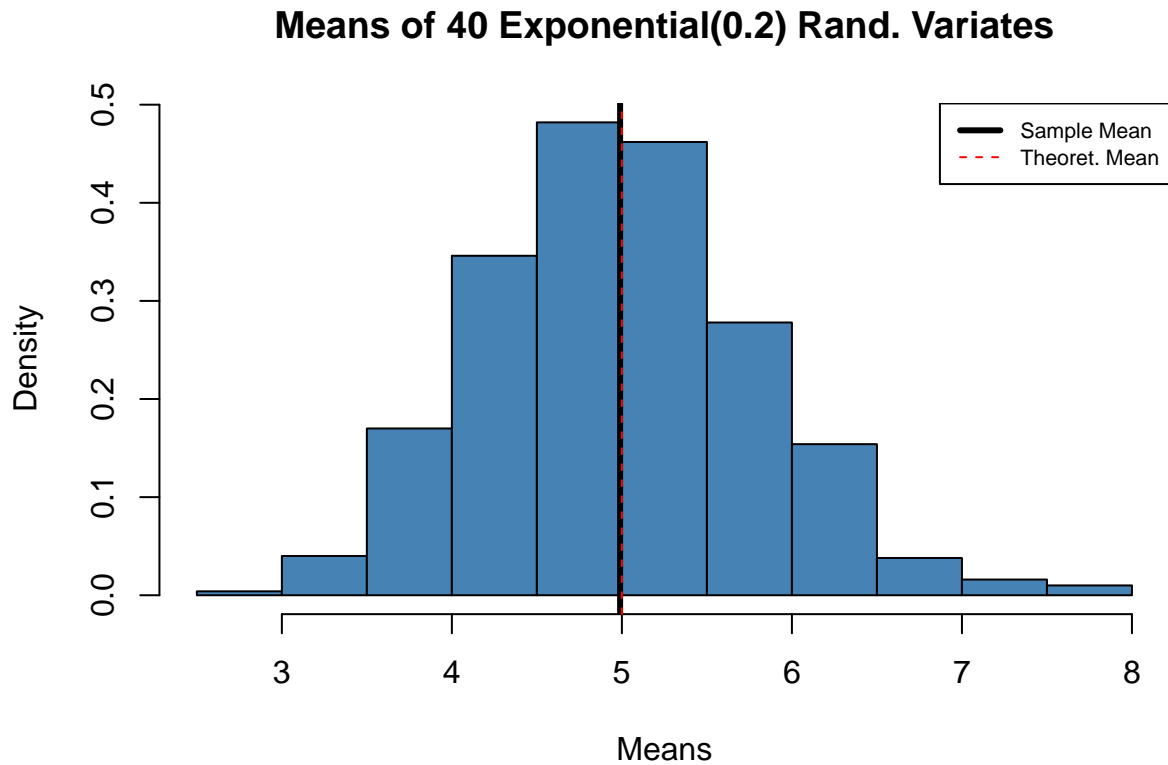
The following code was used to simulate the means of 1000 samples of 40 random exponential variates.

```
set.seed(59)      #Make this reproducible by setting the starting point
N<-1000           #Set number of simulations
means<-c()        #Initialize a vector to hold the means of the samples of 40 variates
for (i in 1:N) {
  p<-rexp(40, rate=.2)
  means[i]<-mean(p)
}
```

Question 1: Show where the distribution is centered at and compare it to the theoretical center of the distribution.

The theoretical center of the distribution of sample means is the same as the center of the distribution that the sample comes from, so it's also $1/\lambda$. In this case that is 5.

The plot below shows a histogram of the means calculated above, together with the mean (center) of the sample distribution as well as the theoretical mean of the sample distribution. (The code for all plots is included at the end)

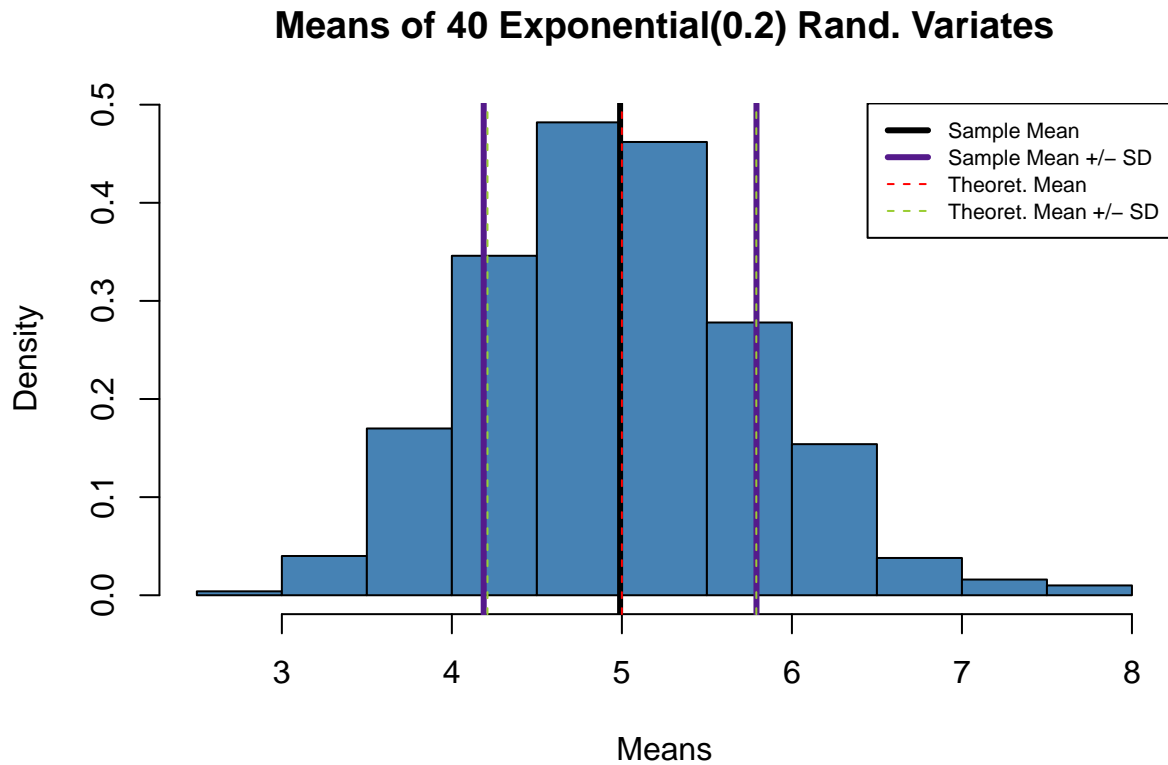


These values are, by visual inspection, quite closely related

Question 2: Show how variable it is and compare it to the theoretical variance of the distribution.

The theoretical variance of the distribution of sample means is σ^2/n , where σ^2 is the variance of the distribution where the sample came from and n is the number of observations in each sample. In this case that is $1/(n\lambda^2)$. Therefore, the theoretical standard deviation of the distribution of sample means is $1/(\sqrt{40} \cdot 0.2)$

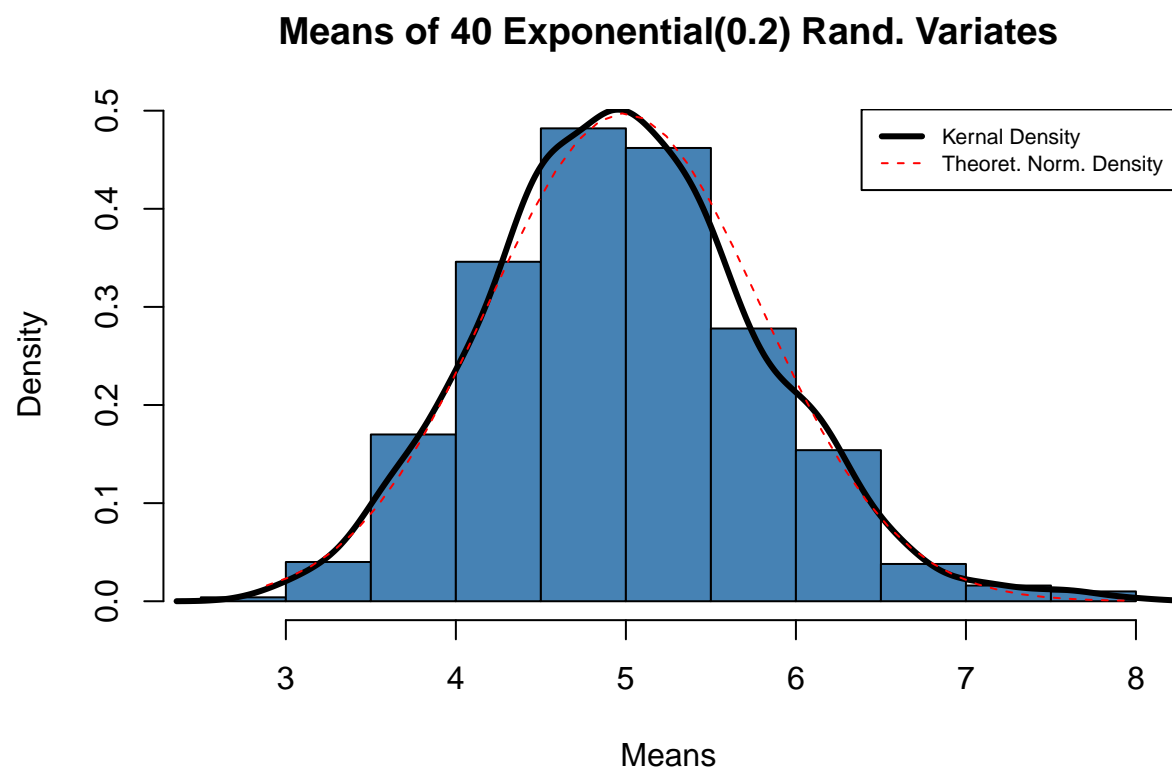
The plot below shows a comparison between the actual and theoretical values of the mean \pm one standard deviation



These values are, by visual inspection, quite closely related

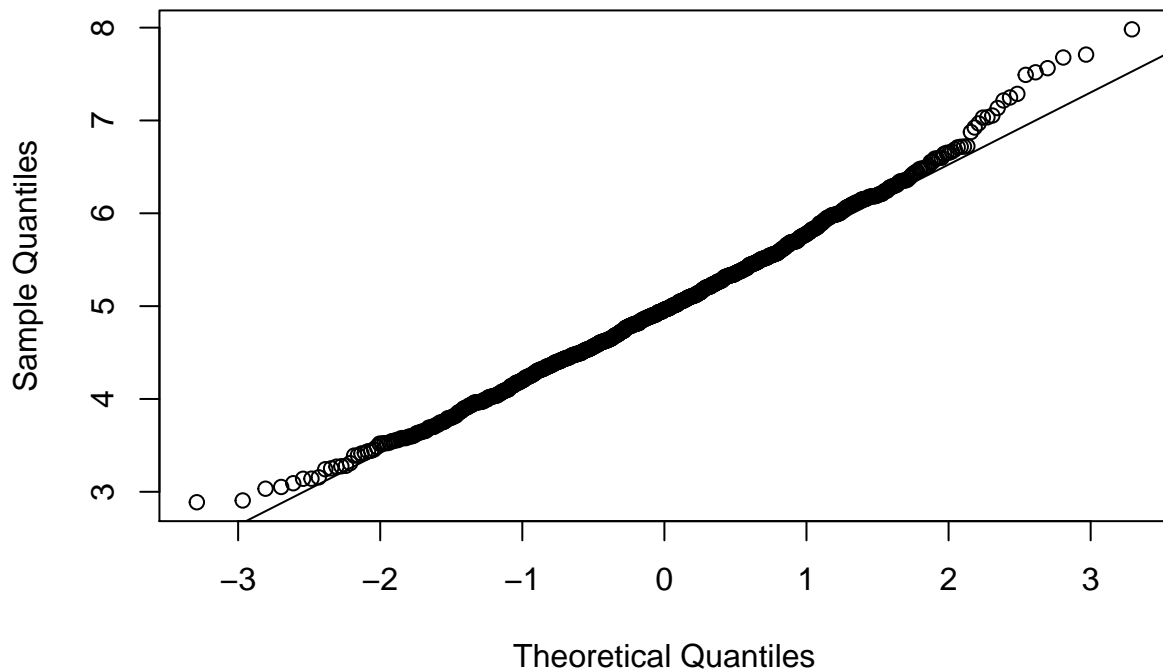
Question 3: Show that the distribution is approximately normal.

I will use two comparisons to show that the distribution is approximately normal. First, a histogram of the mean values overlaid with both the Kernel Density Estimate and a Normal Density with mean and standard deviation equal to the mean and standard deviation of the distribution, shown below.



The second comparison is using a normal q-q plot. This graphs the quantiles of a normal distribution against the quantiles of the distribution of means.

Normal Q-Q Plot



By visual inspection, both of these plots indicate a great deal of similarity between the normal distribution and the distribution of means

Appendix

Code for Plot 1

```
theoretical.mean<-5

hist(means, col="steelblue", main="Means of 40 Exponential(0.2) Rand. Variates", prob=T,
     xlab="Means")
abline(v=mean(means), lty=1, lwd=3)
abline(v=theoretical.mean, lty=2, col="red")
legend(legend=c("Sample Mean", "Theoret. Mean"), x="topright", lty=c(1,2),
      lwd=c(3,1), col=c("black", "red"), cex=.7)
```

Code for Plot 2

```
theoretical.sd<-1/(sqrt(40)*.2)

hist(means, col="steelblue", main="Means of 40 Exponential(0.2) Rand. Variates", prob=T,
     xlab="Means")
abline(v=mean(means), lty=1, lwd=3)
abline(v=theoretical.mean, lty=2, col="red")
abline(v=mean(means)+sd(means), lty=1, lwd=3, col="purple4")
```

```

abline(v=mean(means)-sd(means), lty=1, lwd=3, col="purple4")
abline(v=theoretical.mean+theoretical.sd, lty=2, col="yellowgreen")
abline(v=theoretical.mean-theoretical.sd, lty=2, col="yellowgreen")

legend(legend=c("Sample Mean", "Sample Mean +/- SD" ,"Theoret. Mean",
               "Theoret. Mean +/- SD"),
       x="topright", lty=c(1,1,2,2), lwd=c(3,3,1,1), col=c("black","purple4", "red", "yellowgreen"))

```

Code for Plot 3

```

#Create normal density
xfit<-seq(min(means), max(means), length = 40)
yfit<-dnorm(xfit, mean=mean(means), sd=sd(means))

hist(means, col="steelblue", main="Means of 40 Exponential(0.2) Rand. Variates", prob=T,
     xlab="Means")
lines(density(means), lwd=3)
lines(xfit, yfit, lty=2, col="red")
legend(legend=c("Kernal Density", "Theoret. Norm. Density"),
       x="topright", lty=c(1,2), lwd=c(3,1), col=c("black", "red"), cex=.7)

```

Code for Plot 4

```

qqnorm(means)
qqline(means)

```