Statistical Inference

Samuel Shaw

9/11/2017

## Part I; Simulation

```r
set.seed(301)

lambda<-0.2; # given parameters
n<-40

Ksims<-NULL; # using code from the example provided
for (i in 1 : 1000) Ksims = c(Ksims, mean(rexp(n, lambda)))
```

Now let's look at the data. In the first graph, I show a single sample to demonstrate the skew of an exponential distribution. The point here is that the data are not normally distributed. However, in the second graph, which shows the distribution of the means of 1000 simulated samples, the data appear normally distributed.

```r
par(mfrow=c(1,2), mar=c(4,4,2,1))
singlesample<-rexp(n, lambda)
hist(singlesample, prob=TRUE, breaks=20, main="Single Sample", col="lightblue", xlab="40 exponential ra
lines(density(singlesample), col="magenta", lwd=3, lty=1)

hist(Ksims, col="lightblue", breaks=20, main="Sample Distribution", xlab="Means")
```
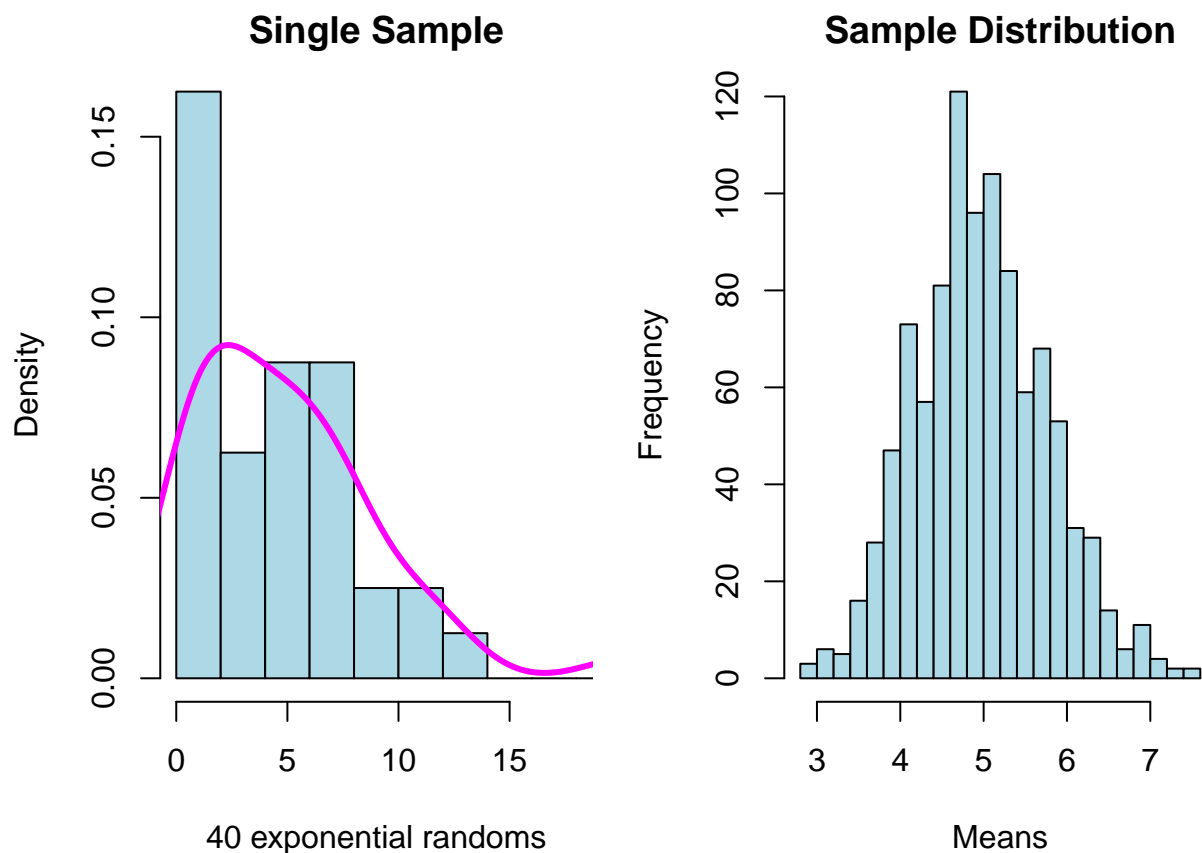
Question 1. Show the sample mean and compare it to the theoretical mean of the distribution.

The theoretical sample mean is 1/lambda. Lambda was given at 0.2, so the theoretical sample mean is 1/.2 = 5.

```
mean(Ksims)
```

```
## [1] 4.970379
```

Indeed, the sample mean is approximately the theoretical mean. The central limit theorem suggests that when n is large enough, the mean of the sample mean should approach the true mean.

```
par(mfrow=c(1,1))
hist(Ksims, col="chartreuse", breaks=20, main="Sample Distribution with Mean", xlab="Means", ylab="frequ
abline(v=mean(Ksims), lwd=3, col="magenta")
```

Sample Distribution with Mean
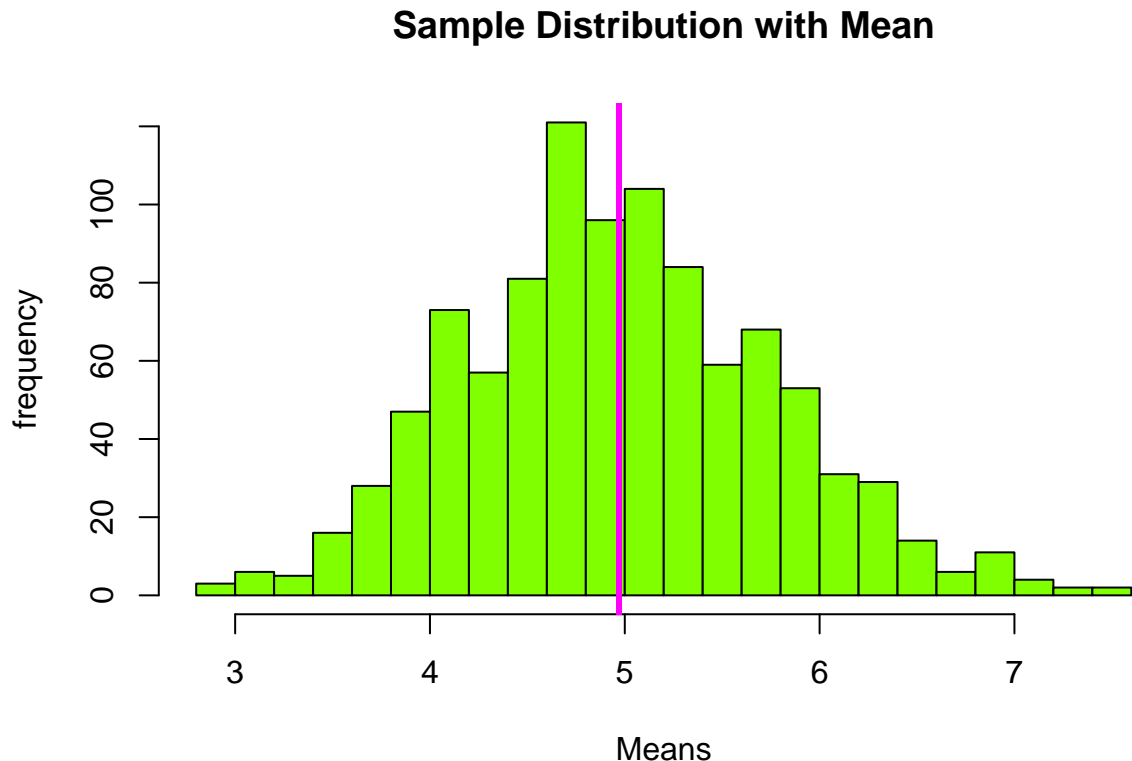
Here, the pink verticle line represents the sample mean; we see that it is approximately 5.

Question 2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

The distribution of means varies widely, but are centered at the sample mean.

range(Ksims) # The theoretical variance of any given random sample is (1/lambda)^2 = 25; the theoretical variance of the sample mean (standard error squared) is ((1/lambda)/sqrt(n))^2

Lets compare this to our simulation

var(Ksims)

We see that these are *vary* similar. Like with the theoretical and observed mean, the central limit theorem would suggest that when n is large enough the observed variance should approach the true variance.

Question 3. Show that the distribution is approximately normal

Finally, the central limit theorem suggests that the sample mean is normally distributed. We can show the density curve of our simulated means vs. a theoretical normal

hist(Ksims, prob=TRUE, col="chartreuse", breaks=20, main="Sample Distribution with Density Curves", xlab="Means", ylab="Frequency")

First With our simulated density curve

lines(density(Ksims), col="magenta", lwd=3, lty=1)

And with a theoretically normal density curve (see http://www. statmethods.net/graphs/density.html)

xfit<-seq(min(Ksims),max(Ksims),length=n^2)   yfit<-dnorm(xfit,   mean=mean(Ksims),sd=sd(Ksims)) lines(xfit, yfit, col="black", lty=2, lwd=2)

It appears our sample distribution is approximately normal

Part 2

library(datasets) data(ToothGrowth)

str(ToothGrowth) summary(ToothGrowth$len); # We'll be interested to test variation in tooth length by type of supplement and by dose

Descriptives

with(ToothGrowth, tapply(len, supp, mean)) with(ToothGrowth, tapply(len, dose, mean))

3-way Crosstab

xtabs(len/10~supp+dose, data=ToothGrowth) # taking advantage of the fact that there are exactly 10 observations in each supp+dose category (i.e, to get the means)

with(ToothGrowth, t.test(len[supp=="OJ"], len[supp=="VC"])) with(ToothGrowth, t.test(len[dose==.5], len[dose==1])) with(ToothGrowth, t.test(len[dose==1], len[dose==2]))

OJ<-ToothGrowth[supp=="OJ",] with(OJ, t.test(len[dose==.5], len[dose==1])) with(OJ, t.test(len[dose==1], len[dose==2])) VC<-ToothGrowth[supp=="VC",] with(VC, t.test(len[dose==.5], len[dose==1])) with(VC, t.test(len[dose==1], len[dose==2]))

low<-ToothGrowth[dose==.5,] with(low, t.test(len[supp=="OJ"], len[supp=="VC"])) mid<-ToothGrowth[dose==1,] with(mid, t.test(len[supp=="OJ"], len[supp=="VC"])) high<-ToothGrowth[dose==2,] with(high, t.test(len[supp=="OJ"], len[supp=="VC"]))