

Homework 4 R Exercise

The text in numerous Chapter 6 problems asks you to sketch a curve when applying the central limit theorem, or asks you to plot the probability histogram of a sampling distribution based on small samples (usually $n=2$) from a discrete random variable. Let us quickly review how we could handle either problem.

E.g., for 6.30, you are asked to describe the shape of the sampling distribution for a sample mean \bar{X} based on a sample of size $n=64$ from a population with mean $\mu = 20$ and $\sigma = 16$. By the Central Limit Theorem, \bar{X} will be approximately normal with mean $\mu = 20$ and $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 16/\sqrt{64} = 2$. Using the empirical rule to select our values for the x-axis, the following commands sketch the curve. We add some options to plot a line, rather than simply points, and print better labels.

```
X_Values=seq(14,26,by=0.1)
plot(X_Values,dnorm(X_Values,mean=20,sd=2),type="l",xlab="x",ylab="p(x)")
```

We reviewed a simple probability distribution in class, and then studied the probability distribution for the mean based on samples of size 2. Code for probability bar graphs for both distributions appears below; would you say that the central limit theorem effect has kicked in for $n=2$?

```
X_Values=c(0,2,8)
pX=c(0.25,0.5,0.25)
Xbar_Values=c(0,1,2,4,5,8)
pXbar=c(0.0625,0.25,0.25,0.125,0.25,0.0625)
plot(X_Values,pX,type="h",lwd=4,ylim=c(0,1),xlab="x",ylab="p(x)",lend=1)
points(Xbar_Values,pXbar,type="h",lwd=4,col="red",lend=1)
```

We have been viewing simulations in class, and will set up a couple here as practice to make sure we understand how the process works. In this first example, we will generate 10000 normal random samples of size 5 with known mean $\mu = 10$, compute their mean and variance using the **apply** command, and then compute a t statistic. We will then overlay a T density function with 4 ($=5-1$) degrees of freedom, and a standard normal curve for comparison.

```
SampleSize15=matrix(rnorm(15*10000,mean=10,sd=2),ncol=15)
Sample_Mean=apply(SampleSize15,1,mean)
Sample_SD=apply(SampleSize15,1,sd)
T_Stat=(Sample_Mean-10)/(Sample_SD/sqrt(15))
hist(T_Stat,freq=F,main="T statistic with 14 df",ylim=c(0,0.40),xlab="x",ylab="f(x)")
X_values=seq(-4,4,by=0.01)
lines(X_values,dt(X_values,14),col="red")
```

I actually had to go back and add the **ylim** statement to adjust the vertical axis, since the height of the histogram was not as high as the density curve. The axis labels were not bad, but I changed those as well with **xlab** and **ylab**.

Try to run the same code, replacing the **rnorm** statement with **rgamma(150000,shape=10,scale=1)**; this is a right-skewed distribution with a mean of 10, just like the normal example. Make any adjustments necessary to the graph scale. How well does the density function follow the histogram? What can you conclude about the robustness of the t statistic?