Mathematics 376 – Probability and Statistics II
Lab Day 1 – A Surprising Regularity
January 28, 2004

Background

Last week we saw that if Y_i are independent samples from a normal $N(\mu, \sigma)$ distribution, then the sample mean $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ is also normally distributed, and moreover

$$U = \sqrt{n}\left(\frac{\overline{Y} - \mu}{\sigma}\right)$$

has a *standard normal* (that is, $N(0, 1)$) distribution.

Today, we will consider the question: What can be said about sample means for samples from *other* distributions? If the Y_i are not normally distributed, then unlike the normal case above the distributions of the random variables

$$U_n = \sqrt{n}\left(\frac{\overline{Y} - \mu}{\sigma}\right) \tag{1}$$

will usually depend on n. However there is an interesting pattern concerning the U_n as we let $n \to \infty$. Today, we will look for this pattern when the Y_i are taken from uniform, exponential, and gamma distributions.

A Maple Procedure

To start off today, you should download the latest version of the course Maple statistics package `MSP.map` from the homepage for this semester's course. Several additional procedures, including some we will need today, have been added since we used it last. Read the package into your current Maple session as usual.

The following Maple procedure can be used to test what happens when the individual samples Y_i have a *uniform* distribution on an interval $[a, b]$:

```
UniformTest:=proc(SampleSize,SimRep,a,b)
local k,U,HP,NP,SS,EP,TP,mu,sigma;
mu:=(a+b)/2; sigma:=(b-a)/sqrt(12);
for k to SimRep do
SS:=UniformSample(a,b,SampleSize);
U[k]:=evalf(sqrt(SampleSize)*(Mean(SS)-mu)/sigma);
end do:
U:=convert(U,list):
EP:=PlotEmpiricalPDF(U,-3,3,20):
TP:=plot(u->NormalPDF(0,1,u),-3..3,color=blue):
display(EP,TP);
end proc:
```

You can either enter this procedure directly in your worksheet (as a single execution group), or you can create a separate file containing the procedure statements, and read it into your worksheet the same way you read in the `MSP.map` file.

The procedure works as follows: You supply the sample size (n above), and the number of simulation repetitions, plus the endpoints of the interval in the procedure call, like this:

```
UniformTest(10,2000,2,5);
```

for 2000 groups of samples of size 10, (that is, 2000 sample means obtained by averaging 10 sample values each time) from a uniform distribution on $[2, 5]$. The output is an "empirical pdf" for the random variable U_n as in (1) above (in red), and for comparison purposes a standard normal pdf plotted on the same set of axes. The `UniformSample`, `Mean` and `PlotEmpiricalPDF` procedures are contained in the Maple stats package. (By the way, the "empirical pdf" is essentially the same as a *probability histogram*, but plotted by "connecting the dots" to make a continuous graph, rather than as a bar chart).

Lab Work

A) Run the `UniformTest` procedure for sample sizes $n = 1, 5, 10, 20, 40$, all with 1000 simulation repetitions, for the uniform distribution on $[a, b] = [3, 7]$. Explain the form of the plot with $n = 1$ (think about the graph of the uniform pdf). What do your graphs seem to indicate about the distributions of the U_n as n increases?

B) Now, we will consider the same question, but for samples Y_i from an exponential distribution. Recall that the exponential pdf from last semester:

$$f(y) = \begin{cases} \frac{1}{\beta}e^{-y/\beta} & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

has a parameter β, which controls the location of the mean and the standard deviation: $\mu = \sigma = \beta$. The procedure given above for the uniform distribution can be modified to do the same sort of thing for samples from an exponential distribution by making a small number of changes. First, the parameter β describes the exponential distribution completely (as do the endpoints a, b for a uniform distribution). So the procedure heading can be changed to

```
ExpTest:=proc(SampleSize,SimRep,beta)
```

Then the statements computing μ and σ need to be adapted for an exponential distribution. Finally, the correct procedure must be used for computing the random samples – `ExponentialSample` from the `MSP.map` package, rather than `UniformSample`. Copy and paste from the `UniformTest` procedure you created before and make these changes to produce a second procedure `ExpTest`. Test it on the exponential distribution with $\beta = 4.3$. Use $n = 1, 5, 10, 20, 40$, all with 1000 simulation repetitions. What does the plot with $n = 1$ show? Think about the form of the graph of an exponential pdf; you can also plot this pdf with a command of the form:

```
plot(u->ExponentialPDF(4.3,u),0..20);
```

What do your graphs seem to indicate about the distributions of the U_n as n increases?

C) Finally, we will consider the same question, but for samples Y_i from a gamma distribution. Recall that the gamma pdf from last semester:

$$f(y) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1} e^{-y/\beta} & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

has two parameters α, β, which control the location of the mean and the standard deviation. The procedure given above for the uniform distribution can also be modified to do the same sort of thing for samples from a gamma distribution by making a small number of changes. The procedure heading can be changed to

```
GammaTest:=proc(SampleSize,SimRep,alpha,beta)
```

The correct procedure for computing the random samples this time is `GammaSample` from the `MSP.map` package, rather than `UniformSample`. Copy and paste from the `UniformTest` procedure you created before once again and make all necessary changes to produce a third procedure `GammaTest`. Test it on the Gamma distribution with $\alpha = 2, \beta = 2$. (This is the same as the $\chi^2(4)$ distribution – chi-squared with 4 degrees of freedom.) Use $n = 1, 5, 10, 20, 40$, all with 1000 simulation repetitions. What do your graphs seem to indicate about the distributions of the U_n as n increases?

Assignment

One worksheet from each group with your test procedures and graphical results. Due in class, Monday February 2.