Mathematics 376 – Probability and Statistics II
Lab Day 1 – The Central Limit Theorem
January 20, 2006

*Background*

Last term we saw that if $Y_i$ are independent samples from any distribution with finite expected value $\mu$ and variance $\sigma^2$, then as $n \to \infty$, the sample means $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ approach a normal distribution. Moreover

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

approaches a *standard normal* (mean 0, standard deviation 1) distribution. The purpose of this first lab session is to "see" this happen as $n \to \infty$ for some distributions described by pdfs that are supported on a finite interval (i.e. $f(y) = 0$ outside an interval $a \le y \le b$).

*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 a version of the sampling procedure `FSample` we discussed in class last time, have been added since we used it last. Read the package into your current Maple session as usual.

The following Maple procedure generates a plot showing the distribution of the means of samples of size $n$ from a given distribution, together with a normal density graph that illustrates the conclusion of the Central Limit Theorem:

```
MP:=proc(PDF,a,b,n)
  local i,Means,MH,NP,CDF,mu,sigma;

mu:=evalf(Int(y*PDF(y),y=a..b));
sigma:=sqrt(evalf(Int(y^2*PDF(y),y=a..b))-mu^2):
CDF:=y -> int(PDF(t),t=a..y);
Means:=[ ];
for i from 1 to 200 do
  Means:=[op(Means),Mean(FSample(CDF,a,b,n))];
end do:
MH:=NormHist(Means,a,b,30):
NP:=plot(NormalPDF(mu,sigma/sqrt(n),y),
     y=mu-3*sigma/sqrt(n)..mu+3*sigma/sqrt(n),color=blue):
display(MH,NP);
end proc:
```

You can either enter this procedure directly in your worksheet (as a single execution group – Shift+Enter for each new line), or you can use a text editor to 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 pdf for the distribution you want (which should be one supported on a finite interval $a \le y \le b$), as a function, the endpoints $a, b$ of the interval, and the sample size ($n$). The procedure computes 200 samples of size $n$ from the given distribution, computes the mean of each one, and plots a normalized histogram showing the distribution of the means, together with a normal density function with mean $\mu$ and standard deviation $\sigma/\sqrt{n}$ – the Central Limit Theorem says that the distribution of the sample means should approach this as $n \to \infty$.

*A First Worked Example*

Enter the following commands to study sample means from the *uniform distribution* on $3 \le y \le 5$. (Put these commands and they output they generate in the worksheet you submit for this assignment, together with answers to the questions posed, in text regions.)

We first need to define the pdf for this as a Maple *function*. In mathematical terms:

$$f(y) = \begin{cases} \frac{1}{5-3} = \frac{1}{2} & \text{if } 3 \le y \le 5 \\ 0 & \text{otherwise} \end{cases}$$

(We say that $f$ is supported on the interval $3 \le y \le 5$, since $f$ is zero outside that interval.) We will build the endpoints into the way the procedure is called, so we can just enter:

```
f:=y->1/2;
MP(f,3,5,1);
```

to compute means of 200 samples of size $n = 1$. What does this plot tell you?

Now enter the `MP` command again several times and compare, for values $n = 5$, 10, 20, 30, and 100. What happens to the red histogram (the sample means), and the blue graph (the normal density) as $n$ increases?

*Lab Work*

A) Part of the surprising nature of the Central Limit Theorem is the way the "shape" of the initial distribution has less and less effect on the distribution of the sample means as $n$ increases. In this question you will study the distribution with pdf:

$$f(y) = \begin{cases} cy^2 & \text{if } -1 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

1) What value of $c$ makes this a legal pdf? (Use Maple to perform the needed integral – see the MP procedure above for some examples of Maple's `int` command for integration if you need a refresher. The Maple online help facility also has a lot of information that may be useful.)
2) Run the `MP` command several times with this new pdf, for for values $n = 1$, 5, 10, 20, 30, and 100. What does the plot for $n = 1$ represent? What happens to the

red histogram (the sample means), and the blue graph (the normal density) as $n$ increases?

B) Repeat question A for the distribution with pdf:

$$f(y) = \begin{cases} cy(2-y)^2 & \text{if } 0 \leq y \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

C) Give a line-by-line breakdown of what each statement (line) in the `MP` procedure does. For the statements that compute values, give the equivalent formulas in usual mathematical notation. (You may want to refer to some of the information in the online documentation for our statistics Maple package and/or Maple's online help for pointers on the Maple := operator, the `for` loop command, etc.)

*Assignment*

One worksheet from each group with your graphical results and answers to part C. Due in class, Friday January 27.