MONT 107N – Understanding Randomness
Lab Day 1 – Demonstration of the Central Limit Theorem
February 8, 2010

Background

Last Friday we saw that for one particular box model, the distribution of the values of the sum of the tickets drawn–as represented in the *probability histogram*–appeared to be getting closer to something like a normal distribution as *the number of draws from the box was increased*. Today, we want to ask: Does this always happen? (for all box models?) We will use a powerful piece of mathematical software called *Maple* to generate some pictures for several different box models and see whether this pattern continues to hold. The underlying mathematical result is called the *Central Limit Theorem*, and it's one of the most important foundational results about random processes.

Launching Maple

To get into Maple, you will need to follow this process.

1) When you start a session, you should see a Novell Client window with prompts for your username and password for the campus network. (If the Windows desktop is showing, that means previous user has probably not logged off. Use START/Shut Down/Log Off Holy Cross to terminate their session and start afresh.) Enter your username (full context) and password. A second login window marked *Windows Workstation* will appear at this point; you should be able to just click OK on this one and proceed to step 2.
2) You should now see the Windows desktop. Double click on Maple 13 shortcut icon.
3) After a few more seconds you should see a new Maple 13 window, with a "subwindow" marked Untitled (1) opened inside it. The Untitled (1) window is a blank *Maple worksheet*.

Maple Worksheets

Worksheets are integrated documents where any or all of the following can be done:

a) you can type in commands from the keyboard to ask Maple to perform many different kinds of calculations,
b) Maple will generate output (numerical values, symbolic formulas, and graphics),
c) you can modify commands, generating new output, in our session today.
d) you can enter text to annotate and explain the results of computations.

Take a few seconds and notice the features of this window – especially the "tool bar" across the top with the icons for various operations, the "scroll bar" on the right that you can use to move around within the worksheet, to see previous input and output lines, etc.

Input and Text Regions in a Worksheet

Maple worksheets can contain explanatory text (for instance answers to questions that are part of an assignment like the one we will be working on today) as well as commands and output. To create a text region, press the `Text` button above the active worksheet window. (It's best to do this right after results of a computation or a graph have been displayed, to put the comment below Maple's output.) Text can now be entered, and that will be treated as an "inert" comment. That is, it will appear as you enter it when you print out the worksheet, but it will not be treated as Maple input. Within a text region, if you want to get "fancy," you can enter special symbols and formulas from the pull-down menus along the left edge of the Maple window.

Saving and Reloading your Maple Worksheets

When you begin working on a worksheet, you will want to *save your work* every once in a while in case a computer problem develops, or in case you need more than one lab session to complete the work you are doing. This can be done most directly by saving to your network P: drive. Follow these directions:

1) In Maple, select the SAVE option from the FILE pull-down menu or press the toolbar icon that looks like a diskette.
2) If you are saving your work for the first time, you will see a SAVE AS dialog box. Make sure your campus network P: drive is showing in the *Save in:* box, then go to the *File name:* box, and type in a name for the file. The "extension" *mw* for Maple work sheet will be supplied by default. For instance, a good choice might be something like *labday1*. Then click the Save button. You only need to type in the filename in once in a session. Subsequent saves (i.e. FILE/SAVE or clicking the diskette toolbar icon) just update the file.
3) When you have the worksheet saved as you want it, you exit Maple, or continue working.
4) To update the worksheet further in a later session, get back into Maple as above, and read the worksheet back into Maple using the OPEN option from the FILE Menu or the "opening folder" toolbar button. Maple will prompt you as above for the name of the worksheet with a dialog box very like the SAVE AS box described above. Make sure your campus network P: drive is showing in the *Look in:* box, then highlight the worksheet you want, and click the Open button.

Printing in HA 136

To print a Maple worksheet in HA 136, click the toolbar icon that looks like a printer, and press OK on the PRINT dialog box (the settings should be set up correctly to print automatically). Your output should appear shortly on the printer at the back of the lab. If you have problems, or if the printer runs out of paper, come and get me or another math faculty member for help. (If several print jobs are sent at the same time, you may need to wait a short time.)

Getting Out

When you leave, quit the Maple window, and log off the campus network.

A Maple Procedure

On the course homepage:

`http://mathcs.holycross.edu/~little/Montserrat/Spring.html`

you will see a link "Maple code for lab day February 8." If you follow that, you will get the code for a Maple procedure that computes and prints out an "empirical probability histogram" for a box model with any number of tickets with whole number values, and any number of draws (with replacement).

Copy and paste the whole contents from the browser window (from the procedure heading line `PH := proc ...` *, down to the* `end proc:` *line into the Maple window. Then press Enter twice in the Maple window with the cursor at the end of the region where you pasted in. If you get any error messages or if nothing happens, be sure to call me over, and I will try to get things to work(!)*

What the procedure does is to simulate making a specified number `nDraws` of draws from the box some given number (`Repetitions`) of times. It keeps track of how many times each possible sum is produced, and it generates and plots a probability histogram based on the simulation. (Note: especially for small numbers of `nDraws` and `Repetitions`, you will see that the output *varies* each time you execute the PH (probability histogram) command. The histogram you are seeing is an *approximation to the true probability histogram* for the different possible values of the sum of the draws.

A First Worked Example

(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.)

Enter the following command to generate the empirical probability histogram for $n = 1$ draw from the box $[1, 2, 3, 4, 5]$, using 500 repetitions of the 40 draws to generate the empirical probability histogram:

```
PH([1,2,3,4,5],1,500);
```

(note the format for the command – the contents of the box are first (as a list), then the number of draws, then the number of repetitions). You should see a probability histogram with 5 bars of approximate height 20 (with areas adding up to 100%). Explain why this is reasonable.

A technical note: If you click the left mouse button with the cursor over a graph, Maple will allow you to resize by dragging and dropping one of the corners or edges. This is especially important to do when you generate lots of graphs (as in this lab!) because when you go to print, the default settings would make each graph take up nearly a whole sheet *of paper. Don't waste paper unnecessarily – please resize all of your graphs to a height of 1.5 to 2 inches on the screen!*

Below the graph generated by the first command, enter another command:

```
PH([1,2,3,4,5],2,500);
```

Note that the number of draws from the box is $n = 2$ now. Describe the new empirical probability histogram: What is the most likely sum for 2 draws? What are the least likely sums for 2 draws?

Now, repeat for $n = 5, 10, 20, 30$. Describe what happens to the probability histogram as n increases.

What are the expected value and standard error for the box with 30 draws and how does that correspond to the probability histogram? Using our formulas, we have

$$\text{EV} = \text{nDraws} \times \text{average of box}$$

and

$$\text{SE} = \sqrt{\text{nDraws}} \times \text{SD of box.}$$

You can have Maple compute these for you, if you first load a package of statistics commands by entering:

```
with(Statistics);
```

(note capital "S," which must be there). Then, for instance

```
EV:=30.0*Mean([1,2,3,4,5]);
```

computes the EV of the box (and saves it under the name `EV`), and

```
SE:=sqrt(30.0)*StandardDeviation([1,2,3,4,5])*sqrt(4./5.);
```

computes the SE by our formula (and saves it under the name `SE`). (Recall, as we mentioned last fall in the Excel regression lab, there are unfortunately two *different* ways that are commonly used to compute SD's of a list, and they yield different results. The `sqrt(5./.4)` is the "correction factor" needed to go from Maple's way of doing it to ours.) *In general*, you will need to multiply by the square root of the fraction

$$(\text{number of tickets in box} - 1)/\text{number of tickets.}$$

to convert.

To plot a (shifted and scaled) normal curve together with the probability histogram, first load a package of plotting routines with

```
with(plots):
```

then enter the following commands:

```
PPlot:=PH([1,2,3,4,5],30,500):
NPlot:=plot(100/(sqrt(2*Pi)*SE)*exp(-(x-EV)^2/(2*SE^2)),x=40..140):
display(PPlot,NPlot);
```

Technical note: The colon at the end of some of the commands suppresses the output because we don't need to see it in those cases. The normal curve will be plotted in red and the histogram will be in blue. They should be pretty close!

Lab Work

A) Generate the empirical probability histograms for `nDraws` $= 1, 5, 10, 20, 30$ for each of the following boxes. For the `nDraws` $= 30$ histogram, also generate an approximating normal curve and combine the plots as above. What happens as `nDraws` gets larger in each case?
   1) Box $= [1, 1, 0, 0, 0, 0, 0]$.
   2) Box $= [9, 1]$. (How are the histograms different in this case, and why?)
   3) A box *of your choice* – anything is OK, but make it different from the ones I listed, and different from everyone else's(!)

Important Note

If you save your worksheet in your campus network P: drive workspace and come back to it later to finish, you will use the FILE/OPEN option to read it back in. All the work you saved will be there, but you will not have actually computed anything in the current session. So at the least you will need to execute the commands to load the plots and Statistics packages again. One quick way to get back to exactly where you were at the end of your last session will be to press the tool bar button with !!! (triple exclamation point). This re-executes the entire worksheet. This will produce somewhat different probability histograms, of course, but that should not be a problem!

Assignment

Print out your worksheet (after resizing all the graphs!) and hand in. Due in class, no later than Monday, February 15.