INTRODUCTION TO COMPUTER SCIENCE II
LAB 5
Reading and processing data files

1 Calculating χ^2 from a histogram file

In Project 2, you are writing code that will generate a histogram of values that we expect to be uniformly distributed. For this lab, which is meant to help Project 2 along, we will assume that each run of your shuffle-testing program will generate a histogram file that looks something like this:

```
0 25
1 23
2 30
3 27
...
51 29
```

That is, each line contains, as text, a pair of numbers:

- The *index*, or position in the histogram, and

- The *value* of the histogram at that index.

If the indices range from 0 to $n-1$ inclusive (where, for what we're doing, typically $n = 52$), and assuming uniform distribution, then the expected value for each entry H_i in the histogram is:

$$E = \frac{\sum_{i=0}^{n-1} H_i}{n}$$

So, once we have E, we could calculate χ^2 as:

$$\chi^2 = \sum_{i=0}^{n-1} \frac{(E - H_i)^2}{E}$$

From there, we can look up, in a table that maps χ^2 values to probabilities, how likely the histogram's values are given our assumed uniformity and expected value.

2 Your assignment

Begin by copying `ChiSquared.java` and `Histogram.java` from my public directory[1] (or by clicking on those filenames here, which are links). When you open these files, you will see that the former is a simple program (containing just `main()` and `showUsageAndExit()`) that relies on creating and using a `Histogram` object.

The work to be done lies in `Histogram.java`. Specifically:

- The *specialized constructor*, `Histogram(File f)`, needs to read in the histogram via a `File` object in order to create the corresponding integer array that will be used to represent the histogram internally.

- The `chiSquared()` method must calculate and return that value from its histogram.

Complete these two methods. Write as many helper methods as you like. You may also want to consult the Java API to look up how use the `File` object to be sure that the given file exists and can be read, and to see how to make a `Scanner` object from that `File` object in order to read the integers within the file (and know when to stop reading).

3 How to submit your work

Use the CS submission systems to submit your `Histogram.java` source code file:

- **Web-based:** Visit the submission system web page.
- **Command-line based:** Use the `~lamcgeoch/submit` command at your shell prompt.

This assignment is due on Sunday, Nov-13, 11:59 pm.

---

[1] `sfkaplan/public/COSC-112/lab-5`