A Quick Evaluation of the Quality of Random Numbers Produced by OpenEpi's Random Module

Andy Dean, EpiInformatics.com

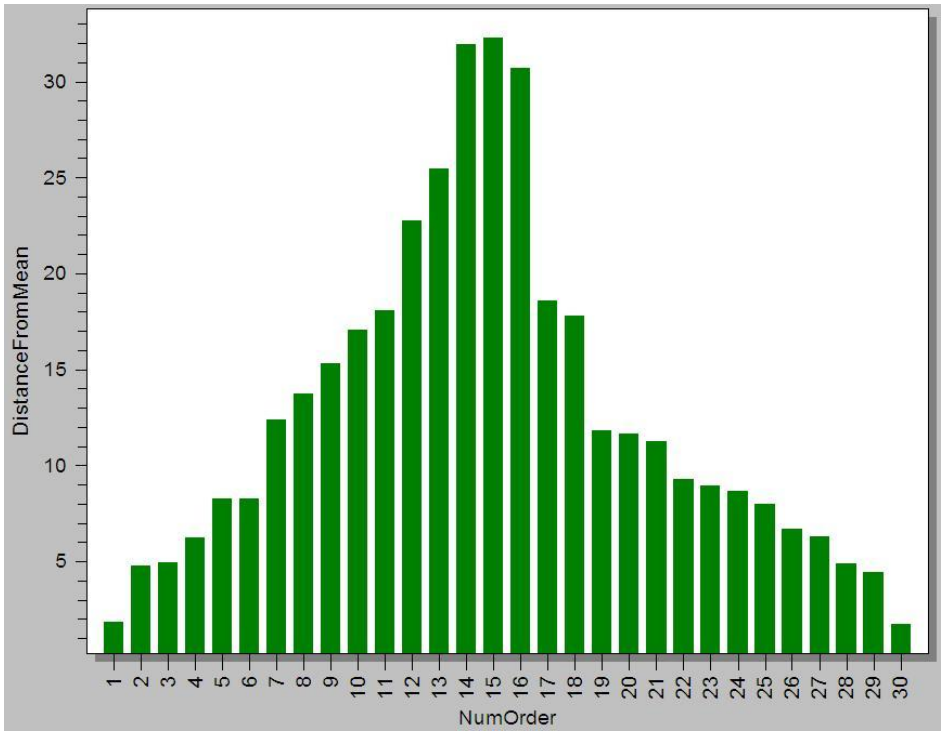
I made some crude attempts to assess the quality of the random numbers produced by OpenEpi's Random module, by generating 10000 numbers between 1 and 999 in a single column with OmitText set to "Yes." Since the file is in HTML, it has a lot of HTML table cells and attributes—more in volume than the numbers. Microsoft Word read the file and I was able to save it as a text file, and then to READ it in Epi Info's Analysis program. It indeed had 10000 numbers, with the following statistics:

Obs	Total	Mean	Variance	Std Dev
10000	5013275.0000	501.3275	82032.8867	286.4138

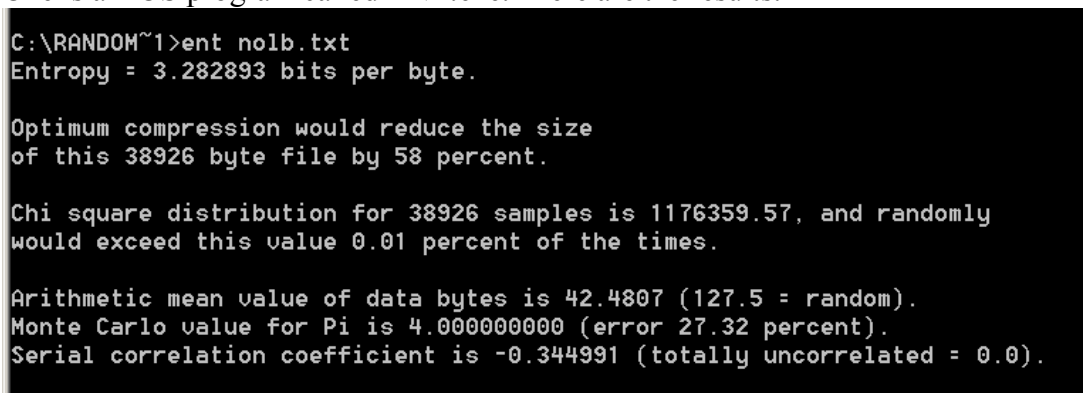
Minimum	25%	Median	75%	Maximum	Mode
1.0000	255.0000	504.0000	747.5000	999.0000	35.0000

After considerable maneuvering, I was able to draw 30 samples, using number mod 30.

The means of the 30 samples plotted as follows, with a little reordering to put the larger values in the middle. Not quite good enough for Dr. Gauss, but a tolerable resemblance to the normal curve. Clever people can take this analysis much further, more as a teaching exercise than an evaluation, since we know the randomness is not perfect.



I downloaded two programs designed specifically to test the quality of random numbers. One is a DOS program called ENT.exe. Here are the results.



Unfortunately, I can't say what they mean, but I know that the average mean value of databytes in an ASCII file filled with digits can never be 127.5, since that would mean including all the ASCII characters, including those above 128. Good for white noise or subatomic particles, perhaps, but not really pertinent to epidemiology.

The second test suite is called RNGmeter 0.9, from ComScire. It has a nice Windows interface. After running it for 5 minutes, it gave results suggesting that my file of numbers did not pass the more rigid tests. Those who enjoy such things will find a lot of software and hardware—generated random numbers on the Internet and may wish to characterize the numbers produced by various browsers. Since the browser suppliers apparently do not make public their tests, it would be a continuing task to keep up with new browsers as they come out.

Evaluation of the Microsoft .NET framework for random bit quality has been done, with excellent results, but it is not clear whether this is the one used by JavaScript in IE.

http://www.atstake.com/research/reports/eval_ms_ibm/analysis/2.3.4.html

I generated 80 numbers from 1 to 100 and entered them in a test available on the net at

http://ubmail.ubalt.edu/~harsham/Business-stat/otherapplets/Randomness.htm

by Professor Hossein Arsham. The module performed the Runs Test with 45 Runs, produced a p-value of 0.18238, and gave the conclusion, "Little or no evidence against randomness."

It is clear, however, that Random.htm produces different numbers each time it is run, that the numbers can be saved to a file by running OpenEpi from OpenEpiSave.hta on the a local disk, or by copying and pasting the output to a word processor or Excel. They should be useful for most epidemiologic purposes.

Comments prepared by Andy Dean, who found that this was fun, but not really a professional evaluation.