A Comparison of Random Number Generators Used in Business – 2004 Update

Wendy Rotz, Eric Falk and Archana Joshee
Ernst and Young LLP, 1225 Connecticut Ave., NW, Washington DC 20036

Key Words: RNG, Excel, SAS, DIEHARD, NIST statistical test suite

Abstract

In the 2001 JSM, Wendy Rotz, et al. presented *the Comparison of Random Number Generators Used in Business*. Based on this study, Microsoft revised Excel's Random Number Generator. Also since the 2001 research, one of the generators reviewed, RAT-STATS, had significant changes, and SAS introduced their new RAND function in version 8.2. This paper was intended to be an update of the current performance of these three packages. However, we found that we could not properly evaluate Excel's random number generator (RNG) because the algorithm has not yet been fully integrated into the package. RAT-STATS had not changed enough to allow direct testing of their package, and we did not find it necessary to repeat tests on their reported algorithm. We did test the new RAND function in SAS 8.2 and found it now passes the DIEHARD battery of RNG tests and the National Institute of Standards and Technology (NIST) statistical test suite for RNGs. This paper documents our findings in SAS and explains the current problems of testing in Excel and RAT-STATS.

Background

The business community increasingly uses computer-generated pseudo random numbers to perform statistical analyses, such as sampling and simulation. While statisticians are generally aware of potential problems with random number generators (RNGs), the business community often is not. Our 2001 study[1] focused on testing the random number generators from software packages used in business settings. This included SAS, Microsoft Excel, Microsoft Access, ACL, and RAT-STATS.

In 2001, using the DIEHARD battery of tests[2] and the statistical test suite from the National Institute of Standards and Technology (NIST),[3] we found some deficiencies in all of the packages in the study except the Wichman and Hill algorithm used by RAT-STATS. It outperformed all other RNGs we tested that year.

Excel's RNG had particularly weak performance in 2001, but our research generated interest at Microsoft. We shared the results of our tests with their software engineers and in the 2003 release of Excel, Microsoft reported[4] it revised its random number generator based on our findings of the Wichman and Hill algorithm.

Our 2001 results for the RANUNI function in SAS version 8.1 were similar to the results found by McCullough for SAS version 6.12.[5] It passed some rather intricate tests in DIEHARD and NIST, but failed a very basic test for a proportionate number of ones and zeros in a bit stream. However, in version 8.2, SAS introduced a new function, RAND, to be used in combination with the STREAMINIT function. The uniform random number generator that the RAND function uses is the Mersenne-Twister (Masumoto and Nishimura 1998). This generator has

[1] Rotz, Falk, Mulrow, and Wood "Comparison of Random Number Generators Used in Business" 2001 Proceedings of the Joint Statistical Meetings, Statistical Computing Section

[2] The tests and a description of them can be downloaded from http://stat.fsu.edu/~geo/diehard.html

[3] The tests and a description of them can be downloaded from http://csrc.nist.gov/rng/

[4] This is reported in the Microsoft Knowledge Base Article – 828795 found at: http://support.microsoft.com/default.aspx?kbid=828795

[5] McCullough, B.D (1998) "Assessing the Reliability of Statistical Software: Part II," *The American Statistician*, May, Vol. 53, No. 2.

a period of 2^{19937}–1 and 623–dimensional equidistribution up to 32-bit accuracy.[6]
Meanwhile, RAT-STATS, a software package used by the U.S. Department of Health and Human Services, finally upgraded from a DOS to a Windows environment. With such significant changes in three packages, we sought to test these newer versions. However, we found that only SAS 8.2 could be tested. This paper documents our problems in testing the new versions and our test results in SAS.

Excel

When we attempted to generate a new sequence of random numbers in the 2003 Windows XP version of Excel using the same seed as our earlier study, we were surprised to find we had ***exactly*** the same sequence of random numbers we had generated three years earlier. It appeared that Excel's RNG had not changed; at least not in the manner we were testing it.

That is, because the worksheet function RAND does not allow specification of a seed, we performed our 2001 and 2003 analyses using the alternative of a seed specification in the menus option *Tools*, *Data Analysis*, *Random Number Generation*. As it turns out, Excel's new RNG algorithm was not incorporated into the *Tools* package.

Yet, RAND was revised and, unfortunately, the worksheet function may now return negative random numbers, especially when it is called multiple times.[7] It is curious how the negative numbers were managed in the original release of Excel 2003. According to Microsoft,[8] they directly used the FORTRAN code described by Wichman and Hill,[9] which is just a linear combination of three modular functions. However the problem was created, Microsoft now has a hot fix to remedy the error.[10]
In the meantime, without the capability of specifying a seed, which is necessary to reproduce data and verify research findings, we did not pursue any further testing of Excel's new RNG at this time.

RAT-STATS

RAT-STATS is a small sampling and estimation package developed by the US Department of Health and Human Services. In 2001, only the RAT-STATS algorithm was tested because this DOS-based software package has limited capabilities and could only generate 1,000 random numbers at a time. This was insufficient for both the DIEHARD and NIST tests.

In our 2003 study, after RAT-STATS converted to a Windows environment, we again attempted to test the RNG inside the package and found the same limitation. Therefore, no further testing of RAT-STATS was done.

SAS

The new SAS RAND function that is used in conjunction with STREAMINT was tested in SAS version 8.2 on a Pentium II processor running Windows 2000 Service Pack 4.

For consistency with our 2001 research, we used the DIEHARD battery of tests and the statistical test suite from NIST in our analyses. Both of these series of tests are described in more detail in our 2001 paper.[11] A brief summary is provided here.

Both packages are comprised of over a dozen sets of hypothesis tests some of which themselves are comprised of hundreds of hypothesis tests on smaller subsets of data. Each test is designed to identify a different type of defect in a data stream. Three million random numbers were needed for the DIEHARD tests and the first one million bits of these numbers were used for the NIST tests.

---

[6] See Uniform Distribution in the description of RAND Function in SAS help http://v9doc.sas.com/cgi-bin/sasdoc/cgihilt?file=/help/lrdict.hlp/a001466748.htm&query=rand

[7] See *The RAND function returns negative numbers in Excel 2003* http://support.microsoft.com/default.aspx?kbid=834520

[8] See a *Description of the RAND function in Excel 2003* http://support.microsoft.com/default.aspx?kbid=828795

[9] Wichman, B.A. and I.D. Hill, *Building a Random-Number Generator*, BYTE, pp. 127-128, March 1987.

[10] See *The Excel 2003 hotfix package: February 29, 2004* http://support.microsoft.com/default.aspx?kbid=834520

[11] Rotz, Falk, Mulrow, and Wood "Comparison of Random Number Generators Used in Business" 2001 Proceedings of the Joint Statistical Meetings, Statistical Computing Section

We used SAS to convert the sequence of random numbers into the hexadecimal format required by the DIE-HARD tests and the binary sequence required by NIST. We repeated the tests for five trials, because with sheer quantity of hypothesis testing, it is possible by chance to get an occasional false failure. Erroneous conclusions are less likely with multiple trials.

We found the new SAS RAND function in conjunction with the STREAMINT function performed well and passed[12] all of the NIST and DIEHARD tests in all five trials. The RANUNI failures found in 2001 were not detected with the Version 8.2 RAND function.

Conclusions

Users of SAS should convert to RAND function instead of the old RANUNI, especially when thousands of calls to the RNG are being used for simulations. To the best we can discern, RAT-STATS, has not changed. Microsoft Excel continues to be plagued with difficulties. We have not tested whether the Microsoft's hot fix remedied the problem with the 2003 release with the RAND worksheet function. The menu's Random Number Generator under Tools does not appear to have been changed.

Next Steps

We will closely monitor Excel for new updates and retest when it is possible to do so with a random number seed. In the meantime, S-Plus is gaining popularity in business and should be considered in our next round of testing.

References

1. Dwyer, Jr. G. P. and Williams, K. B. "Portable Random Number Generators," Working Paper 99-14, Federal Reserve Bank of Atlanta http://www.frbatlanta.org/publica/work_papers/wp99/wp9914.pdf.

2. Gustafson, H. et al. (1994) "A Computer Package for Measuring Strength of Encryption Algorithms," *Journal of Computers & Security*, Vol. 13, No. 8.

3. Knuesel L. *On the reliability of Microsoft Excel XP for Statistical Purposes*, University of Munich http://www.stat.uni-muenchen.de/~knuesel/elv/excelxp.pdf.

4. Knuesel L. *On the Accuracy of Statistical Distributions in Microsoft Excel 97.* Computational Statistics and Data Analysis V 26 pp 375-377.

5. Knuth, G. (1998) *The Art of Computer Programming, Seminumerical Algorithms*, Vol. 2, 3rd Edition, Addition Wesley, Reading, Massachusetts.

6. Marsaglia, G. (1968) "DIEHARD: A Battery of Tests of Randomness," http://stat.fsu.edu/~geo.

7. McCullough, B.D (1998) "Assessing the Reliability of Statistical Software: Part I," *The American Statistician*, November, Vol. 52, No. 4.

8. McCullough, B.D (1998) "Assessing the Reliability of Statistical Software: Part II," *The American Statistician*, May, Vol. 53, No. 2.

9. McCullough, B.D, Wilson Berry *On the Accuracy of Statistical Procedures in Microsoft Excel 2000 and Excel XP* Computational Statistics and Data Analysis 40 2002 pp 713 – 721.

10. Microsoft (2004), *Description of the RAND function in Excel 2003* Microsoft Knowledge Base article 828795 http://support.microsoft.com/default.aspx?kbid=828795.

11. Microsoft (2004) , *Description of the effects of the improved statistical functions for the Analysis ToolPak in Excel 2003*, Microsoft Knowledge Base article 829208 http://support.microsoft.com/default.aspx?scid=kb;en-us;829208.

12. Microsoft (2004), *The RAND function returns negative numbers in Excel 2003 Microsoft Knowledge Base article 834520* http://support.microsoft.com/default.aspx?kbid=834520.

[12] By "passed" we mean that the tests did not fail an unusual number of times considering probability theory, the p-value, used and the number of hypothesis tests in the set.

13. Microsoft (2004) , *Excel 2003 hotfix package: February 29, 2004* Microsoft Knowledge Base article 833855 http://support.microsoft.com/default.aspx?kbid=833855.

14. Menezes, A. et al. (1997) *Handbook of Applied Cryptography*, CRC Press.

15. NIST Statistical Test Suite is available on http://csrc.nist.gov/rng/.

16. RAT-STATS is available on http://oig.hhs.gov/oas/ratstat.html.

17. Rotz, W. and E. Falk, D. Wood, and J. Mulrow, *A Comparison of Random Number Generators Used in Business*, 2001 Proceedings of the Joint Statistical Meetings, Section on Computational Statistics.

18. Soto, J. (1999) "Statistical Testing of Random Number Generators," *Proceedings of the 22nd National Information Systems Security Conference*. http://csrc.nist.gov/rng/nissc-paper.pdf.

19. Wichmann, B. and Hill, D. (1987) "Building a Random-Number Generator", *BYTE*, March 1987.

20. Wichmann, B A. and I.D. Hill, *Algorithm AS 183: An Efficient and Portable Pseudo-Random Number Generator*, Applied Statistics, 31, 188-190, 1982.