HISTORY OF RAND'S RANDOM DIGITS - SUMMARY

George W. Brown

P-113

June 1949

The RAND Corporation · SANTA MONICA · CALIFORNIA

HISTORY OF RAND'S RANDOM DIGITS - SUMMARY

The present speaker is reporting on work of several other people, dating to early days of Project RAND. Experimental probability problems arose quite early in RAND's history, in connection with a diversity of applications, not restricted exclusively to random walk problems of the Monte Carlo type. The variety of applications led to a desire for a large supply of random digits, of sufficiently high quality so that the user wouldn't have to question whether they were good enough for his particular application in the case of every different application.

Thus motivated, the RAND people, with the assistance of Douglas Aircraft Company engineering personnel, designed an electronic roulette wheel based on a variation of a proposal made by Cecil Hastings. For purposes of this talk a brief description will suffice. A random frequency pulse source was gated by a constant frequency pulse, about once a second, providing on the average about 100,000 pulses in one second. Pulse standardization circuits passed the pulses to a five place binary counter, so that in principle the machine is like a roulette wheel with 32 positions, making on the average about 3000 revolutions on each turn. A binary to decimal conversion was used, throwing away 12 of the 32 positions, and the resulting random digit was fed to an I.B.M. punch, yielding punched card tables of random digits. A detailed analysis of the randomness to be expected from such a machine was made by the designers and indicated that the machine should yield very high quality output.

This analysis leaned heavily on the assumption of ideal pulse standardization to overcome natural preferences among the counter positions; later experience showed that this assumption was the weak point, and much of the later fussing with the machine was concerned with troubles originating at this point.

From this point on much of the history comes from memoranda by Bernice Brown. Production of random digits really began on April 29, 1947, and by May 21 there were half a million digits generated. Some of the early digits had had some analysis and looked generally very good, at least there was no evidence of anything serious. The first half million digits had some exhaustive tests performed, for example frequency tests were made on every tenth block of 1000 digits, serial tests were made on three sets of 2000 digits each, various run tests were made, etc. These tests included the standard tests used by Kendall and Smith, the tests that almost everybody uses on random digits. Not only were chi-square tests made within blocks, but chi-square tests were also performed to test the fit of the distribution of the chi-square criteria against the hypothetical chi-square test. Everything seemed fine at this point. There was still no evidence of anything wrong.

By July 7 there were a million digits available, and subsequent analysis of a block of 125,000 digits showed the first results that could be considered disturbing, if one is sufficiently critical. There were 62,881 odd digits and 62,119 even digits in the sample of 125,000. The deviation from expected number was 381, of the order of 1/2 per cent of the expected number, leading to a

χ^2 value of 4.04 on one degree of freedom, which lies between the five per cent and two per cent points on the χ^2 distribution. All other tests were satisfactory and it is clear that there would be considerable doubt about how much importance to attach to the odd-even result. Since this one degree of freedom did stick out, there seemed to be good reason for making analyses of some more of the million digits.

The next analyses were made on two blocks of 125,000 digits each, block number one obtained in the period June 4 – 10, which came just after the machine had had a thorough tuneup, and block number two obtained July 7 – 8, after over one month of continuous operation without adjustment. The results were like this:

		Block 1		Block 2	
		χ^2	P	χ^2	P
Frequency χ^2	(9 d.f.)	6.0	.76	21.0	.03
Odd-even χ^2	(1 d.f.)	3.0	.15	7.0	<.01
Serial χ^2	(81 d.f.)	78.7	.52	105.6	.03

Most interesting is the fact that the excesses in both blocks 1 and 2 are in favor of odds over evens, and that Block 1 by itself looks satisfactory, even if the odd-even χ^2 might be considered suggestive, while Block 2 is comparatively very bad. The evidence seems to be, therefore, that the machine had been running down in the month since its tune up. It should be remembered that none of the biases which cause these extreme χ^2 values are large, in an absolute sense. The largest bias anywhere would not be estimated at more than 1-1/2 per cent or 2 per cent of the hypothetical frequency. The biases are, however, quite large compared with the predictions in the design analysis. The indications are that this particular machine required excessive maintenance to keep it in tiptop shape, but on the basis of what was learned it seems possible to build a machine of this type so that it will yield extremely high quality output.

At this point we had our original million digits, 20,000 I.B.M. cards with 50 digits to a card, with the small but perceptible odd-even bias disclosed by the statistical analysis. It was now decided to rerandomize the table, or at least alter it, by a little roulette playing with it, to remove the odd-even bias. We added (mod 10) the digits in each card, digit by digit, to the corresponding digits of the previous card. The derived table of one million digits was then subjected to the various standard tests, frequency tests, serial tests, poker tests, etc. These million digits have a clean bill of health and have been adopted as RAND's table of random digits.

There was, of course, good reason to believe that the addition process would do some good. In a general way, the underlying mechanism is the limiting approach of sums of random variables modulo the unit interval to the rectangular distribution, in the same way that unrestricted sums of random variables approach normality. This method has been used by Horton and Smith, of the Interstate Commerce Commission, to obtain some good batches of apparently random numbers from larger batches of badly non-random numbers.

My own personal hope for the future is that we won't have to build any more random digit generators. It was an interesting experiment, it fulfilled a useful purpose, and one can do it again that way, if necessary, but it may not be asking too much to hope that this addition property, perhaps, or some other numerical process will permit us to compute our random numbers as we need them. The advantages of such a method are fairly obvious in large-scale computation where extensive tabling operations are relatively clumsy.

References:

Kendall and Smith, Journal Royal Statistical Society, 1938, Supplement 1939

Bernice Brown, Some Tests of the Randomness of a Million Digits, RAND RAOP-44, October 19, 1948

Horton and Smith, Annals of Math. Stat., March, 1949.