Research Journal of Mathematical and Statistical Sciences ISSN 2320-6047
Vol. **6(3),** 1-5, March **(2018)** *Res. J. Mathematical and Statistical Sci.*

Short Review Paper

A statistical study of randomness among the first 5,00,000 digits of Pi (π)

Yogesh G. Raut
Maharashtra University of Health Sciences, Nashik, MS, India
yogesh.raut2@rediffmail.com

Available online at: www.isca.in, www.isca.me
Received 5th December 2017, revised 27th February 2018, accepted 11th March 2018

Abstract

A large amount of work has been done on the randomness of the digits of Pi (π) with various statistical tests of randomness which are used to distinguish good from not-so-good random number generators when applied to the digits of Pi (π). Sampling and simulation are the vital areas in statistics. In both of these areas random sample is the basic requirement to arrive at correct decision. To draw random sample, lots of methods ranging from lottery method to computer based random number generation are readily available. The digits of Pi (π) have to pass the tests as well as from the good random number generator (RNG) can be easily and rapidly generate in the computer. I have made an interesting study in the statement in which first 5,00,000 digits of Pi (π) were divided into various consecutive blocks and each block was tested for randomness by using Chi-square test goodness of fit. A statistical analysis of the first 5,00,000 digits of Pi (π) was carried out with a view to examine, in close detail, the degree of randomness in the frequency and in the order of appearance of the various digits therein. Frequency counts were done for single digits within blocks of 5000; 10,000 and 20,000 of digits of Pi (π). Calculation of the various statistical quantities shows that the sets of digits under analysis confirm closely to the hypothesis of perfect randomness.

Keywords: Frequency test, random number generator, Pi (π), Chi-square.

Introduction

Randomness testing or evaluating the quality of different pseudorandom number generators (PRNG) is becoming more crucial for satisfying the safety of conveyance. Use of random number was very critical; it was very difficult to generate a random number in exercise. Random number generators which generate numbers belong to entirely random sequences of need. In view of that, study needs to introduce few methods to identify the efficiency of random number generator. Whereas, there were different definitions for randomness and also different tools and criteria to assess the randomness of a generator[1-4]. The most popular technique for this purpose is the use of appropriate statistical tests for testing random sequence. The nature of a statistical test was statistical analysis of a random number. The purpose of generating random number has certain methods for such few statistical techniques were available[5-7]. The very common method for statistical analysis of these random number was chi-square test for goodness of fit and the run test[8-14]. There were also a number of individual tests[4,15,16]. The different statistical techniques were available to identify the sequence of random number. To study randomness of a generator normally maximum numbers were considered for its testing of sequence and it is necessary that it should pass the test of random sequence. It is also stated that pass or fail of a random sequence under any statistical testing does not mean that given sequence is random or nor. Though it was statistical methods which show the behavior of the data which can pass the test. Whereas, Run test is one of the most powerful statistical test for testing the random sequence of random number generator[8,12].

In connection with application of Monte Carlo methods to different problems in mathematical physics and the generating of random samples in the purpose of statistics there appear a need for the so-called random digits. The quick improvement was made in these fields of consideration the demand has more considerably during recent years. Accordingly, number of standard sets of such digits have been created and were being put to repeated use by researcher busy in this fields[17].

In this paper, the randomness of digits of Pi(π) among certain blocks in entirely random sequence was measure and was analyzed through statistical software. Accordingly, a correct technique for calculating the randomness chi-square test for goodness of fit results was presented for various sizes of blocks for digits of Pi (π).

Frequency Test: First of all the 5,00,000 digits of Pi (π)were considered as divided into 100; 50 and 25 consecutive blocks of 5,000; 10,000 and 20,000 each respectively. These blocks were taken to identify the randomness among the digits of Pi (π) with different block sizes. Because as the size of blocks increases non-random patches was observed. The frequencies (fi) with which the different digits i (i=0, 1,…..,9) noticeable in these blocks were noted. To save space, the actual values of these frequencies are not being given here. However, it was listed in

Table-1, Table-2 and Table-3 the values of the chi square statistics (χ^2), calculated (on the basis of the hypothesis of perfect randomness) for each one of the 100; 50 and 25 consecutive blocks and the corresponding probabilities (P) for nine (9) degrees of freedom. Table-1 shows that the values of P was significant in the case of the 5th, 11th, 43th, 53th, 83th, 92th and 100th blocks for the size of 5,000 digits of Pi (π). In Table-2, 27th block was highly significant; in both these cases the deviations of the actual frequencies from the expected ones were too high. Moreover, of all the blocks it was for the 8th block of Table-1, 28th block of Table-2 and 23rd block of Table-3 shows that the overall statistical deviations are the lowest and therefore the statistical P-value was high. The degree of non-randomness present in this block may not, however, remain tolerable in a more important approach of these digits for instance, if the rejection levels are at 5% and 95% confidence interval.

Objectives: i. To Study of randomness among the first 5,00,000 digits of Pi (π). ii. To Study of randomness among the digits of Pi (π) within various size of blocks of 5,000; 10,000 and 20,000.

Table-1: Values of the χ^2 statistic and the corresponding probabilities as obtained by applying the first tests to the 100 consecutive blocks of 5,000 digits each.

Block No.	Frequency Test		Block No.	Frequency Test		Block No.	Frequency Test		Block No.	Frequency Test	
	χ^2 (9 d.f.)	P		χ^2 (9 d.f.)	P		χ^2 (9 d.f.)	P		χ^2 (9 d.f.)	P
1	10.768	0.292	26	10.188	0.335	51	4.188	0.899	76	10.284	0.328
2	7.440	0.591	27	9.700	0.375	52	3.624	0.934	77	3.880	0.919
3	6.808	0.657	28	8.720	0.464	53*	17.964	0.036	78	9.872	0.361
4	4.900	0.843	29	8.344	0.500	54	12.108	0.207	79	7.284	0.608
5*	19.156	0.024	30	3.448	0.944	55	12.520	0.186	80	3.996	0.912
6	5.768	0.763	31	3.920	0.917	56	4.908	0.842	81	7.768	0.558
7	9.752	0.371	32	5.688	0.771	57	5.216	0.815	82	10.936	0.280
8	1.032	0.999	33	13.924	0.125	58	7.484	0.587	83*	17.916	0.036
9	8.540	0.481	34	5.500	0.789	59	6.924	0.645	84	3.516	0.940
10	3.944	0.915	35	5.012	0.833	60	4.564	0.871	85	10.056	0.346
11*	17.896	0.036	36	5.832	0.757	61	3.064	0.962	86	4.700	0.860
12	4.888	0.844	37	5.252	0.812	62	3.836	0.922	87	5.984	0.742
13	11.236	0.260	38	4.984	0.836	63	5.672	0.772	88	4.004	0.911
14	4.664	0.863	39	10.524	0.310	64	2.968	0.966	89	11.564	0.239
15	8.208	0.513	40	8.164	0.518	65	10.556	0.307	90	4.532	0.873
16	7.284	0.608	41	9.992	0.351	66	4.600	0.868	91	3.192	0.956
17	8.072	0.527	42	11.048	0.272	67	6.372	0.702	92*	17.244	0.045
18	3.660	0.932	43*	20.220	0.017	68	11.592	0.237	93	6.600	0.679
19	9.332	0.407	44	13.088	0.159	69	3.444	0.944	94	10.204	0.334
20	5.476	0.791	45	7.932	0.541	70	12.092	0.208	95	15.632	0.075
21	7.708	0.564	46	7.004	0.637	71	10.656	0.300	96	2.352	0.985
22	4.736	0.857	47	5.412	0.797	72	7.716	0.563	97	13.904	0.126
23	5.288	0.809	48	13.032	0.161	73	9.368	0.404	98	5.724	0.767
24	8.888	0.448	49	3.756	0.927	74	9.076	0.430	99	9.320	0.408
25	5.632	0.776	50	16.236	0.062	75	4.228	0.896	100*	20.456	0.015

*Blocks that fail the frequency test.

Table-2: Values of the χ^2 statistic and the corresponding probabilities as obtained by applying the first tests to the 50 consecutive blocks of 10,000 digits each.

Block No.	Frequency Test		Block No.	Frequency Test		Block No.	Frequency Test		Block No.	Frequency Test	
	χ^2 (9 d.f.)	P		χ^2 (9 d.f.)	P		χ^2 (9 d.f.)	P		χ^2 (9 d.f.)	P
1	9.318	0.408	14	12.088	0.208	27*	23.696	0.005	40	7.128	0.624
2	5.272	0.810	15	5.570	0.782	28	2.786	0.972	41	7.966	0.538
3	14.430	0.108	16	3.100	0.960	29	4.986	0.836	42	6.324	0.707
4	6.196	0.720	17	9.584	0.385	30	9.206	0.418	43	7.188	0.618
5	3.538	0.939	18	2.410	0.983	31	3.260	0.953	44	4.014	0.910
6	6.584	0.680	19	4.592	0.868	32	3.746	0.927	45	6.818	0.656
7	10.222	0.333	20	15.216	0.085	33	3.720	0.929	46	7.944	0.540
8	11.826	0.223	21	9.496	0.393	34	5.894	0.750	47	7.904	0.544
9	8.308	0.503	22	6.006	0.739	35	9.984	0.352	48	8.224	0.512
10	8.580	0.477	23	3.294	0.951	36	9.840	0.364	49	12.834	0.170
11	7.780	0.556	24	9.358	0.405	37	6.518	0.687	50	10.830	0.288
12	3.968	0.914	25	8.922	0.445	38	8.210	0.513			
13	9.220	0.417	26	5.278	0.809	39	4.920	0.841			

* Blocks that fail the frequency test.

Table-3: Values of the χ^2 statistic and the corresponding probabilities as obtained by applying the first tests to the 25 consecutive blocks of 20,000 digits each.

Block No.	Frequency Test		Block No.	Frequency Test		Block No.	Frequency Test	
	χ^2 (9 d.f.)	P		χ^2 (9 d.f.)	P		χ^2 (9 d.f.)	P
1	7.740	0.561	10	12.788	0.172	19	8.523	0.482
2	11.514	0.242	11	8.998	0.437	20	8.099	0.524
3	3.157	0.958	12	7.942	0.540	21	5.037	0.831
4	8.397	0.495	13	7.865	0.548	22	6.696	0.669
5	5.410	0.797	14	13.457	0.143	23	1.305	0.998
6	8.513	0.483	15	8.289	0.505	24	10.596	0.304
7	15.144	0.087	16	3.633	0.934	25	15.895	0.069
8	3.351	0.949	17	6.192	0.721			
9	4.749	0.856	18	15.217	0.085			

Methodology

Study Design: A descriptive analytical study was carried out.

Study Setting: The current study was carried out in Maharashtra University of Health Sciences, Nashik, Maharashtra, India from July2016 to January 2017.

Statistical Analysis: From website[18] first 5,00,000 digits of Pi(π) were taken for this study. The blocks of 5,000; 10,000 and 20,000 were made with the help computer program. The data was analyzed using SPSS version 17.0 statistical package. Chi-square test was used to study the randomness among the digits of Pi (π) and assessed the statistical significance of the differences noticed.

Results and discussion

It was clear from Table-1 that the P-value for the 5^{th}, 11^{th}, 43^{th}, 53^{th}, 83^{th}, 92^{th} and 100^{th} blocks was significant and leads to their rejection outright. In fact, these blocks, too, when taken singly, may not be considered as sufficiently suitable for a practical application. It is essential to first dilute them with a number of neighboring blocks before they can be considered fit for practical purposes.

Next, this study test to the 50 consecutive blocks of 10,000 digits each. The results obtained through a statistical test χ^2 analysis were given in Table-2. The 27^{th} block was highly significant that fails the frequency test. As first sight it appears quite significant that a block of as many as 10,000 digits should fail to comply with requirements of randomness.

After that, study applied to test to the 25 consecutive blocks of 20,000 digits each. The results obtained through a statistical test χ^2 analysis are given in Table-3. There was no block who fails frequency test. However, the result was hardly surprising because in an infinitely large set of digits, non-random patches of this, or ever a much larger size were quite expected, although the set on the whole may confirm to the hypothesis of randomness.

Discussion: The current analysis has been carried out mostly in blocks of 5,000; 10,000 and 20,000 digits of Pi (π)each with a vision to find 'patches', if any of that suffer from lack of randomness. In comparing the actual class frequencies with the ones expected on the basis of the hypothesis of perfect randomness, the χ^2 test for goodness of fit has been applied; the rejection level, follows to the Kendall and Smith[19] have been retain at 1% and 99% confidence interval.

Conclusion

In this paper, the randomness of blocks, for a first 5,00,000 digits of Pi (π) were analyzed. Chi square tests for goodness of fit were used for testing the random sequence within different blocks. This study highlighted chi-square tests for goodness of fit based on the probability distribution function to study randomness among the first 5,00,000 digits of Pi(π) within different sizes of blocks. It was observed that this test seems to be very difficult to pass the randomness of digits within blocks. Using the proper combination of digits of Pi (π), study found randomness in digits of Pi (π) within defined blocks. But simultaneously it was also found that, within few blocks a digit of Pi does not follow the characteristics of randomness. The result of this study was hardly surprising because in an infinitely large set of digits, non-random patches of this, or ever a much larger size were quite expected, although the set on the whole may confirm to the hypothesis of randomness.

Recommendation: The study had some limitation, as digits of Pi (π) were used for this study was limited. The blocks were made in multiple of thousands to study randomness among the digits of Pi (π). Further studies, will be carry out by using large no. of digits of Pi (π) and with a different size of blocks. It will help to study more regarding randomness among patches of larger size of the blocks of the digits of Pi (π).

Acknowledgement

I would like to show my gratitude to the Dr. K.D. Chavan (Registrar) and Mr. Nitin Kavede (Deputy Registrar) of Maharashtra University of Health Sciences, Nashik for sharing their pearls of wisdom with us during the course of this research. I thank my colleague Mr. Pravin P. Ghatekar and Mr. Ganesh E. Gaikwad for assistant with data handling and methodology for comments that greatly improved the study.

References

1. Sourabh S.K., Chakraborty S. and Das B.K. (2009). Are Subsequences of Decimal Digits of PI Random?. Annals. *Computer Science Series*, 7(2).
2. L'Ecuyer P. (1992). Testing random number generators. In Winter Simulation Conference: Proceedings of the 24 th conference on Winter simulation, 13(16), 305-313.
3. Marsaglia G. (1985). A current view of random number generators. *Computer Science and Statistics*, 16th Symposium on the Interface. Elsevier Science Publishers, North-Holland, Amsterdam, 3-10.
4. Maurer U.M. (1992). A universal statistical test for random bit generators. *Journal of cryptology*, 5(2), 89-105. IJCSNS International Journal of Computer Science and Network Security, 90, 9(9), September 2009.
5. Peizari B. and Dakhilalian M. (2003). Improvement of runs test and its application on sub blocks. 2nd Iranian Society of Cryptography Conference, Sharif University, September 2003.
6. Golomb S.W., Sequences S.R. and Holden-Day S.F. (1982). CA, 1967. Aegean Park.

7. Beker H. and Piper F. (1982). Cipher systems: the protection of communications. Northwood Books., London.

8. Rukhin A., Soto J., Nechvatal J., Smid M. and Barker E. (2001). A statistical test suite for random and pseudorandom number generators for cryptographic applications. *Booz-Allen and Hamilton Inc Mclean Va.* , NIST Special Publication 800-22, 15 May 2001. See http://csrc.nist.gov/rng/.

9. Anderson T.W. and Darling D.A. (1954). A test of goodness of fit. *Journal of the American statistical association*, 49(268), 765-769.

10. Gustafson H., Dawson E., Nielsen L. and Caelli W. (1994). A computer package for measuring the strength of encryption algorithms. *Computers & Security*, 13(8), 687-697. See http://www.isi.qut.edu.au/resources/cryptx.

11. Knuth D.E. (1998). The Art of Computer Programming. Seminumerical Algorithms, 3rd ed., Addison-Wesley, Reading, Mass, 2.

12. L'Ecuyer P., Simard R. and Test U01 (2001). A Software Library in ANSI C for Empirical Testing of Random Number Generators. Software user's guide. See http://www.iro.umontreal.ca/ simardr/testu01/tu01.html.

13. Marsaglia G. (2015). DIEHARD: a battery of tests of randomness (1996). See http://stat.fsu.edu/geo/diehard.html.

14. Marsaglia G. and Tsang W.W. (2002). Some difficult-to-pass tests of randomness. *Journal of Statistical Software*, 7(3), 1-9. See http://www.jstatsoft.org/v07/i03/tuftests.pdf.

15. Wegenkittl S. (1995). Empirical Testing of Pseudorandom Number Generators. Master of Science Thesis, Salzbufg University.

16. Wegenkittl S. (1998). Generalized-divergence and Frequency Analysis in Markov Chains. Ph.D. thesis, University of Salzburg. See http://random.mat.sbg.ac.at/team/.

17. Pathria R.K. (1962). A statistical study of randomness among the first 10,000 digits of π. *Mathematics of Computation*, 16(78), 188-197.

18. https://www.angio.net/pi/digits.html (http://www.angio.net/pi/digits/pi1000000.txt)

19. Kendall M.G. and Smith B.B. (1939). Random Sampling Numbers, Tracts for Computers. No. 24, *R. Statitical Soc.*, 101, 147-66, Cambridge Univ. Press

20. Kasture Madhukar, Pandharikar Nanda and Mathankar Mayura (2012). Index of First Order Independence within the set up of Markov Dependence. *International Journal of Management Studies, Statistics and Applied Economics*, 2(1), 15-20.

21. Patharia R.K. (1963). Study of Randomness Among The First 60,000 Digits of e *Department of Physics*, University of Delhi, Delhi 6, *(Received September 2, 1963) Communicated by P. V. Krishna Iyer, F.N.I.*

22. Pathria R.K. (1961). A Statistical Analysis of the First 2,500 Decimal Places of e and 1/e. *Department of Physics*, University of Delhi, Delhi 6*, (Received February 2, 1963) Communicated by P. V. Krishna Iyer, F.N.I.,* 27, 270-282.

23. Ruhkin A.L. (2001). Testing randomness: A suite of statistical procedures. *Theory of Probability & Its Applications*, 45(1), 111-132.