A Statistical Study of Randomness Among the First 10,000 Digits of π

By R. K. Pathria

1. Introduction. In connection with the application of Monte Carlo methods to various problems in mathematical physics and the drawing of random samples in statistics there arose a demand for the so-called random digits. As a result of the rapid progress made in these fields of investigation this demand has increased considerably during recent years. Consequently, a number of standard sets of such digits have been produced and are being put to frequent use by workers engaged in these fields [1]–[4].

At this stage it appears worthwhile to investigate, as has recently been suggested by the author [5], the extent to which one can utilize the digits appearing in the decimal development of the various constants of mathematical analysis, such as e, π, etc., for the purposes mentioned above. It is obvious that such a suggestion would have been hardly of any practical interest if it had been made at a time when the values of these constants were not yet available to a reasonably large number of decimal places. However, certain computations of this type have been carried out during recent years and in the near future they are to be extended to the point where they will surely provide sets of digits as large as the existing ones.*

Obviously, the question of randomness of the digits to be studied here cannot be decided on *a priori* grounds. One has to subject them to various tests and obtain internal evidence for their randomness before they can be declared fit for practical use. It appears worthwhile to mention here that apart from the specific purposes indicated above, a study of this type is fascinating also because of its intrinsic interest. It was apparently for this latter reason that Reitwiesner [6], at the suggestion of von Neumann, computed the values of π and e to more than 2,000 decimal places and Metropolis, Reitwiesner and von Neumann [7] carried out a statistical treatment thereof by studying the frequency distribution of the various digits. This study was extended to about 3,000 decimal places by Gruenberger [8] in the case of e and by Nicholson and Jeenel [9] in the case of π.

In the present paper a report is given of the results obtained by applying the four classical tests of Kendall and Smith (the frequency test, the serial test, the poker test, and the gap test) [10] and a fifth one due to Yule (the five-digit sum

---

Received February 27, 1961.

* Dr. D. B. Gillies of the Digital Computer Laboratory, University of Illinois, has kindly informed the author that in a year or so they will probably compute one million digits of e. At present their computation extends to 60,000 decimal places. A statistical study of these digits is also being carried out by the author and will be reported shortly.

Very recently Dr. Shanks and Dr. Wrench have made an IBM 7090 calculation of both π and e to 100,000 decimal places. The frequency distribution of the decimal digits of both the constants has also been computed. The author is highly grateful to Dr. Wrench for illuminating communications on this subject.

test) [11] to the first 10,000 digits of π.† The value investigated here is the one computed by Genuys [13] using the formula

$$\pi = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)} \left\{ \frac{16}{5^{2k+1}} - \frac{4}{239^{2k+1}} \right\},$$

and computing its terms on the IBM 704. The present analysis has been carried out mostly in blocks of 1,000 digits each, with a view to discover 'patches,' if any, that suffer from lack of local randomness. Of course, blocks which are found patchy are not suitable for drawing a random sample when used by themselves. They have to be suitably diluted by combining them with some of the neighboring blocks in order to obtain larger ones which could safely be employed in a statistical investigation.

In comparing the actual frequencies with expectations the χ^2 test has mostly been employed; the rejection levels, following Kendall and Smith [2], have been kept at 1 and 99 per cent.

2. The Frequency Tests. The 10,000 digits of $\pi - 3$ have been divided into ten consecutive blocks of 1,000 digits each and the frequencies f_i with which the various digits $i(= 0, 1, \cdots, 9)$ appear in these blocks have been recorded. These frequencies, along with the respective values of the statistic χ^2 and the corresponding probabilities P for nine degrees of freedom, are given in Table 1. It is only in the case of the third and the ninth blocks that the value of P is found to be significant; in the former case the deviations from the expected frequencies are too high, while in the latter they are too low.

Taking the table as a whole, of the 100 frequencies recorded 34 deviate from the expected value of 100 by more than the standard deviation $\sigma(= \sqrt{90} = 9.487)$ and 6 by more than 2σ. These figures compare well with the corresponding ones, namely, 31.73 and 4.55 per cent, for a normal distribution. Further, in the case of total frequencies the χ^2 value (9.318 for 9 d.f.) may be partitioned into three components, with the following obviously satisfactory results:

Classification	χ^2	d.f.	P
Odd versus even digits	0.360	1	~55 %
Within groups of odd digits	4.358	4	~35 %
Within groups of even digits	4.602	4	~35 %

3. The Serial Tests. These tests are employed with a view of looking for any evidence of serial association among the digits under study. The relevant test here consists in classifying the digit pairs (ij) with respect to the members i and j comprising a pair and comparing the frequencies thus obtained with expectations. We have tabulated the frequencies for the 10,000 overlapping pairs, formed by the first

† Gruenberger [12] has shown how the tests given by Kendall and Smith can be applied to any set of digits, punched on IBM cards, mechanically and without regard to the order of the digits on the cards, using standard IBM equipment. In the absence of such a facility, however, the author has made the various tabulations by hand and has satisfied himself about their correctness by applying suitable cross-checks.

TABLE 1

Frequency Distribution Among the First 10,000 Digits of $\pi - 3$

Block \ Digit	0	1	2	3	4	5	6	7	8	9	χ^2	P(%)
1	93	116	103	102	93	97	94	95	101	106	4.74	~85
2	89	96	104	86	102	108	106	102	101	106	4.94	~85
3	77	97	96	77	123	110	102	90	108	120	22.80	<1
4	103	120	105	103	87	102	96	90	95	99	7.58	~60
5	104	103	88	91	103	108	115	111	87	90	9.38	~40
6	91	94	98	113	105	97	106	118	90	88	9.28	~40
7	100	107	98	114	89	108	89	88	98	109	7.84	~55
8	97	100	119	95	107	104	108	92	84	94	8.80	~45
9	101	103	100	103	101	99	98	97	90	108	1.98	>99
10	113	90	110	90	102	113	107	87	94	94	9.32	~40
(1-10)*	968	1026	1021	974	1012	1046	1021	970	948	1014	9.318	~40

* The cumulative frequencies obtaining in this row are in complete agreement with the ones given by Dr. Wrench (private communication). See also J. W. Wrench, Jr., "The evolution of extended decimal approximations to π," *The Math. Teacher*, v. 53, 1960, p. 644–650; v. 55, 1962, p. 129–130.

10,001 digits of π, in Table 2. The following relations exist among these frequencies:

$$\sum_{i,j} f_{ij} = N$$

and

$$\sum_{l} f_{lm} = \sum_{n} f_{mn} + \epsilon_m ,$$

where $N = 10{,}000$ and ϵ_m which represents the "end effects" is equal to zero if the digit m appears either at both the ends of the set or at none; it is equal to -1 if the set opens with m and $+1$ if the set closes with m. In the case under study, we have $\epsilon_3 = -1$ and $\epsilon_8 = +1$. As a final check on the entries in this table, one verifies that the sum $\sum_{i,j} f_{ij}(i - j)$, which should obviously be equal to the difference between the first and the last digits of the set, is really equal to -5.

Now, the overall expectation m_{ij} of f_{ij} is, for each of the pairs, equal to Np^2, where p is the probability of occurrence of a particular digit. The variance of f_{ij} is, however, given by

$$\sigma_{ij}^2 = Np^2(1 + 2p\delta_{ij} - 3p^2),$$

where δ_{ij} is the Kronecker delta. Thus, whereas the expectation for each of the hundred elements of the array is 100 on the basis of perfect randomness, the standard deviation for the diagonal elements is 10.82 and that for the non-diagonal ones is 9.85. The observed values of the root-mean-square deviation are 9.76 and 8.78, respectively. Comparing the differences with the standard error in the dispersion one finds that none of these values is significant.

Several essentially equivalent values of χ^2 have been computed from Table 2. First, assuming all the hundred types of pairs to be equally likely (expected value of 100 for each cell), a χ^2 of 78.84 is obtained which, for 90 d.f., is at about 80 per cent probability level. Second, given the row sums and assuming the ten digits to be equally likely to follow (e.g., expected value of 96.8 for each of the cells in the first row), a χ^2 of 69.39 is obtained which, for 90 d.f., is at about 95 per cent prob-

TABLE 2

Frequency Distribution Among the First 10,000 Overlapping Pairs (ij) of $\pi(= 3.14 \cdots 78)$

$i \backslash j$	0	1	2	3	4	5	6	7	8	9	Total
0	85	103	98	103	98	89	101	93	83	115	968
1	99	99	103	102	121	95	106	90	98	113	1026
2	101	115	110	99	82	118	100	101	100	95	1021
3	102	92	86	94	114	100	90	102	97	98	975
4	95	100	100	89	102	110	103	108	101	104	1012
5	92	117	110	96	108	96	115	107	96	109	1046
6	107	95	117	97	101	124	91	101	90	98	1021
7	89	105	99	91	92	101	95	97	103	98	970
8	86	97	99	93	96	106	114	83	80	93	947
9	112	103	99	110	98	107	106	88	100	91	1014
Total	968	1026	1021	974	1012	1046	1021	970	948	1014	10000

ability level. Third, assuming the expectation of a particular cell to be one-tenth of the corresponding column sum, we get $\chi^2 = 69.26$ which, again for 90 d.f., gives $P \simeq 95$ per cent. Fourth, fitting all the expectations to both the row sum and the column sum, a value of 59.83 results which, for 81 d.f., is at about 96 per cent probability level. All these figures are obviously satisfactory.

Next, we have computed from Table 2 the value of the quantity $\overline{ij}$ whose theoretical expectation and standard deviation are given by

$$E(\overline{ij}) = (\bar{\imath})^2$$

and

$$\sigma(\overline{ij}) = \{\bar{\imath}^2 - (\bar{\imath})^2\} \cdot N^{-1/2}.$$

The actual value of this quantity turns out to be 20.062 which deviates from the expectation by an amount -1.1 times the S.D. The probability of equal or greater divergence of either sign is about 27 per cent—a result that is not significant.

So far we have been discussing the question of serial association between the neighboring digits comprising the whole set of 10,000 digits. We shall now study the various blocks one by one and see if they are individually also locally random. For this purpose, we give below the results of the χ^2 test, carried out on the assumption of equal *a priori* probability for each of the hundred cells:

Block	1	2	3	4	5	6	7	8	9	10
χ^2	96.6	82.2	115.2	101.4	96.2	135.4	90.8	93.0	80.6	100.2
P(%)	30	71	4	20	31	0.1	46	40	75	22

The P value in the case of the sixth block is too low and leads to its rejection outright.‡ The only other block for which the P value is rather low is the third one; this, however, has already failed to meet the frequency test.

‡ It may be noted that this block passed the frequency test very well. The failure here is mainly due to an essentially non-random arrangement of the digits in the block. For instance, the pair (77) appears 28 times (including 2 triplets and 3 quartets). Such an extreme pattern is dangerous even if diluted by one of its neighboring blocks. It can only be made harmless by combining with many other blocks.

TABLE 3

Classified Distribution of the First 2,000 Poker Hands of π

Classes	Actual Frequencies in Blocks										Expected Values	Actual Frequencies in the Whole Set	Expected Values
	1	2	3	4	5	6	7	8	9	10			
Busts (abcde)	63	54	74	72	68	58	51	71	60	58	60.48	629	604.8
Pairs (aabcd)	97	100	89	88	93	108	98	90	105	110	100.80	978	1008.0
Two pairs (aabbc)	23	30	25	27	22	18	32	17	15	18	21.60	227	216.0
Threes (aaabc)	14	13	11	10	13	14	15	19	18	13	14.40	140	144.0
Full house (aaabb)	2	3	1	3	1	0	3	2	2	1	1.80	18	18.0
Fours (aaaab)	1	0	0	0	3	2	1	1	0	0	0.90	8	9.0
Fives (aaaaa)	0	0	0	0	0	0	0	0	0	0	0.02	0	0.2
Total	200	200	200	200	200	200	200	200	200	200	200.00	2000	2000.0
χ^2(3 d.f.) →	0.34	4.04	6.47	6.16	1.55	1.29	6.78	5.36	2.68	2.11		2.69 (5 d.f.)	
$P(\%)$ →	95	25	10	10	50	50	8	15	45	55		75	

4. The Poker Test. The 10,000 digits of π are printed in 2,000 hands of five digits each. Among these hands we have noted the frequency of occurrence of those hands whose digits, with respect to their values, are either (i) all different, or (ii) one pair and the other three different, or (iii) two pairs and one different, or (iv) one triplet and two different, or (v) one triplet and one pair, or (vi) one quartet and one different, or finally (vii) one quintet. As usual, the frequencies thus obtained are compared with expectations. The results are given in Table 3. None of the χ^2 values is found to be significant.

An interesting observation may, however, be made here. Since the deviations in the third and the fourth blocks are, on the whole, in the same direction, a grouping of these two consecutive blocks results in a P value of about 1.5 per cent which is pretty low, though not below the rejection level. If, however, the rejection level were at 5 per cent, as might be the case in a more serious application of these digits, this combined sample of 2,000 digits would no longer be considered locally random. In that case it would be essential to combine this sample with one of a sufficiently large strength before one could employ its digits in an investigation.

5. The Gap Test. Next, a frequency count has been made of the lengths of the gaps between successive zeros of the set. This frequency distribution is compared with the expected one only for the whole set and not for the individual blocks, because the frequencies in the latter case are too small unless, of course, a very coarse grouping of the classes is adopted.

TABLE 4

Length Distribution Among the Gaps Between Successive Zeros of π

Length of the Gap	Actual Frequency	Expected Frequency	Length of the Gap	Actual Frequency	Expected Frequency
0	85	100.00	16	18	18.53
1	78	90.00	17	13	16.68
2	87	81.00	18	14	15.01
3	68	72.90	19	13	13.51
4	80	65.61	20	10	12.16
5	61	59.05	21	13	10.94
6	47	53.14	22	10	9.85
7	49	47.83	23	9	8.86
8	38	43.05	24	7	7.98
9	42	38.74	25–29	32[1]	29.39
10	29	34.87	30–34	9[2]	17.36
11	30	31.38	35–39	8[3]	10.24
12	25	28.24	40–49	13[4]	9.64
13	29	25.42	$\geqq 50$	11[5]	5.15
14	25	22.88			
15	14	20.59	Total	967	1000.00

[1] 6, 10, 3, 7 and 6, in order of increasing length.
[2] 3, 3, 0, 1 and 2, in order of increasing length.
[3] 1, 2, 3, 2 and 0, in order of increasing length.
[4] 2 each, of lengths 40, 41, 43, 44, 46 and 47; 1 of length 48.
[5] Actual lengths are 51, 51, 54, 55, 60, 62, 63, 65, 65, 65 and 67.

We have 968 zeros in our set and hence 967 gaps; their length distribution§, along with the one expected on the basis of perfect randomness, is given in Table 4. The χ^2 value for the grouping as indicated in the table is 28.06 which, for 30 degrees of freedom, gives $P \simeq 55$ per cent—a result that is not significant.

The mean length of a gap (excluding those of length zero) is found to be 10.190 which deviates from the expected value of 10 by 0.601 times the standard deviation (viz., 0.316). The result is obviously satisfactory.

6. The Five-Digit Sum Test. This test, as applied here, consists in taking the sum of the five digits comprising a (poker) hand as the variable, denoted by the symbol i, say, and comparing its distribution over the various hands with the one expected theoretically. The latter may be obtained in an elegant manner as follows (refer to the alternative approach of Yule [11]).

If Ω_i denotes the number of ways in which the five digits of the hand can give a sum i, it will be enumerated by the generating function

$$\begin{aligned}\sum_{i=0}^{45} \Omega_i x^i &= [\sum_{k=0}^{9} x^k]^5 \\ &= (1 - x^{10})^5 \cdot (1 - x)^{-5} \\ &= f(x), \qquad \text{say.}\end{aligned}$$

It immediately follows that

$$\Omega_i = \sum_{r=0}^{5} (-1)^r \binom{5}{r} \binom{i - 10r + 4}{4}.$$

Moreover,

$$\sum_i \Omega_i = \underset{x \to 1}{Lt}\, f(x) = 10^5,$$

being the total number of ways in which a hand of five digits can be formed out of the digits of ten kinds. The probability p_i for the value i of the variable is then given by

$$p_i = \frac{\Omega_i}{\sum_i \Omega_i} = 10^{-5}\Omega_i\,,$$

which leads to the expected distribution. This is the same as the one given in Table I of reference [11]. The mean value of i is 22.5 and its standard deviation is

$$(41.25)^{1/2} = 6.4226.$$

The standard error of the mean of n observations is, therefore, equal to 6.4226 $n^{-1/2}$. Further, the standard error of the standard deviation turns out to be

$$(18.1/n)^{1/2} = 4.2544\, n^{-1/2}.$$

---

§ As a check, it has been verified that the total length of the 967 gaps, as tabulated here, is 8988 which, together with the 31 digits preceding the first zero and the 13 digits following the last one, makes 9,032—the number of non-zeros in the set.

TABLE 5

Five-Digit Sum Distribution Among the First 2,000 Hands of π

i	Expected Frequency in a Block of 400 Hands	Actual Frequencies in Blocks of 400 Hands Each					Actual Frequency in the Whole Set	Expected Frequency in the Whole Set
		I	II	III	IV	V		
0	0.004	0	0	0	0	0	0	0.02
1	0.020	0	0	0	0	0	0	0.10
2	0.060	0	1	0	0	0	1	0.30
3	0.140	0	0	0	0	0	0	0.70
4	0.280	0	0	0	1	0	1	1.40
5	0.504	1	0	0	1	0	2	2.52
6	0.840	0	2	2	0	1	5	4.20
7	1.320	2	0	2	1	4	9	6.60
8	1.980	1	5	1	1	3	11	9.90
9	2.860	5	0	1	1	4	11	14.30
10	3.984	2	3	1	1	2	9	19.92
11	5.360	7	7	9	8	4	35	26.80
12	6.980	3	7	2	10	6	28	34.90
13	8.820	11	7	9	9	13	49	44.10
14	10.840	8	13	12	12	7	52	54.20
15	12.984	21	12	17	17	17	84	64.92
16	15.180	13	10	16	18	15	72	75.90
17	17.340	13	15	19	17	12	76	86.70
18	19.360	17	17	25	23	23	105	96.80
19	21.120	21	24	18	26	24	113	105.60
20	22.524	17	23	20	19	19	98	112.62
21	23.500	21	19	23	22	32	117	117.50
22	24.000	29	29	28	20	29	135	120.00
23	24.000	23	22	25	28	23	121	120.00
24	23.500	20	17	24	26	21	108	117.50
25	22.524	24	29	24	20	22	119	112.62
26	21.120	26	19	19	15	19	98	105.60
27	19.360	19	26	15	21	22	103	96.80
28	17.340	18	15	15	23	13	84	86.70
29	15.180	20	14	10	12	12	68	75.90
30	12.984	12	13	15	11	9	60	64.92
31	10.840	10	17	10	11	12	60	54.20
32	8.820	13	10	17	4	7	51	44.10
33	6.980	5	7	4	7	9	32	34.90
34	5.360	6	7	6	5	8	32	26.80
35	3.984	5	5	3	3	2	18	19.92
36	2.860	4	1	3	3	1	12	14.30
37	1.980	2	2	4	3	2	13	9.90
38	1.320	0	0	1	1	1	3	6.60
39	0.840	0	2	0	0	1	3	4.20
40	0.504	1	0	0	0	0	1	2.52
41	0.280	0	0	0	0	1	1	1.40
42	0.140	0	0	0	0	0	0	0.70
43	0.060	0	0	0	0	0	0	0.30
44	0.020	0	0	0	0	0	0	0.10
45	0.004	0	0	0	0	0	0	0.02
Total	400.000	400	400	400	400	400	2000	2000.00

TABLE 6

Mean Values of the Sum i, with Differences from Expectation, Etc.

Block	Mean Value	Difference from Expectation	Divided by Standard Error	Square of the Preceding Column
I	22.7450	+0.2450	+0.7629	0.5820
II	22.7550	+0.2550	+0.7941	0.6306
III	22.4625	−0.0375	−0.1168	0.0136
IV	22.0925	−0.4075	−1.2690	1.6104
V	22.1800	−0.3200	−0.9965	0.9930
The whole set	22.4470	−0.0530	−0.3690	

TABLE 7

Standard Deviations of the Sum i, with Differences from Expectation, Etc.

Block	Standard Deviation	Difference from Expectation	Divided by Standard Error	Square of the Preceding Column
I	6.3945	−0.0281	−0.1321	0.0174
II	6.4475	+0.0249	+0.1169	0.0137
III	6.2919	−0.1307	−0.6143	0.3773
IV	6.2241	−0.1986	−0.9334	0.8713
V	6.3716	−0.0510	−0.2398	0.0575
The whole set	6.3524	−0.0702	−0.7379	

The actual frequency distribution obtained from the 2,000 hands of the set is given in Table 5, where the results are also given for consecutive blocks of 400 hands each, i.e., comprising 2,000 digits each. The actual distribution is compared with the expected one through the mean values of the variable and its dispersions. In Table 6 we have listed for each of the five blocks, I to V, and for the whole set, the mean values and their deviations from the expectation in terms of the standard errors of the mean. None of the various deviations is found to be significant. In fact, the chance of equal or greater divergence, of either sign, in the case of the whole set is about 70 per cent. Moreover, even if we group together the last three blocks (each having a deviation of the same sign) the corresponding result comes out to be about 17 per cent. Still worse, if we take the last two blocks, for which the deviations are not only of the same sign but also of the greatest magnitude, the result is still about 11 per cent. Further, we note that the sum of the entries in the last column of the table is 3.83. Entering the χ^2 table with this value of χ^2 and 5 degrees of freedom, we find P to be about 60 per cent.

Finally, we study the standard deviations in the value of the variable as obtained from the frequencies tabulated above and compare them with the corresponding theoretical expectations. The relevant figures are given in Table 7. Expressing the deviations in terms of the standard errors of the standard deviation, we obtain results which do not exceed unity. Further, entering the χ^2 table with the sum of the squares of these numbers, namely, 1.34, and 5 degrees of freedom, we find that P lies between 90 and 95 per cent. For the whole set, the deviation of the actual standard deviation from the expected value is equal to -0.74 times the corresponding standard error; the chance of an equal or greater deviation of either sign is about 46 per cent.

Department of Physics
University of Delhi
Delhi-6, India

1. L. H. C. Tippett, *Random Sampling Numbers (41,600)*, *Tracts for Computers, No. 15*, Cambridge University Press, 1927.
2. M. G. Kendall & B. B. Smith, *Random Sampling Numbers (100,000)*, *Tracts for Computers, No. 24*, Cambridge University Press, 1939.
3. *Table of 105,000 $\bar{x}$ Random Decimal Digits*, Interstate Commerce Commission, Bureau of Transport Economics and Statistics, Statement No. 4914, Washington, D. C., May 1949.
4. Rand Corporation. *A Million Random Digits with 100,000 Normal Deviates.* The Free Press, Glencoe, Illinois, 1955.
5. R. K. Pathria, "A statistical analysis of the first 2,500 decimal places of e and $1/e$," *Proc. Nat. Inst. Sci. India*, v. A27, 1961, p. 270–282. See also *Math. Gaz.*, London, v. 45, 1961, p. 142–143.
6. G. W. Reitwiesner, "An ENIAC determination of π and e to more than 2000 decimal places," *MTAC*, v. 4, 1950, p. 11–15.
7. N. C. Metropolis, G. Reitwiesner & J. von Neumann, "Statistical treatment of values of first 2000 decimal places of e and π calculated on the ENIAC," *MTAC*, v. 4, 1950, p. 109–111.
8. F. Gruenberger, "Further statistics on the digits of e," *MTAC*, v. 6, 1952, p. 123–124.
9. S. C. Nicholson & J. Jeenel," Some comments on a NORC computation of π," *MTAC*, v. 9, 1955, p. 162–164.
M. G. Kendall & B. B. Smith, "Randomness and random sampling numbers," *J. Roy. Statist. Soc.*, v. 101, 1938, p. 147–166.
G. U. Yule, "A test of Tippett's random sampling numbers," *J. Roy. Statist. Soc.*, v. 101, 1938, p. 167–172.
12. F. Gruenberger, "Tests on random digits," *MTAC*, v. 4, 1950, p. 244–245.
13. F. Genuys, "Dix mille decimales de π," *Chiffres*, v. 1, 1958, p. 17–22.