A STUDY OF THE DIGITS OF π, e AND CERTAIN OTHER IRRATIONAL NUMBERS

by

BRUCE R. JOHNSON & DAVID J. LEEMING

DM–430–IR

DECEMBER, 1986
(Revised January, 1988)

Short Title: DIGITS OF CERTAIN IRRATIONAL NUMBERS

A STUDY OF THE DIGITS OF π, e AND CERTAIN OTHER IRRATIONAL NUMBERS

by

BRUCE R. JOHNSON and DAVID J. LEEMING

Department of Mathematics
University of Victoria
Victoria, B.C.
Canada V8W 2Y2

AMS 1980 Subject Classification (1985 revision): 65C10

A STUDY OF THE DIGITS OF π, e AND CERTAIN OTHER IRRATIONAL NUMBERS

Short Title: DIGITS OF CERTAIN IRRATIONAL NUMBERS

by

BRUCE R. JOHNSON and DAVID J. LEEMING

Abstract. The first 100,000 digits in the decimal expansions of π, e, $\sqrt{2}$, $\sqrt{3}$, $\sqrt{5}$, $\sqrt{7}$, $\sqrt{11}$ and $\sqrt{13}$ were investigated for properties of randomness. Using a measure of randomness based on several different runs statistics, the decimal expansions of these irrational numbers behaved very much like random sequences when compared to the outputs of two popular random number generators. Also, for a better understanding of power, the measure of randomness was evaluated for several different kinds of nonrandom digit sequences.

AMS 1980 Subject Classification (1985 revision): 65C10.

1. **INTRODUCTION.** The digits in the decimal expansions of such irrational numbers as π, e, $\sqrt{2}$, $\sqrt{3}$, $\sqrt{5}$, $\sqrt{7}$, $\sqrt{11}$, and $\sqrt{13}$, are of continuing interest. In 1987 Yasumasa Kanada and his colleagues at the University of Tokyo computed π to more than 134 million decimals on a NEC SX–2 supercomputer, eclipsing the record set in 1986 by David H. Bailey using a Cray–2 supercomputer at the NASA Ames Research Center (see Peterson (1987)). Shanks and Wrench (1969) have calculated e to 100,000 decimals, and Dutka (1971) has computed $\sqrt{2}$ to one million decimals. Also, Beyer *et al.* (1969) have tabled $\sqrt{n}$ for $n = 2,3,5,6,7,10,11,13,14$, and 15 to 29,354 octal digits.

Many statistical investigations of these digits have been completed. Pathria (1962) carried out a statistical study of the first 10,000 digits of π, applying frequency tests, serial tests, poker tests, gap tests and five–digit sum tests. From the multitude of statistical tests for randomness, Pathria found very little statistical evidence to disprove the hypothesis that the digits of π well–approximate a table of random digits. Guilloud and Bouyer (1974) extended Pathria's study to the first million decimal digits of π. Stoneham (1965) completed a similar study on the first 60,000 digits of e finding frequency, poker, serial, and gap tests in support of the hypothesis of pure chance selection in the sequence of decimal digits for e, with only a few anomalies. In 1970, Beyer *et al.* extended the work of Good and Gover (1967) by applying the generalized serial test for randomness to the digits of irrational $\sqrt{n}$ in bases b where $2 \le n,b \le 15$. Again the results were consistent, except for a few aberrations, with the hypothesis of randomness of the digits. However, all statistical tests for randomness are very inadequate for proving randomness on statistical grounds. Their approach is to weight the strength of statistical evidence against the hypothesis of randomness H_0. This is accomplished by focusing on an appropriate test statistic and computing, under the assumption that H_0 is true, the probability that a rerun of the experiment would produce evidence against H_0 at least as strong as the observed result. A decision to retain H_0 is based not on statistical proof of randomness but on the lack of sufficiently strong evidence against randomness. That is, randomness is assumed unless nonrandomness is proved beyond reasonable doubt.

In this paper we undertake a statistical investigation of the first 100,000 digits in the decimal expansions of π, e, $\sqrt{2}$, $\sqrt{3}$, $\sqrt{5}$, $\sqrt{7}$, $\sqrt{11}$, and $\sqrt{13}$ using a new approach. In our study we derive a measure of randomness based on Run Statistics and the Strong Law of Large Numbers. We use this measure to compare the digits in the decimal expansions of these eight irrational numbers with the digits generated by two of the most popular random number generating routines. It will be seen that for randomness, the irrational numbers compare very well with the outputs of these routines.

Apart from the purely mathematical interest, since it is now posssible to compute the value of these constants accurately to a large number of decimal places, our work supports the suggestion that the digits of π, e, etc. might provide convenient sets of random digits for such applications as Monte Carlo studies in statistics and in mathematical physics (e.g. nuclear scattering problems).

2. **COMPUTATIONS.** The computation of π to 100,000 decimal places was done using the method of Shanks and Wrench (1962). This method involves computing π_1, and π_2 as a check, where

$$\pi_1 = 8 \arctan\left[\frac{1}{57}\right] + 4 \arctan\left[\frac{1}{239}\right] + 24 \arctan\left[\frac{1}{8}\right]$$

and

$$\pi_2 = 48 \arctan\left[\frac{1}{18}\right] + 32 \arctan\left[\frac{1}{57}\right] - 20 \arctan\left[\frac{1}{239}\right].$$

The number of correct digits is given by the integer part of the negative of the common logarithm of the absolute value of $\pi_1 - \pi_2$. The computing time for π to 100,000 digits was two hours and 50 minutes on an IBM 3083 using a multiple precision floating point arithmetic package obtained from the Share Program Library Agency, Triangle Universities Computation Center, Research Triangle Park, North Carolina, U.S.A. The computation of e was done using the obvious series expansion (e_1) with a check on accuracy provided by

$$e_2 = \left[\sum_{j=0}^{k} \frac{(-1)^j}{j!} \right]^{-1}$$

for sufficiently large k. The number of correct digits is obtained by converting e_1 and e_2 to character strings and comparing the two until a difference is detected.

The digits of $\sqrt{2}$, $\sqrt{3}$, $\sqrt{5}$, $\sqrt{7}$, $\sqrt{11}$ and $\sqrt{13}$ were all obtained using Newton's method with the usual accuracy check of comparing successive iterates.

The two random integer generators used in this study for comparison purposes were URAND and the NAGFLIB routine GO5DYF (see Fortran Library Manual (1984)). The URAND routine is described and listed in Forsythe *et al.* (1977). The GO5DYF routine is designed to return a pseudo–random integer taken from a uniform distribution between integers M and N (inclusive). In our study we set $M = 0$ and $N = 9$.

3. **MEASURE OF RANDOMNESS.** For a sequence of Bernoulli (success/failure) trials, any sequence of like observations bounded by observations of the other type is called a *run.* For example, the sequence

S F F F F S S F S F

results in 6 runs (3 runs of successes and 3 runs of failures).

The probability distribution of the number of runs, R, based on a sequence of n independent Bernoulli trials with success probability = p is given by:

$$p_{2k} = P(R=2k) = \sum_{m=k}^{n-k} 2 \begin{bmatrix} m-1 \\ k-1 \end{bmatrix} \begin{bmatrix} n-m-1 \\ k-1 \end{bmatrix} p^m (1-p)^{n-m},$$

and

$$p_{2k+1} = P(R=2k+1) = \sum_{m=k}^{n-k} \left\{ \begin{bmatrix} m-1 \\ k \end{bmatrix} \begin{bmatrix} n-m-1 \\ k-1 \end{bmatrix} + \begin{bmatrix} m-1 \\ k-1 \end{bmatrix} \begin{bmatrix} n-m-1 \\ k \end{bmatrix} \right\} p^m (1-p)^{n-m}.$$

See Conover (1971), pp. 349–356. This probability distribution is tabulated below for the two cases $(n = 10, p = .1)$ and $(n = 10, p = .5)$.

	$p_r = P(R=r)$	
r	p = .1	p = .5
1	.348678440	.001953125
2	.087169610	.017578125
3	.348678439	.070312500
4	.073549361	.164062500
5	.110324041	.246093750
6	.018004259	.246093750
7	.012002839	.164062500
8	.001264961	.070312500
9	.000316240	.017578125
10	.000011810	.001953125

TABLE 1. Probability Distributions of R.

A sequence of n independent Bernoulli trials with success probability $= p$ can be viewed as a multinomial trial with cell probabilities $p_1, p_2, \cdots, p_n$, where $p_r = P(R=r)$, $r = 1,2,\cdots,n$. The frequency, $f_r(N)$, of occurrence of the event $\{R=r\}$ resulting from N such independent multinomial trials should be near the expected frequency Np_r when N is large. By the Strong Law of Large Numbers, the sum

$$(1) \qquad \sum_{r=1}^{n} |f_r(N)-Np_r|/Np_r$$

of fractional absolute deviations of observed cell frequencies from expected cell frequencies will, with probability one, converge to zero as $N \to \infty$.

One way to evaluate the relative suitability of sequences of digits for use as tables of random numbers is suggested by the following scheme. First, partition a given sequence of digits into N consecutive strings of n digits each. Divide the set of possible digits $\{0,1,2,\cdots,9\}$ into two nonempty subsets S and F. View each digit in the given sequences as a success or failure depending on whether it belongs to the set S or F, respectively. Then for a truly random sequence, each digit is the outcome of an independent Bernoulli trial with success probability $p = P(S)$. Therefore, when N is large the statistic (1) should be small. Among several different equal–size sequences of digits, the one having the smallest value of (1) is in some sense the most suitable as a table of random numbers. A more comprehensive measure of randomness can be achieved by summing the separate realizations of (1) for several representative designations of the success set S.

4. **RESULTS AND CONCLUSIONS.** In our study each measure of randomness is based on the sum of twelve realizations of (1) obtained from the following twelve different designations of S: $\{0\}$, $\{1\}$, $\{2\}$, $\cdots$, $\{9\}$, $\{5,6,7,8,9\}$, and $\{0,2,4,6,8\}$. Note that the success probability is $p = .1$ for the first ten designations of S, and $p = .5$ for the last two. All sequences of digits considered here consist of $N = 10{,}000$ strings of $n = 10$ digits each. The eight irrational number expansions included in this study are the first 100,000 digits of the decimal expansions of the fractional parts of π, e, $\sqrt{2}$, $\sqrt{3}$, $\sqrt{5}$, $\sqrt{7}$, $\sqrt{11}$ and $\sqrt{13}$. For comparison, ten different runs of 100,000 digits were generated by each of two widely used random number generator routines URAND and GO5DYF. The results

are summarized in Table 2. Of course, the smaller the tabled entry, the higher the randomness rating. When computing the statistic (1) for the cases where $p = .1$, the run–length = 10 cell was combined with the run–length = 9 cell because the run–length = 10 probability is so small (.000011810) when $p = .1$.

	π	e	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{5}$	$\sqrt{7}$	$\sqrt{11}$	$\sqrt{13}$
Measure of Randomness	7.89	11.63	11.78	9.69	10.08	8.09	9.22	10.78

	Measure of Randomness		
Random # Generator	Smallest of 10 runs	Largest of 10 runs	Average
URAND	9.20	11.73	10.30
G05DYF	8.12	11.82	9.87

TABLE 2: Randomness Ratings. The smaller the tabled number, the higher the randomness rating.

From these results it is apparent that, as tables of random digits, the digits from the decimal expansions of the eight irrational numbers in this study compare very favorably with the outputs from the two random number generator routines. In fact, both π and $\sqrt{7}$ achieved higher randomness ratings than the best of 10 independent 100,000–digit runs from each of the two random number generator routines.

The realm of nonrandomness is so vast that a comprehensive study of the measure's power (ability to detect nonrandomness) is not feasible. However, to gain some understanding regarding the power question, the measure was evaluated for four different kinds of nonrandom sequences (NRS) of digits as described below.

NRS 1: A sequence of digits was generated using random number generator GO5DYF. Each time a zero digit was generated, a random decision–digit was generated independently, and the zero was excluded from the sequence whenever the decision–digit was low (i.e. 0, 1, 2, 3, or 4). Each decision–digit was used only once and then discarded. Therefore, NRS 1 sequences contained about 5% zero digits and $10.\overline{5}\%$ each of the nine nonzero digits.

NRS 2: A sequence of digits was generated using random number generator GO5DYF. Each time a nonzero digit was generated, a random decision–digit was generated independently, and the nonzero digit was changed to a zero digit whenever the decision–digit was zero. Each decision–digit was used only once and then discarded. Therefore, NRS 2 sequences contained about 19% zero digits and about 9% each of the nine nonzero digits.

NRS 3: A sequence of digits was generated using random number generator GO5DYF, except that a generated digit was recorded in the sequence only when it was different from the immediately preceding digit. Therefore, NRS 3 sequences contained about 10% each of digits zero through nine, but contained no consecutively repeated digits.

NRS 4: A sequence of digits was generated using random number generator GO5DYF, except that each generated digit was recorded twice in the sequence whenever the immediately preceding digit was even. Therefore, NRS 4 sequences contained about 10% each of digits zero through nine, but contained more consecutively repeated digits than would have been expected under true randomness.

Ten different runs of 100,000 digits were generated by each of the four procedures NRS 1–4. The results are summarized in Table 3. For each of these nonrandom sequences, the measure of randomness gave a much lower randomness rating (higher value) as compared to the ratings in Table 2.

	Measure of Randomness		
Generating Procedure	Smallest of 10 runs	Largest of 10 runs	Average
NRS 1	16.31	20.77	19.32
NRS 2	38.77	47.88	41.68
NRS 3	41.19	48.00	44.36
NRS 4	95.04	98.31	96.62

TABLE 3: Randomness Ratings. The smaller the tabled number, the higher the randomness rating.

REFERENCES

Beyer, W.A., Metropolis, N. and Neergaard, J.R. (1969): Square roots of integers 2 to 15 in various bases 2 to 10: 88062 binary digits or equivalent. *Math. Comp.*, **23**, 679.

________________________ (1970): Statistical study of digits of some square roots of integers in various bases. *Math. Comp.*, **24**, 455–473.

________________________ (1970): The generalized serial test applied to expansions of some irrational square roots in various bases. *Math. Comp.*, **24**, 745–747.

Conover, W.J. (1971): *Practical Nonparametric Statistics*, Wiley, New York.

Dutka, J. (1971): The square root of 2 to 1,000,000 decimals. *Math. Comp.*, **25**, 927–930.

Forsythe, G.E., Malcolm, M.A. and Moler, C.B. (1977): *Computer Methods for Mathematical Computations*, Prentice–Hall, Englewood Cliffs, N.J.

Fortran Library Manual, Mark II (1984): Random number generators (GO5DYF), v. 5, Numerical Algorithms Group, Oxford.

Good, I.J. and Gover, T.N. (1967): The generalized serial test and the binary expansion of $\sqrt{2}$. *J. Roy. Statist. Soc. Ser. A*, **130**, 102–107.

________________________ (1968): Corrigendum. *J. Roy. Statist. Soc. Ser. A.*, **131**, 434.

Guilloud, J. and Bouyer, M. (1974): *Un Million de Décimales de π*, Technical Report, Commissariat a l'Energie Atomique, Paris.

Pathria, R.K. (1962): A statistical study of randomness among the first 10,000 digits of π. *Math. Comp.*, **16**, 188–197.

Peterson, I. (1987): Pi wars: dueling supercomputers. *Sci. News*, **131**, 118.

Shanks, D.E. and Wrench, Jr., J.W. (1962): Calculation of π to 100,000 decimals. *Math. Comp.*, **16**, 76–99.

________________________ (1969): Calculation of e to 100,000 decimals. *Math. Comp.*, **23**, 679–680.

Stoneham, R.G. (1965): A study of the 60,000 digits of the transcendental "e". *Amer. Math. Monthly*, **72**, 483–506.