International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 4, 2005, pp 13-24

Revisiting the Digits of π and Their Randomness

Herbert C. Heien
Minnesota State University,
Mankato, MN 56001, USA
Email:Herbert.Heien@i3Magnifi.com

Mezbahur Rahman
Minnesota State University,
Mankato, MN 56001, USA
Email:mezbahur.rahman@mnsu.edu

[Received May 17, 2005; Revised August 4, 2005; Accepted August 16, 2005]

Abstract

The number that has been studied longer than any other number is π, the ratio of the circumference of a circle to its diameter. Starting with Archimedes, the first theoretical analysis of π has grown from 3 or 4 digits of accuracy to billions of digits of accuracy. Here we explore the recent developments in retrieving the digits of π. We extend the statistical analysis regarding randomness of the digits in π. A Discrete version of the Anderson Darling goodness-of-fit test is used along with the Normal and the Chi-square tests in testing randomness of the digits in π.

Keywords and Phrases: BBP Formula, Binomial Probability, Greek Numbers, Quantiles.

AMS Classification: 62-07.

1 Introduction

If one asks a person, "think of a number that has been studied longer than any other number on Earth" they would most likely answer π, the ratio of the circumference of a circle to its diameter. Mathematicians have often wondered how an elementary ratio can have such an incredibly complex structure. This ratio has fascinated mathematicians for many centuries. Many early mathematicians spent days calculating digits of π. Following Ramanujan (see Bailey (1988) and Borwein et. al. (1989) for details), a famous experiment completed by Bailey et.al. (1997) known as the BBP formula (BBP stands for David H. Bailey, Peter B. Borwein and Simon Plouffe) which can give the d-th digit of π without calculating the $d-1$ previous digits. The formula requires little memory and no multiple-precision arithmetic software. This formula is truly unique since before its discovery no mathematician thought that it was possible to calculate the d-th digit without first calculating the $d-1$ preceding digits. This demonstrates that currently there are still some new formulas of π being found.

Currently, we know a lot about π, but questions still arise when thinking about the decimal expansion. For instance, in the decimal expansion of π, is there a place where a thousand consecutive digits are all zeros? There is strong evidence that the digits of π are random, but such questions have not yet been proven precisely. A popular tool used in answering this age old question lies in statistical tests for randomness. Although these statistical tests have been around for a while they are dwarfed when compared with how long people have thought of the decimal expansion of π. This implies that rigorous statistical analysis is fairly new and many possibilities for new testing procedures can still be found.

1.1 Past Ideas of π

The early Babylonians, Egyptians, and Greeks wrestled with π in a slow process of geometry. This tedious process influenced people to come up with faster methods for finding π. Various formulas evolved through the imagination of some of the most prominent mathematicians. These formulas were faster, but again became tedious after very few digits were calculated by hand. Many pioneers of math thought that someone would find the exact ratio of π. The mystery of πs' digit expansion lasted until when Johann Lambert in 1791 showed that π was an irrational number. However, this did not satisfy the ever present hunger to find out more about the decimal expansion of π. The first theoretical calculation seems to have been carried out by Archimedes (287-212 BC). He obtained the approximation $223/71 < \pi < 22/7$. This calculation was found by geometry. For nearly the next two centuries variations of this geometrical scheme were the basis for all high-accuracy calculations of π. Discovered in the 19^{th} century was a simple recursion that can be stated as follows:

$$\begin{array}{l} \text{Set:} \\ a_0 = 2\sqrt{3} \; and \; b_0 = 3 \\ \text{Then define:} \\ a_{n+1} = \frac{2a_n b_n}{a_n + b_n} \\ b_{n+1} = \sqrt{a_{n+1} b_n} \;. \end{array} \tag{1}$$

An important observation to be made is that Archimedes did not have this notational advantage and had to derive all calculation from a geometrical perspective. Even more impressive is the fact that Archimedes did not have the knowledge of our current

Hindu Arabic decimal system. This forced Archimedes to write his calculations using lengthy strings of Greek numbers. His calculations stopped at $a_6 < \pi < b_6$ which is a rather impressive achievement to complete using Greek Numbers manually.

1.2 Present Ideas of π

Presently, the advent of the computer has taken away the tedious manual calculations and has led to fast computer algorithms for present day experimental mathematics. Calculating a 100 digits of π would take a person in the early 1700's months and possibly years, but now it takes less than a second. One of the most fruitful implementations of calculating π was discovered in 1914 by the mathematician Ramanujan in his work with modular equations. His famous notebook contains many theorems and formulas with no rigorous proofs, but very few of these formulas have been flawed. After his death in 1920, efforts were made to demystify his simple yet complex theorems and identities. His 1914 paper "Modular Equations and Approximations to π" gives the famous equation:

$$\frac{1}{\pi} = \frac{\sqrt{8}}{9801} \sum_{n=0}^{\infty} \frac{(4n!)}{(n!)^4} \frac{[1103 + 26390n]}{396^{4n}}. \tag{2}$$

Equation (2) rests on a modular identity of order 58 and appears, like most of his work, without proof. This formula also shows that Ramanujan was the first to find a connection between the transformation theory for elliptic integrals and the rapid convergence to π.

In the 1980's Chudnovsky and Chudnovsky (1988) found the following variation of (2),

$$\frac{1}{\pi} = 12 \sum_{k=0}^{\infty} \frac{(-1)^k (6k)! (13591409 + 545140134k)}{(3k)! (k!)^3 640320^{3k+3/2}}. \tag{3}$$

Equation (3) is implemented in the software of Mathematica for calculating π.

The development of computers in the 1950's was clearly the turning point in numerical mathematics. No longer did it take mathematicians months or even years to calculate a thousand decimal places to π. Computers also give mathematicians the freedom to search massive vectors, such as $X = (x_1, x_2, \cdots, x_n)$ made up of real or complex numbers, to detect an integer relation. Currently, the best known algorithm for this is the PSLQ algorithm. The name "PSLQ" comes from its usage of the partial sum of squares vector and a LQ (lower-diagonal-orthogonal) matrix factorization. Bailey (2000) gives more information on integer relation detection algorithms. Bailey, et.al. (1997) performed a computer experiment using the PSLQ algorithm and found an extraordinary property about the log(2), where 'log' stands for logarithm. The log(2) has an uncomplicated infinite-sum-formula that can reveal the millionth binary digit of log(2) with no need to compute the 999,999 preceding digits. They had an inclination that π may also contain this same property. By using the PSLQ algorithm and several inspired guesses they found the BBP formula:

$$\pi = \sum_{n=0}^{\infty} \left(\frac{4}{8n+1} - \frac{2}{8n+4} - \frac{1}{8n+5} - \frac{1}{8n+6} \right) \left(\frac{1}{16} \right)^n . \tag{4}$$

This infinite sum is not as quick to converge to π, but it is an essential identity for finding, say, the d-th digit of π given in base 16. The identity is of a special form,

$$\sum_{k=1}^{\infty} \frac{p(k)}{b^{ck} q(k)}, \tag{5}$$

where p and q are polynomials with integer coefficients and c is a positive integer. Bailey, et.al. (1997)showed that this formula can be computed in polynomially logarithmic space with either polynomial or linear time. It is denoted respectfully by SC or SC*. The class denoted as SC has space = $\log^{O(1)}(d)$ and time $d^{O(1)}$, the class SC* has space = $\log^{O(1)}(d)$ and time $O(d \log^{O(1)}(d))$. It is not known whether division is possible in SC, nor is it known if base changes are possible in SC. Also, it is not known whether multiplication is possible in SC*. This problem will not be discussed in this paper, but more details are in: Borwein (1988); Brent (1974); Cook (1985); Crandall and Buhler (1995); Knuth (1981). The proof of the identity (4) is fairly straightforward, and can be found in Bailey, et.al. (1997). Identity (4) can be coded into Mathematica to give the d-th hex digit of π in base 16. This code and the general method for finding similar such formulas can be found in Adamchik and Wagon (1996).

Shortly after the announcement of the BBP formula, colleagues of Borwein, Bailey, and Plouffe began to discover similar formulas for π such as:

$$\pi = \sum_{i=0}^{\infty} \frac{(-1)^i}{4^i} \left(\frac{2}{4i+1} + \frac{2}{4i+2} + \frac{1}{4i+3} \right) \tag{6}$$

$$\pi\sqrt{3} = \frac{9}{32} \sum_{k=0}^{\infty} \frac{1}{64^k} \left(\frac{16}{6k+1} - \frac{8}{6k+2} - \frac{2}{6k+4} - \frac{1}{6k+5} \right) \tag{7}$$

$$\pi^2 = \frac{1}{8} \sum_{k=0}^{\infty} \frac{1}{64^k} \left(\frac{144}{(6k+1)^2} - \frac{216}{(6k+2)^2} - \frac{72}{(6k+3)^2} - \frac{54}{(6k+4)^2} + \frac{9}{(6k+5)^2} \right) \tag{8}$$

$$\pi^2 = \frac{2}{27} \sum_{k=0}^{\infty} \frac{1}{729^k} \left(\begin{array}{l} \frac{243}{(12k+1)^2} - \frac{405}{(12k+2)^2} - \frac{81}{(12k+4)^2} - \frac{27}{(12k+5)^2} \\ -\frac{72}{(12k+6)^2} - \frac{9}{(12k+7)^2} - \frac{9}{(12k+8)^2} - \frac{5}{(12k+10)^2} + \frac{1}{(12k+11)^2} \end{array} \right). \tag{9}$$

The BBP formula may hold the key in answering the question of whether or not π is truly random. It may be that π is chaotic and acts randomly but is not truly random. Currently, the BBP formula is being studied extensively to see if there is a connection between π and chaotic behavior.

The BBP formula is a perfect example of how the computer can help find new formulas that were once thought of as nonexistent. The experimenters who found the BBP formula ran the PSLQ algorithm on a super computer for several days at a time. Each time they ran the algorithm they made educated guesses of what parameters to implement. With this in mind, it is not surprising why mathematicians avoided this type of algorithm before the advent of the computer. Why would anyone spend years exhaustively searching for something based on an educated guess? Thanks to the computer, a risk like this is cheap and not as time consuming.

2 Statistical Analysis of the Digits in π

The incentive for the computation of π is to answer the question of whether or not the digits of π are truly random. Before Lambert's proof in 1766 that π was irrational, there was great interest in examining whether or not its decimal expansion ultimately repeats. π is known to be irrational but there still remains an unanswered question. Is the digit expansion of π statistically random?

Statistically, if the digits of π are truly random then a sequence, or string, of n-digits occur with a limiting relative frequency 10^{-n}. The evidence gathered through numerous computer experiments on the decimal expansion of π strongly suggests that the decimal expansion is statistically random. However, the proof still remains unfound.

The constant π is not random and its digits are from a deterministic sequence. The mathematicians failed to justify that π is not an irrational number, the statisticians failed to show that the appearances of the digits of π are not random. In literature, randomness of π's digits are described by many authors, here we give some references of recent works, for example, Peterson (2001), and Preuss (2001), to name a few.

2.1 Normal Approximation to Binomial Probability

In this paper the use of Mathematica and Matlab provided n-long string frequency counts for analysis of the first million digits in π. The analysis was completed for n-digit strings ranging from 1 through 6. Table 1 describes the frequency count, Deviation, Z-score, and gives the P-value for each of the one-digit strings. The Z-scores are computed as the deviation from the mean ($np = 1,000,000(1/10) = 100,000$) divided by the standard deviation $\sqrt{1,000,000(1/10)(9/10)} = 300$. These statistics should follow an approximate normal distribution with mean zero and variance one.

Table 1				
One-Digit Statistics				
Digit	Count	Dev.	Z-sc	P-value
0	99959	-41	-0.14	0.8913
1	99757	-243	-0.81	0.4179
2	100026	26	0.09	0.9309
3	100229	229	0.76	0.4453
4	100230	230	0.77	0.4433
5	100358	358	1.19	0.2327
6	99548	-452	-1.51	0.1319
7	99800	-200	-0.67	0.5050
8	99985	-15	-0.05	0.9601
9	100106	106	0.35	0.7238

In Table 1, all the p-values are high and hence all the digits are random. Table 2 holds the same information as Table 1 for the two-digit strings. For two-digit strings the mean is 10,000 occurrences and the standard deviation is 99.4987.

Table 2				
Significant Strings From Two Digit Strings				
String	Count	Dev.	Z-sc	P-value
12	9721	-279	-2.80	0.0050
27	10224	224	2.25	0.0244
55	10232	232	2.33	0.0197
94	10239	239	2.40	0.0163

Based on Table 2, only 4% of the two-digit strings appear significantly different than the expected frequency under a 5% level of significance. However, for simultaneous testing α (the level of significance) should be lower, or more precisely $\alpha/m = 0.05/100 = 0.0005$, where m is the number of simultaneous tests (Weisberg 1985). No p-values are less than 0.0005, therefore the two digit strings are random. In Table 2 the lower p-values were displayed only to point out the strings with smaller p-values.

Table 3 displays the strings for the smaller p-values, but none of these p-values are smaller than $0.05/1000 = 0.00005$ hence the three digit strings are random. Similar computations could be completed for higher n-digit strings, but the focus is turned

Table 3				
Significant Strings From Three Digit Strings				
String	Count	Dev.	Z-sc	P-value
013	929	-71	-2.25	0.0247
027	1075	75	2.37	0.0176
050	1081	81	2.56	0.0104
067	898	-102	-3.23	0.0013
075	923	-77	-2.44	0.0149
077	921	-79	-2.50	0.0124
126	911	-89	-2.82	0.0049
136	1078	78	2.47	0.0136
160	1089	89	2.82	0.0049
166	926	-74	-2.34	0.0192
172	919	-81	-2.56	0.0104
195	1077	77	2.44	0.0148
238	927	-73	-2.31	0.0209
244	1073	73	2.31	0.0209
270	1073	73	2.31	0.0209
451	915	-85	-2.69	0.0072
453	1089	89	2.82	0.0049
523	910	-90	-2.845	0.0044
654	1092	92	2.92	0.0036
685	923	-77	-2.44	0.0148
735	1078	78	2.47	0.0136
750	919	-81	-2.56	0.0104
934	1086	86	2.72	0.0065
968	925	-75	-2.37	0.0176

towards combined tests such as the chi-square goodness-of-fit test.

2.2 Chi-Square Goodness of Fit Test

A more appropriate statistical procedure for testing whether or not n-digit strings follow a discrete uniform distribution is the chi-square test. The chi-square statistic for k observations $X_1,\ X_2,\ ...,\ X_k$ is defined as:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{10}$$

where E_i is the expected frequency of the random variable X_i, and O_i is the observed frequency of X_i. Here $k = 10^n$ and $E_i = d \cdot 10^{-n}$ for all i, where $d = 1,000,000$. The statistic χ^2 has a chi-square distrution with $k - 1$ degrees of freedom. The results of the χ^2 analysis are shown in Table 4.

Table 4

String Length	Chi-square	df.	P-value
1	5.50676	9	0.7881
2	94.2672	99	0.6157
3	958.121	999	0.8192
4	9978.53	9999	0.5557
5	100378.5	99999	0.1979
6	1001628	999999	0.1247

Table 4 confirms the usual conclusion that the n-digit strings in π appear to be uniformly distributed. In the analysis of the one-digit string, there is no significant difference between the observed frequency and the expected mean frequency. The same conclusion can be found for the other n-digit strings given in Table 4. It should be noted that the p-values for five-digit and six-digit strings are fairly close to 0.05 compared with the other p-values. This is due to the large amount of combinations that can arise from five-digit and six-digit strings. If the workstation used in calculating π could calculate a substantially larger amount of digits, then the p-values would become larger and less significant. The conclusion is that the digits in π appear to be uniformly distributed, and hence random.

3 Anderson-Darling Goodness of Fit Test

Another goodness of fit test is the Anderson-Darling test (1954). The test will consider a sample that has been drawn from a population with a specified continuous cumulative distribution function $F(x)$. The test procedure proposed by Anderson-Darling is the following: Let $x_1, x_2, \cdots x_n$ be n ordered from the smallest to the largest observations from the sample, and let $u_i = F(x_i)$. Then compute:

$$W_n^2 = -n - \frac{1}{n}\sum_{j=1}^{n}(2j-1)\left[\ln u_i + \ln(1-u_{n-j+1})\right] . \tag{11}$$

If the number is too large, the hypothesis that the sample is from the specified distribution is to be rejected. This test can be used when a researcher wishes to reject the hypothesis whenever the true distribution differs from the claimed distribution. The quantiles for W_n are given in Anderson and Darling (1954).

Several test procedures are based on comparing the specified cumulative distribution function $F(x)$ with its empirical cumulative distribution function $F_n(x)$. Anderson-Darling suggested the use of:

$$W_n^2 = n\int_{-\infty}^{\infty}\left[F_n(x) - F(x)\right]^2 \Psi(F(x))\, dF(x), \tag{12}$$

where $\Psi(u)$ is some nonnegative weight function chosen by the researcher to emphasize the variation of $F_n(x) - F(x)$, where the test is preferred to have sensitivity. Formula (11) is obtained by writing (12) as:

$$\frac{1}{n}W_n^2 = \int_{-\infty}^{x_1}\frac{F^2(x)}{F(x)[1-F(x)]}\, dF(x) + \int_{x_1}^{x_2}\frac{[F_n(x)-F(x)]^2}{F(x)[1-F(x)]}\, dF(x) + \cdots + \int_{x_1}^{x_2}\frac{[1-F(x)]^2}{F(x)[1-F(x)]}\, dF(x) .$$

Rahman and Chakrobartty (2004) showed that the Anderson-Darling test is more powerful for the goodness-of-fit testing for the continuous uniform distribution compared to different versions of the Cui-square tests, the Cramer von-misses test, and the Watson test.

3.1 Anderson-Darling Test for Discrete Random Variable

In conducting a test to determine the uniformity of the digits in π, a continuous cumulative distribution function is no longer applicable. The test statistic must be discretized, or changed in a way to accommodate discrete data. The sum of W_n^2 for π where $\Psi(u)$ is a constant function is completed for each discrete value using a test statistic similar to (12),

$$W_d^2 = n\sum_{i=1}^{d}\left[F_d(x) - F(x)\right]^2 f(x) \tag{13}$$

where d stands for the number of discrete values, $F_d(x)$ is the empirical cumulative distribution function, $F(x)$ is the claimed c.d.f. (cumulative distribution function), and $f(x)$ is the claimed p.m.f. (probability mass function).

3.1.1 Quantiles of W_d^2

The computation of the W_d^2 quantiles was done using a simulation in the Matlab software. Since the p.m.f. will always be a constant and hence omitted during simulation; however, in practice the p.m.f. should be used in the calculation of the quantiles for a non-uniform null distribution. One hundred thousand random samples were generated for sample sizes of 20, 30, 40, 50(50)1000 from a discrete uniform distribution with possible single digit values of $\{0, 1, 2, 3, ..., 9\}$. These simulated values were then ordered and placed into their respective quantile values to compute the P-values of the tests.

3.1.2 Application: Checking Randomness for the Digits in π

Here the W_d^2 statistic is calculated for π as described above. However, since the simulated test statistic is dependent upon the sample size new quantiles for the test statistic will have to be computed for analyzing the digits in π. For this reason the random number generator in Matlab is utilized, and the quantiles are computed as above with the respective sample sizes. Table 5 shows the the W_d^2 statistic for n-digit strings one through six. In all cases, the results show that the digits in π are uniformly distributed.

Table 5

String Length	W_d^2
1	0.53636
2	7.81901
3	79.12841
4	793.24277
5	7945.20887
6	79275.77655

4 Concluding Remarks and Future Research

One can notice that all three tests; normal, chi-square, and Anderson-Darling give the same conclusions about the n-digit string distribution in π. One would probably find that the larger the sample size the more uniformly the n-digit strings would be distributed. The chi-square test is overall more powerful than the Anderson-Darling test. However, since the sample size taken is large, the Anderson-Darling test gave reasonable results. In all three cases it was shown that the digits in π are random.

The Anderson-Darling statistic holds a lot of information about the data being sampled and therefore, should give better results. However, as shown in this paper, the Anderson-Darling test is not as powerful as the chi-square test when analyzing discrete data. It has been shown that for continuous data, the Anderson-Darling test is a more powerful test than the standard chi-square test (Rahman and Chakrobartty (2004)). Given this information, it can be said that more research needs to be completed for discretizing the Anderson-Darling test.

The age old question of whether the digit expansion of π is random, or only appears random, is still unanswered. However, thanks to the BBP formula (4) this question may be close to being answered. Before the discovery of this formula it was thought that no such formula could exist for π. Thanks to the BBP formula it is now suspected that π could in fact be chaotic and not random at all. This would fit all conclusions which arise from any statistical tests, since anything following a chaotic pattern will most likely appear to be random. Perhaps, better statistical tests will arise that help conclude whether data follows a chaotic pattern.

References

[1] Anderson, T.W. and Darling, D.A. (1954). *A Test of Goodness of Fit*, Journal of the American Statistical Association, **49**(268), 765-769.

[2] Adamchik, V. and Wagon, S. (1996). π *A 2000-Year Search Changes Direction*, Mathematica in Education and Research, **5**(1), 11-19.

[3] Bailey, D.H. (2000). *Integer relation detection,* Computing in Science and Engineering, 24-28.

[4] Bailey, D.H., Borwein, P.B. and Plouffe, S. (1997). *On the rapid computation of various polylogarithmic constants*, Mathematics of Computation, **66**(218), 903-913.

[5] Bailey, D.H. (1988). *The Computation of* π *to 29,360,000 decimal digits using Borweins'quartically convergent algorithm*, Mathematics of Computation, **50**, 283-96.

[6] Borwein, J.M., Borwein, P.B. and Bailey, D.H. (1989). *Ramanujan, Modular Equations, and Approximations to Pi or How to Compute One Billion Digits of Pi*, The American Mathematical Monthly, **96**(3), 201-219.

[7] Borwein, J.M. and Borwein, P.B. (1988). *On the complexity of familiar functions and numbers*, SIAM Review, **30**, 589-601.

[8] Brent, R. (1974). *The parallel evaluation of general arithmetic expressions*, Journal of the Association for computing machinery, **21**, 201-206.

[9] Chudnovsky, D.V. and Chudnovsky, G.V. (1988). *Approximations and complex multiplication according to Ramanujan, in Ramanujan Revisited*, Academic Press Inc., Boston, 375-396 and 468-472.

[10] Cook, S. (1985). *A taxonomy of problems with fast parallel algorithms*, Information and Control, **64**, 2-22.

[11] Crandall, R. and Buhler, J. (1995). *On the evaluation of Euler sums*, Experimental Mathematics, **3**, 275-285.

[12] Klarreich, E. (April 24, 2004). *Computer experiments are transforming mathematics*, Science News, Math Lab, **165**, 266-268.

[13] Knuth, D.E. (1981). *The Art of Computer Programming. Seminumerical Algorithms*, Volume 2, Addison-Wesley, Reading, MA.

[14] Peterson, I. (2001). *Mathematicians tackle the seeming randomness of π's digits,* Science News, **160** (9), 136.

[15] Preuss, P. (2001). *Are the digits of π random?* Lab researcher may hold the key. Lawrence Berkeley National Laboratory press release, July 2003. Available at http://www.lbl.gov/Science-Articles/Awchive/pi-random.html.

[16] Rahman, M. and Chakrobartty, S. (2004). *Tests for Uniformity : A Conparative Study,* Journal of the Korean Data & Information Science Society, **15**(1), 211-218.