Searching for primes in the digits of π

D. Byatt, M. L. Dalrymple and R. M. Turner

Department of Mathematics and Statistics
University of Canterbury
Private Bag 4800
Christchurch, New Zealand

Report Number: UCDMS2003/23 December 2003

Keywords: Pi, primes, random numbers, mathematical constants

To appear: Computers and Mathematics with Applications, 2004

SEARCHING FOR PRIMES IN THE DIGITS OF π

D. BYATT, M. L. DALRYMPLE, AND R. M. TURNER

ABSTRACT. Many people are fascinated by π. Vast amounts of human and computer resources have been spent producing billions of its digits. Similarly, many people are intrigued by primes. How many primes are there within the digits of π? How many can we expect to find? We present the results of some computer searches and develop a theory to predict how many of these primes extensive computer searches are likely to find.

1. INTRODUCTION

Certain mathematical constants hold special appeal for a great many people, probably none more so than π. Over the years considerable time and effort has been spent developing new formulae for, and calculating ever more digits of π. Once of purely academic interest large-scale calculations of π have gained increasing respectability. Nowadays such computations are routinely used as a quality control check for new super-computers for example. Similar computational effort has also been spent on searches for certain types of prime numbers. Primes also hold special appeal for many people. This article was motivated in part by an interest in both these areas. We present the results of some computer searches for primes within the digits of π (and a few other well-known constants) and develop a theory for calculating the expected number of primes to be found by such searches.

The procedure for discovering these so-called "π-primes" is straightforward: simply check the primality of $\pi_k = \lfloor \pi \cdot 10^{k-1} \rfloor$ (where $\lfloor x \rfloor$ represents the integer part of x), for each $k \in \mathbb{N}$. For example, $\pi_1 = 3$ is prime, $\pi_2 = 31$ is also prime, but $\pi_3 = 314$ is composite. The first four π-primes are π_1, π_2, π_6 and π_{38}.

It is important to note that many modern computer-implemented primality tests are probabilistic rather than deterministic in nature. As such they do not guarantee a number that passes such a test is prime, rather, it is (very) probably prime (see for example [7], and the references therein). Such numbers are sometimes called *pseudo-primes*. Conversely however, numbers which fail pseudo-prime tests are composite. Probabilistic primality tests are used because they tend to be significantly faster than deterministic tests and, for most purposes, just as good. Details of a new deterministic prime test which runs in polynomial time can be found in [1]. Similarly, the "random" numbers generated by a computer (as in Section 3) are calculated by a deterministic process and so are not truly random. Consequently, they are often referred to as *pseudo-random*.

Before discussing the number of π-primes, a theory for the expected number of primes when positive integers are chosen at random is presented. After some numerical experiments, this theory is extended to predict the number of π-primes. We then show that these results apply to just about any constant, not just π. The resulting theory is used to predict the number of these primes that extensive computer searches are likely to find. The results of some computer searches are presented for comparison with the predicted values.

2. Primes at random — theory

If a natural number m has exactly k digits[1] then $10^{k-1} \leqslant m < 10^k$. Let $N(k)$ be the number of positive integers with exactly k digits so that

$$N(k) = 9 \cdot 10^{k-1}.$$

Suppose $\Pi(m)$ represents the number of primes[2] less than m. The number of primes with exactly k digits, $P(k)$, is given by

$$P(k) = \Pi(10^k) - \Pi(10^{k-1}). \tag{1}$$

The probability that m, a randomly chosen positive integer with exactly k digits, is prime is therefore

$$p(k) = \frac{P(k)}{N(k)}.$$

Values of $P(k)$ and $p(k)$ for $k \in \{1, 2, \ldots, 22\}$ are presented in Table 1 [5].

As m increases, it becomes increasingly difficult (time consuming) to calculate $\Pi(m)$. However there are several well known approximating functions. For example,

$$\Pi_1(m) = \frac{m}{\log m - 1.08366} \tag{2}$$

is the approximation suggested by Legendre in 1798 and

$$\Pi_2(m) = \frac{m}{\log m} \tag{3}$$

is the approximation, perhaps first considered by Gauss in 1791. In both equations (2) and (3) $\log m$ represents the natural logarithm of m. In 1896 Hadamard and de la Vallée-Poussin (independently) proved that Π_2 is an arbitrarily good approximation for Π, in the sense that, $\Pi_2(m)/\Pi(m) \to 1$ as $m \to \infty$. This subsequently became known as the

[1]When referring to a k-digit integer, only the significant digits are counted. For example, 001230 is a four digit integer.

[2]The usual notation used by number theorists is $\pi(m)$, however $\Pi(m)$ will be used here to avoid confusion with the constant π and the integers π_k.

k	$P(k)$	$p(k)$
1	4	0.444
2	21	0.233
3	143	0.159
4	1061	0.118
5	8363	0.093
6	68906	0.077
7	586081	0.065
8	5096876	0.057
9	45086079	0.050
10	404204977	0.045
11	3663002302	0.041
12	33489857205	0.037
13	308457624821	0.034
14	2858876213963	0.032
15	26639628671867	0.030
16	249393770611256	0.028
17	2344318816620308	0.026
18	22116397130086627	0.025
19	209317712988603747	0.023
20	1986761935284574233	0.022
21	18906449883457813088	0.021
22	180340017203297174362	0.020

Table 1. Number of primes with k-digits ($P(k)$), and the probability they are chosen at random ($p(k)$).

Prime Number Theorem. Note that the constant 1.08366 in equation (2) was based on Legendre's limited table of values for $\Pi(m)$; de la Vallée-Poussin showed that 1 is the best value to use for large m.

For the values of k given in Table 1, Π_1 is a very accurate approximation of Π (within about 0.5%) when $k > 4$ and Π_2 is within 3% of Π when k is larger than about 15.

The probability that a randomly selected positive integer with exactly k digits is prime, $p(k)$, can be approximated by either

$$p_1(k) = \frac{P_1(k)}{N(k)} \quad \text{or} \quad p_2(k) = \frac{P_2(k)}{N(k)}$$

where P_1 and P_2 are defined as in equation (1). The calculation of p_2 can be further simplified. Since

$$p_2(k) = \frac{P_2(k)}{N(k)} = \frac{1}{9 \log 10}\left(\frac{10}{k} - \frac{1}{k-1}\right),$$

$p_2(k)$ can be approximated by

$$p_3(k) = \frac{1}{k \log 10},$$

and the accuracy of the approximation improves as k increases.

Now suppose that n positive integers are chosen at random. How many primes would be expected amongst the n numbers? Since each number is chosen at random the expected value for the number of primes is simply the sum of the probabilities that the individual numbers are prime. If the jth number has k_j digits then the expected number of primes is $\sum_{j=1}^{n} p(k_j)$. If Π (and therefore p) is unavailable, the expected number of primes can be approximated by using either p_1, p_2 or p_3 in place of p.

For the remainder of this section, the special case where n positive integers are chosen at random so that the kth such number has k digits (similar to the π-prime situation) is discussed.

Suppose X_k is the Bernoulli random variable representing whether a single randomly chosen k-digit positive integer is prime ($X_k = 1$) or composite ($X_k = 0$) so that

$$X_k = \begin{cases} 1 & \text{with probability } p(k) \\ 0 & \text{with probability } 1 - p(k). \end{cases}$$

Define $T = \sum_{k=1}^{n} X_k$ to be the random variable representing the total number of primes found when randomly choosing n positive integers so that the kth such number has k digits. The expected number of primes $E(T) = \sum_{k=1}^{n} E(X_k)$ where $E(X_k) = \sum_{j=0}^{1} \Pr(X_k = j) \cdot j$, hence

$$E(T) = \sum_{k=1}^{n} p(k). \tag{4}$$

Since Table 1 only goes as far as 22 digits, equation (4) can be approximated using p_1, p_2 or p_3 whenever $n > 22$. For example, using p_3 gives

$$\begin{aligned} E(T) &\approx \sum_{k=1}^{22} p(k) + \sum_{k=23}^{n} p_3(k) \\ &= 1.6791 + \sum_{k=23}^{n} \frac{1}{k \log 10}. \end{aligned} \tag{5}$$

Figure 1 shows the ratios p_1/p, p_2/p and p_3/p (as percentages) for $k \in \{1, 2, \ldots, 22\}$. Once again, p_1 is a good approximation for p when $k > 4$. Furthermore, the simplification leading to p_3 reduces some of the error in approximation p_2. Although the approximations for p may be reasonably accurate, summing a series of such approximations may lead to the accumulation of a large error. However, Figure 2 shows that the expected values obtained by each of the approximations are

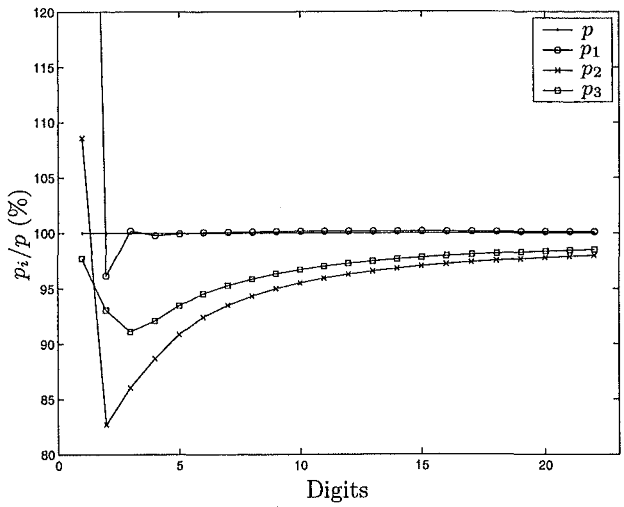

Figure 1. Relative accuracy of approximate probabilities.

sufficiently close to the theoretical value for all practical purposes. In fact, since the difference $p_3(k) - p_2(k)$ is

$$\delta_e(k) = \frac{1}{9 \log 10 \cdot k(k-1)},$$

the total accumulated error in using p_3 instead of p_2 whenever $k > 22$, for the calculation of the expected value $E(T)$ in equation (5), is bounded above by

$$\sum_{k=23}^{\infty} \delta_e(k) = \frac{1}{198 \log 10} \approx 0.0022.$$

Furthermore, the approximation p_3 allows the expected value to be calculated relatively easily for very large numbers of digits (as in Table 2).

Numerical experiments using equation (5) show that

$$E(T) - 0.33 \approx \lfloor E(T) \rfloor = j \tag{6}$$

whenever $n = 10^j$ for some $j \in \mathbb{N}$. In order to find at least t primes, it would be expected that $E(T) > t - 0.5$. If the conditions for equation (6) are met, this occurs whenever $n > 1.5 \times 10^{t-1}$. Note that n is the number of digits in the final number tested for primality, therefore this final number will be greater than $10^{10^{t-1}}$.

Let $\sigma^2(T)$ be the population variance of the total number of primes found when randomly choosing n positive integers so that the kth such number has k digits. Then $\sigma^2(T) = \sum_{k=1}^{n} \sigma^2(X_k)$, where

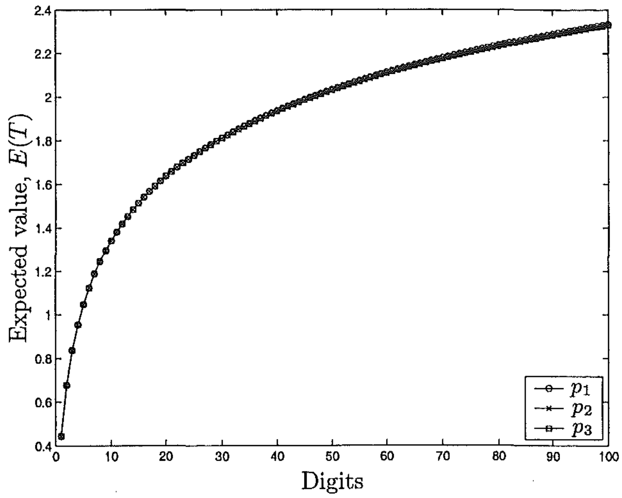

Figure 2. Expected value calculated using each of the approximations (exact up to 22 digits).

$\sigma^2(X_k) = E(X_k^2) - E(X_k)^2$, so that

$$\sigma^2(T) = \sum_{k=1}^{n} p(k)(1 - p(k)). \tag{7}$$

Equation (7) can be approximated using p_1, p_2 or p_3 whenever $n > 22$. Using p_3, for example gives

$$\begin{aligned}
\sigma^2(T) &\approx \sum_{k=1}^{22} p(k)(1-p(k)) + \sum_{k=23}^{n} p_3(k)(1-p_3(k)) \\
&= 1.3514 + \sum_{k=23}^{n} \frac{k\log 10 - 1}{(k\log 10)^2} \\
&\approx 1.3514 + \sum_{k=23}^{n} p_3(k) \\
&\approx 1.3514 + E(T) - \sum_{k=1}^{22} p(k) \\
&\approx E(T) - 0.3276.
\end{aligned} \tag{8}$$

Since the difference between $p_3(k)$ and $p_3(k)(1 - p_3(k))$ is

$$\delta_{\mathrm{v}}(k) = \frac{1}{k^2 \log^2 10},$$

Digits	$E(T)$	$\lfloor E(T) \rfloor$	$E(T) - 0.3276$	$\sigma^2(T)$
10^1	1.3	1	1.01	1.02
10^2	2.3	2	2.00	1.99
10^3	3.3	3	3.00	2.99
10^4	4.3	4	4.00	3.99
10^5	5.3	5	5.00	4.99
10^6	6.3	6	6.00	5.99

Table 2. Comparison of the expected value $E(T)$, and the variance $\sigma^2(T)$.

the total accumulated error in using $p_3(k)$ instead of $p_3(k)(1 - p_3(k))$ whenever $k > 22$, for the calculation of the variance $\sigma^2(T)$ in equation (8), is bounded above by

$$\sum_{k=23}^{\infty} \delta_{\text{v}}(k) = \frac{1}{\log^2 10}\left(\frac{\pi^2}{6} - \sum_{k=1}^{22} \frac{1}{k^2}\right) \approx 0.0084.$$

Some numerical values for the expected value $E(T)$, $\lfloor E(T) \rfloor$, $E(T) - 0.3276$ and the variance $\sigma^2(T)$ are presented in Table 2.

3. Primes at random — practice

We now discuss the number of primes found when k-digit positive integers were tested for primality for $k \in \{1, 2, \ldots, n\}$. Two methods are used for generating the numbers. In the *Random Method*, each k-digit number is chosen at random from all possible k-digit numbers. The *Construction Method* generates successive numbers by appending a randomly chosen digit (0–9) at each iteration. Numbers generated by the Construction Method are no longer chosen at random, nor are they independent — the number generated at the next iteration is highly dependent on the current one.

The theory developed in the previous section describes numbers generated by the Random Method, however numbers generated when searching for π-primes are more similar to those of the Construction Method. Computer experiments were performed to see how closely numbers generated by the Construction Method are modelled by the Random Method theory. In each experiment, k-digit positive integers were tested for primality for $k \in \{1, 2, \ldots, n\}$ where $n \in \{10, 100, 1000\}$. The experiments were repeated 10000 times for $n \in \{10, 100\}$ but only 100 times for $n = 1000$ due to the dramatic increase in the amount of computer time required. All of the computer experiments were performed using Maple 7 on a Sun Enterprise 450 machine.

The results for the Random and Construction Methods are presented in Tables 3 and 4, where n is the number of digits in the final number

n	n_r	$E(T)$	$\bar{x}_r$	$\sigma^2(T)$	s_r	95% CI
10	10000	1.34	1.35	1.01	1.01	(1.33,1.37)
100	10000	2.33	2.32	1.41	1.41	(2.30,2.35)
1000	100	3.33	3.54	1.73	1.63	(3.2,3.9)

Table 3. Results for the Random Method.

n	n_c	$E(T)$	$\bar{x}_c$	$\sigma^2(T)$	s_c	95% CI
10	10000	1.34	1.34	1.01	1.02	(1.32,1.36)
100	10000	2.33	2.32	1.41	1.43	(2.29, 2.34)
1000	100	3.33	3.57	1.73	1.83	(3.2,3.9)

Table 4. Results for the Construction Method.

tested, $E(T)$ and $\sigma^2(T)$ are the theoretical expected value and variance, $n_{r,c}$, $\bar{x}_{r,c}$ and $s_{r,c}$ are the number of times the experiments were repeated (sample size), the sample mean and the sample variance for the Random and Construction Methods. By the Central Limit Theorem the sample means are normally distributed. Tables 3 and 4 include 95% confidence intervals about each of the sample means.

In every case, the Construction Method performed similarly to the Random Method. To determine if the means for the two methods (μ_r and μ_c) are significantly different, the null hypothesis $H_0\colon \mu_r = \mu_c$ was tested against the alternative hypothesis $H_a\colon \mu_r \neq \mu_c$ for each $n \in \{10, 100, 1000\}$. The appropriate test statistic for examining the difference between the means is

$$z = \frac{\bar{x}_r - \bar{x}_c}{\left(\frac{s_r^2}{n_r} + \frac{s_c^2}{n_c}\right)^{\frac{1}{2}}}.$$

The corresponding p-values are $p_{10} = 0.6912$, $p_{100} = 0.6837$ and $p_{1000} = 0.2212$. Hence there is insufficient evidence to reject the null hypotheses. With such large p-values we can be confident that the theory developed in Section 2 models the number of primes for numbers generated by both the Random and Construction Methods.

4. Results

When searching for π-primes the successive digits are not appended at random — they must come from the digits of π. However, as the known digits of π pass tests for randomness [2, 4, 6], the successive digits *appear* to have been chosen at random. As such, the theory presented in the previous section is expected to model the distribution of primes within the digits of π. Furthermore, it is also expected to apply to any real number whose successive digits appear random — which is just about all of them [3].

Digits	$E(T)$	90%	π	e	$\sqrt{2}$	γ	e^π	π^e	i^i	ϕ
10^1	1.3	$(0,3)$	3	3	0	2	2	2	1	1
10^2	2.3	$(0,5)$	4	4	2	3	3	2	1	2
10^3	3.3	$(1,7)$	4	4	3	4	6	2	2	4
10^4	4.3	$(1,8)$								
10^5	5.3	$(2,9)$								
10^6	6.3	$(3,11)$								

Table 5. Numbers of primes within a selection of constants.

As the expected value $E(T)$ and the variance $\sigma^2(T)$ are approximately equal we use the Poisson distribution which is characterised by a single parameter λ, and is often used to model rare events, to model the expected number of π-primes. The parameter $\lambda = E(T) \approx \sigma^2(T)$, in this case, represents the expected number of primes per n-digit search.

The results of searches for primes in the digits of some well known constants are presented in Table 5. For completeness, the constants e, i, γ and the so-called "golden ratio" ϕ are defined as

$$\begin{aligned} e &= \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \\ i &= \sqrt{-1} \\ \gamma &= \lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - \log n\right) \\ \phi &= \frac{\sqrt{5}+1}{2}. \end{aligned}$$

For each of the constants in Table 5, the number of primes found when searching from one digit, up to the number of digits in the *Digits* column are tabulated, as is a 90% confidence interval based on a Poisson distribution with parameter $\lambda = E(T)$ (that is, there is a 5% chance of finding fewer primes than the lower limit of the confidence interval and a 5% chance of finding more primes than the upper limit). The results presented in Table 5 are by no means exhaustive, in fact some of the entries have been left blank for you to fill in yourself when you have a bit of spare time.

Based on a Poisson distribution with parameter $\lambda = E(T)$, and given that it is infeasible to determine the primality of numbers with more than about 10^4 digits, there is only a 5% chance of finding more than 8 primes. If it is possible to test numbers with up to 10^6 digits, then there is a 5% chance of finding more than 11 primes. With the rather optimistic limit of up to 10^9 digits there is a 5% chance of finding more than 15 primes.

It is worth pointing out that it is not necessary to run a (computationally expensive) primality test at every iteration. Half the numbers generated are expected to be even and so need not be tested. Similarly, one third of the numbers are expected to be divisible by three. Hence performing a primality test only on the numbers that are equivalent to ± 1 modulo 6 removes the need to test about two-thirds of all the numbers generated. For example, if 1000 numbers are generated, only about 300 need to be tested for primality.

5. Conclusion

Although the expected number of primes, $E(T)$, is unbounded, we conjecture (despite the notorious unreliability of such predictions) that there is little point spending vast amounts of computer time searching for say, the next 10 π-primes. Although having said that, maybe it would be possible to find just *one* more ...

References

[1] M. Agrawal, N. Kayal, and N. Saxena. Primes is in P. Research report, Department of Computer Science and Engineering, Indian Institute of Technology, Kanpur, India, August 2002. Submitted to Annals of Mathematics.

[2] D. H. Bailey and R. E. Crandall. On the random character of fundamental constant expansions. *Experimental Mathematics*, 10(2):175–190, 2001.

[3] C. Calude. *Information and randomness: An algorithmic perspective.* Monographs in Theoretical Computer Science. Springer-Verlag, Berlin, 2nd edition, 2002. Revised and extended.

[4] G. H. Choe and D. H. Kim. Entropy and the randomness of the digits of π. *Communications of the Korean Mathematical Society*, 15(4):683–689, 2000.

[5] X. Gourdon and P. Sebah. Counting the number of primes, 2001. `http://numbers.computation.free.fr/Constants/Primes/countingPrimes.html`.

[6] B. R. Johnson and D. J. Leeming. A study of the digits of π, e and certain other irrational numbers. *Sankhya, The Indian Journal of Statistics*, 52(2):183–189, 1990. Series B.

[7] S. Y. Yan. Primality testing of large numbers in Maple. *Computers and Mathematics with Applications*, 29(12):1–8, 1995.

Department of Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand

E-mail address: `d.byatt@math.canterbury.ac.nz`

E-mail address: `m.dalrymple@math.canterbury.ac.nz`

E-mail address: `r.turner@math.canterbury.ac.nz`