Combinatorial Proof of the Hot Spot Theorem

Ernie Croot

May 30, 2006

1 Introduction

A problem which has perplexed mathematicians for a long time, is to decide whether the digits of π are random-looking, or whether they contain some hidden regularity. This question is usually formalized as

Question. Is π normal base 10 ?

Recall that a real number α is said to be normal base b if for every $k \geq 1$, every string of digits

$$d_1 d_2 ... d_k, \ d_i \in \{0, ..., b-1\},$$

appears among the digits of α with the expected frequency; that is, if we write α in base b, truncated at the Mth digit, as

$$\alpha(M) \ = \ \alpha_1 \alpha_2 ... \alpha_t \ . \ \alpha_{t+1} \alpha_{t+2} ... \alpha_M,$$

(here we have assumed $\alpha > 0$, though it need not be), and if we let $C(\alpha(M))$ count the number of times that $d_1 ... d_k$ appears as a substring of these digits, which is

$$C(\alpha(M)) \ = \ |\{1 \leq j \leq M-k+1 \ : \ \alpha_j = d_1, \alpha_{j+1} = d_2, ..., \alpha_{j+k-1} = d_k\}|,$$

then

$$\lim_{M \to \infty} \frac{C(\alpha(M))}{M} \ = \ d^{-k}.$$

It is not difficult to show that this is equivalent to the following, easier to work with definition: α is normal to the base b if for every interval $I \subseteq [0,1]$ we have that

$$\lim_{N\to\infty} \frac{|\{n \leq N \ : \ \{b^n\alpha\} \in I\}|}{N} \ = \ |I|,$$

where $\{x\}$ denotes the fractional part of x, and $|I|$ denotes the length of I.

Currently, we do not appear to be close to answering the **Question** above, and for that matter, it is not known whether π is normal to any integral base 2 or higher, or even whether the same is true of $\sqrt{2}$, $\ln(2)$, e, and many other constants. However, Bailey and Crandall [2] showed that the numbers

$$\alpha_{b,c}(r) \ = \ \sum_{k=1}^{\infty} \frac{1}{c^k b^{c^k + r_k}}$$

are normal base b when $b, c > 1$ are coprime, and r_k is the kth binary digit of an arbitrary real number $r \in [0,1)$, which, among other things, provides a constructive proof of the uncountability of normal numbers to the base b.

Although it is not known whether π has an expansion of this form (of the α constants), it does have some remarkable expansions somewhat similar to the one above, involving falling powers of c. One such expansion was found by Bailey, Borwein and Plouffe [4], and is given as follows

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

This expansion allowed BBP to determine the nth digit of π in base 16 using only $(\log n)^{O(1)}$ bits of computer memory.

Following the original proof of Bailey and Crandall [2] that the $\alpha_{b,c}(r)$ constants are normal base b, Bailey [1] developed an alternate proof of this fact using a theorem due to Kuipers and Niederreiter [5] (see Lemma 8.1 on page 71, and exercise 8.8 on page 77), which Bailey calls the 'Weak Hot spot Theorem'. This theorem is given as follows:

Theorem 1 *If there exists a constant D such that for every interval $I \subseteq [0,1]$*

$$\limsup_{N\to\infty} \frac{|\{n \leq N \ : \ \{b^n\alpha\} \in I\}|}{N} \ \leq \ D|I|,$$

then α is normal to the base b.

The theorem is asserting that if α fails to be normal to some base b, then there are intervals $I \subseteq [0,1]$ such that $\{b^n\alpha\}$ visits I arbitrarily more often than expected. The fact that the theorem only requires an upper bound on this limsup to prove normality is what makes it highly useful, because upper bounds of this sort (especially if proved using methods such as exponential sums) are much easier to prove.

Following the success of using this theorem to prove normality of the α constants, Bailey and Misiurewicz [3] generalized this result of Kuipers and Niederreiter, and proved a theorem which they term the 'Strong Hot spot Theorem'. We will not state this result here.

The method of proof employed by Kuipers and Niederreiter involves measure theory and some elementary functional analysis, and the method of Bailey and Misiurewicz involves ergodic theory; and, from both of these approaches, one is not left with a clear, intuitive idea of why such a result should be true. In this note we give an elementary combinatorial proof of Theorem 1 above, which provides a more natural and satisfying explanation of the 'hot spot' phenomena than that provided by the measure theory and ergodic theory approaches (at least more satisfying to a combinatorialist).

2 Proof of Theorem 1

First, let us define some terms. Suppose we have an alphabet Σ, consisting of b symbols. A *finite string* s over this alphabet will be a finite sequence of symbols

$$s_1\ s_2\ ...\ s_N,\ \text{where } s_i \in \Sigma;$$

infinite strings are defined analogously. A *substring* of s will be a finite sequence of symbols

$$a_1\ a_2\ ...\ a_j,\ \text{such that } a_1 = s_n,\ a_2 = s_{n+1},\ ...,\ a_j = s_{n+j-1},$$

for some integer $n \leq N - j + 1$.

The following claim implies Theorem 1.

Claim. Suppose

$$s(N_1),\ s(N_2),\ s(N_3),\ ...$$

is an infinite sequence of strings over an alphabet Σ, having lengths

$$N_1 \ < \ N_2 \ < \ N_3 \ < \ \cdots,$$

respectively. [1] Further, suppose w_0 is some string of length k, and let $C(s)$ denote the number of times that w_0 appears as a substring in a finite string s. Then, for every $D > 0$, and every $i = 1, 2, ...$, if

$$\frac{C(s(N_i))}{N_i} \ < \ (1-\epsilon)b^{-k}, \text{ or } \frac{C(s(N_i))}{N_i} \ > \ (1+\epsilon)b^{-k},$$

then there is a finite string w_1, say of length $\ell = \ell(D, \epsilon)$, and a subsequence

$$M_1 \ < \ M_2 \ < \ M_3 \ < \ \cdots$$

of the integers N_i, such that w_1 occurs more than $Db^{-\ell}M_i$ times as a substring of $s(M_i)$.

Basically, to use this theorem to prove Theorem 1 we just take $\Sigma = \{0, ..., b-1\}$, and $s(N)$ will be a truncation of α in base b at N digits. Now, if α fails to be normal base b, then it means that there is some $\epsilon > 0$ and some finite string w_0 (of length k) such that for infinitely many N, w_0 either appears more than $(1+\epsilon)b^{-k}N$ times, or at most $(1-\epsilon)b^{-k}N$ times, as a substring of $s(N)$. Let

$$N_1 \ < \ N_2 \ < \ N_3 \ < \ \cdots$$

be a sequence of integers N where either of these possibilities occurs. The hypotheses of the claim above are now met, and therefore we have our string w_1 and sequence of integers M_i with the claimed properties. Now suppose that $d_1...d_\ell$ is our string w_1, and let $x \in [0, 1]$ be given as

$$x \ = \ 0 \ . \ d_1 \ d_2 \ ... \ d_\ell.$$

Next, let $I \subseteq [0, 1]$ be all numbers whose base-b expansion begins like x; that is, $I \ = \ [x, x+b^{-\ell})$. Then, from the conclusion of our claim above, we deduce that

$$\liminf_i \frac{|\{n \leq M_i \ : \ \{b^n\alpha\} \in I\}|}{M_i} \ \geq \ Db^{-\ell} \ = \ D|I|,$$

which clearly proves Theorem 1.

[1] To understand what this claim is saying, it is perhaps helpful to think of $s(N_i)$ as the truncation of some infinite string to N_i symbols.

2.1 The Idea of the Proof of the Claim

Basically, the proof will consist of combining a certain counting lemma with an averaging lemma.

The counting lemma shows that for k fixed and for ℓ tending to infinity, there are very few – in fact, $o(b^\ell)$ – strings of length ℓ, where some string w_0 of length k occurs an aberrant number of times. [2]

The averaging lemma will then show that if w_0 occus an aberrant number of times in $s(N_i)$, then there are "lots" of windows of consecutive ℓ symbols (in fact, at least cN_i, $c = c(\epsilon)$) appearing in $s(N_i)$ that themselves contain an aberrant number of copies of w_0. But by the counting lemma, this will mean that these windows of length ℓ must lie in a set of only $o(b^\ell)$ strings. Thus, there is some string of length ℓ which appears at least $cN_i/o(b^\ell)$ times as a substring of $s(N_i)$.

At this point we have that each string $s(N_i)$ contains a highly aberrant substring of length ℓ. By a pigeonhole argument, we will have a sequence of strings $s(M_1), s(M_2), ...$ having the *same* highly aberrant pattern of length ℓ, which then would finish the proof of the claim.

2.2 Proof of the Claim

Suppose there is some string w_0 of length k which appears at most $(1-\epsilon)b^{-k}N_i$ times, or at least $(1+\epsilon)b^{-k}N_i$ times, as a substring of $s(N_i)$ for all $i = 1, 2, ...$

We will require the following two lemmas:

Lemma 1 (Averaging Lemma) *Let $s_j(N_i)$ denote the substring of $s(N_i)$ of length ℓ starting at the jth symbol.*

There exists $c = c(\epsilon)$ such that the following holds for i sufficiently large: There is a set of at least cN_i of the substrings $s_j(N_i)$ in which w_0 appears at most $(1 - \epsilon/2)b^{-k}\ell$ times, or more than $(1 + \epsilon/2)b^{-k}\ell$ times.

Proof of the lemma. Consider a copy of w_0 occurring in $s(N_i)$ starting at the tth symbol. If $t \in [\ell, N_i - \ell]$, then there are exactly ℓ of the substrings $s_1(N_i), s_2(N_i), ...$ that also contain that particular copy of w_0; on the other hand, if $t < \ell$ or $t > N_i - \ell$, at most ℓ of these substrings contain that copy of w_0. Thus, if we define $C(S)$ to be the number of times that the string w_0

[2] By 'aberrant' we mean at most $(1-\delta)b^{-k}\ell$ or at least $(1+\delta)b^{-k}\ell$ times as a substring.

appears as a substring of a string S, then we deduce that

$$\sum_{j=1}^{N_i-\ell+1} C(s_j(N_i)) \;=\; \ell C(s(N_i)) + O(\ell^2).$$

Now suppose $C(s(N_i)) > (1+\epsilon)b^{-k}N_i$. Then, from this equation we deduce that the average value of $C(s_j(N_i))$, over $j = 1, ..., N_i - \ell + 1$, is $(1+\epsilon-o(1))b^{-k}\ell$. This clearly implies that there are at least cN_i, $c = c(\epsilon)$ values of $j = 1, ..., N_i - \ell + 1$ such that $C(s_j(N_i)) > (1-\epsilon/2)b^{-k}\ell$, once N_i is sufficiently large. We get the analogous conclusion if $C(s(N_i)) < (1-\epsilon)b^{-k}N_i$, and so the lemma is proved. ■

Lemma 2 (Counting Lemma) *Fix $0 < \delta \leq 1$, fix a string w_0 of length k, and let $f(\ell)$ denote the number of strings of length ℓ that contain either at most $(1-\delta)b^{-k}\ell$, or at least $(1+\delta)b^{-k}\ell$, copies of w_0 as a substring. Then, $f(\ell) = o(b^\ell)$.*

Proof. We prove this using a probabilistic argument (equivalently, one can use the second moment method). First, let Z be a random string with ℓ symbols, where each string is equally likely to be chosen. Then, let $g(Z)$ denote the number of times that w_0 appears as a substring of Z. We may express

$$g(Z) \;=\; X_1 + X_2 + \cdots + X_{\ell-k+1},$$

where $X_i = 1$ if the string w_0 appears as a substring of Z starting at the ith symbol, and is 0 if the substring of Z starting at the ith symbol does not equal w_0. These X_i are therefore Bernoulli random variables which take on the value 1 with probability b^{-k}, and take on the value 0 with probability $1 - b^{-k}$. They are not all independent of one another, but they almost are, as we will see.

Clearly,

$$\mathbb{E}(g(Z)) \;=\; b^{-k}(\ell - k + 1) \;\sim\; b^{-k}\ell.$$

To compute the variance of $S(Z)$ we instead calculate

$$\mathbb{E}(g(Z)^2) \;=\; \sum_{1\leq i,j\leq \ell-k+1} \mathbb{E}(X_iX_j). \tag{1}$$

Now, if $|i-j| \geq k$, then X_i and X_j are independent, in which case

$$\mathbb{E}(X_i X_j) \;=\; \mathbb{E}(X_i)\mathbb{E}(X_j);$$

otherwise, we have the simple upper bound

$$\mathbb{E}(X_i X_j) \;\leq\; \mathbb{E}(X_i) \;=\; b^{-k}.$$

Thus, from (1) we have that

$$\begin{aligned} \mathbb{E}(g(Z)^2) \quad &\leq \quad \left(\sum_{1 \leq i,j \leq \ell-k+1} \mathbb{E}(X_i)\mathbb{E}(X_j) \right) \;+\; b^{-k}k(\ell-k+1) \\ &= \quad \mathbb{E}(g(Z))^2 + b^{-k}k(\ell-k+1). \end{aligned}$$

So, the variance satisfies

$$V(g(Z)) \;=\; \mathbb{E}(g(Z)^2) - \mathbb{E}(g(Z))^2 \;\leq\; b^{-k}k(\ell-k+1).$$

Chebychev's inequality then gives

$$\mathbb{P}(|g(Z) - \mathbb{E}(g(Z))| > \delta b^{-k}\ell) \;\leq\; \frac{V(g(Z))}{\delta^2 b^{-2k}\ell^2} \;\leq\; \frac{kb^k}{\delta^2 \ell},$$

which, for a fixed k, clearly tends to 0 as $\ell \to \infty$. This is saying that with probability $1 - o(1)$, $g(Z) \in [(1-\delta)b^{-k}\ell,\ (1+\delta)b^{-k}\ell]$, which clearly proves the lemma. ■

Now, if we apply Lemma 2 with $\delta = \epsilon/2$ we find that the cN_i substrings $s_j(N_i)$ given by Lemma 1, must lie in a set of only $o(b^\ell)$ strings – thus, one of these few strings of length ℓ appears as a substring $s_j(N_i)$ at least $cN_i/o(b^\ell)$ times. This proves that for any $k, C \geq 1$, and $0 < \epsilon < 1$, there exists ℓ such that for N_i sufficiently large, there is a string $w_1(N_i)$ of length ℓ that appears at least $Cb^{-\ell}N_i$ times as a substring of $s(N_i)$. By the pigeonhole principle, there is a string w_1 of length ℓ, and an infinite sequence $M_1 < M_2 < \cdots$, such that $w_1 = w_1(M_1) = w_1(M_2) = \cdots$, which proves our theorem.

References

[1] D. Bailey, *A Hot Spot Proof of Normality for the Alpha Constants*, available at

http://crd.lbl.gov/~dhbailey/dhbpapers/alpha-normal.pdf

[2] D. Bailey and R. Crandall, *Random Generators and Normal Numbers*, Experimental Mathematics **11** (2002), no. 4, 527-546.

[3] D. Bailey and M. Misiurewicz, *A Strong Hot spot Theorem*, To appear in Proc. of the Amer. Math Soc.

[4] D. Bailey, P. Borwein, and S. Plouffe, *On the Rapid Computation of Various Polylogarithmic Constants*, Math. Comp. **66** (1997), 73-88.

[5] L Kuipers and H. Niederreiter, Uniform Distribution of Sequences, Wiley-Interscience, New York, 1974.