THE NORMALITY OF DIGITS IN ALMOST CONSTANT ADDITIVE FUNCTIONS

J. VANDEHEY

ABSTRACT. We consider numbers formed by concatenating some of the base b digits from additive functions $f(n)$ that closely resemble the prime counting function $\Omega(n)$. If we concatenate the last

$$\left\lceil y \frac{\log\log\log n}{\log b} \right\rceil$$

digits of each $f(n)$ in succession, then the number so created will be normal if and only if $0 < y \le 1/2$. This provides insight into the randomness of digit patterns of additive function after the Erdős-Kac theorem becomes ineffective.

1. INTRODUCTION

We say a real number z with fractional part $\{z\} = 0.z_1 z_2 z_3 \ldots$ written in base b is normal if every digit string $a_1 a_2 \ldots a_k$ consisting of k base b digits occurs with relative frequency b^{-k}—in other words,

$$\lim_{x \to \infty} \frac{\#\left\{1 \le n \le x - k \mid z_{n-1+i} = a_i, 1 \le i \le k\right\}}{x} = b^{-k}.$$

(We shall, for the remainder of the paper, assume that the base b is always the same unless otherwise specified, so we may save the reader from constant repetitions of the phrase "in base b.")

As a simple consequence of Birkhoff's pointwise ergodic theorem, almost all real numbers are normal [5, Ch. 3]; however, despite their omnipresence among the reals, all numbers currently known to be normal have been explicitly constructed to be normal. We do not know if any of the common mathematical constants such as π, e, or even $\sqrt{2}$ is normal.

The first and simplest construction of a normal number is that of Champernowne [3]: in base 10, the number

$$0.(1)(2)(3)(4)(5)(6)(7)(8)(9)(10)(11)\ldots$$

composed by concatenating all natural numbers sequentially is normal. Here we use the notation (n) to refer to the string of digits that compose n, and the notation $(n)(m)$ to denote the concatenation of the two digit strings.

Not long thereafter, Copeland and Erdős [4] proved that, in base 10, the number

$$0.(2)(3)(5)(7)(11)(13)(17)\ldots$$

composed by concatenating all primes is normal. The proof of this fact is one of the few places in mathematics where the phrase, "Because the primes are sufficiently dense in $\mathbb{N}$," is used in earnest.

From here, the study of normal numbers has flourished in a number of different settings. Many mathematicians have sought to generalize the notion of normality to other settings and these have provided interesting results for continued fraction expansions [1], expansions with respect to the bases $-n \pm i$ [14], matrix number systems [13], Q-Cantor set expansions [16, 17, 18], Markov shifts and intrinsically ergodic subshifts [21], and higher-dimensional sequences (see [12] and the papers cited therein).

Studies into constructions for real normal numbers have tended towards one of two primary paths. In one direction, several papers have investigated the base b normality of numbers of the form

$$\sum_{i=1}^{\infty} \frac{1}{b^{m_i} c^{n_i}}$$

for special sequences $\{m_i\}$ and $\{n_i\}$. The most powerful results of this type known to the author are those of Bailey and Crandall [2]; in their paper, they include a general conjecture that, if true, would prove the normality of a wide swath of common constants at once, including π.

We are more interested in the second direction, which follows in the footsteps of Champernowne, Copeland, and Erdős. These results are concerned with the normality of the numbers

$$\theta_f = 0.(f(1))(f(2))(f(3))(f(4))(f(5))\dots$$
$$\tau_f = 0.(f(2))(f(3))(f(5))(f(7))(f(11))\dots$$

for some function $f(n)$. Davenport and Erdős [6] showed that θ_f is normal when $f(n)$ is a positive, integer-valued polynomial. Nakai and Shiokawa [19] showed that both θ_f and τ_f are normal when $f(n)$ is the floor of a polynomial with real coefficients. Madritsch, Thuswaldner, and Tichy [15] likewise showed that both θ_f and τ_f are normal when $f(n)$ is the floor of an entire function with small logarithmic order. De Konick and Kátai [7, 8] have investigated more number-theoretic functions $f(n)$ and have proved the normality of θ_f and τ_f for $f(n) = P(n+1)$, where $P(n)$ is the largest prime divisor of n.

In this paper, we want to extend these results to better understand the normality of θ_f when we let $f(n)$ be an additive function. A function $f(n)$ is said to be additive if $f(mn) = f(m) + f(n)$ for relatively prime m and n; such a function is defined entirely by its values on prime powers. We will focus on additive functions closely related to the prime counting function $\Omega(n)$ and distinct prime counting function $\omega(n)$, which are defined by $\Omega(p^k) = k$ and $\omega(p^k) = 1$.

However, it should come as no surprise that the constants θ_Ω and θ_ω are not normal. The famous Erdős-Kac theorem states that the values of $\omega(n)$ for $n \le x$ are normally distributed with mean and variance $\log\log x$ (and various generalizations state the same for $\Omega(n)$); thus, for almost all $n \le x$, given large x, the first digits of $\omega(n)$ should closely resemble the first half of the digits of $\log\log x$. So if, say, $(\log\log x)$ contains an overabundance of zeros in the first half, we should expect the same to be true of $(\omega(n))$ for most $n \le x$.

The Erdős-Kac theorem does not say anything about what happens in the last

$$\approx \frac{\log \log \log x}{2 \log b}$$

digits of $\omega(n)$ (for $n \leq x$). To consider this further, we need a few extra definitions. Let the truncation function $T_b(z, m)$ denote the string of the last m base b digits of $\lfloor z \rfloor$; we assume z is always at least 0. If $\lfloor z \rfloor$ has fewer than m digits, the truncation function $T_b(z, m)$ will include additional zeros at the head of the string to make sure the string returned is m digits long. Therefore,

$$T_{10}(151, 2) = (51) \qquad T_{10}(1, 2) = (01) \qquad T_{10}(.5, 2) = (00).$$

Also, for $y > 0$, let

$$K_y(x) = \begin{cases} \left\lceil y \dfrac{\log \log \log x}{\log b} \right\rceil, & x > e^e, \\ 1, & \text{otherwise.} \end{cases}$$

and let $\mathcal{K}(x) = K_{1/2}(x)$. Alternately, the number $\mathcal{K}(x)$ may be roughly interpreted as half the digits of $\log \log x$ in base b, rounded up.

Now, given an additive function $f(n)$ and a fixed base b, let $f_y(n) = T_b(f(n), K_y)$ and let

$$\theta_{f,y} = 0.(f_y(1))(f_y(2))(f_y(3))(f_y(4))(f_y(5)) \ldots .^1$$

By the same argument we made above, Erdős-Kac suggests that $\theta_{\omega,y}$ should not be normal for any $y > 1/2$, but says nothing about any smaller y. For that, we present the following new result.

Theorem 1.1. *For any y satisfying $0 < y \leq 1/2$, the numbers $\theta_{\Omega,y}$ and $\theta_{\omega,y}$ are normal. For any y satisfying $1/2 < y$, the numbers $\theta_{\Omega,y}$ and $\theta_{\omega,y}$ are not normal.*

So while the Erdős-Kac theorem states roughly that the first half of the digits of $\omega(n)$ for $n \leq x$ follow a very clear pattern, Theorem 1.1 states that the second half are statistically random.

We wish to extend these results to a slightly larger class of functions. We say a non-negative additive function f is almost constant on primes if the following holds: there exist $c > 0$, $1/2 > \delta > 0$, such that for all small $\epsilon > 0$, we have

$$\exp\left(\sum_{p \leq x} \frac{B_\epsilon(x, p)}{p^{1-\delta}} \right) = o\left(\log x \right), \qquad x \to \infty,$$

where

$$B_\epsilon(x, p) = \min\left\{ 2, \frac{1}{(\log \log x)^{\epsilon/4}} |f(p) - c| \right\}$$

For example, any additive function that satisfies $|f(p) - c| < p^{-\delta}$ for some $c, \delta > 0$ is almost constant on primes.

---

[1]Even though $f_y(n)$ is already considered to be just a string of digits, we include the extra parenthesis to prevent confusion.

Likewise, we say that a non-negative additive function is weakly additive to mean the following. For all small $\epsilon > 0$, we have

$$\prod_{p \leq x} \left(1 - \sum_{p^k \leq x, \ k \geq 2} \frac{C_\epsilon(x, p, k)}{p^k}\right) = 1 + o(1), \qquad x \to \infty,$$

where

$$C_\epsilon(x, p, k) = \min\left\{2, \frac{1}{(\log\log x)^{1+\epsilon/4}}|f(p^k) - f(p^{k-1}) - f(p)|\right\}.$$

All completely additive functions, such as $\Omega(n)$, are weakly additive, as are all additive functions where $|f(p^k) - f(p^{k-1}) - f(p)|$ always belongs to a bounded set, such as $\omega(n)$.

With these definitions we can state a general theorem that contains Theorem 1.1 as a specific case.

Theorem 1.2. *Let $f(n)$ be a non-negative additive function that is almost constant on primes. If $0 < y \leq 1/2$, then $\theta_{f,y}$ is normal. If $1/2 < y$ and $f(n)$ is also weakly additive, then $\theta_{f,y}$ is not normal.*

Somewhat surprisingly, in the case $0 < y \leq 1/2$, when $\theta_{f,y}$ is normal, we have no restrictions whatsoever on the values of f at prime powers (outside of being non-negative).

The reason why $\theta_{f,y}$ fails to be normal is again due to the digits of $f(n)$ for $n \leq x$ too closely resembling the digits of $\log\log x$, so one may wonder whether an Erdős-Kac type distribution is the culprit. Given an additive function, let

$$A_n = \sum_{p<n} \frac{f(p)}{p} \quad \text{and} \quad B_n = \sum_{p<n} \frac{f(p)^2}{p}.$$

There are various different conditions on $f(n)$ that guarantee it will be normally distributed about A_n with variance B_n. The simplest such conditions merely require that B_n goes to infinity as n goes to infinity and that $f(p)$ is uniformly bounded [10]. Alternately, one may require that B_n goes to infinity as n goes to infinity and that

$$\lim_{n\to\infty} B_n^{-1} \sum_{\substack{p<n \\ |f(p)|>\epsilon B_n^{1/2}}} \frac{f(p)^2}{p} = 0$$

for every $\epsilon > 0$ [20].

However, there is nothing in the definition of $f(n)$ being almost constant on primes which implies either of these sets of conditions: on a very sparse set of primes, $f(p)$ could be enormous, and these few primes might be the primary contributers to B_n. It is certainly possible that by ignoring such large values, we could show that being constant on primes would imply some type of Erdős-Kac distribution on average.

For more on these types of results, see [9].

2. Bounds on exponential sums

In the proof of normality results dealing with θ_f and τ_f, a key technique is to reduce the problem to one concerning exponential sums, often of the form

$$\sum_n e\left(\frac{\nu}{b^m} f(n)\right)$$

for some integer ν and positive integer m, where the function $e(z)$ denotes $e^{2\pi i z}$. When Davenport and Erdős examined the case $f(n)$ is a polynomial, they used estimates on Weyl sums. In our case, we want to examine the case when $f(n)$ is a non-negative additive function and will in turn use the Selberg-Delange method to estimate these sums.

We use results and terminology from Tenenbaum's book [22, §II.5.3], mildly simplified to remove an unneeded parameter. (In particular, we take $\delta = 1$ in Tenenbaum's notation and replace his c_0 with our δ.)

Let $z \in \mathbb{C}$, $\delta > 0$, $M > 0$. Then the Dirichlet series

$$F(s) = \sum_{n=1}^{\infty} a_n n^{-s}$$

is said to possess the property $\mathcal{P}(z; \delta, M)$ if the Dirichlet series

$$G(s; z) := F(s)\zeta(s)^{-z}$$

can be extended to a holomorphic function on

$$\sigma \geq 1 - \delta/(1 + \log^+ |\tau|), \quad (s = \sigma + i\tau)$$

which is bounded on this domain in the following way

$$|G(s; z)| \leq M.$$

Suppose $F(s)$ has the property $\mathcal{P}(z; \delta, M)$ and there exist real positive numbers $\{b_n\}_{n=1}^{\infty}$ such that $|a_n| \leq b_n$ for all n and the series

$$\sum_{n=1}^{\infty} b_n n^{-s}$$

satisfies the property $\mathcal{P}(w; \delta, M)$ for a complex number w, then $F(s)$ is said to be of the type $\mathcal{T}(z, w; \delta, M)$.

Theorem 2.1. ([22, §II.5.3, Theorem 3] with $N = 0$) Suppose the Dirichlet series $F(s) := \sum_{n=1}^{\infty} a_n n^{-s}$ is of the type $\mathcal{T}(z, w; \delta, M)$. Then for $x \geq 3$, $A > 0$, $|z| \leq A$, and $|w| \leq A$, we have that

$$\sum_{n \leq x} a_n = x(\log x)^{z-1}\left(\frac{G(1; z)}{\Gamma(z)} + O\left(M\left(e^{-c_1\sqrt{\log x}} + \frac{c_2}{\log x}\right)\right)\right)$$

where the positive constants c_1, c_2 and the implicit constant in the big-O term depend on at most δ and A.

We now apply this result to the case we will need in our proof.

Proposition 2.2. *Let $f(n)$ be a non-negative additive function that is almost constant on primes, with constant c. Let η and ϵ be fixed positive real numbers strictly less than 1, and let a be a non-zero integer satisfying $|a| < \eta^{-2}$.*

(1) *For $\epsilon\mathcal{K}(x) \leq m \leq (1-\epsilon)\mathcal{K}(x)$, we have*

$$\sum_{n\leq x} e\left(\frac{a}{b^m}f(n)\right) = o(x), \quad x \to \infty,$$

where the rate of decay in the little-o term is uniform for fixed η and ϵ.

(2) *If $f(n)$ is also weakly additive, then for $(1+\epsilon)\mathcal{K}(x) \leq m$, we have*

$$\sum_{n\leq x} e\left(\frac{a}{b^m}f(n)\right) = xe\left(\frac{a}{b^m}c\log\log x\right)(1+o(1)), \quad x \to \infty$$

where the rate of decay in the little-o term is uniform for fixed η and ϵ.

Recall, by the definition of $K_y(x)$ that

$$\mathcal{K}(x) := \begin{cases} \left\lceil \dfrac{\log\log\log x}{2\log b} \right\rceil, & x > e^e, \\ 1, & \text{otherwise.} \end{cases}$$

Proof. For a fixed x, we shall consider the function $F(s)$ with coefficients a_n, which are assumed to be multiplicative and defined on prime powers by

$$a_{p^k} := \begin{cases} e\left(\dfrac{a}{b^m}f(p^k)\right) & p^k \leq x \\ e\left(\dfrac{a}{b^m}(f(p^{k-1}) + f(p))\right) & p \leq x < p^k \\ e\left(\dfrac{a}{b^m}kc\right) & p > x \end{cases}.$$

Let $c' = e(ac/b^m)$. Then to the right of the line $\sigma = 1$, we have

$$G(s; c') = F(s)\zeta(s)^{-c'}$$

$$= \prod_p \left(1 + \frac{a_p}{p^s} + \frac{a_{p^2}}{(p^2)^s} + \dots\right)\left(1 - \frac{1}{p^s}\right)^{c'}$$

$$= \exp\left(\sum_p \log\left(1 + \frac{a_p}{p^s} + \frac{a_{p^2}}{(p^2)^s} + \dots\right) + c'\log\left(1 - \frac{1}{p^s}\right)\right)$$

$$= \exp\left(\sum_{p\leq x} \frac{a_p - c'}{p^s} + \sum_p O(p^{-2s})\right).$$

This shows that $G(s, c')$ can be extended to a holomorphic function on $\sigma > 1/2$.

The difference $a_p - c'$ can be bounded in one of two ways: first, by

$$|a_p - c'| = \left| e\left(\frac{ac}{b^m}\right)\left(e\left(\frac{a(f(p) - c)}{b^m}\right) - 1\right)\right|$$
$$\leq \frac{2\pi a|f(p) - c|}{b^m}$$
$$\leq \frac{2\pi\eta^{-2}|f(p) - c|}{(\log\log x)^{\epsilon/2}},$$

using $|e^{ix} - 1| \leq |x|$, and, second, by 2 since $|a_p| = |c'| = 1$.

Provided x is sufficiently large so that $(\log\log x)^{\epsilon/4} \geq 2\pi\eta^{-2}$, we have that

$$\exp\left(\sum_{p \leq x}\frac{a_p - c'}{p^s}\right) = o(\log x)$$

for $\sigma \geq 1 - \delta$ by the definition of f being almost constant on primes. Thus, for some $M = M(x)$ that is $o(\log x)$, we have that $G(s; c') \leq M$ for $\sigma \geq 1 - \delta$, and so $F(x)$ has property $\mathcal{P}(c'; \delta, M)$.

Since each $|a_n| = 1$, we can let $b_n = 1$ and note that the zeta function itself has property $\mathcal{P}(1; \delta, M)$ rather trivially, so that $F(s)$ has property $\mathcal{T}(c', 1; \delta, M)$. We can therefore apply Theorem 2.1.

Thus, we have that

$$\sum_{n \leq x} e(af(n)/b^m) = x(\log x)^{c'-1}\left(\frac{G(1; c')}{\Gamma(c')} + o(1)\right)$$

by the bound on M. The function bounded by $o(1)$ here decays to zero as x tends to infinity and depends only on ϵ and η.

Suppose we are in the case where $\epsilon\mathcal{K}(x) \leq m \leq (1 - \epsilon)\mathcal{K}(x)$. Then the real part of $c' - 1$ is bounded by

$$\mathrm{Re}(e(ac/b^m) - 1) \leq -\frac{1}{2!}\left(\frac{2\pi ac}{b^m}\right)^2 + \frac{1}{4!}\left(\frac{2\pi ac}{b^m}\right)^4$$
$$\leq -\frac{1}{2}\left(\frac{2\pi ac}{b^{(1-\epsilon)\mathcal{K}}}\right)^2(1 + O((\log\log x)^{-\epsilon}))$$
$$\leq -e^{O(1)}(\log\log x)^{-(1-\epsilon)}(1 + O((\log\log x)^{-\epsilon})).$$

Therefore, we have

$$x(\log x)^{c'-1} \leq x\exp\left(-e^{O(1)}(\log\log x)^{\epsilon}(1 + O((\log\log x)^{-\epsilon}))\right) = o(x).$$

The function $1/\Gamma(z)$ is uniformly bounded in any sufficently small compact neighborhood of 1, and c' is in a small neighborhood of 1 for suffficiently large x.

To finish the proof of the first half of the proposition, it suffices to show that $G(1; c') = O(1)$, which itself would be implied by

$$\sum_{p \leq x}\frac{B_\epsilon(x, p)}{p} = O(1)$$

for all sufficiently small ϵ. By the definition of $f(n)$ being almost constant on primes, we know that

$$\sum_{p \leq x} \frac{B_\epsilon(x,p))}{p^{1-\delta}} = O(\log \log x),$$

and we also have that $B_\epsilon(x,p)$ are positive, decreasing functions in x. Therefore, by partial summation, we have

$$\sum_{p \leq x} \frac{B_\epsilon(x,p)}{p} = x^{-\delta} \sum_{p \leq x} \frac{B_\epsilon(x,p)}{p^{1-\delta}} + \delta \int_1^x t^{-\delta-1} \sum_{p \leq t} \frac{B_\epsilon(x,p)}{p^{1-\delta}} \, dt$$

$$\leq O(x^{-\delta} \log \log x) + \delta \int_1^x t^{-\delta-1} \sum_{p \leq t} \frac{B_\epsilon(t,p)}{p^{1-\delta}} \, dt$$

$$= O(x^{-\delta} \log \log x) + O\left(\int_1^x t^{-\delta-1} \log \log t \, dt\right)$$

$$= O(1),$$

which completes the proof of statement (1).

Now suppose we are in the case where $(1+\epsilon)\mathcal{K}(x) \leq m$ and $f(n)$ is weakly additive. Then, in this case, we have

$$c' - 1 = 2\pi i \frac{a}{b^m} c + O\left(\frac{1}{b^{2m}}\right),$$

so that

$$(\log x)^{c'-1} = e\left(\frac{a}{b^m} c \log \log x\right) \cdot \exp\left(O\left(\frac{\log \log x}{b^{2m}}\right)\right)$$

$$= e\left(\frac{a}{b^m} c \log \log x\right) (1 + O((\log \log x)^{-\epsilon})).$$

Also, since $c' - 1 = O((\log \log x)^{-1})$, we have

$$\frac{1}{\Gamma(c')} = 1 + O((c' - 1)) = 1 + O((\log \log x)^{-1}).$$

Therefore, to finish the proof in this case, it suffices to show that $G(1; c') = 1 + o(1)$ uniformly in x. Let $G(z)$ denote the function

$$\prod_p \left(1 - \frac{z}{p}\right)^{-1} \left(1 - \frac{1}{p}\right)^z.$$

We have $G(1) = 1$ and $G(z)$ is analytic at 1, so that $G(c') = 1 + O((\log \log x)^{-1})$. Moreover, we can factor such a term out of $G(1; c')$ to leave

$$G(1; c') = G(c') \prod_{p \leq x} \left(1 + \sum_{k=1}^{\infty} \frac{a_{p^k}}{p^k}\right) \left(1 - \frac{a_p}{p}\right).$$

(We do not need that the product is absolutely convergent since we are at most changing finitely many factors.) Now we examine this latter product in greater detail. We have,

first, that

$$\prod_{p\leq x}\left(1+\sum_{k=1}^{\infty}\frac{a_{p^k}}{p^k}\right)\left(1-\frac{a_p}{p}\right)=\prod_{p\leq x}\left(1+\sum_{k\geq2}\frac{1}{p^k}(a_{p^k}-a_{p^{k-1}}a_p)\right)$$

$$=\prod_{p\leq x}\left(1+\sum_{p^k\leq x,\ k\geq2}\frac{1}{p^k}(a_{p^k}-a_{p^{k-1}}a_p)\right),$$

and provided x is sufficiently large, we have also, by a similar argument to earlier in the proof, that this is bounded above and below by

$$\prod_{p\leq x}\left(1+\sum_{p^k\leq x,\ k\geq2}\frac{C_\epsilon(x,p,k)}{p^k}\right)\quad\text{and}\quad\prod_{p\leq x}\left(1-\sum_{p^k\leq x,\ k\geq2}\frac{C_\epsilon(x,p,k)}{p^k}\right)$$

respectively. The latter equals $1+o(1)$, since $f(n)$ is weakly additive, and this in turn implies that the former is $1+o(1)$ as well, since the logarithm of each factor in the latter product is larger in norm than the logarithm of the corresponding factor in the former product. Thus we have $G(1;c')=1+o(1)$ and this completes the proof. $\qquad\square$

If we only care about the case $\epsilon\mathcal{K}(x)\leq m\leq(1-\epsilon)\mathcal{K}(x)$, then the above proof still works under weaker conditions on $f(n)$. For example, we could replace $o(\log x)$ in the definition of $f(n)$ being almost constant on primes with any function that would satisfy $o((\log x)^{2-c'})$, such as $O(\log x\log\log x)$.

Similarly, if we consider an alternate definition of the truncation function $T_\epsilon^*(z,x)$ which returns the string of the last $(1-\epsilon)\mathcal{K}(x)$ through $\epsilon\mathcal{K}(x)$ digits of $\lfloor z\rfloor$, and let $f_\epsilon^*(n)=T_\epsilon^*(f(n),n)$, then we would expect the number

$$\theta_{f,\epsilon}^*=0.(f_\epsilon^*(1))(f_\epsilon^*(2))(f_\epsilon^*(3))(f_\epsilon^*(4))\ldots$$

to be normal while only requiring the definition of being almost constant on primes to hold for this specific ϵ. This suggests that the "easiest" place for randomness to occur in the digits of an additive function is around the

$$\frac{1}{4}\frac{\log\log\log x}{\log b}\text{th}$$

place.

3. Proof of Theorem 1.2

Consider a string of k base b digits $a_1a_2\cdots a_k$. Let $N^*(x)$ denote the number of times the string occurs in the first x digits of $\theta_{f,y}$ after the decimal point. To prove Theorem 1.2, it suffices to prove

$$\lim_{x\to\infty}\frac{1}{x}N^*(x)\begin{cases}=b^{-k},&\text{if }0<y\leq1/2,\\\neq b^{-k},&\text{if }1/2<y.\end{cases}$$

The xth digit of $\theta_{f,y}$ may occur in the middle of some string $(f_y(x'))$. If we let $N(x)$ denote the number of times we see the string occur within the digits of

$$(f_y(1))(f_y(2))(f_y(3))\ldots(f_y(x')),$$

then it is clear that $N^*(x) = N(x') + O(K_y(x'))$, since the latter function adds at most $K_y(x')$ new places for the string to occur. Moreover, the string $(f_y(n))$ is exactly $K_y(n)$ digits long, so, by a simple application of partial summation, the entire string

$$(f_y(1))(f_y(2))(f_y(3))\dots(f_y(x')),$$

is $x'K_y(x')(1 + o(1))$ digits long. Therefore, we have

$$\frac{1}{x}N^*(x) = \frac{1}{x'K_y(x')}N(x')(1 + o(1))$$

and so it suffices now to prove

$$\lim_{x\to\infty}\frac{1}{xK_y(x)}N(x) = b^{-k}, \qquad \text{if } 0 < y \le 1/2,$$

$$\limsup_{x\to\infty}\frac{1}{xK_y(x)}N(x) > b^{-k}, \qquad \text{if } 1/2 < y.$$

We can rewrite $N(x)$ in the following way:

$$(1) \qquad N(x) = \sum_{n=1}^{x}\sum_{m=k}^{K_y(n)}\theta(b^{-m}f(n)) + O(x)$$

where

$$\theta(z) := \begin{cases} 1, & 0.a_1a_2\cdots a_k \le \{z\} < 0.a_1a_2\cdots a_k + b^{-k}, \\ 0, & \text{otherwise.} \end{cases}$$

The term $\theta(b^{-m}f(n))$ is an indicator of whether the string $a_1a_2\dots a_k$ occurs in $f(n)$ starting at the mth place (to the left of the decimal). The big-O term accounts for occurences of the string that start in a given $f_y(n)$ and finish in another.

As noted above, the whole string

$$(f_y(1))(f_y(2))(f_y(3))\dots(f_y(x')),$$

contains $xK_y(x)(1 + o(1))$ digits, so in (1) we may replace $K_y(n)$ with $K_y(x)$ with an error of $o(xK_y(x))$ and switch the order of summation to obtain

$$(2) \qquad N(x) = \sum_{m=k}^{K_y(x)}\sum_{n=1}^{x}\theta(b^{-m}f(n)) + o(xK_y(x)).$$

Let $Y = \min\{y, 1/2\}$. Then for any small, fixed $\epsilon > 0$, let

$$U(x) = U(x, \epsilon) = \sum_{m=\epsilon\mathcal{K}(x)}^{(1-\epsilon)K_Y(x)}\sum_{n=1}^{x}\theta(b^{-m}f(n))$$

$$V(x) = V(x, \epsilon) = \sum_{m=(1+\epsilon)K_Y(x)}^{K_y(x)}\sum_{n=1}^{x}\theta(b^{-m}f(n)).$$

Note that

$$\mathcal{K}(x) = \frac{1/2}{y}K_y(x) + O_y(1),$$

and that unless $1/2 < y$, the sum $V(x)$ will be identically zero.

Thus we may write $N(x)$ as

$$N(x) = U(x) + V(x) + O(\epsilon x K_y(x)) + o(x K_y(x)).$$

We estimate $U(x)$ first. Following Davenport and Erdős, we pick a small positive constant η and consider two functions $\theta_1(z)$ and $\theta_2(z)$ both periodic in z with period 1, with $\theta_1(z) \leq \theta(z) \leq \theta_2(z)$ and having Fourier expansions

$$\theta_1(z) = (b^{-k} - \eta) + \sum_{\nu \neq 0} A_\nu^{(1)} e(vz)$$

$$\theta_2(z) = (b^{-k} + \eta) + \sum_{\nu \neq 0} A_\nu^{(2)} e(vz)$$

where the coefficients are bounded by

$$|A_\nu^{(*)}| \leq \min\left(\frac{1}{|\nu|}, \frac{1}{\eta \nu^2}\right).$$

Note that this makes the sums absolutely convergent and hence we can change the order of summation without trouble. (For the existence of such functions, see [11, pp. 91–92,99].)

In particular, this gives, for $Y = \min\{y, 1/2\}$ as above,

$$U(x) \geq (b^{-k} - \eta)x((1 - \epsilon)K_Y(x) - \epsilon\mathcal{K}(x)) + \sum_{m=\epsilon\mathcal{K}(x)}^{(1-\epsilon)K_Y(x)} \sum_{n=1}^{x} \sum_{\nu \neq 0} A_\nu^{(1)} e\left(\frac{\nu}{b^m} f(n)\right)$$

$$= b^{-k} x K_Y(x)(1 + O(\epsilon) + O(\eta)) + \sum_{\nu \neq 0} A_\nu^{(1)} \sum_{m=\epsilon\mathcal{K}(x)}^{(1-\epsilon)K_Y(x)} \sum_{n=1}^{x} e\left(\frac{\nu}{b^m} f(n)\right)$$

$$= b^{-k} x K_Y(x)(1 + O(\epsilon) + O(\eta)) + O(x K_y(x)\eta)$$

$$+ \sum_{\substack{\nu \neq 0 \\ |\nu| \leq \eta^{-2}}} A_\nu^{(1)} \sum_{m=\epsilon\mathcal{K}(x)}^{(1-\epsilon)K_Y(x)} \sum_{n=1}^{x} e\left(\frac{\nu}{b^m} f(n)\right).$$

Applying Proposition 2.2, we then obtain

$$U(x) \geq b^{-k} x K_Y(x)(1 + O(\epsilon) + O(\eta)) + \sum_{\substack{\nu \neq 0 \\ |\nu| \leq \eta^{-2}}} A_\nu^{(1)} \sum_{m=\epsilon\mathcal{K}(x)}^{(1-\epsilon)K_Y(x)} o(x)$$

$$= b^{-k} x K_Y(x)(1 + O(\epsilon) + O(\eta) + o_{\eta,\epsilon}(1)),$$

where the little-o term tends to zero uniformly for fixed η and ϵ. By applying the same idea with $\theta_2(z)$, we obtain

$$U(x) \leq b^{-k} x K_Y(x)(1 + O(\epsilon) + O(\eta) + o_{\eta,\epsilon}(1))$$

as well.

We can now complete the proof in the case $0 < y \le 1/2$. In this case, we have

$$\frac{1}{xK_y(x)}N(x) = b^{-k}(1 + O(\epsilon) + O(\eta) + o_\eta(1)).$$

We have chosen ϵ and η to be fixed as x varies, so this does not immediately give us that the limit is b^{-k} (because we have no guarantee that the limit exists). However, it does show that the lim sup and lim inf of $N(x)/xK_y(x)$ are both equal to $b^{-k}(1 + O(\epsilon) + O(\eta))$; by taking ϵ and η arbitrarily small, we see that the lim sup is at most b^{-k} and the lim inf is at least b^{-k}, so the full limit itself must exist and equal b^{-k}.

Now we investigate the case $1/2 < y$. Consider a function $\theta_3(z)$ that satisfies the following properties:

(1) $0 \le \theta_3(z) \le \theta(z)$;
(2) $\theta_3(z) = \theta(z)$ except on the intervals

$$[0.a_1a_2 \dots a_k - b^{-2K_y(x)}, 0.a_1a_2 \dots a_k + b^{-2K_y(x)}]$$

and

$$[0.a_1a_2 \dots a_k + b^{-k} - b^{-2K_y(x)}, 0.a_1a_2 \dots a_k + b^{-k} + b^{-2K_y(x)}];$$

and,
(3) $\theta_3(z)$ is continuous and piecewise smooth.

In particular, since $\theta_3(z)$ is assumed to be continuous and piecewise smooth, its Fourier series converges absolutely [23, p. 81], so the order of summation doesn't matter. We shall again refer to the Fourier coefficients by $A_\nu^{(3)}$.

Therefore, given some small η, we have

$$V(x) \ge \sum_{m=(1+\epsilon)\mathcal{K}(x)}^{K_y(x)} \sum_{n=1}^{x} \sum_\nu A_\nu^{(3)} e\left(\frac{\nu}{b^m}f(n)\right)$$

$$= \sum_\nu A_\nu^{(3)} \sum_{m=(1+\epsilon)\mathcal{K}(x)}^{K_y(x)} \sum_{n=1}^{x} e\left(\frac{\nu}{b^m}f(n)\right)$$

$$= \sum_{|\nu|<\eta^{-2}} A_\nu^{(3)} \sum_{m=(1+\epsilon)\mathcal{K}(x)}^{K_y(x)} \sum_{n=1}^{x} e\left(\frac{\nu}{b^m}f(n)\right) + O(\eta x(K_y(x) - (1+\epsilon)\mathcal{K}(x))).$$

Applying Proposition 2.2 again and noting that we can add the $|\nu| \ge \eta^{-2}$ terms back in at the cost of another copy of

$$O(\eta x(K_y(x) - (1+\epsilon)\mathcal{K}(x))),$$

this becomes

$$V(x) \geq \sum_{|\nu|<\eta^{-2}} A_\nu^{(3)} \sum_{m=(1+\epsilon)\mathcal{K}(x)}^{K_y(x)} xe\left(\frac{\nu}{b^m}\log\log x\right)(1+o(1))$$
$$+ O(\eta x(K_y(x) - (1+\epsilon)\mathcal{K}(x)))$$
$$= \sum_\nu A_\nu^{(3)} \sum_{m=(1+\epsilon)\mathcal{K}(x)}^{K_y(x)} xe\left(\frac{\nu}{b^m}\log\log x\right)$$
$$+ x(K_y(x) - (1+\epsilon)\mathcal{K}(x))(o(1) + O(\eta)).$$

The sum

$$\sum_\nu A_\nu^{(3)} \sum_{m=(1+\epsilon)\mathcal{K}(x)}^{K_y(x)} e\left(\frac{\nu}{b^m}\log\log x\right) = \sum_{m=(1+\epsilon)\mathcal{K}(x)}^{K_y(x)} \sum_\nu A_\nu^{(3)} e\left(\frac{\nu}{b^m}\log\log x\right)$$
$$= \sum_{m=(1+\epsilon)\mathcal{K}(x)}^{K_y(x)} \theta_3\left(\frac{1}{b^m}\log\log x\right)$$

approximates the number of times the string occurs in the digits of $\log\log x$ to the left of the $(1+\epsilon)\mathcal{K}$th place. It will, in fact, equal and not just approximate the number of occurences of the string if we make two assumptions: that x is sufficiently large and that $\log\log x$ is an integer ending in a non-zero digit base b. In order for θ_3 evaluated at $\log\log x/b^m$ to differ from θ at the same point, the fractional part of $\log\log x/b^m$ must be within $b^{-2K_y(x)}$ of a number in $[0,1]$ that has all zeros after the kth place, but since we have assumed that $\log\log x$ is an integer ending in a non-zero digit and since m is at least $\mathcal{K}(x)$ but at most $K_y(x)$, the fractional part of $\log\log x/b^m$ has a non-zero digit after the kth place, but none after the $K_y(x) + 1$th place.

Thus, for x sufficiently large and $\log\log x$ an integer ending on a non-zero digit, we have that

$$V(x) \geq x \sum_{m=(1+\epsilon)\mathcal{K}(x)}^{K_y(x)} \theta\left(\frac{1}{b^m}\log\log x\right) + x(K_y(x) - (1+\epsilon)\mathcal{K}(x))(o(1) + O_\eta(1)).$$

Now let M be a large integer and consider x_M so that $\log\log x_M$ is an integer whose digit string is composed of M copies of the string $a_1 a_2 \ldots a_k$ concatenated together followed by exactly $\mathcal{K}(x)$ digits, the last of which is non-zero. In particular, this implies that $M = (K_y(x_M) - \mathcal{K}(x_M))/k$. In this case, we have

$$N(x_M) \geq b^{-k} x_M \mathcal{K}(x_M)(1 + O(\epsilon) + O(\eta) + o_\eta(1)) + x_M \frac{K_y(x_M) - \mathcal{K}(x_M)}{k}$$
$$+ x_M K_y(x_M)(O(\epsilon) + o(1) + O_\eta(1))$$
$$= x_M K_y(x_M)\left(\frac{1}{k}\left(1 - \frac{1}{2y}\right) + \frac{1}{b^k}\frac{1}{2y}\right)(1 + O(\epsilon) + O(\eta) + o_\eta(1) + O_\eta(1)).$$

By choosing ϵ and η sufficiently small, we see that the lim sup of $N(x)/xK_y(x)$ must exceed b^{-k}.

Thus, the limit of $N(x)/xK_y(x)$ cannot be b^{-k}, so $\theta_{f,y}$ is not normal in this case.

4. Acknowledgements

The author wishes to thank Heini Halberstam for his helpful comments.

References

[1] R. Adler, M. Keane, and M. Smorodinsky. *A construction of a normal number for the continued fraction transformation*, J. Number Theory 13 (1981), no. 1, 95–105.

[2] D.H. Bailey and R.E. Crandall. *Random Generators and Normal Numbers*. Exper. Math. 11 (2002), 527–546.

[3] D.G. Champernowne. *The construction of decimal normal in the scale of ten*, J. London Math. Soc. 8 (1933), no. 3, 254–260.

[4] A.H. Copeland and P. Erdős. *Note on Normal Numbers*, Bull. Amer. Math. Soc. 52 (1946), 857–860.

[5] K. Dajani and C. Kraaikamp. *Ergodic Theory of Numbers*, The Carus Mathematical Monographs, 29. Mathematical Association of America, Washington, DC, 2002.

[6] H. Davenport and P. Erdős. *Note on normal decimals*, Can. J. Math, 4 (1952), 58–63.

[7] J.-M. De Konick and I. Kátai. *Construction of normal numbers by classified prime divisors of integers*, Funct. Approx. Comment. Math., 45 (2011), no. 2, 231–253.

[8] J.-M. De Konick and I. Kátai. *On a problem of normal numbers raised by Igor Shparlinski*, Bull. Aust. Math. Soc., 84 (2011), 337–349.

[9] H. Delange and H. Halberstam *A note on additive functions*, Pacific J. Math. 7 (1957), 1551–1556.

[10] P. Erdős and M. Kac. *The Gaussian law of errors in the theory of additive number-theoretic functions*, Amer. J. Math. 62 (1940), 738–742.

[11] J.F. Koksma. *Diophantische Approximationen*, Ergebnisse der Math., IV, 4. Berlin, 1936.

[12] M. Levin and M. Smorodinsky. *Explicit construction of normal lattice configurations*, Colloq. Math. 102 (2005), no. 1, 33–47.

[13] M. Madritsch. *A note on normal numbers in matrix number systems*, Math. Pannon. 18 (2007), no. 2, 219227.

[14] M. Madritsch. *Generating normal numbers over Gaussian integers*, Acta Arith. 135 (2008), no. 1, 63–90.

[15] M. Madritsch, J. Thuswaldner, and R. Tichy. *Normality of numbers generated by values of entire functions*, J. of Number Theory 128 (2008), 1127–1145.

[16] B. Mance. *Construction of normal numbers with respect to the Q-Cantor series expansion for certain Q*, Acta Arith. 148 (2011), no. 2, 135–152.

[17] B. Mance *Normal numbers with respect to the Cantor series expansion*, Thesis (Ph.D.)—The Ohio State University. 2010. 290 pp.

[18] B. Mance. *Typicality of normal numbers with respect to the Cantor series expansion*, New York J. Math. 17 (2011), 601–617

[19] Y. Nakai and I. Shiokawa. *Normality of numbers generated by the values of polynomials at primes* Acta Arith. 81 (1997), no. 4, 345–356.

[20] H.N. Shapiro. *Distribution functions of additive arithmetic functions*, Proc. Nat. Acad. Sci. U.S.A. 42 (1956), 426–430.

[21] M. Smorodinsky and B. Weiss *Normal sequences for Markov shifts and intrinsically ergodic subshifts*, Israel J. Math., 59 (1987), no. 2, 225–233.

[22] G. Tenenbaum. *Introduction to analytic and probabilistic number theory*, Translated from the second French edition (1995) by C. B. Thomas. Cambridge Studies in Advanced Mathematics, 46. Cambridge University Press, Cambridge, 1995.

[23] G. Tolstov *Fourier Series* Second English translation. Translated from the Russian and with a preface by Richard A. Silverman. Dover Publications, Inc., New York, 1976.