ON THE DENSITY OF HAPPY NUMBERS

J. Gilmer

Department of Mathematics, Rutgers, Piscataway, New Jersey
jmgilmer@math.rutgers.edu

Abstract

The happy function $H : \mathbb{N} \to \mathbb{N}$ sends a positive integer to the sum of the squares of its digits. A number x is said to be happy if the sequence $\{H^n(x)\}_{n=1}^{\infty}$ eventually reaches 1 (here $H^n(x)$ denotes the nth iteration of H on x). A basic open question regarding happy numbers is what bounds on the density can be proved. This paper uses probabilistic methods to reduce this problem to experimentally finding suitably large intervals containing a high (or low) density of happy numbers as a subset. Specifically, we show that $\bar{d} > .18577$ and $\underline{d} < .1138$, where $\bar{d}$ and $\underline{d}$ denote the upper and lower density of happy numbers respectively. We also prove that the asymptotic density does not exist for several generalizations of happy numbers.

1. Introduction

It is well known that if you iterate the process of sending a positive integer to the sums of the squares of its digits, you eventually arrive at either 1 or the cycle

$$4 \to 16 \to 37 \to 58 \to 89 \to 145 \to 42 \to 20 \to 4.$$

If we change the map, instead sending an integer to the sum of the cubes of its digits, then there are 9 different possible cycles (see Section 5.2.1). Many generalizations of these kinds of maps have been studied. For instance, [3] considered the map which sends an integer n to the sum of the eth power of its base-b digits. In this paper, we study a more general class of functions.

Definition 1.1. Let $b > 1$ be an integer, and let h be a sequence of b non-negative integers such that $h(0) = 0$ and $h(1) = 1$. Define $H : \mathbb{Z}^+ \to \mathbb{Z}^+$ to be the following function: for $n \in \mathbb{Z}^+$, with base-b representation $n = \sum_{i=0}^{k} a_i b^i$, $H(n) := \sum_{i=0}^{k} h(a_i)$. We say H is the b-happy function with digit sequence h.

As a special case, the b-happy function with digit sequence $\{0, 1, 2^e, \ldots, (b-1)^e\}$ is called the (e, b)-function.

Definition 1.2. Let H be any b-happy function and let $C \subseteq \mathbb{N}$. We say $n \in \mathbb{N}$ is type-C if there exists k such that $H^k(n) \in C$.

For the (e, b)-function we refer to type-$\{1\}$ integers as (e, b)-happy.

Fix a b-happy function H and let $\alpha := \max\limits_{i=0,\ldots,b-1} \big(H(i)\big)$. If n is a d-digit integer in base-b, then $H(n) \leq \alpha d$. If d^* is the smallest $d \in \mathbb{N}$ such that $\alpha d < b^{d-1}$, then for all n with $d \geq d^*$ digits, $H(n) \leq \alpha d < b^{d-1} \leq n$. This implies the following

Fact 1.3. For all $n \in \mathbb{N}$, there exists an integer i such that $H^i(n) < b^{d^*} - 1$.

Moreover, to find all possible cycles for a b-happy function, it suffices to perform a computer search on the trajectories of the integers in the interval $[0, b^{d^*} - 1]$.

Richard Guy asks a number of questions regarding $(2, 10)$-happy numbers and their generalizations, including the existence (or not) of arbitrarily long sequences of consecutive happy numbers and whether or not the asymptotic density exists [4, problem E34]. To date, there have been a number of papers in the literature addressing the former question ([1],[3],[6]). This paper addresses the latter question. Informally, our main result says that if the asymptotic density exists, then the density function must quickly approach this limit.

Theorem 1.4. *Fix a b-happy function H. Let I be a sufficiently large interval and let $S \subseteq I$ be a set of type-C integers. If $\frac{|S|}{|I|} = d$, then the upper density of type-C integers is at least $d\,(1 - o(1))$.*

Note as a corollary we can get an upper bound on the lower density by taking C to be the union of all cycles except the one in which we are interested. In Sections 3 and 4 we will define explicitly what constitutes a sufficiently large interval and provide an expression for the $o(1)$ term. Using Theorem 1.4, one can prove the asymptotic density of (e, b)-happy numbers (or more generally type-$\{C\}$ numbers) does not exist by finding two large intervals I_1, I_2 for which the density in I_1 is large and in I_2 is small. In the case of $(2, 10)$-happy numbers, taking $I_1 = [10^{403}, 10^{404} - 1]$ and $I_2 = [10^{2367}, 10^{2368} - 1]$, we show that $\bar{d} \geq .185773(1 - 10^{-49})$ and $\underline{d} \leq .11379(1 + 10^{-100})$ respectively.

We also show that the asymptotic density does not exist for 8 of the cycles for the $(3, 10)$-function (see Section 5). It should be noted that our methods only give one sided bounds. In an earlier version of this manuscript, we asked if $\bar{d} < 1$ for $(2, 10)$-happy numbers. Recently, [5] has announced a proof of this. Specifically, he proves that $.1962 < \bar{d} < .38$, and $0.002937 < \underline{d} < .1217$.

2. Preliminaries

Unless otherwise noted, we regard an interval $I = [a, b]$ as the integer interval $\{n \in \mathbb{Z}^+ : a \leq n \leq b\}$ where, in general, $a, b \in \mathbb{R}$. We denote $|I|$ to be the

cardinality of this set. We also denote $[n]$ as the set $\{0, 1, \ldots, n\}$. Throughout this section let H denote an arbitrary b-happy function.

Definition 2.1. Let I be an interval and Y the random variable uniformly distributed amongst the set of integers in I. Then we say the random variable Y is induced by the interval I.

Definition 2.2. The type-C density of an integer interval I is defined to be the quantity

$$d_C(I) := \frac{|\{n \in I : n \text{ is type-}C\}|}{|I|}.$$

Observation 2.3. If Y is the random variable induced by an interval I, then

$$d_C(I) = \mathbb{P}\big(H(Y) \text{ is type-C}\big).$$

Usually, we take C to be one of the cycles arising from a b-happy function H. However, if we wish to upper bound the lower density of type-C integers, then we study the density of type-C' integers, where C' is the union of all cycles except C.

2.1. The Random Variable $H(Y_m)$

Consider the random variable Y_m induced by the interval $[0, b^m - 1]$, i.e., Y_m is a random m-digit number. If X_i is the random variable corresponding to the coefficient of b^i in the base-b expansion of Y_m, then

$$H(Y_m) = \sum_{i=0}^{m-1} H(X_i). \tag{1}$$

In this paper, we will be interested in the mean and variance of $H(Y_1)$ (i.e., the image of a random digit) which we refer to as the *digit mean* (μ) and *digit variance* (σ^2) of H. The random variables $H(X_i)$ in (1) are all independent and identically distributed (i.i.d.), thus,

$$\mathbf{E}[H(Y_m)] = \mu m \qquad \mathbf{Var}[H(Y_m)] = \sigma^2 m. \tag{2}$$

The random variable $H(Y_m)$ is equivalent to rolling m times a b-sided die with faces $0, 1, H(2), \ldots, H(b-1)$ and taking the sum. Since it is a sum of m i.i.d. random variables, it approaches a normal distribution as m gets large. Also, the distribution of $H(Y_m)$ is concentrated near the mean. This implies the following key insight:

Observation 2.4. The density of happy numbers amongst m-digit integers depends almost entirely on the distribution of happy numbers near μm.

2.2. Computing Densities

Let $P_{m,i} := \mathbb{P}\big(H(Y_m) = i\big)$. Then

$$P_{m,i} = \frac{\left|\{(a_1, a_2, \ldots, a_m) : a_k \in H([b-1]) \text{ and } \sum_{k=1}^{m} a_k = i\}\right|}{b^m}.$$

For fixed m, the sequence $\{P_{m,i}\}_{i=1}^{\infty}$ has generating function

$$f_m(x) = \sum_{i=0}^{\infty} P_{m,i} x^i = \left(\frac{1 + x + x^{H(2)} + \cdots + x^{H(b-1)}}{b}\right)^m. \tag{3}$$

This implies the following recurrence relation with initial conditions $P_{0,0} = 1$, and $P_{0,i} = 0$ for $i \in \mathbb{Z} - \{0\}$.

$$P_{m,i} = \frac{P_{m-1,i} + P_{m-1,i-1} + P_{m-1,i-H(2)} + \cdots + P_{m-1,i-H(b-1)}}{b}. \tag{4}$$

To see this, write $f_m(x) = \left(\frac{1+x+x^{H(2)}+\cdots+x^{H(b-1)}}{b}\right)^{m-1} \left(\frac{1+x+x^{H(2)}+\cdots+x^{H(b-1)}}{b}\right)$ and consider the coefficient of x^i.

If $\alpha = \max\limits_{i=0,\ldots,b-1} \big(H(i)\big)$, then $H(Y_m) \subseteq [0, m\alpha]$. In particular, $P_{m,i} = 0$ if $i > m\alpha$. Using this fact combined with (4), we can implement the following simple algorithm for quickly calculating the type-C density of the interval $[0, b^m - 1]$.

1. First, using the recurrence (4), calculate $P_{m,i}$ for $i = 0, \ldots, m\alpha$.

2. Using brute force, find the type-C integers in the interval $[0, m\alpha]$.

3. Output $\sum\limits_{\substack{i \in [0,m\alpha] \\ i \text{ type-C}}} P_{m,i}$.

Using this algorithm, calculating the density for large n becomes computationally feasible. Figure 1 graphs the density of $(2, 10)$-happy numbers $< 10^n$ for n up to 8000.

The peak near 10^{400} and valley near 10^{2350} will be used to imply the bounds obtained in this paper.

2.3. A Local Limit Law

The random variable $H(Y_m)$ approaches a normal distribution as m becomes large. The following theorem[1], presented in [2, p. 593], gives a bound.

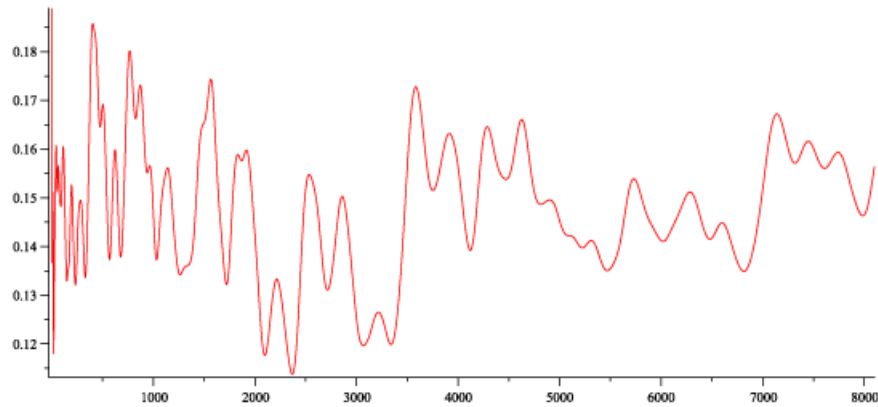[1]We quote a simpler version, with a minor typo corrected

Figure 1: Relative Density of $(2,10)$-Happy Numbers $< 10^n$

Theorem 2.5. (Local limit law for sums) *Let* $X_1, \ldots, X_n$ *be i.i.d. integer-valued variables with probability generating function (PGF)* $B(z)$, *mean* μ, *and variance* σ^2, *where it is assumed that the* X_i *are supported on* $\mathbb{Z}^+$. *Assume that* $B(z)$ *is analytic in some disc that contains the unit disc in its interior and that* $B(z)$ *is aperiodic with* $B(0) \neq 0$. *Then the sum,*

$$S_n := X_1 + X_2 + \cdots + X_n$$

satisfies a local limit law of the Gaussian type: for t in any finite interval, one has

$$\mathbb{P}(S_n = \lfloor \mu n + t\sigma\sqrt{n} \rfloor) = \frac{e^{-t^2/2}}{\sqrt{2\pi n}\sigma} \left(1 + O(n^{-1/2})\right).$$

Here aperiodic means that the $\gcd\{j : b_j > 0, j > 0\} = 1$, where $B(z) = \sum\limits_{j=0}^{\infty} b_j z^j$ (or more informally, the digit sequence for H cannot all be divisible by some integer larger than 1). In our case, the PGF of the $H(X_i)$ is the polynomial

$$p(x) = \frac{x^{H(0)} + x^{H(1)} + \cdots + x^{H(b-1)}}{b}.$$

It is important in our definition of b-happy functions that we assume that $H(0) = 0$, and $H(1) = 1$. This guarantees that $p(x)$ is aperiodic and in particular that the above theorem applies for the sum $H(Y_m)$. As a consequence, for a fixed interval $[-T, T]$, if $i = \mu m + t\sigma\sqrt{m}$ for some $t \in [-T, T]$, then

$$P_{m,i} = \frac{e^{-t^2/2}}{\sqrt{2\pi m}\sigma} \left(1 + O\left(m^{-1/2}\right)\right).$$

The above error term, $O\left(m^{-1/2}\right)$, will prove to a technical difficulty which will be discussed later.

2.4. Overview of the Proof

The following heuristic will provide the general motivation for the proofs. Recall that the random variable Y_m is concentrated near its mean μm.

Observation 2.6. Suppose I is a large interval with type-C density d. Consider the choices of m such that the mean of $H(Y_m)$ is in the interval I; then for some choices of m we likely have

$$\mathbb{P}\big(H(Y_m) \text{ is type-}C\big) \geq d.$$

The key idea to turn this heuristic into a proof is to average over all reasonable choices of m in order to imply there is an m with the desired property.

We will use Theorem 2.5 to show that, for small k, $H(Y_m)$ and $H(Y_{m+k})$ have essentially the same distribution only shifted by a factor of μk. Thus, as k varies, the distributions $H(Y_{m+k})$ should uniformly cover the interval I. It is crucial here to use the fact that $H(Y_m)$ is locally normal, otherwise the proof will fail. For example, suppose all the happy numbers in I are odd. In this case, if $H(Y_m)$ is not locally normal and instead is supported on the even numbers for all m, then every shift $H(Y_{m+k})$ will miss all of the happy numbers in I.

Unfortunately, the fuzzy term in the local limit law prevents us from obtaining explicit bounds on the error (and any explicit bounds seem unsatisfactory for our purposes). Section 3 adds a necessary step, which is to construct an interval within $[b^{n-1}, b^n]$ with high type-C density for n arbitrarily large. The trick is to consider intervals of the form $I_k := [1^k 0^{n-k}, 1^k 0^m (b-1)^{n-k-m}]$ (here 1^k denotes k consecutive 1's). This solves the issue where $H(Y_m)$ and $H(Y_{m+k})$ are not exact shifts of each other, as the distributions induced by the I_k *are* exact shifts under the image of H. These distributions will uniformly cover the base interval I with much better provable bounds. The main result is presented in Section 4, the proof uses the local limit law with the result from Section 3.

3. Constructing Intervals

Throughout this section, if Y is a random variable and k is an integer, let $\tau_k(Y)$ denote the random variable $Y + k$.

Definition 3.1. We say an integer interval I is n-strict if $I \subseteq [b^{n-1}, b^n - 1]$ and $|I| = b^{3n/4}$.

The primary goal of this section is to construct n-strict intervals of high type-C density for arbitrarily large n.

Our choice of the definition of n-strict is only for the purpose of simplifying calculations, there is nothing special about the value $\frac{3}{4}$. In fact, any ratio $> \frac{1}{2}$ would work. Note if 4 does not divide n, then no n-strict intervals exist.

For the entirety of this section we will make the following assumptions:

- H is a b-happy function with digit mean μ and digit variance σ^2.

- We wish to lower bound the upper density of type-C integers for some $C \subset \mathbb{N}$.

- We have found, by computer search, an appropriate starting interval I_1, which is n_1-strict and has suitably large type-C density $d_C(I_1)$.

The results in this section apply only if this n_1 is sufficiently large, so we state here exactly how large n_1 must be so one knows where to look for the interval I_1. In particular, we say an integer n satisfies the bounds (**B**) if

B1: $4\left(1 + 3\mu + \sqrt{2}\sigma b^{5n/8}\right) \le b^{n-1}$,

B2: $\sqrt{3\mu b}\sigma \le b^{3n/8}$,

B3: $4\mu\left(3\mu + 1 + b^{3n/4} + 2\sigma\mu^{-1/2}b^{5n/8}\right) \le b^{n-1}$.

Generally, n need not be too large to satisfy these bounds. For example, if H is a $(2,10)$-happy function, assuming $n > 13$ is enough to guarantee that it satisfies bounds (**B**). This is well within the scope of the average computer as it is possible to compute the density of type-C integers in $[0, b^n - 1]$ for n up to (and beyond) 1000 using the algorithm in Section 2. These bounds are necessary in the proof of Theorem 3.5.

Our first goal is to use an arbitrary n-strict interval I to construct a second interval, I_2, which is n_2-strict for some n_2 much larger than n and contains a similar density of type-C integers as I. The next lemma will be a helpful tool.

Lemma 3.2. *Let $I := [i_1, i_2]$, $J := [j_1, j_2]$ be integer intervals. Let $S \subseteq I$ and Y be an integer-valued random variable whose support is in J. For $k \in \mathbb{Z}$, denote the random variable $Y + k$ as $\tau_k(Y)$. Then there exists an integer $k \in [i_1 - j_2, i_2 - j_1]$ such that $\mathbb{P}(\tau_k(Y) \in S) \ge \frac{|S|}{|I|+|J|-1}$.*

Proof. The idea of the proof is that by averaging over all appropriate k, the distributions of $\tau_k(Y)$ should uniformly cover I. More formally, let $k_1 := i_1 - j_2$, $k_2 := i_2 - j_1$, and let K be the set of integers in the interval $[k_1, k_2]$. Note that $|K| = |I| + |J| - 1$. Pick k uniformly at random from K and consider the random variable $Z := \mathbb{P}(\tau_k(Y) \in S)$. Then

$$\mathbf{E}[Z] = \frac{1}{|K|} \sum_{k=k_1}^{k_2} \mathbb{P}(\tau_k(Y) \in S)$$

$$= \frac{1}{|I| + |J| - 1} \sum_{k=k_1}^{k_2} \sum_{i \in S} \mathbb{P}(\tau_k(Y) = i)$$

$$= \frac{1}{|I| + |J| - 1} \sum_{i \in S} \sum_{k=k_1}^{k_2} \mathbb{P}(\tau_k(Y) = i). \tag{5}$$

Note that $\mathbb{P}(\tau_k(Y) = i) = \mathbb{P}(Y = i - k)$ and for $i \in S \subseteq I$ we have,

$$J \subseteq [i - k_2, i - k_1].$$

Thus, for all $i \in S$,

$$\sum_{k=k_1}^{k_2} \mathbb{P}(\tau_k(Y) = i) = \mathbb{P}(Y \in [i - k_2, i - k_1]) = 1.$$

Therefore,

$$(5) = \frac{|S|}{|I| + |J| - 1}.$$

So there exists k such that $\mathbb{P}(\tau_k(Y) \in S) \geq \mathbf{E}[Z] = \frac{|S|}{|I|+|J|-1}$.

$\square$

Using Lemma 3.2, we will not lose much density assuming that $|I|$ is much larger than $|J|$. However, if Y_m is induced by the interval $[0, b^m - 1]$, then the random variable $H(Y_m)$ will be supported on a set J that is much too large. As a result, it will be more useful to consider a smaller interval where the bulk of the distribution lies.

Lemma 3.3. *Let Y be an integer-valued random variable with mean μ_Y and variance σ_Y^2, and let $\lambda > 0$. Let $S \subseteq [i_1, i_2] = I$ be a set of integers where $|S|/|I| = d$. Then there exists an integer $k \in [i_1 - (\mu_Y + \sigma_Y \lambda), i_2 - (\mu_Y - \sigma_Y \lambda)]$ such that*

$$\mathbb{P}(\tau_k(Y) \in S) \geq \left(1 - \frac{1}{\lambda^2}\right) \left(\frac{d}{1 + \frac{2\sigma_Y \lambda}{|I|}}\right).$$

Proof. By Chebyshev's Inequality[2] we have $\mathbb{P}(|Y - \mu_Y| < \sigma_Y \lambda) > 1 - \frac{1}{\lambda^2}$. Let Y' be Y conditioned on being in the interval $J := [\mu_Y - \sigma_Y \lambda, \mu_Y + \sigma_Y \lambda]$. Note that

$$|J| \leq 2\sigma_Y \lambda + 1.$$

Then, by Lemma 3.2, there exists $k \in [i_1 - (\mu_Y + \sigma_Y \lambda), i_2 - (\mu_Y - \sigma_Y \lambda)]$ such that $\mathbb{P}(\tau_k(Y') \in S) \geq \frac{|S|}{|I|+|J|-1} = \frac{d}{1+\frac{2\sigma_Y \lambda}{|I|}}$. Therefore, we have

$$\mathbb{P}(\tau_k(Y) \in S) \geq \mathbb{P}(Y \in J)\mathbb{P}(\tau_k(Y') \in S) \geq \left(1 - \frac{1}{\lambda^2}\right) \left(\frac{d}{1 + \frac{2\sigma_Y \lambda}{|I|}}\right).$$

$\square$

---

[2]We certainly could do better than Chebyshev's Inequality here. However, the bounds it gives will suit our purposes fine.

It is possible to construct sets of intervals which, under the image of H, act as shifts of each other. For example, in base-10 (recall $H(0) = 0, H(1) = 1$) if the random variable X_1 is induced by $[1100, 1199]$ and X_2 is induced by $[0, 99]$, then $H(X_1) = H(X_2) + 2$.

We will now further expand on the example above. Let $n \in \mathbb{N}$ be divisible by 4. Let $B_0 := [0, b^{3n/4} - 1]$ and, for $k = 1, \ldots, \frac{n}{4}$, consider the interval

$$B_k := [b^{n-1} + b^{n-2} + \cdots + b^{n-k}, b^{n-1} + b^{n-2} + \cdots + b^{n-k} + b^{3n/4} - 1].$$

Then the intervals B_k will all be n-strict (with exception of B_0), and a random integer $x \in B_k$ will have the following base-b expansion:

$$x = \underbrace{11\ldots1}_{k \text{ digits}} \underbrace{00\ldots0}_{\frac{n}{4} - k \text{ digits}} \underbrace{X_i X_{i-1} \ldots X_1}_{\frac{3n}{4} \text{ digits}}.$$

That is, x will have its first k digits equal to 1, the next $\frac{n}{4} - k$ digits equal to 0, and the remaining $\frac{3n}{4}$ digits will be i.i.d. random variables X_i taking values uniformly in the set $\{0, 1, \ldots, b-1\}$.

Let Y_0 be the random variable induced by B_0, and Y_k be induced by B_k. Then

$$H(Y_k) = H(Y_0) + k = \tau_k\left(H(Y_0)\right).$$

Recall $H(Y_0)$ has mean $\frac{3n}{4}\mu$, and variance $\frac{3n}{4}\sigma^2$. Consider an interval $I = [i_1, i_2]$ containing a set of type-C integers S, and let $\lambda > 0$. By Lemma 3.3, there exists $k' \in \left[i_1 - \left(\frac{3n}{4}\mu + \sqrt{\frac{3n}{4}}\lambda\sigma\right), i_2 - \left(\frac{3n}{4}\mu - \sqrt{\frac{3n}{4}}\lambda\sigma\right)\right]$ such that

$$\mathbb{P}\left(\tau_{k'}\left(H(Y_0)\right) \in S\right) \geq \left(1 - \frac{1}{\lambda^2}\right)\left(\frac{d_C(I)}{1 + \frac{\sqrt{3n}\lambda\sigma}{|I|}}\right). \tag{6}$$

Thus, if $I \subseteq \left[1 + \frac{3}{4}n\mu + \lambda\sigma\sqrt{\frac{3}{4}n}, \frac{1}{4}n + \frac{3}{4}n\mu - \lambda\sigma\sqrt{\frac{3}{4}n}\right]$, then $1 \leq k' \leq \frac{n}{4}$. Setting $k := k'$ produces the interval B_k, which will be n-strict with

$$d_C(B_k) \geq \left(1 - \frac{1}{\lambda^2}\right)\left(\frac{d}{1 + \frac{\sqrt{3n}\lambda\sigma}{|I|}}\right).$$

In fact, we have proven the following

Theorem 3.4. *Let $n \in \mathbb{N}$ be divisible by 4 and let $C \subset \mathbb{N}$. For $\lambda > 0$, define $J_{n,\lambda} := \left[1 + \frac{3}{4}n\mu + \lambda\sigma\sqrt{\frac{3}{4}n}, \frac{1}{4}n + \frac{3}{4}n\mu - \lambda\sigma\sqrt{\frac{3}{4}n}\right]$. Fix an interval $I \subseteq J_{n,\lambda}$. Then there exists an n-strict interval, I_2, such that $d_C(I_2) \geq d_C(I)\left(1 - \frac{1}{\lambda^2}\right)\left(\frac{1}{1 + \frac{\sqrt{3n}\sigma\lambda}{|I|}}\right)$.*

The goal for the rest of the section is to use the previous theorem iteratively to construct a sequence of intervals $\{I_i\}_{i=1}^{\infty}$, each with high type-C density, such that

each I_i is n_i-strict and the sequence $\{n_i\}_{i=1}^\infty$ grows quickly. One technical issue to worry about is that $d_C(I_{i+1}) < d_C(I_i)$ for all i. How much smaller $d_C(I_{i+1})$ is depends on how large we choose λ_i to be in each step. We wish to choose λ_i as large as possible, but choose λ_i too large and two bad things can happen: First, I_i will not be contained in J_{n_{i+1},λ_i} for any choice of n_{i+1}. Second, $\frac{\sqrt{3n}\sigma\lambda_i}{|I|}$ will not be small. We are helped by the fact that the sequence $\{n_i\}_{i=1}^\infty$ will grow super exponentially (in fact, $n_{i+1} = \Omega(b^{n_i})$). Choosing $\lambda_i = b^{n_i/8}$ in each step will work well; however, we will need the initial n_1 to be sufficiently large. The next theorem gives precise calculations. The proof follows from a number of routine calculations and estimations, some of which we have left for the appendix.

Theorem 3.5. *Suppose I is n-strict, where n satisfies bounds ($\boldsymbol{B}$). Then there exists $n_2 \geq \frac{b^{n-1}}{\mu}$, and an n_2-strict interval I_2 such that*

$$d_C(I_2) \geq d_C(I) \left(1 - b^{-n/4}\right) \left(1 - \frac{2\sigma}{\sqrt{\mu}} b^{-n/8}\right).$$

Proof. As before, let $J_{m,\lambda} := \left[1 + \frac{3}{4}m\mu + \lambda\sigma\sqrt{\frac{3}{4}m}, \frac{1}{4}m + \frac{3}{4}m\mu - \lambda\sigma\sqrt{\frac{3}{4}m}\right]$. We assumed that I is n-strict, so $|I| = b^{3n/4}$. Write I as $[a, a + b^{3n/4} - 1]$. Setting $\lambda := b^{n/8}$, we attempt to find an n_2 divisible by 4 such that $I \subseteq J_{n_2,\lambda}$. It would be prudent to consider $f(m) := 1 + \frac{3}{4}m\mu + \lambda\sigma\sqrt{\frac{3}{4}m}$, which is the left endpoint of $J_{m,\lambda}$. We first find an integer n_2 such that $f(n_2) \leq a$ and $a - f(n_2)$ is small. By Lemma 6.2 in the Appendix, assuming n satisfies bounds ($\boldsymbol{B}$), it follows that there exists n_2 such that:

- $4 \mid n_2$,
- $\frac{b^{n-1}}{\mu} \leq n_2 \leq \frac{4}{3\mu}b^n$,
- $0 \leq a - f(n_2) \leq 3\mu + 1$.

We now check that $I \subseteq J_{n_2,\lambda}$ in order to invoke Theorem 3.4. We already have that the left endpoint $f(n_2) \leq a$. It remains to check the right endpoints of I and J_{n_2}. We need to show that

$$a - 1 + b^{3n/4} \leq \frac{n_2}{4} + \frac{3n_2}{4}\mu - \lambda\sigma\sqrt{\frac{3n_2}{4}}. \tag{7}$$

The above is equivalent to

$$a - \left(\frac{3n_2}{4}\mu + \lambda\sigma\sqrt{\frac{3n_2}{4}} + 1\right) + b^{3n/4} \leq \frac{n_2}{4} - 2\lambda\sigma\sqrt{\frac{3n_2}{4}}.$$

Simplifying, the above follows from showing that

$$a - f(n_2) + b^{3n/4} + 2\lambda\sigma\sqrt{\frac{3n_2}{4}} \leq \frac{n_2}{4}.$$

Now let

$$\text{LHS} := a - f(n_2) + b^{3n/4} + 2\lambda\sigma\sqrt{\frac{3n_2}{4}}.$$

Then

$$\text{LHS} \le 3\mu + 1 + b^{3n/4} + \lambda\sigma\sqrt{3n_2}.$$

Using the facts that $\lambda = b^{n/8}$ and $n_2 \le \frac{4b^n}{3\mu}$, we get that

$$\text{LHS} \le 3\mu + 1 + b^{3n/4} + 2\frac{\sigma}{\sqrt{\mu}}b^{5n/8}.$$

Now consider $\text{RHS} := \frac{n_2}{4}$. By the assumptions on n_2, we have

$$\text{RHS} \ge \frac{b^n}{4b\mu}.$$

So (7) follows from showing that

$$3\mu + 1 + b^{3n/4} + 2\frac{\sigma}{\sqrt{\mu}}b^{5n/8} \le \frac{b^n}{4b\mu}.$$

The above is exactly the bound (**B3**). Therefore, $I \subseteq J_{n_2,\lambda}$. Thus, by applying Theorem 3.4 with $\lambda = b^{n/8}$, there exists an n_2-strict interval I_2 such that

$$d_C(I_2) \ge d_C(I)\Big(1 - \frac{1}{b^{n/4}}\Big)\Big(\frac{1}{1 + \frac{\sqrt{3n_2}\sigma b^{n/8}}{b^{3n/4}}}\Big).$$

Since $n_2 \le \frac{4}{3\mu}b^n$, it follows that

$$\frac{1}{1 + \frac{\sqrt{3n_2}\sigma b^{n/8}}{b^{3n/4}}} \ge \frac{1}{1 + \frac{2\sigma}{\sqrt{\mu}}b^{-n/8}} \ge 1 - \frac{2\sigma}{\sqrt{\mu}}b^{-n/8}.$$

Thus, we conclude that $d_C(I_2) \ge d_C(I)\left(1 - b^{-n/4}\right)\left(1 - \frac{2\sigma}{\sqrt{\mu}}b^{-n/8}\right)$. $\square$

Apply the previous theorem to our starting n_1-strict interval I_1 to get an n_2-strict interval I_2. Since $n_2 > n_1$, we can apply Theorem 3.5 again on I_2. Continuing in this manner produces a sequence of integers $\{n_i\}_{i=1}^{\infty}$ and n_i-strict intervals $\{I_i\}_{i=1}^{\infty}$ such that, for all i:

- $n_{i+1} \ge \frac{b^{n_i-1}}{\mu}$,

- $d_C(I_{i+1}) \ge d_C(I_i)\left(1 - b^{-n_i/4}\right)\left(1 - \frac{2\sigma}{\sqrt{\mu}}b^{-n_i/8}\right).$

The second condition implies that, for all i,

$$d_C(I_i) \ge d_C(I_1)\prod_{i=1}^{\infty}\left(\left(1 - b^{-n_i/4}\right)\left(1 - \frac{2\sigma}{\sqrt{\mu}}b^{-n_i/8}\right)\right). \tag{8}$$

The following fact will help simplify the above expression. For positive real numbers x and α, if $x \geq 2\alpha > 0$, then

$$1 - \alpha x^{-1} \geq \frac{1}{1 + 2\alpha x^{-1}} \geq e^{-2\alpha x^{-1}}.$$

Therefore, (8) implies that

$$d_C(I_i) \geq d_C(I_1) \cdot \exp\left(\sum_{i=1}^{\infty} -2b^{-n_i/4} - \frac{4\sigma}{\sqrt{\mu}} b^{-n_i/8}\right).$$

For all $i \in \mathbb{N}$, it holds that $n_i \geq in_1$ (it may happen that $n_2 < 2n_1$ if μ is very large, but assuming the bounds (**B**) this will not be the case). The sum in the previous inequality is the sum of two geometric series, one with ratio $r = b^{-n_1/4}$ and first term $a = -2b^{-n_1/4}$. The second has $r = b^{-n_1/8}$ and $a = \frac{-4\sigma}{\sqrt{\mu}} b^{-n_1/8}$. Recall that an infinite geometric series with $|r| < 1$, and first term a sums to

$$\frac{a}{1 - r}.$$

Therefore, the first series sums to $\frac{-2b^{-n_1/4}}{1 - b^{-n_1/4}}$, the second sums to $\frac{-4\sigma b^{-n_1/8}}{\sqrt{\mu}(1 - b^{-n_1/8})}$. After simplifying we conclude that, for all i,

$$d_C(I_i) \geq d_C(I_1) \cdot \exp\left(\frac{-2}{b^{n_1/4} - 1} + \frac{-4\sigma}{\sqrt{\mu}(b^{n_1/8} - 1)}\right).$$

Thus, we have proven the following

Theorem 3.6. *Assume there exists n_1 satisfying the bounds (**B**) and an n_1-strict interval I_1. Then, for all $N \in \mathbb{N}$, there exists $n > N$ and an n-strict interval I such that*

$$d_C(I) \geq d_C(I_1) \cdot \exp\left(\frac{2}{1 - b^{n_1/4}} + \frac{4\sigma}{\sqrt{\mu}(1 - b^{n_1/8})}\right).$$

4. Main Result

As in the previous section we continue to assume that H is a b-happy function with digit mean μ and digit variance σ^2. Also, we assume that we have experimentally found a suitable starting n_1-strict interval, I_1, with large type-C density for some $C \subset \mathbb{N}$. As in Section 2, for positive integers m, let Y_m denote the random variable induced by the interval $[0, b^m - 1]$.

In this section we give a proof of the following

Theorem 4.1. *Suppose I_1 is n_1-strict, where n_1 satisfies bounds ($\mathbf{B}$). Let $\bar{d}$ denote the upper density of the set of type-C integers. Then*

$$\bar{d} \geq d_C(I_1) \cdot exp\left(\frac{2}{1 - b^{n_1/4}} + \frac{4\sigma}{\sqrt{\mu}(1 - b^{n_1/8})}\right).$$

The digit mean and digit variance for the case $(e, b) = (2, 10)$ are 28.5 and 721.05 respectively. In this case, if $n > 13$, then it satisfies bounds ($\mathbf{B}$). After performing a computer search we find that the density of happy numbers in the interval $[10^{403}, 10^{404} - 1]$ is at least .185773; thus, there exists a 404-strict interval containing at least this density of happy numbers as a subset. Consider

$$\delta(n) := \left(\frac{2}{1 - b^{n/4}} + \frac{4\sigma}{\sqrt{\mu}(1 - b^{n/8})}\right).$$

Plugging in the value for n, we find that $e^{\delta(404)} > 1 - 10^{-49}$. Thus, by Theorem 4.1, the upper density of type-$\{1\}$ integers is at least .1857729. For the lower density, the type-$\{1\}$ density of $[10^{2367}, 10^{2368} - 1]$ is at most .11379. This implies that there is a 2368-strict interval with type-$\{4\}$ density at least $1 - .11379$. We can then apply the main result to conclude that the upper density of type-$\{4\}$ integers is at least $1 - .1138$. This gives the following

Corollary 4.2. *Let $\underline{d}$ and $\bar{d}$ be the lower and upper density of $(2, 10)$-happy numbers respectively. Then $\underline{d} < .1138$ and $\bar{d} > .18577$.*

The proof of Theorem 4.1 is somewhat technical despite having a rather intuitive motivation. For the sake of clarity we first give a sketch of how to use Theorems 3.6 and 2.5 in order to prove a lower bound on the upper density of type-C numbers.

Given our starting interval I_1, apply Theorem 3.6 to construct an n-strict interval I, where

$$d_C(I) \geq (1 - o(1))d_C(I_1).$$

Do this with n large enough as to make all the following error estimations arbitrarily small. Pick m_1 such that μm_1 (i.e., the mean of $H(Y_{m_1})$) lands in the interval I. Since $I \subseteq [b^{n-1}, b^n]$, we have that $m_1 = \Theta(b^n)$. This implies that the standard deviation of $H(Y_{m_1})$ is roughly $b^{n/2}$. This will be much less than $|I| = b^{3n/4}$ and thus the bulk of the distribution of $H(Y_{m_1})$ will lie in the interval I.

Next, use Lemma 3.3 with a large λ to find an integer k for which

$$\mathbb{P}\left(\tau_k\big(H(Y_{m_1})\big) \text{ is type-C}\right) \geq (1 - o(1))d_C(I_1).$$

Note that this k will be smaller than $|I| = b^{3n/4}$ and that the mean of $\tau_k\big(H(Y_{m_1})\big)$ is equal to $\mu m_1 + k$. Clearly, there exists an integer m_2 such that

$$|\mu m_2 - (\mu m_1 + k)| \leq \mu.$$

Consider the random variable $H(Y_{m_2})$. The means of $H(Y_{m_2})$ and $\tau_k\left(H(Y_{m_1})\right)$ are almost equal. Since k is much smaller relative to m_1 and m_2, the variance of these two distributions will be close. Furthermore, the distributions of $H(Y_{m_2})$ and $\tau_k\left(H(Y_{m_1})\right)$ are asymptotically locally normal, so we may apply the local limit law to conclude that the distributions are point-wise close near the means. Thus,

$$\mathbb{P}\left(H(Y_{m_2}) \text{ is type-C}\right) \geq (1 - o(1))d_C(I_1).$$

This implies that the interval $[0, b^{m_2} - 1]$ has type-C density at least $d_C(I_1)(1 - o(1))$. Note that in the above analysis, we may take n (and therefore m_2) to be arbitrarily large. This lower bounds the upper density of type-C integers by $d_C(I_1)(1 - o(1))$. In fact, the only contribution to the error term is from the application of Theorem 3.6 (the rest of the error tends to 0 as n tends to infinity).

4.1. Some Lemmas

We will now begin to prove the main result. We have broken some of the pieces down for 3 lemmas. The proofs primarily consist of calculations and we leave them for after the proof of the main result. Note that Lemma 4.5 (part 1) is the only place where the local limit law is used.

Lemma 4.3. *There exists a sufficiently large N such that, if $n > N$ and I is an n-strict interval, then there exists $m \in \mathbb{N}$ with the property that*

$$[\mu m - \sigma m^{5/8}, \mu m + \sigma m^{5/8}] \subseteq I.$$

Lemma 4.4. *Let $\epsilon > 0$ be given (assume as well that $\epsilon \leq 1$). Let $\lambda := \sqrt{\frac{6}{\epsilon}}$. Then there exists a sufficiently large N such that, if n, m_1, and I satisfy:*

- $n > N$,
- $m_1 \in [\frac{b^{n-1}}{\mu}, \frac{b^n}{\mu}]$,
- *I is n-strict,*

then the following hold:

1. *$\lambda \leq m_1{}^{1/8}$,*
2. *$\left| 1 - \frac{1}{1 + \frac{2\lambda\sigma\sqrt{m_1}}{b^{3n/4}}} \right| \leq \frac{\epsilon}{6}$.*

Lemma 4.5. *Let $\epsilon > 0$ be given (assume as well that $\epsilon \leq 1$). Let $T := \frac{2\sqrt{6}}{\sqrt{\epsilon}}$, and $\lambda := \frac{\sqrt{6}}{\sqrt{\epsilon}}$. Then there exists a sufficiently large N such that, if n, m_1, m_2, k, and I satisfy:*

- $n > N$,

- $m_1 \in [\frac{b^{n-1}}{\mu}, \frac{b^n}{\mu}], m_2 \in [\frac{b^{n-2}}{\mu}, \frac{b^{n+1}}{\mu}]$,

- $|k| \leq b^{3n/4}$,

- $|\mu m_1 + k - \mu m_2| \leq \mu$,

- I is n-strict,

then the following hold:

1. *For $i \in \{1, 2\}$,* $\displaystyle\max_{|t| \leq T} \left| 1 - \frac{\mathbb{P}(H(Y_{m_i}) = \lfloor \mu m_i + t\sigma\sqrt{m_i} \rfloor)}{\frac{e^{t^2/2}}{\sqrt{2\pi m_i}\sigma}} \right| \leq \frac{\epsilon}{6}$.

2. $\left| 1 - \sqrt{\frac{m_1}{m_2}} \right| \leq \frac{\epsilon}{6}$.

3. *For any real numbers t_1 and t_2, where $t_1 \in [\frac{-T}{2}, \frac{T}{2}]$ and*

$$\mu m_1 + k + t_1 \sigma \sqrt{m_1} = \mu m_2 + t_2 \sigma \sqrt{m_2},$$

   it holds that $t_2 \in [-T, T]$ and $|1 - e^{t_1^2 - t_2^2/2}| \leq \frac{\epsilon}{6}$.

4.2. Proof of Theorem 4.1

Proof. In order to lower bound the upper density of type-C integers, it suffices to show that, for all $\epsilon > 0$ and $N_1 \in \mathbb{N}$, there exists $m > N_1$ such that

$$d_C([0, b^m - 1]) \geq d_C(I_1) \cdot \exp\left(\frac{2}{1 - b^{n/4}} + \frac{4\sigma}{\sqrt{\mu}(1 - b^{n/8})} \right)(1 - \epsilon).$$

Let ϵ and N_1 be arbitrary (with $\epsilon \leq 1$). Set $T := \frac{2\sqrt{6}}{\sqrt{\epsilon}}$. Also, in anticipation of applying Lemma 3.3, set $\lambda := \frac{\sqrt{6}}{\sqrt{\epsilon}}$.

First, pick $N > N_1$ large enough to apply Lemmas 4.3, 4.4, and 4.5. By Theorem 3.6, there exists an n-strict interval I, where $n > N$ and

$$d_C(I) \geq d_C(I_1) \cdot \exp\left(\frac{2}{1 - b^{n_1/4}} + \frac{4\sigma}{\sqrt{\mu}(1 - b^{n_1/8})} \right). \tag{9}$$

For $m \in \mathbb{N}$, let

$$J_m := [\mu m - \sigma m^{5/8}, \mu m + \sigma m^{5/8}].$$

Recall that $\mathbf{E}[H(Y_m)] = \mu m$ and $\mathbf{Var}[H(Y_m)] = \sigma^2 m$. Hence, J_m is where the bulk of the distribution of $H(Y_m)$ lands. Pick m_1 such that $J_{m_1} \subseteq I$ (the existence of such m_1 is guaranteed by Lemma 4.3). Note that $m_1 \in [\frac{b^{n-1}}{\mu}, \frac{b^n}{\mu}]$ since I is n-strict.

Let S be the set of type-C integers in I. Apply Lemma 3.3 on the random variable $H(Y_{m_1})$ to find an integer k such that

$$\mathbb{P}\Big(\tau_k\big(H(Y_{m_1})\big) \in S\Big) \geq d_C(I)\left(1 - \frac{1}{\lambda^2}\right)\left(\frac{1}{1 + \frac{2\lambda\sigma\sqrt{m_1}}{|I|}}\right). \tag{10}$$

Since $J_{m_1} \subseteq I$ and $|I| = b^{3n/4}$, it follows that $k \leq b^{3n/4}$.

Let $\tau_k(J_{m_1}) := [a + k, b + k]$, where $J_{m_1} = [a, b]$. Let S' be the set of type-C integers in interval $\tau_k(J_{m_1})$. Recall the proof of Lemma 3.3. In particular, we applied Lemma 3.2 after ignoring the tails of the distribution of $H(Y_{m_1})$ outside of $\lambda\sigma\sqrt{m_1}$ from the mean. Since $\lambda \leq m_1^{1/8}$ (by Lemma 4.4, part 1), we may replace (10) by the stronger conclusion that

$$\sum_{i \in S'} \mathbb{P}\Big(\tau_k\big(H(Y_{m_1})\big) = i\Big) \geq d_C(I)\left(1 - \frac{1}{\lambda^2}\right)\left(\frac{1}{1 + \frac{2\lambda\sigma\sqrt{m_1}}{|I|}}\right).$$

Using the assumption that $\lambda = \sqrt{\frac{6}{\epsilon}}$ and part 2 of Lemma 4.4, we simplify the above as

$$\sum_{i \in S'} \mathbb{P}\Big(\tau_k\big(H(Y_{m_1})\big) = i\Big) \geq d_C(I)\left(1 - \frac{\epsilon}{6}\right)^2. \tag{11}$$

Now pick $m_2 \in \mathbb{N}$ such that $|m_1\mu + k - m_2\mu| \leq \mu$. Since $|k| \leq b^{3n/4}$, it follows that $m_2 \in [\frac{b^{n-2}}{\mu}, \frac{b^{n+1}}{\mu}]$. In particular m_1, m_2, n, k, and I now all satisfy the conditions of Lemma 4.5. It remains to show that near the mean of $\tau_k\big(H(Y_{m_1})\big)$, the distributions of $\tau_k\big(H(Y_{m_1})\big)$ and $H(Y_{m_2})$ are similar. This will imply that the interval $[0, b^{m_2} - 1]$ contains a large density of type-C integers. Making this precise, we prove the following

Claim 1. For integers $i \in \tau_k\big(J_{m_1}\big)$,

$$\frac{\mathbb{P}\big(H(Y_{m_2}) = i\big)}{\mathbb{P}\big(\tau_k\big(H(Y_{m_1})\big) = i\big)} \geq \left(1 - \frac{\epsilon}{6}\right)^4.$$

Proof. Let $i \in \tau_k(J_{m_1})$ be fixed and pick t_1, t_2 such that

$$i = \mu m_1 + k + t_1\sigma\sqrt{m_1} = \mu m_2 + t_2\sigma\sqrt{m_2}.$$

It is important now that we had chosen $\lambda = \frac{T}{2}$, this implies that $|t_2| \leq T$ (see Lemma 4.5 part 3). We can use the local limit law to estimate the distributions of $\tau_k\big(H(Y_{m_1})\big)$ and $H(Y_{m_2})$. By Lemma 4.5 part 1,

$$\mathbb{P}\big(H(Y_{m_2}) = i\big) = \mathbb{P}\big(H(Y_{m_2}) = \mu m_2 + t_2\sigma\sqrt{m_2}\big) \geq \frac{e^{-t_2^2/2}}{2\pi\sigma\sqrt{m_2}}\left(1 - \frac{\epsilon}{6}\right)$$

and

$$\mathbb{P}\Big(\tau_k\big(H(Y_{m_1})\big) = i\Big) = \mathbb{P}\big(H(Y_{m_1}) = \mu m_1 + t_1\sigma\sqrt{m_1}\big) \leq \frac{e^{-t_1{}^2/2}}{2\pi\sigma\sqrt{m_1}}\left(1 + \frac{\epsilon}{6}\right).$$

Hence,

$$\frac{\mathbb{P}\big(H(Y_{m_2}) = i\big)}{\mathbb{P}\Big(\tau_k\big(H(Y_{m_1})\big) = i\Big)} \geq \exp\big((t_1{}^2 - t_2{}^2)/2\big)\,\frac{\sqrt{m_1}}{\sqrt{m_2}}\frac{(1 - \frac{\epsilon}{6})}{(1 + \frac{\epsilon}{6})}.$$

The above, by Lemma 4.5 parts 2 and 3, is at least $\left(1 - \frac{\epsilon}{6}\right)^4$. $\qquad\square$

Putting it all together, we have shown that

$$\begin{aligned}
d_C([0, b^{m_2} - 1]) &\geq \sum_{i \in S'} \mathbb{P}\big(H(Y_{m_2}) = i\big) \\
&= \sum_{i \in S'} \frac{\mathbb{P}\big(H(Y_{m_2}) = i\big)}{\mathbb{P}\Big(\tau_k\big(H(Y_{m_1})\big) = i\Big)}\mathbb{P}\Big(\tau_k\big(H(Y_{m_1})\big) = i\Big). \\
&\geq \left(1 - \frac{\epsilon}{6}\right)^4 \sum_{i \in S'} \mathbb{P}\Big(\tau_k\big(H(Y_{m_1})\big) = i\Big) && \text{(Claim 1)} \\
&\geq d_C(I)\left(1 - \frac{\epsilon}{6}\right)^6 && \text{(Equation 11)} \\
&\geq d_C(I)(1 - \epsilon).
\end{aligned}$$

To conclude the proof, equation (9) implies that

$$d_C([0, b^{m_2} - 1]) \geq d_C(I_1) \cdot \exp\left(\frac{2}{1 - b^{n_1/4}} + \frac{4\sigma}{\sqrt{\mu}(1 - b^{n_1/8})}\right)(1 - \epsilon).$$

$\qquad\square$

We conclude this section with the proofs of the lemmas used in the previous theorem.

Proof of Lemma 4.3

Proof. For $m \in \mathbb{N}$, let $J_m := [\mu m - \sigma m^{5/8}, \mu m + \sigma m^{5/8}]$. If I is an n-strict interval, then $I \subseteq [b^{n-1}, b^n - 1]$. Note that $\mu m \in I$ implies that $m = O(b^n)$. This in turn shows that

$$|J_m| = O(b^{5n/8}) \ll |I| = b^{3n/4}.$$

Comparing the growth rates of $|J_m|$ and $|I|$ it is clear that we can pick N_1 large enough such that $n > N_1$ implies that there exists m with $J_m \subseteq I$. $\qquad\square$

Proof of Lemma 4.4

Proof. We find N_1, N_2 for the two parts respectively and then choose $N = \max(N_1, N_2)$.

1. λ is a fixed constant here and it is assumed that $m_1 \geq \frac{b^{n-1}}{\mu}$, so the result is trivial (this gives N_1).

2. For $x > 0$, to show that $\left|1 - \frac{1}{1+x}\right| \leq \frac{\epsilon}{6}$, it is equivalent to prove that

$$\left(1 - \frac{\epsilon}{6}\right)(1 + x) \leq 1 \leq \left(1 + \frac{\epsilon}{6}\right)(1 + x).$$

The above follows if $x \leq \frac{\epsilon}{6}$. Thus, the result will follow by finding N large enough such that $\frac{2\sigma\lambda\sqrt{m_1}}{b^{3n/4}} \leq \frac{\epsilon}{6}$. Using the assumption that $m_1 \leq \frac{b^n}{\mu}$, we get

$$\frac{2\sigma\lambda\sqrt{m_1}}{b^{3n/4}} \leq \frac{2\sigma\lambda}{\sqrt{\mu}b^{n/4}}.$$

This is equivalent to

$$\frac{12\sigma\lambda}{\sqrt{\mu}\epsilon} \leq b^{n/4}.$$

Hence, picking $N_2 \geq 4\log_b(\frac{12\sigma\lambda}{\sqrt{\mu}\epsilon})$ suffices.

$\square$

Proof of Lemma 4.5

Proof. We first find $N_1, N_2,$ and N_3 for the three parts respectively, and then define $N := \max(N_1, N_2, N_3)$.

1. For each m, we have

$$H(Y_m) = \sum_{i=1}^{m} H(X_i),$$

where each X_i is uniform in the set $\{0, 1, \cdots, b - 1\}$. Recall that it is assumed that $H(0) = 0$ and $H(1) = 1$. In particular, the random variables $H(X_i)$ satisfy the aperiodic condition required by Theorem 2.5. Thus, the result follows from applying Theorem 2.5 to the sum $\sum_{i=1}^{m} H(X_i)$ with finite interval $[-T, T]$. Fix M large enough such that $m > M$ implies that the $O(m^{-1/2})$ term in Theorem 2.5 is less than $\frac{\epsilon}{6}$. By assumption, we have that both m_1 and m_2 are larger than $\frac{b^{n-2}}{\mu}$. Hence, setting $N_1 = \log_b(\mu M) + 2$ suffices.

2. Ignoring the square root, it suffices to show that

$$\left|1 - \frac{m_1}{m_2}\right| \leq \frac{\epsilon}{6}. \tag{12}$$

By assumption

$$|\mu m_1 + k - \mu m_2| \leq \mu.$$

Dividing through by μm_2, it follows that

$$\left| 1 - \frac{m_1}{m_2} - \frac{k}{\mu m_2} \right| \le \frac{1}{m_2}.$$

This implies that

$$\frac{k}{\mu m_2} - \frac{1}{m_2} \le 1 - \frac{m_1}{m_2} \le \frac{1}{m_2} + \frac{k}{\mu m_2}.$$

Thus, (12) follows from showing that $\frac{1}{m_2} + \frac{k}{\mu m_2} \le \frac{\epsilon}{6}$. Using the assumption that $m_2 \ge \frac{b^{n-2}}{\mu}$ and $|k| \le b^{3n/4}$, it follows that

$$\frac{1}{m_2} + \frac{k}{\mu m_2} \le \mu b^{-(n-2)} + b^{2-(n/4)}.$$

Therefore, picking $N_2 := \max(\log_b(\frac{12\mu}{\epsilon}) + 2, 4\log_b(\frac{12}{\epsilon}) + 2)$ suffices.

3. We first find N' such that $n > N'$ implies that $t_2 \in [-T, T]$. We start with the assumption that

$$\mu m_1 + k + t_1 \sigma \sqrt{m_1} = \mu m_2 + t_2 \sigma \sqrt{m_2}.$$

Using the facts that $|\mu m_1 + k - \mu m_2| \le \mu$ and $|t_1| \le \frac{T}{2}$, this implies that

$$|t_2| \le \frac{\mu}{\sigma \sqrt{m_2}} + \frac{T\sqrt{m_1}}{2\sqrt{m_2}}.$$

We assumed that $m_2 \ge \frac{b^{n-2}}{\mu}$. Also, in part (2) we showed that there exists N_2 such that $n > N_2$ implies that $\frac{\sqrt{m_1}}{\sqrt{m_2}} \le \left(1 + \frac{\epsilon}{6}\right) \le \frac{7}{6}$. Hence, if we take $N' > N_2$, it follows that

$$|t_2| \le \frac{\mu^2}{\sigma b^{n-2/2}} + \frac{7T}{12}.$$

Pick $N' > N_2$ large enough such that $n > N'$ implies that $\frac{\mu^2}{\sigma b^{n-2/2}} \le \frac{5T}{12}$. This will take care of the size of t_2.

Now we must show that there exists N'' large enough such that $n > N''$ implies that

$$\left| 1 - e^{(t_1{}^2 - t_2{}^2)/2} \right| \le \frac{\epsilon}{6}.$$

For a real number x, if we wish to show that $|1 - e^x| \le \frac{\epsilon}{6}$, it is equivalent to prove that

$$\ln\left(1 - \frac{\epsilon}{6}\right) \le x \le \ln\left(1 + \frac{\epsilon}{6}\right).$$

Set $\epsilon^* := \min\left(\ln\left(1 + \frac{\epsilon}{6}\right), \left|\ln\left(1 - \frac{\epsilon}{6}\right)\right|\right)$. We find N'' such that $n > N''$ implies that

$$\left| \frac{t_2{}^2 - t_1{}^2}{2} \right| \le \epsilon^*.$$

It was assumed that

$$\mu m_1 + k + t_1 \sigma \sqrt{m_1} = \mu m_2 + t_2 \sigma \sqrt{m_2}.$$

Equivalently

$$\mu m_1 + k - \mu m_2 = t_2 \sigma \sqrt{m_2} - t_1 \sigma \sqrt{m_1}.$$

Applying the assumption that the left hand side is at most μ and dividing both sides by $\sigma \sqrt{m_2}$, we get

$$\left| t_2 - t_1 \sqrt{\frac{m_1}{m_2}} \right| \leq \frac{\mu}{\sqrt{m_2} \sigma}.$$

Rearranging, this gives

$$\left| t_2 - t_1 + t_1 (1 - \sqrt{\frac{m_1}{m_2}}) \right| \leq \frac{\mu}{\sqrt{m_2} \sigma}.$$

This implies that

$$|t_2 - t_1| \leq \frac{\mu}{\sqrt{m_2} \sigma} + \left| t_1 (1 - \sqrt{\frac{m_1}{m_2}}) \right|.$$

We assumed that $m_2 \geq \frac{b^{n-2}}{\mu}$ and $|t_1| \leq T$. By part (2), if we chose $N'' > N_2$, then

$$\left| 1 - \sqrt{\frac{m_1}{m_2}} \right| \leq \mu b^{-(n-2)} + b^{-(n-2)/4}.$$

Putting this together, it follows that

$$\left| \frac{t_2^2 - t_1^2}{2} \right| = \left| \left(\frac{t_2 + t_1}{2} \right) (t_2 - t_1) \right| \leq T \left(\frac{\mu^{3/2} b^{-(n-2)/2}}{\sigma} + T(\mu b^{-(n-2)} + b^{-(n-2)/4}) \right).$$

Now, since T, μ, σ, and b are all constants, it follows that the right hand side tends to 0 as n goes to infinity. Therefore, there exists N'' such that $n > N''$ implies that the right hand side is at most ϵ^*. Finally, set $N_3 := \max(N', N'')$. $\qquad \square$

5. Experimental Data

The data[3] presented in this section is the result of short computer searches, so the bounds surely can be improved with more computing time. Floating point approximation with conservative rounding was used.

---

[3]Data generated by fellow graduate student, Patrick Devlin.

5.1. Finding an Appropriate n-Strict Interval

If n is divisible by 4 and the interval $[b^{n-1}, b^n - 1]$ has type-C density d, then there exists an n-strict interval with type-C density at least d to which we may apply Theorem 4.1. The type-C density of $[b^{n-1}, b^n - 1]$ for various n can be quickly calculated by first computing the densities of intervals of the form $[0, b^n - 1]$; the algorithm which does this was discussed in Section 2. After an appropriate n-strict interval is found, we check to see that n satisfies bounds (**B**), compute the error term, and find the desired bound. Our results show that in almost all cases, the asymptotic density of type-C numbers does not exist.

5.2. Explanation of Results

The following information is given in tables (in the order of column in which they appear):

1. The cycle C in which type-C densities are being computed.

2. The lower bound on the upper density (UD) implied by Theorem 4.1.

3. The upper bound on the lower density (LD) implied by Theorem 4.1.

4. The n such that the interval $[b^{n-1}, b^n - 1]$ is used to find the bound (denoted as UD n or LD n).

5. The $\delta(n) = \left(\frac{2}{1 - b^{n/4}} + \frac{4\sigma}{\sqrt{\mu}(1 - b^{n/8})} \right)$ part of the error term for Theorem 4.1 (we only present an upper bound on $|\delta(n)|$, the true number is always negative). In all cases the error is small enough not to affect the bounds as we only give precision of about 5 or 6 decimal places.

5.2.1. Cubing the Digits in Base-10

In this case, if $n > 16$, then it satisfies bounds (**B**). Table 1 shows the results for the cycles when studying the $(3, 10)$-happy function. There are 9 possible cycles. Figure 2 graphs the density of type-$\{1\}$ integers less than 10^n. It is easy to prove, in this case, that $3 \mid n$ if an only if n is type-$\{153\}$.

5.2.2. A More General Function

In order to emphasize the generality of Theorem 4.1, we consider the function in base-7 with digit sequence $[0, 1, 7, 4, 17, 9, 13]$. There are only two cycles for this function, both are fixed points. Written in base-10 the cycles are $\{1\}$ and $\{20\}$. Figure 3 graphs the relative density of type-$\{1\}$ numbers. Table 2 shows the bounds derived. As there are only two cycles, we focus on the cycle $\{1\}$. In this case, if $n > 12$, then it satisfies bounds (**B**).

Cycle	UD	LD	UD n	LD n	UD $\delta(n)$	LD $\delta(n)$
$\{1\}$	$> .028219$	$< .0049761$	10^{864}	10^{132}	$< 10^{-106}$	$< 10^{-14}$
$\{55,250,133\}$	$> .06029$	$< .0447701$	10^{208}	10^{964}	$< 10^{-24}$	$< 10^{-118}$
$\{136,244\}$	$> .024909$	$< .006398$	10^{204}	10^{420}	$< 10^{-23}$	$< 10^{-51}$
$\{153\}$	$= \frac{1}{3}$	$= \frac{1}{3}$	N/A	N/A	N/A	N/A
$\{160,217,352\}$	$> .050917$	$< .03184$	10^{160}	10^{456}	$< 10^{-18}$	$< 10^{-56}$
$\{370\}$	$> .19905$	$< .16065$	10^{276}	10^{560}	$< 10^{-32}$	$< 10^{-68}$
$\{371\}$	$> .30189$	$< .288001$	10^{836}	10^{420}	$< 10^{-102}$	$< 10^{-50}$
$\{407\}$	$> .04532$	$< .0314401$	10^{420}	10^{836}	$< 10^{-50}$	$< 10^{-103}$
$\{919,1459\}$	$> .04425$	$< .01843$	10^{916}	10^{120}	$< 10^{-112}$	$< 10^{-13}$

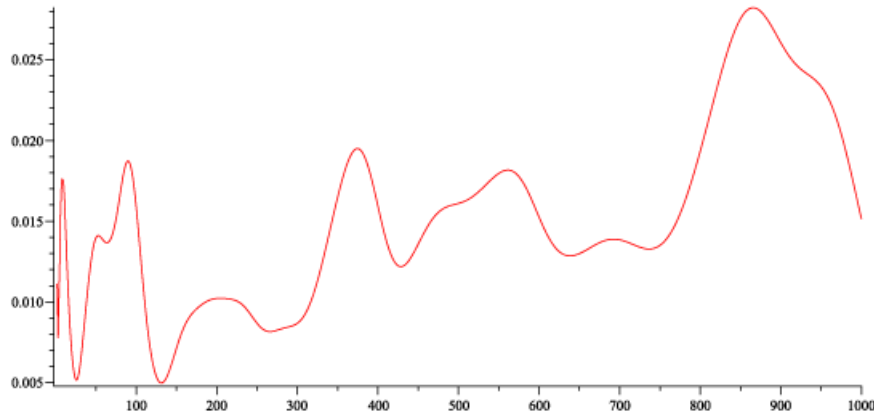Table 1: Bounds for the cycles appearing for the $(3, 10)$-function

Figure 2: Density of type-$\{1\}$ integers in the interval $[0, 10^n - 1]$ for the $(3, 10)$-function

6. Appendix

Lemma 6.1. *Fix $a > 0$. Assume that $f : \mathbb{R}^+ \to \mathbb{R}^+$ has continuous first and second derivatives such that, for all $x \in \mathbb{R}^+$, $f'(x) > 0$ and $f''(x) < 0$. Also, assume that $\lim_{x\to\infty} f(x) = \infty$. Furthermore, suppose we have $x^* \in \mathbb{R}^+$ such that $f(x^* + 1) \leq a$. Then there exists $n \in \mathbb{N}$ such that $n \geq x^*$ and $0 \leq a - f(n) \leq f'(x^*)$.*

Proof. This follows from a first order Taylor approximation of the function f. Let x^* such that $f(x^*+1) \leq a$ be given. Set $n := \sup\{m \in \mathbb{N} | f(m) \leq a\}$. Since f is strictly

Cycle	UD	LD	UD n	LD n	UD $\delta(n)$	LD $\delta(n)$
$\{1\}$	$> .9858$	$< .94222$	7^{176}	7^{384}	$< 10^{-17}$	$< 10^{-40}$

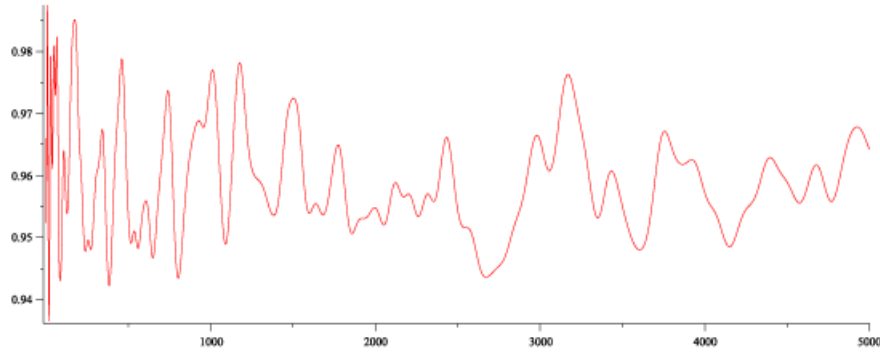Table 2: Bounds for the cycles appearing for the function with digit sequence $[0, 1, 7, 4, 17, 9, 13]$

Figure 3: Density of type-$\{1\}$ integers in the interval $[0, 7^n - 1]$ for digit sequence $[0, 1, 7, 4, 17, 9, 13]$

increasing and unbounded, this n exists. Note that $f(n) \leq a$ and $f(n + 1) > a$. It also follows that $n \geq x^*$, as otherwise $\lceil x^* \rceil$ would be the supremum. By the concavity of f, we have

$$f(n + 1) - f(n) \leq f'(n) \leq f'(x^*).$$

However, $f(n + 1) > a$, so we conclude that $0 \leq a - f(n) \leq f'(x^*)$. $\square$

Lemma 6.2. *Let n be a positive integer, $\lambda = b^{n/8}$, and $a \in [b^{n-1}, b^n]$. Let μ and σ be the digit mean and variance of some b-happy function H. Also, assume that n satisfies bounds (**B**). Let $f(n) := 1 + \frac{3}{4}\mu n + \lambda\sigma\sqrt{\frac{3}{4}n}$. Then there exists an integer n_2 such that:*

- $\frac{b^{n-1}}{\mu} \leq n_2 \leq \frac{4}{3\mu}b^n$,

- $4 \mid n_2$,

- $0 \leq a - f(n_2) \leq 3\mu + 1$.

Proof. Since we require that $4 \mid n_2$, we apply Lemma 6.1 on the function

$$g(m) = f(4m) = 1 + 3\mu m + \lambda\sigma\sqrt{3m}.$$

Let $x^* := \frac{b^{n-1}}{4\mu}$. We first check that $g(x^* + 1) \leq a$. By assumption, $a \geq b^{n-1}$. Therefore, we need to show that

$$1 + 3\mu\Big(\frac{b^{n-1}}{4\mu} + 1\Big) + b^{n/8}\sigma\sqrt{3\Big(\frac{b^{n-1}}{4\mu} + 1\Big)} \leq b^{n-1}.$$

Simplifying the above, it suffices to show that

$$1 + 3\mu + b^{5n/8}\sigma\sqrt{\frac{3}{4b\mu} + 3b^{-n}} \leq \frac{b^{n-1}}{4}. \tag{13}$$

To keep the results of this paper as general as possible, we only assumed that $\mu \geq \frac{1}{b}$ (this would correspond to the quite uninteresting b-happy function H which maps all digits to 0 except for the digit 1). Also it is clear that $b^n \geq 3$, and therefore

$$\frac{3}{4b\mu} + 3b^{-n} \leq \frac{3}{4} + 1 \leq 2.$$

Plugging this in and rearranging, we see that equation (13) follows if

$$4(1 + 3\mu + \sqrt{2}\sigma b^{5n/8}) \leq b^{n-1}.$$

This is exactly the bound (**B1**) and is true by assumption. Therefore, by Lemma 6.1, there exists $m \in \mathbb{N}$ such that

$$0 \leq a - g(m) \leq g'(x^*).$$

Also,

$$g'(x^*) = 3\mu + \frac{\sqrt{3}\sigma b^{n/8}}{2\sqrt{\frac{b^{n-1}}{4\mu}}} = 3\mu + \sqrt{3\mu b}\sigma b^{-3n/8}.$$

Again, by the assumption (**B2**) on n, the previous statement is bounded above by $3\mu + 1$. Set $n_2 := 4m$. Then $4 \mid n_2, n_2 \geq \frac{b^{n-1}}{\mu}$, and $0 \leq a - f(n_2) \leq 3\mu + 1$. Finally, we note that $f(\frac{4}{3\mu}b^n) > a$ and, since f is strictly increasing, we conclude that $n_2 \leq \frac{4}{3\mu}b^n$. $\qquad\square$

References

[1] E. El-Sedy and S. Siksek, On happy numbers. Rocky Mountain J. Math. **30** (2000), 565-570.

[2] P. Flajolet and R. Sedgewick, Analytic Combinatorics, Cambridge University Press, 2009.

[3] H.G. Grundman, E.A. Teeple, Sequences of generalized happy numbers with small bases, J. Integer Seq. (2007), Article 07.1.8.

[4] R.K. Guy, Unsolved Problems in Number Theory, Springer-Verlag, New York, 2nd Edition, 2004.

[5] D. Moews, Personal communication, November 2011.

[6] Hao Pan, On consecutive happy numbers, J. Number Theory **128 (6)** (2008), 1646-1654.