



DIGITAL PROBLEMS IN THE THEORY OF PARTITIONS

Joseph Vandehey

Department of Mathematics, University of Texas at Tyler, Texas
 jvandehey@uttyler.edu

Received: 5/31/23, Accepted: 3/6/24, Published: 5/27/24

Abstract

Let $\mathcal{P}(n)$ denote the multi-set of all parts of all partitions of n . We study various digital questions concerning $\mathcal{P}(n)$, and show that it satisfies the generalized Benford's law and is also asymptotically normal regardless of base.

1. Introduction

Let $p(n)$ denote the usual partition function, counting all possible partitions of n , and let $\mathcal{P}(n)$ denote the multi-set of all parts in partitions of n , with appropriate multiplicity. For example, if $n = 5$, then the partitions of n are

5, 4+1, 3+2, 3+1+1, 2+2+1, 2+1+1+1, 1+1+1+1+1.

So we would have $p(5) = 7$ and

$$\mathcal{P}(5) = \{5, 4, 3, 3, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1\}.$$

We will be interested in studying the following two digital properties.

- We say that a sequence of multi-sets $\mathcal{A}(n)$ satisfies *the generalized Benford's law* if the proportion of elements of $\mathcal{A}(n)$ that start with a number c in base b approaches $\log_b(1 + 1/c)$ as n goes to infinity.¹
- We say that a sequence of multi-sets $\mathcal{A}(n)$ is *asymptotically normal* if the number of times a base- b string s appears within the digits of the elements of $\mathcal{A}(n)$ is proportional to $b^{-|s|}$ (relative to all substrings of length $|s|$) as n goes to infinity.

Throughout this paper, we will assume we are working with a fixed base $b \in \mathbb{N}$, $b \geq 2$, and we will often elide the difference between positive integers and strings

DOI: 10.5281/zenodo.11352985

¹The standard version of Benford's law assumes that c is a one-digit number.

(which are sequences of base- b digits) as there is a natural bijection between them, provided one ignores strings that start with 0. Also, $|s|$ refers to the number of digits in the string s .

Our main result will be the following.

Theorem 1. *The sequence of multi-sets $\mathcal{P}(n)$ satisfies the generalized Benford’s law and is asymptotically normal.*

The fact that the numbers $p(n)$ themselves satisfy Benford’s law was proved in [1]. Another Benford’s law result relating to partitions was covered in [12].

Our investigations were inspired by a question regarding normal numbers that Nathan Jones asked of the author at a conference. A number $x \in [0, 1)$ with base- b expansion

$$x = 0.a_1a_2a_3 \cdots, \quad a_i \in \{0, 1, \dots, b - 1\}$$

is said to be base- b normal if every base- b string s (including those beginning with 0) appears in $a_1a_2 \cdots a_n$ with proportion approaching $b^{-|s|}$ as n goes to infinity. While the normality of many commonly used irrationals is still unknown, there have been a variety of constructions of normal numbers. One common method is to take a sequence of strings s_n related to some function or arithmetic data and concatenate them to form a number $x = 0.(s_1)(s_2)(s_3) \cdots$. Most famously, Champernowne [2] showed that concatenating the integers in order resulted in a normal number, so

$$x = 0.12345678910111213141516 \cdots$$

is base-10 normal. Other constructions of this type include concatenating primes [3], concatenating values of polynomials [5], concatenating values of the totient function [15], and concatenating values of the largest prime divisor function [6].

Jones asked if one could concatenate the partition function $p(n)$ and obtain a normal number. That question remains open, due to the rate of growth of $p(n)$.² Instead, we will show the following related result.

Theorem 2. *Let $\overline{\mathcal{P}(n)}$ denote the string formed by concatenating the elements of $\mathcal{P}(n)$ in order from greatest to least. Then $0.(\overline{\mathcal{P}(1)})(\overline{\mathcal{P}(2)})(\overline{\mathcal{P}(3)}) \cdots$ is normal.*

The choice of ordering elements from greatest to least is not necessary; any order would work, but we have chosen a specific one to give a compactly stated theorem.

In base 10, the number constructed by Theorem 2 would look like

$$0.121132111143222111111154332221111111111 \cdots$$

Due to the prevalence of 1’s, this does not appear normal from the digits so far given, and indeed even after a decillion digits, there is still a noticeably larger number of 1’s relative to all other digits. See Section 2 for more numerical data.

²Currently, constructions of normal numbers via concatenating a function require that the growth rate of that function does not exceed polynomial.

Theorem 2 is not an immediate consequence of Theorem 1. Instead we must make use of the following idea.

Proposition 1. *Given a sequence of multi-sets of integers $\mathcal{A}(n)$, let $L(\mathcal{A}(n))$ denote the sum of the (base- b) lengths of elements of $\mathcal{A}(n)$:*

$$L(\mathcal{A}(n)) = \sum_{a \in \mathcal{A}(n)} (\lfloor \log_b a \rfloor + 1).$$

Let $\overline{\mathcal{A}(n)}$ denote a string formed by concatenating the elements of $\mathcal{A}(n)$ in any order.

If

$$\lim_{n \rightarrow \infty} \frac{L(\mathcal{A}(n))}{|\mathcal{A}(n)|} = \infty \quad \lim_{n \rightarrow \infty} \frac{L(\mathcal{A}(n+1))}{\sum_{i=1}^n L(\mathcal{A}(i))} = 0$$

and $\mathcal{A}(n)$ is asymptotically normal, then the number

$$0.\overline{(\mathcal{A}(1))}(\overline{\mathcal{A}(2)})\overline{(\mathcal{A}(3))} \dots$$

is normal.

This proposition is a simple variation of the combinatorial method of normality proofs (see [14, Theorem 1]). As such, we will not provide a full proof of the proposition, but instead include a sketch of the proof.

Our main tool in proving Theorem 1 will be the function $Q_k(n)$, which will represent the number of times k appears in $\mathcal{P}(n)$. Using our example from the start of the paper, $Q_3(5) = 2$ and $Q_1(5) = 12$. The On-Line Encyclopedia of Integer Sequences (OEIS) contains several entries related to $Q_k(n)$ [13]. Entry A066633 gives a triangle of values of $Q_k(n)$, where n gives the row and k the column. Entry A000070 gives $Q_1(n)$ (with offset 1) and entries A024786–A024794 give $Q_2(n)$ through $Q_{10}(n)$. The functions $Q_1(n)$ (and more generally $Q_k(n)$) have been studied for their application to Stanley’s Theorem (resp., Elder’s Theorem) [4, 9].

We will need the following asymptotic results on $Q_k(n)$.

Lemma 1. *Let $\epsilon \in (0, 1/4)$. Then we have*

$$Q_k(n) = \frac{p(n)}{e^{\pi k / \sqrt{6\tilde{n}}} - 1} (1 + O(\tilde{n}^{-2\epsilon})) = \frac{e^{\pi \sqrt{2\tilde{n}/3}}}{4\sqrt{3}\tilde{n}(e^{\pi k / \sqrt{6\tilde{n}}} - 1)} (1 + O(\tilde{n}^{-2\epsilon})), \quad (1)$$

uniformly for $1 \leq k \leq n^{3/4-\epsilon}/3$ as n goes to infinity. Here $\tilde{n} = n - 1/24$.

Moreover, if $\epsilon > 0$, then we have that

$$Q_k(n) = p(n - k) \left(1 + O\left(e^{-n^\epsilon}\right) \right)$$

uniformly in $n^{1/2+\epsilon} \leq k \leq n$.

The full range of k is crucial for our work. Both Fristedt [8] and Grabner, Knopfmacher, and Wagner [10] provide strong asymptotics for $Q_k(n)$ but they only hold when $k = o(n^{1/2})$. On OEIS, Vaclav Kotesovec gives some asymptotics for $Q_k(n)$ for a few specific values of k , although the method used is not clear.

We will frequently make use of asymptotic notations in this paper. By $f(x) = O(g(x))$ or $f(x) \ll g(x)$ we mean that $|f(x)| \leq C|g(x)|$ for all sufficiently large x . The constant C is known as the implicit constant. If we write $f(x) = O_y(g(x))$, then we mean that the implicit constant may depend on the variable y . We also write $f(x) = o(g(x))$ for $\lim_{x \rightarrow \infty} f(x)/g(x) = 0$ and $f(x) \asymp g(x)$ for $g(x) \ll f(x) \ll g(x)$. We will make use of the fact that $\int_a^b O(f(x)) dx = O(\int_a^b |f(x)| dx)$.

The paper is arranged as follows. In Section 2, we provide some computer calculations on the digits of $\mathcal{P}(n)$. In Section 3, we will provide a proof of Lemma 1. In Section 4, we will prove the first half of Theorem 1, that $\mathcal{P}(n)$ satisfies the generalized Benford’s law. In Section 5, we will prove the second half of Theorem 1, that $\mathcal{P}(n)$ is asymptotically normal. In Section 6, we will prove Theorem 2 after a sketched proof of Proposition 1.

2. Numerical Data

The rate of convergence of many of these results is quite slow. For example, consider

$$\overline{(\mathcal{P}(1))(\mathcal{P}(2))(\mathcal{P}(3))} \cdots \overline{(\mathcal{P}(1000))} \tag{2}$$

in base 10. In order for the infinite concatenation of $\overline{\mathcal{P}(n)}$ to be normal, we would expect every digit from 0 to 9 to appear one-tenth of the time. However, (2) contains over $74 \cdot 10^{33}$ digits of which the digit 1 occurs approximately 34% of the time, while the digit 0 occurs approximately 3% of the time. The frequency of the digits appears more like a Benford distribution than an equidistribution expected by normality. However, this should not be surprising, as the numbers being concatenated are so small, a large chunk of the digits we are considering are still leading digits.

Part of the issue is the choice of base. By choosing a smaller base, we can see the convergence occurring quicker. For the remaining examples, we will consider base 4. As a result of this new base, in (2), the digit 1 occurs approximately 44% of the time, the digit 2 approximately 25% of the time, the digit 3 approximately 18% of the time, and the digit 0 approximately 14% of the time. This is still not equidistributed, but much closer to it. In Tables 1 and 2, we have calculated frequencies of leading digits or the frequencies of all digits across various $\mathcal{P}(4^n)$. The choice of testing at 4^n was used for ease of calculation. We see a fairly steady (although not monotonic) approach to the expected frequency.

Frequency of leading digit	1	2	3
$\mathcal{P}(4^1)$	0.667	0.25	0.083
$\mathcal{P}(4^2)$	0.640	0.240	0.120
$\mathcal{P}(4^3)$	0.607	0.251	0.142
$\mathcal{P}(4^4)$	0.588	0.259	0.153
$\mathcal{P}(4^5)$	0.572	0.265	0.163
$\mathcal{P}(4^6)$	0.562	0.269	0.169
$\mathcal{P}(4^7)$	0.553	0.272	0.174
$\mathcal{P}(4^8)$	0.548	0.275	0.178
Expected frequency	0.5	0.292	0.208

Table 1: Frequency of base-4 leading digits in various $\mathcal{P}(n)$.

Frequency of digit	0	1	2	3
$\mathcal{P}(4^1)$	0.077	0.615	0.231	0.077
$\mathcal{P}(4^2)$	0.076	0.573	0.230	0.121
$\mathcal{P}(4^3)$	0.101	0.514	0.239	0.147
$\mathcal{P}(4^4)$	0.121	0.472	0.244	0.163
$\mathcal{P}(4^5)$	0.138	0.439	0.247	0.176
$\mathcal{P}(4^6)$	0.151	0.414	0.249	0.186
$\mathcal{P}(4^7)$	0.162	0.394	0.250	0.194
$\mathcal{P}(4^8)$	0.171	0.379	0.251	0.200
Expected frequency	0.25	0.25	0.25	0.25

Table 2: Frequency of base-4 digits in various $\mathcal{P}(n)$.

3. Facts about $p(n)$ and $Q_k(n)$

Hardy and Ramanujan [11] proved the following asymptotic for the partition function:

$$p(n) = \frac{1}{2\pi\sqrt{2}} \frac{d}{dx} \left(\frac{e^{\pi\sqrt{\frac{2}{3}(x-\frac{1}{24})}}}{\sqrt{x-\frac{1}{24}}} \right) \Bigg|_{x=n} + O\left(e^{D\sqrt{n}}\right), \quad n \geq 1,$$

where D is any number larger than $\frac{\pi}{2}\sqrt{\frac{2}{3}}$. If we analyze this formula further, we get

$$p(n) = \frac{e^{\pi\sqrt{\frac{2}{3}(n-\frac{1}{24})}}}{4\sqrt{3}(n-\frac{1}{24})} - \frac{e^{\pi\sqrt{\frac{2}{3}(n-\frac{1}{24})}}}{4\sqrt{2}\pi(n-\frac{1}{24})^{3/2}} + O\left(e^{D\sqrt{n}}\right), \quad n \geq 1. \quad (3)$$

Since it will come up frequently, we will write $\tilde{n} := n - \frac{1}{24}$. We will also assume that $D = \frac{3\pi}{4}\sqrt{\frac{2}{3}}$, so that the error term here is significantly smaller than any other term.

The simple form of this asymptotic will be used as well:

$$p(n) \asymp \frac{e^{\pi\sqrt{2n/3}}}{n}, \quad n \geq 1. \tag{4}$$

We first show how to write $Q_k(n)$ as a sum of values of the partition function. This result is not new to this paper and can be found, for example, in [4].

Lemma 2. *We have that*

$$Q_k(n) = \sum_{j \geq 1} p(n - jk).$$

Proof. We begin by showing that $Q_k(n) = p(n - k) + Q_k(n - k)$. There is a natural bijection between those partitions of n that contain at least one part of size k and all partitions of size $n - k$: simply remove the part of size k from the partition of n to get a partition of $n - k$. Note that all remaining parts are unchanged by this. Thus $Q_k(n)$ counts the total number of parts of size k in all partitions of n (which we can assume have at least one part of size k , since the other partitions do not contribute to this term), while $Q_k(n - k)$ counts the total number of parts of size k in all partitions of $n - k$. The only difference between these are the parts that are removed by the bijection, and there must be one for every such partition, namely there must be $p(n - k)$ of them, which completes the argument.

Iterating this argument repeatedly we get

$$Q_k(n) = p(n - k) + Q_k(n - k) = p(n - k) + p(n - 2k) + Q_k(n - 2k) = \dots,$$

which leads to our desired formula. Note that since $Q_k(n) = 0$ whenever $k > n$, this process will eventually terminate. We also use the standard definition of $p(0) = 1$, which matches our expectation of $Q_k(k) = 1$. □

We now prove Lemma 1.

Proof of Lemma 1. To prove this, we begin by applying Lemma 2 together with the Hardy-Ramanujan asymptotic Equation (3):

$$Q_k(n) = O(1) + \sum_{1 \leq j < n/k} \left(\frac{e^{\pi\sqrt{\frac{2}{3}(\tilde{n}-jk)}}}{4\sqrt{3}(\tilde{n}-jk)} - \frac{e^{\pi\sqrt{\frac{2}{3}(\tilde{n}-jk)}}}{4\sqrt{2}\pi(\tilde{n}-jk)^{3/2}} + O\left(e^{D\sqrt{n-jk}}\right) \right). \tag{5}$$

The $O(1)$ comes from the possibility that $j = n/k$, as here $p(n - jk) = 1$, but the Hardy-Ramanujan asymptotic does not apply.

Let $\epsilon \in (0, 1/4)$, $N = \tilde{n}^{3/4-\epsilon}$, and assume $k \leq N/3$. Consider a sum of the form

$$\sum_{1 \leq j < N/k} \frac{e^{\pi\sqrt{\frac{2}{3}(n-jk-\frac{1}{24})}}}{(n-jk-\frac{1}{24})^\lambda}, \tag{6}$$

for λ equal to 1 or $3/2$. Applying Taylor's Theorem, we have that

$$\begin{aligned} \sqrt{\tilde{n} - jk} &= \sqrt{\tilde{n} \left(1 - \frac{jk}{\tilde{n}}\right)} \\ &= \sqrt{\tilde{n}} \left(1 - \frac{jk}{2\tilde{n}} + O\left(\left(\frac{jk}{\tilde{n}}\right)^2\right)\right) \\ &= \sqrt{\tilde{n}} - \frac{jk}{2\sqrt{\tilde{n}}} + O(n^{-2\epsilon}), \end{aligned} \tag{7}$$

provided we are in the range $1 \leq j < N/k$. Similarly, we have

$$(\tilde{n} - jk)^{-\lambda} = \frac{1}{\tilde{n}^\lambda} \left(1 + O\left(\frac{jk}{\tilde{n}}\right)\right) = \frac{1}{\tilde{n}^\lambda} (1 + O(n^{-2\epsilon})). \tag{8}$$

Thus, applying Equation (7) and equation (8) to Equation (6) and using that $e^x = 1 + O(x)$ for bounded x , we get

$$\sum_{1 \leq j < N/k} \frac{e^{\pi\sqrt{\frac{2}{3}(\tilde{n}-jk)}}}{(\tilde{n} - jk)^\lambda} = \frac{e^{\pi\sqrt{2\tilde{n}/3}}}{\tilde{n}^\lambda} (1 + O(n^{-2\epsilon})) \sum_{1 \leq j < N/k} e^{-\frac{\pi k}{\sqrt{6\tilde{n}}}j}.$$

Since $\sum_{1 \leq j \leq \ell} e^{-cj} = (1 - e^{-c\lfloor \ell \rfloor})/(e^c - 1)$ and since

$$e^{-\frac{\pi k}{\sqrt{6\tilde{n}}} \left\lfloor \frac{N}{k} \right\rfloor} \leq e^{-\frac{\pi}{2\sqrt{6}} \tilde{n}^{1/4+\epsilon}} = O(n^{-2\epsilon})$$

for sufficiently large n , we obtain

$$\sum_{1 \leq j \leq N/k} \frac{e^{\pi\sqrt{\frac{2}{3}(\tilde{n}-jk)}}}{(\tilde{n} - jk)^\lambda} = \frac{e^{\pi\sqrt{2\tilde{n}/3}}}{\tilde{n}^\lambda} \cdot \frac{1}{e^{\pi k/\sqrt{6\tilde{n}}} - 1} (1 + O(\tilde{n}^{-2\epsilon})). \tag{9}$$

This estimate shrinks as k increases, so since $k \leq N/3$, at a minimum we have that

$$\sum_{1 \leq j \leq N/k} \frac{e^{\pi\sqrt{\frac{2}{3}(\tilde{n}-jk)}}}{(\tilde{n} - jk)^\lambda} \gg \frac{e^{\pi\sqrt{2\tilde{n}/3}}}{\tilde{n}^\lambda} \cdot \frac{1}{e^{\pi N/3\sqrt{6\tilde{n}}} - 1} \gg n^{-\lambda} e^{\pi\sqrt{2\tilde{n}/3} - \pi\tilde{n}^{1/4-\epsilon}/3\sqrt{6}}. \tag{10}$$

For the remaining terms, first note that we can bound the sum by the largest term times the number of terms, which becomes the following:

$$\sum_{N/k \leq j < n/k} \frac{e^{\pi\sqrt{\frac{2}{3}(\tilde{n}-jk)}}}{(\tilde{n} - jk)^\lambda} \ll \sum_{N/k \leq j < n/k} e^{\pi\sqrt{\frac{2}{3}(\tilde{n}-jk)}} \ll ne^{\pi\sqrt{\frac{2}{3}(\tilde{n}-N)}}.$$

Applying Equation (7), we see that $\sqrt{\tilde{n}-N} = \sqrt{\tilde{n}} - n^{1/4-\epsilon}/2 + O(n^{-2\epsilon})$. Thus we can bound the remainder sum by

$$\sum_{N/k \leq j < n/k} \frac{e^{\pi\sqrt{\frac{2}{3}(\tilde{n}-jk)}}}{(\tilde{n}-jk)^\lambda} \ll ne^{\pi\sqrt{\frac{2}{3}(\sqrt{\tilde{n}}-n^{1/4-\epsilon}/2)}}. \tag{11}$$

Similarly, we have that

$$\sum_{1 \leq j < n/k} e^{D\sqrt{n-jk}} \leq ne^{D\sqrt{n}}. \tag{12}$$

We see that the upper bounds of Equations (11) and (12) are both orders of magnitude smaller than the lower bound in Equation (10); in fact, they are both less than $n^{-2\epsilon}$ times the lower bound.

Therefore, combining Equations (5), (9), (11), and (12), we get

$$Q_k(n) = \frac{e^{\pi\sqrt{2\tilde{n}/3}}}{4\sqrt{3}\tilde{n}} \cdot \frac{1}{e^{\pi k/\sqrt{6\tilde{n}}} - 1} (1 + O(\tilde{n}^{-2\epsilon})),$$

as desired.

Now suppose that $\epsilon > 0$ and $n^{1/2+\epsilon} \leq k < n/2$. Then

$$Q_k(n) = \sum_{j \geq 1} p(n-jk) = p(n-k) + O(np(n-2k)).$$

We use the fact that $\sqrt{a-b} \leq \sqrt{a} - \frac{b}{2\sqrt{a}}$ for $0 < b \leq a$ to obtain

$$\sqrt{n-2k} \leq \sqrt{n-k} - \frac{k}{2\sqrt{n-k}} \leq \sqrt{n-k} - \frac{k}{2\sqrt{n}} \leq \sqrt{n-k} - \frac{n^\epsilon}{2}.$$

Since $n-2k \geq 1$, we can apply the simple asymptotic Equation (4) to $p(n-2k)$ together with the bound just shown and obtain

$$\begin{aligned} np(n-2k) &\ll \frac{ne^{\pi\sqrt{2(\tilde{n}-2k)/3}}}{\tilde{n}-2k} \ll ne^{\pi\sqrt{2(\tilde{n}-k)/3}-\pi n^\epsilon/\sqrt{6}} \\ &\ll \frac{e^{\pi\sqrt{2(\tilde{n}-k)/3}}}{(\tilde{n}-k)} e^{-n^\epsilon} \ll p(n-k)e^{-n^\epsilon}. \end{aligned}$$

Thus $Q_k(n) = p(n-k)(1 + O(e^{-n^\epsilon}))$ as desired. When $k = n/2$, then

$$Q_k(n) = p(n-k) + p(n-2k) = p(n/2) + 1,$$

and when $n/2 < k \leq n$, then $Q_k(n) = p(n-k)$. These trivially satisfy the asymptotic $p(n-k)(1 + O(e^{-n^\epsilon}))$ and complete the proof. \square

4. Proof of Theorem 1 – Benford’s Law

In this section, we will prove that the sequence of multi-sets $\mathcal{P}(n)$ satisfies the generalized Benford’s law. To do this we will make use of the indicator function $I_c(k)$, which, for a given string c not starting with 0, is 1 if the expansion of k starts with c and is 0 otherwise.

We begin with the following useful lemma.

Lemma 3. *We have*

$$\sum_{k \leq x} \frac{I_c(k)}{k} = \log_b \left(1 + \frac{1}{c} \right) \log x + O \left(\frac{1}{c} \right),$$

where the implicit constant is only dependent on the number of digits in c .

Proof. We will prove the desired asymptotic holds for $\sum_{k < x} I_c(k)/k$, as the difference between these sums is negligible compared to the error term $O(1/c)$.

Recall the formula

$$\sum_{k < x} \frac{1}{k} = \log x + \gamma + O(1/x),$$

where γ is the Euler–Mascheroni constant.

Let J denote the number of base- b digits in c . Note that for any $j \in \mathbb{N}_{\geq 0}$, we have

$$\begin{aligned} \sum_{k \in [b^{J+j-1}, b^{J+j})} \frac{I_c(k)}{k} &= \sum_{k \in [0, b^j)} \frac{1}{b^j c + k} \\ &= \log \left(\frac{b^j c + b^j}{b^j c} \right) + O \left(\frac{1}{b^j c} \right) \\ &= \log \left(1 + \frac{1}{c} \right) + O \left(\frac{1}{b^j c} \right). \end{aligned}$$

This entire sum could also be bounded by $O(1/c)$.

Let b^K be the smallest power of b with $x \leq b^K$. Then

$$\begin{aligned} \sum_{k < x} \frac{I_c(k)}{k} &= \sum_{k < b^K} \frac{I_c(k)}{k} + O \left(\sum_{k \in [b^{K-1}, b^K)} \frac{I_c(k)}{k} \right) \\ &= \sum_{k < b^K} \frac{I_c(k)}{k} + O \left(\frac{1}{c} \right). \end{aligned}$$

We can now break up this first sum into b -adic intervals and apply our earlier estimation to obtain

$$\sum_{k < x} \frac{I_c(k)}{k} = \sum_{j=0}^{K-J} \sum_{k \in [b^{J+j-1}, b^{J+j})} \frac{I_c(k)}{k} + O \left(\frac{1}{c} \right)$$

$$\begin{aligned}
 &= \sum_{j=0}^{K-J} \left(\log \left(1 + \frac{1}{c} \right) + O \left(\frac{1}{b^j c} \right) \right) + O \left(\frac{1}{c} \right) \\
 &= (K - J + 1) \log \left(1 + \frac{1}{c} \right) + O \left(\frac{1}{c} \right) \\
 &= (\log_b x + O_J(1)) \log \left(1 + \frac{1}{c} \right) + O \left(\frac{1}{c} \right) \\
 &= \log_b \left(1 + \frac{1}{c} \right) \log x + O_J \left(\frac{1}{c} \right).
 \end{aligned}$$

This completes the proof. □

We are ready to show that $\mathcal{P}(n)$ satisfies the generalized Benford’s Law.

Proof of the first half of Theorem 1. It suffices to prove that

$$\sum_{k \leq n} I_c(k) Q_k(n) = \log_b \left(1 + \frac{1}{c} \right) \frac{\log(\tilde{n}) e^{\pi \sqrt{2\tilde{n}/3}}}{4\pi \sqrt{2\tilde{n}}} \left(1 + O \left(\frac{1}{\log n} \right) \right), \tag{13}$$

as the sum on the left will count the number of $k \in \mathcal{P}(n)$ with $I_c(k) = 1$, and the main term on the right will be identical over all c except for the factor of $\log_b(1 + 1/c)$, as desired.

We will split the sum $\sum_{k \leq n} I_c(k) Q_k(n)$ into two parts, one over $k \leq n^{3/5}$ and $k > n^{3/5}$. Selecting $\epsilon = 1/10$, we may apply the first half of Lemma 1 uniformly to all k in the first interval and the second half of Lemma 1 uniformly to all k in the second interval. In fact, for the second interval, we will use the much simpler bound $Q_k(n) \ll p(n - k)$.

For the sum over $k > n^{3/5}$, we have

$$\sum_{n^{3/5} < k \leq n} I_c(k) Q_k(n) \leq \sum_{n^{3/5} < k \leq n} Q_k(n) \ll \sum_{n^{3/5} < k \leq n} p(n - k) \ll np(n - n^{3/5}).$$

Again using the fact that $\sqrt{a - b} \leq \sqrt{a} - \frac{b}{2\sqrt{a}}$ for $0 \leq b \leq a$, together with Equation (4), we have that

$$np(n - n^{3/5}) \ll \frac{ne^{\pi \sqrt{\frac{2}{3}(\tilde{n} - n^{3/5})}}}{4\sqrt{3}\tilde{n}} \leq e^{\pi \sqrt{\frac{2}{3}\tilde{n}} - \frac{\pi}{\sqrt{6}} n^{1/10}}. \tag{14}$$

Due to the $e^{-\pi n^{1/10}/\sqrt{6}}$, the sum over $k > n^{3/5}$ is orders of magnitude smaller than even the error term in Equation (13).

For the remaining terms, we make use of Equation (1) for the $Q_k(n)$, and note that the $(1 + O(\tilde{n}^{-1/5}))$ can be factored out of the sum due to being uniform over

all the terms. Thus we have

$$\sum_{k \leq n^{3/5}} I_c(k) Q_k(n) = \left(\sum_{k \leq n^{3/5}} \frac{I_c(k)}{e^{\pi k/\sqrt{6\tilde{n}}}-1} \right) \frac{e^{\pi\sqrt{2\tilde{n}/3}}}{4\sqrt{3} \cdot \tilde{n}} (1 + O(\tilde{n}^{-1/5})). \tag{15}$$

We will evaluate this inner sum using partial summation as follows, making use of Lemma 3 as we proceed:

$$\begin{aligned} \sum_{k \leq n^{3/5}} \frac{I_c(k)}{e^{\pi k/\sqrt{6\tilde{n}}}-1} &= \sum_{k \leq n^{3/5}} \frac{I_c(k)}{k} \cdot \frac{k}{e^{\pi k/\sqrt{6\tilde{n}}}-1} \\ &= \left(\sum_{k \leq n^{3/5}} \frac{I_c(k)}{k} \right) \frac{n^{3/5}}{e^{\pi n^{3/5}/\sqrt{6\tilde{n}}}-1} \\ &\quad - \int_1^{n^{3/5}} \left(\sum_{k \leq u} \frac{I_c(k)}{k} \right) \frac{d}{du} \left(\frac{u}{e^{\pi u/\sqrt{6\tilde{n}}}-1} \right) du \\ &= \left(\log_b \left(1 + \frac{1}{c} \right) \log(n^{3/5}) + O\left(\frac{1}{c}\right) \right) \frac{n^{3/5}}{e^{\pi n^{3/5}/\sqrt{6\tilde{n}}}-1} \tag{16} \end{aligned}$$

$$- \log_b \left(1 + \frac{1}{c} \right) \int_1^{n^{3/5}} \log u \frac{d}{du} \left(\frac{u}{e^{\pi u/\sqrt{6\tilde{n}}}-1} \right) du \tag{17}$$

$$+ O\left(\frac{1}{c} \int_1^{n^{3/5}} \frac{d}{du} \left(\frac{u}{e^{\pi u/\sqrt{6\tilde{n}}}-1} \right) du\right). \tag{18}$$

In (18), we could pull the big-O out of the integral because the derivative is of constant sign. The terms in Equations (16) and (18) are bounded by

$$O\left(\frac{n^{3/5} \log n}{c(e^{\pi n^{3/5}/\sqrt{6\tilde{n}}}-1)} + \frac{1}{c(e^{\pi/\sqrt{6\tilde{n}}}-1)}\right) = O\left(\frac{\sqrt{\tilde{n}}}{c}\right).$$

When multiplied by $e^{\pi\sqrt{2\tilde{n}/3}}/\tilde{n}$ from Equation (15), the resulting order of magnitude is precisely that of the error term in Equation (13).

This leaves the integral in Equation (17), which can be evaluated precisely using the following formula:

$$\begin{aligned} &\int \log u \frac{d}{du} \left(\frac{u}{e^{\pi u/\sqrt{6\tilde{n}}}-1} \right) du \\ &= \frac{u \log u}{e^{\pi u/\sqrt{6\tilde{n}}}-1} - \frac{\sqrt{6\tilde{n}} \log(e^{\pi u/\sqrt{6\tilde{n}}}-1)}{\pi} + u + C. \tag{19} \end{aligned}$$

Note that $\log(x-1) = \log x + O(1/x)$ for large x . Thus, when evaluated at $u = n^{3/5}$, the second and third term of Equation (19) will almost entirely cancel to leave just

the first term and a small remainder:

$$\frac{n^{3/5} \log(n^{3/5})}{e^{\pi n^{3/5}/\sqrt{6\tilde{n}}} - 1} + O\left(\frac{n}{e^{\pi n^{3/5}/\sqrt{6\tilde{n}}}}\right). \tag{20}$$

As this integral is multiplied by $-\log_b(1 + 1/c) = O(1/c)$ in Equation (17), everything in Equation (20) is negligible compared with the $O(\sqrt{n}/c)$ term already found. When Equation (19) is evaluated at $u = 1$, we use the fact that $\log(e^x - 1) = \log x + O(x)$ when x is close to 0 to obtain

$$\begin{aligned} & \frac{1 \log 1}{e^{\pi/\sqrt{6\tilde{n}}} - 1} - \frac{\sqrt{6\tilde{n}} \log(e^{\pi/\sqrt{6\tilde{n}}} - 1)}{\pi} + 1 \\ &= -\frac{\sqrt{6\tilde{n}}}{\pi} \left(\log\left(\frac{\pi}{\sqrt{6\tilde{n}}}\right) + O\left(\frac{1}{\sqrt{\tilde{n}}}\right) \right) + 1 \\ &= \frac{\sqrt{\tilde{n}} \log \tilde{n}}{\pi\sqrt{2/3}} \left(1 + O\left(\frac{1}{\log n}\right) \right). \end{aligned}$$

Together this gives

$$\sum_{k \leq n^{3/5}} \frac{I_c(k)}{e^{\pi k/\sqrt{6\tilde{n}}} - 1} = \log_b\left(1 + \frac{1}{c}\right) \frac{\sqrt{\tilde{n}} \log \tilde{n}}{\pi\sqrt{2/3}} \left(1 + O\left(\frac{1}{\log n}\right) \right). \tag{21}$$

Inserting Equation (21) into Equation (15) gives the desired result. □

5. Proof of Theorem 1 – Asymptotic Normality

As in the previous section, we will need some initial lemmas about an ancillary function. Let c denote an arbitrary string, now including strings that start with 0, and let $N_c(k)$ denote the number of times the string c appears as a substring of k . The following lemma is due to Copeland and Erdős [3].

Lemma 4. *Let c be an arbitrary string. For any $\epsilon > 0$ and all k in $[1, x]$, we have that*

$$N_c(k) = \frac{\log_b(k) + O(1)}{b^{|c|}} (1 + O(\epsilon)) \tag{22}$$

with at most $O(x^{1-\delta})$ exceptions, where $\delta > 0$ is dependent on ϵ , b , and $|c|$. The implicit constants are all uniform provided x is sufficiently large (where the bound on x may depend on ϵ , b , and $|c|$ again).

From this we derive the following estimate.

Lemma 5. *For all strings c of the same length, we have*

$$\sum_{k \leq x} N_c(k) = \frac{x \log_b x}{b^{|c|}} (1 + o(1)) \tag{23}$$

uniformly.

Proof. We make use of Equation (22) and note that for the $O(x^{1-\delta})$ exceptions, we still have that $N_c(k) = O(\log_b(x))$, as the number of times c appears in k is bounded by the number of base- b digits k has, which in turn is bounded by the number of base- b digits x has. Thus

$$\begin{aligned} \sum_{k \leq x} N_c(k) &= \sum_{k \leq x} \frac{\log_b(k) + O(1)}{b^{|c|}} (1 + O(\epsilon)) + O(x^{1-\delta} \log_b x) \\ &= \left(\frac{x(\log_b x + O(1))}{b^{|c|}} \right) (1 + O(\epsilon)) + O(x^{1-\delta} \log_b x). \end{aligned}$$

Provided x is large enough, we have that

$$\sum_{k \leq x} N_c(k) = \frac{x \log_b x}{b^{|c|}} (1 + O(\epsilon)).$$

However, since this holds uniformly for any $\epsilon > 0$, we may replace $O(\epsilon)$ with $o(1)$. □

We will proceed directly into the proof.

Proof of the second half of Theorem 1. We will look at $\sum_{k \leq n} N_c(k) Q_k(n)$ as we did in the previous proof. It suffices to show that

$$\sum_{k \leq n} N_c(k) Q_k(n) = \frac{1}{b^{|c|}} F(n) (1 + o(1)), \tag{24}$$

for some function $F(n)$ that is independent of c . (The $o(1)$ may depend on c .) While we will not find a simple form for $F(n)$, we will prove that it is of order at least

$$M(n) = \frac{\log n \cdot e^{\pi \sqrt{2n/3}}}{\sqrt{n}}.$$

In particular, any contribution to $\sum_{k \leq n} N_c(k) Q_k(n)$ that we can show is $o(M(n))$ will be swept into the $o(1)$ term of Equation (24). In the following, we will assume that $|c|$ is fixed, but otherwise c may range freely.

We will, as before, split the sum $\sum_{k \leq n} N_c(k) Q_k(n)$ into two parts, one over $k \leq n^{3/5}$ and one over $k > n^{3/5}$, in order to apply Lemma 1 with $\epsilon = 1/10$.

For $k > n^{3/5}$, we will use the fact that $N_c(k) \ll \log n$ for $k \leq n$ and also Equation (14) to obtain

$$\sum_{n^{3/5} < k \leq n} N_c(k)Q_k(n) \ll n \log n \cdot p(n - n^{3/5}) \ll \frac{n \log n \cdot e^{\pi\sqrt{\frac{2}{3}\tilde{n}} - \frac{\pi}{6}n^{1/10}}}{4\sqrt{3\tilde{n}}},$$

which is $o(M(n))$ due to the $e^{-\frac{\pi}{6}n^{1/10}}$ term.

For the remaining terms, we apply our asymptotic from Lemma 1, to get

$$\sum_{k \leq n^{3/5}} N_c(k)Q_k(n) = \left(\sum_{k \leq n^{3/5}} \frac{N_c(k)}{e^{\pi k/\sqrt{6\tilde{n}}} - 1} \right) \frac{e^{\pi\sqrt{2\tilde{n}/3}}}{4\sqrt{3\tilde{n}}} (1 + O(n^{-1/5})).$$

We apply partial summation to the resulting sum:

$$\begin{aligned} \sum_{k \leq n^{3/5}} \frac{N_c(k)}{e^{\pi k/\sqrt{6\tilde{n}}} - 1} &= \frac{1}{e^{\pi n^{3/5}/\sqrt{6\tilde{n}}} - 1} \sum_{k \leq n^{3/5}} N_c(k) \\ &\quad - \int_1^{n^{3/5}} \left(\sum_{k \leq u} N_c(k) \right) \frac{d}{du} \left(\frac{1}{e^{\pi u/\sqrt{6\tilde{n}}} - 1} \right) du \\ &= \frac{1}{e^{\pi n^{3/5}/\sqrt{6\tilde{n}}} - 1} \sum_{k \leq n^{3/5}} N_c(k) \end{aligned} \tag{25}$$

$$- \int_1^{\log \log n} \left(\sum_{k \leq u} N_c(k) \right) \frac{d}{du} \left(\frac{1}{e^{\pi u/\sqrt{6\tilde{n}}} - 1} \right) du \tag{26}$$

$$- \int_{\log \log n}^{n^{3/5}} \left(\sum_{k \leq u} N_c(k) \right) \frac{d}{du} \left(\frac{1}{e^{\pi u/\sqrt{6\tilde{n}}} - 1} \right) du. \tag{27}$$

We assume that n is large enough so that we can apply Equation (23) to Equations (25) and (27).

Applying Equation (23) to Equation (25) shows that it is $o(1)$. For the integral in Equation (26) we have

$$\sum_{k \leq u} N_c(k) \ll u \log u \ll \log^2 \log n \quad \text{for } 1 \leq u \leq \log \log n$$

and that $1/(e^{\pi u/\sqrt{6\tilde{n}}} - 1)$ is a strictly decreasing function, so that

$$\int_1^{\log \log n} \left(\sum_{k \leq u} N_c(k) \right) \frac{d}{du} \left(\frac{1}{e^{\pi u/\sqrt{6\tilde{n}}} - 1} \right) du$$

$$\begin{aligned} &\ll \log^2 \log n \int_1^{\log \log n} \left| \frac{d}{du} \left(\frac{1}{e^{\pi u/\sqrt{6\tilde{n}}} - 1} \right) \right| du \\ &\ll \log^2 \log n \int_1^{\log \log n} \frac{d}{du} \left(\frac{1}{e^{\pi u/\sqrt{6\tilde{n}}} - 1} \right) du \\ &\ll \frac{\log^2 \log n}{e^{\pi/\sqrt{6\tilde{n}}} - 1} \ll \sqrt{n} \log^2 \log n, \end{aligned}$$

which even after multiplication by $e^{\pi\sqrt{2\tilde{n}/3}/\tilde{n}}$ will still be $o(M(n))$.

Finally let us consider Equation (27). Applying Equation (23), and noting that the asymptotic is uniform over the entire integral, we obtain

$$-\frac{1 + o(1)}{b^{|c|}} \int_{\log \log n}^{n^{3/5}} u \log u \cdot \frac{d}{du} \left(\frac{1}{e^{\pi u/\sqrt{6\tilde{n}}} - 1} \right) du.$$

When $u \asymp n^{1/2}$, we have that

$$u \log u \cdot \frac{d}{du} \left(\frac{1}{e^{\pi u/\sqrt{6\tilde{n}}} - 1} \right) \asymp \log n.$$

Thus the entire integral must be at least a constant times $n^{1/2} \log n$, which, after multiplication by $e^{\pi\sqrt{2\tilde{n}/3}/\tilde{n}}$, is on the order of $M(n)$. Thus, if we set

$$F(n) = -\frac{e^{\pi\sqrt{2\tilde{n}/3}}}{4\sqrt{3\tilde{n}}} \int_{\log \log n}^{n^{3/5}} u \log u \cdot \frac{d}{du} \left(\frac{1}{e^{\pi u/\sqrt{6\tilde{n}}} - 1} \right) du,$$

then Equation (24) is true. □

6. Proof of Theorem 2

Sketch of the proof of Proposition 1. If we let $x = 0.\overline{(\mathcal{A}(1))}(\overline{(\mathcal{A}(2))})(\overline{(\mathcal{A}(3))}) \cdots$, then we want to show that the frequency with which a string c appears in the first k digits of x tends to $b^{-|c|}$ as $k \rightarrow \infty$. Suppose that the k th digit of x appears in $\overline{(\mathcal{A}(n))}$ and let $l(k) = |\overline{(\mathcal{A}(1))}(\overline{(\mathcal{A}(2))}) \cdots (\overline{(\mathcal{A}(n))})|$. The assumption that

$$\lim_{n \rightarrow \infty} \frac{L(\mathcal{A}(n+1))}{\sum_{i=1}^n L(\mathcal{A}(i))} = 0$$

implies that as k gets large the proportion with which c appears in the first k digits will be negligibly different from the proportion with which c appears in the first $l(k)$ digits. So we can focus on the latter instead. The assumption that

$$\lim_{n \rightarrow \infty} \frac{L(\mathcal{A}(n))}{|\mathcal{A}(n)|} = \infty$$

guarantees that the occurrences of c that appear starting in one of the concatenated strings and ending in another are negligible. Normality of x then follows immediately from the asymptotic normality of the sets $\mathcal{A}(n)$. \square

Proof of Theorem 2. We will make use of Proposition 1. Since we have already shown that $\mathcal{P}(n)$ is asymptotically normal in Theorem 1, it suffices to show that

$$\lim_{n \rightarrow \infty} \frac{L(\mathcal{P}(n))}{|\mathcal{P}(n)|} = \infty \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{L(\mathcal{P}(n+1))}{\sum_{i=1}^n L(\mathcal{P}(i))} = 0.$$

By summing Equation (13) over all possible $c \in \{1, 2, \dots, b-1\}$, we see that

$$|\mathcal{P}(n)| = \sum_{k \leq n} Q_k(n) = \frac{\log \tilde{n} \cdot e^{\pi\sqrt{2\tilde{n}/3}}}{4\pi\sqrt{2\tilde{n}}} (1 + o(1)). \tag{28}$$

(This is also a well-known result of Erdős and Lehner [7].) Let us consider the contribution to this sum of those terms with $k \leq \log n$. Since $(e^x - 1)^{-1} = x^{-1}(1 + O(x))$ for small x , we have

$$\begin{aligned} \sum_{k \leq \log n} Q_k(n) &= \left(\sum_{k \leq \log n} \frac{1}{e^{\pi k/\sqrt{6\tilde{n}}} - 1} \right) \frac{e^{\pi\sqrt{2\tilde{n}/3}}}{4\sqrt{3\tilde{n}}} (1 + O(n^{-1/6})) \\ &= \left(\sum_{k \leq \log n} \frac{\sqrt{6\tilde{n}}}{\pi k} \right) \frac{e^{\pi\sqrt{2\tilde{n}/3}}}{4\sqrt{3\tilde{n}}} (1 + O(n^{-1/6})) \\ &= \frac{\log \log n \cdot e^{\pi\sqrt{2\tilde{n}/3}}}{2\pi\sqrt{2\tilde{n}}} (1 + o(1)), \end{aligned}$$

which is $o(1)$ of the full sum Equation (28). Thus

$$\begin{aligned} L(\mathcal{P}(n)) &= \sum_{k \leq n} ([\log_b k] + 1) Q_k(n) \\ &\geq \sum_{\log n < k \leq n} [\log_b \log n] Q_k(n) \\ &= [\log_b \log n] \sum_{k \leq n} Q_k(n) + O\left(\frac{\log^2 \log n \cdot e^{\pi\sqrt{2\tilde{n}/3}}}{2\pi\sqrt{2\tilde{n}}}\right) \\ &= \frac{\log \tilde{n} \log_b \log n \cdot e^{\pi\sqrt{2\tilde{n}/3}}}{4\pi\sqrt{2\tilde{n}}} (1 + o(1)). \end{aligned}$$

Thus $L(\mathcal{P}(n))/|\mathcal{P}(n)| \rightarrow \infty$ as $n \rightarrow \infty$.

On the other hand, since

$$|\mathcal{P}(n)| = \sum_{k \leq n} Q_k(n) \quad \text{and} \quad L(\mathcal{P}(n)) = \sum_{k \leq n} ([\log_b k] + 1) Q_k(n),$$

we have that by bounding $1 \ll \lfloor \log_b k \rfloor \ll \log n$ that

$$\frac{\log n \cdot e^{\pi\sqrt{2\tilde{n}/3}}}{4\pi\sqrt{2\tilde{n}}} \ll L(\mathcal{P}(n)) \ll \frac{\log^2 n \cdot e^{\pi\sqrt{2\tilde{n}/3}}}{4\pi\sqrt{2\tilde{n}}}.$$

Utilizing this lower bound, we have

$$\begin{aligned} \sum_{i=1}^n L(\mathcal{P}(n)) &\gg \sum_{i=1}^n \frac{\log n \cdot e^{\pi\sqrt{2n/3}}}{4\pi\sqrt{2n}} \gg \sum_{i=2}^n \frac{e^{\pi\sqrt{2n/3}}}{\sqrt{n}} \\ &\geq \int_2^n \frac{e^{\pi\sqrt{2x/3}}}{\sqrt{x}} dx \asymp e^{\pi\sqrt{2n/3}}. \end{aligned}$$

Note that the integrand is strictly increasing on the given range, which is why the comparison applies. Thus

$$L(\mathcal{P}(n+1)) \ll \frac{\log^2(n+1) \cdot e^{\pi\sqrt{2(\tilde{n}+1)/3}}}{\sqrt{\tilde{n}+1}} = o\left(e^{\pi\sqrt{2n/3}}\right) = o\left(\sum_{i=1}^n L(\mathcal{P}(n))\right)$$

as desired. □

Acknowledgement. The author wishes to thank Maddie Dawsey, Ken Ono, and Wei-Lun Tsai for their assistance and patience in answering questions.

References

- [1] T. Anderson, L. Rolin, and R. Stoehr, Benford’s law for coefficients of modular forms and partition functions, *Proc. Amer. Math. Soc.* **139** (2011), 1533–1541.
- [2] D. G. Champernowne, The construction of decimals normal in the scale of ten, *J. Lond. Math. Soc. (2)* **1** (1933), 254–260.
- [3] A. H. Copeland and P. Erdős, Note on normal numbers, *Bull. Amer. Math. Soc. (N.S.)* **52** (1946), 857–860.
- [4] M. G. Dastidar and S. S. Gupta, Generalization of a few results in integer partitions, *Notes Number Theory Discrete Math.* **19** (2013), 69–76.
- [5] H. Davenport and P. Erdős, Note on normal decimals, *Canad. J. Math.* **4** (1952), 58–63.
- [6] J.-M. De Koninck and I. Katai, On a problem on normal numbers raised by igor shparlinski, *Bull. Aust. Math. Soc.* **84** (2011), 337–349.
- [7] P. Erdős and J. Lehner, The distribution of the number of summands in the partitions of a positive integer, *Duke Math. J.* **8** (1941), 335–345.
- [8] B. Fristedt, The structure of random partitions of large integers, *Trans. Amer. Math. Soc.* **337** (1993) 703–735.

- [9] R. A. Gilbert, A fine rediscovery, *Amer. Math. Monthly* **122** (2015), 322–331.
- [10] P. J. Grabner, A. Knopfmacher, and S. Wagner, A general asymptotic scheme for the analysis of partition statistics, *Combin. Probab. Comput.* **23** (2014)(6), 1057–1086.
- [11] G. H. Hardy and S. Ramanujan, Asymptotic formulæ in combinatory analysis, *Proc. Lond. Math. Soc. (3)* **2** (1918), 75–115.
- [12] J. R. Iafrate, S. J. Miller, and F. W. Strauch, Equipartitions and a distribution for numbers: A statistical model for benford’s law, *Phys. Rev. E* **91** (2015), 062138.
- [13] O. F. Inc., The On-Line Encyclopedia of Integer Sequences, 2022, URL <http://oeis.org>.
- [14] P. Pollack and J. Vandehey, Besicovitch, bisection, and the normality of $0.(1)(4)(9)(16)(25)\dots$, *Amer. Math. Monthly* **122** (2015), 757–765.
- [15] P. Pollack and J. Vandehey, Some normal numbers generated by arithmetic functions, *Canad. Math. Bull.* **58** (2015), 160–173.