WALKING TO INFINITY ALONG SOME NUMBER THEORY SEQUENCES

STEVEN J. MILLER, FEI PENG, TUDOR POPESCU, JOSHUA M. SIKTAR, NAWAPAN WATTANAWANICHKUL, AND THE POLYMATH REU PROGRAM

Abstract. An interesting open conjecture asks whether it is possible to walk to infinity along primes, where each term in the sequence has one digit more than the previous. We present different greedy models for prime walks to predict the long-time behavior of the trajectories of orbits, one of which has similar behavior to the actual backtracking one. Furthermore, we study the same conjecture for square-free numbers, which is motivated by the fact that they have a strictly positive density, as opposed to primes. We introduce stochastic models and analyze the walks' expected length and frequency of digits added. Lastly, we prove that it is impossible to walk to infinity in other important number-theoretical sequences or on primes in different bases.

Contents

Date: October 27, 2020.

Key words and phrases. Walking to infinity, stochastic models, prime walks, square-free walks, Appending unbounded digits.

The Polymath REU Walking to Infinity Group is: William Ball, Corey Beck, Aneri Brahmbhatt, Alec Critten, Michael Grantham, Matthew Hurley, Jay Kim, Junyi Huang, Bencheng Li, Tian Lingyu, Adam May, Saam Rasool, Daniel Sarnecki, Jia Shengyi, Ben Sherwin, Yiting Wang, Lara Wingard, Chen Xuqing, and Zheng Yuxi.

This work was partially supported by NSF grant DMS1561945, Carnegie Mellon University, and Williams College.

1. Introduction

1.1. **Background.** An interesting, open question is whether or not it is possible to walk to infinity through primes, where each term in the sequence has one digit more than the previous. If we drop the restriction that we can only add one digit at a time, the answer is yes, and follows immediately from Dirichlet's theorem for primes in arithmetic progression. Specifically, given any prime p other than 2 or 5, by adding enough digits *to the left*, we can ensure that our new number is prime.[1]

We consider appending one digit at a time, and, in order not to have Dirichlet's theorem available, we append digits to the right. For example, $3, 31, 317$ are all primes, but we cannot append a digit to the right of 317 and remain prime. However, if we started our sequence $3, 31, 311$, then we could continue with 3119 (and we can only append a 9 and nothing else to 311). We could go one step further, to 31193, but nothing else. Conversely, we could run the sequence backward: 31193, 3119, 311, 31, 3. In particular, a straightforward computation shows that the longest such sequence starting with 3 is 3, 37, 373, 3733, 37337, 373379, 3733799, 37337999, which has length 8.

This leads to the notion of a *right truncatable prime*, which is a prime that remains prime after removing the rightmost digits successively. It is known that there are exactly 83 right truncatable primes, with the largest one being 73939133 [TP]. Notice that any right truncatable prime with d digits generates a prime walk of length d, and thus 73939133 yields a prime walk of length 8. However, it is possible to have longer walks if we do not require starting (or, if you take the opposite viewpoint, ending) at a one-digit prime. For example,

$$\{19, 197, 1979, 19793, 197933, 1979339, 19793393, 197933933, 1979339333\}$$

is a walk with step sizes always one and it has length 9, while

$$\{409, 4099, 40993, 409933, 4099339, 40993391, 409933919, 4099339193,$$
$$40993391939, 409933919393, 4099339193933\}$$

is one of length 11. In particular, an exhaustive search shows that the above is the longest prime walk with a starting point less than 1,000,000, tied with

$$\{68041, 680417, 6804173, 68041739, 680417393, 6804173939, 68041739393,$$
$$680417393939, 6804173939393, 68041739393933, 680417393939333\}.$$

On the other hand, some small primes do not have long walks, such as 11, whose longest walk is $\{11, 113\}$ while appending any digit to 53 yields a composite, hence a prime walk starting at 53 always has length 1.

This suggests several questions.

[1]Given p, choose m so that $10^m > p$. As 10^m and p are relatively prime, there are infinitely many primes congruent to p modulo 10^m, and thus we obtain the next prime in our sequence.

- Is it possible to walk to infinity among the primes, where each prime in the sequence is the result of appending one digit to the right of the previous? From the last observation, we are no longer starting with a one-digit prime. If there is an infinite prime walk in base 10, we eventually only append digits 3 and 9.[2]

- What if, instead of appending just one digit, we append at most a bounded number of digits to the right? More generally, what if the number of digits we append in moving from p_n to p_{n+1} is at most $f(p_n)$ for some function f tending to infinity? Unlike the case of appending to the left, we cannot immediately deduce the answer by appealing to Dirichlet's theorem for primes in arithmetic progressions.

1.2. **Stochastic models.** Like most problems in number theory, the above questions are easy to state but resist progress. We thus consider instead related random problems to try and get a sense of what might be true. Such models have been used elsewhere with great success, from suggesting there are only finitely many Fermat primes to the veracity of the Twin Prime and Goldbach conjectures.

For example, recall the n^{th} Fermat number is $F_n = 2^{2^n} + 1$. The Prime Number Theorem says that the number of primes up to x is about $x/\log x$, and thus one often models a randomly chosen number of order x as being prime with probability $1/\log x$. This is the famous Cramér model; while it is known to have some issues (see [MS]), it gives reasonable answers for many problems. If we let $\{X_n\}$ be independent Bernoulli random variables where $X_n = 1$ with probability $1/\log F_n$, then the expected number of X_n's that are 1 (and thus the expected number of Fermat primes) is

$$\mathbb{E}\left[\sum_{n=0}^{\infty} X_n\right] \;=\; \sum_{n=0}^{\infty} \frac{1}{\log(2^{2^n}+1)} \;\approx\; \sum_{n=0}^{\infty} \frac{1}{\log 2} \cdot \frac{1}{2^n} \;=\; \frac{2}{\log 2} \;\approx\; 2.89, \tag{1.1}$$

which is reasonably close to the number of known Fermat primes, five, coming from $n \in \{0, 1, 2, 3, 4\}$.

As primes are difficult to work with, we ask related questions of other sequences, such as square-free numbers. From our heuristic model and numerical explorations, we do not believe one can walk to infinity through the primes by adding a bounded number of digits to the right; however, we believe it is possible for square-free numbers. For example, starting with 2, we can get a walk of length 17 just by always appending the smallest digit that yields a square-free number:

$$\{2, 21, 210, 2101, 21010, 210101, 2101010, 21010101, 210101010, 2101010101,$$
$$21010101010, 210101010101, 2101010101010, 2101010101010102,$$
$$210101010101021, 2101010101010210, 21010101010102101\}.$$

[2] In base 10, we can never append a 5, save to the empty string. Consider primes exceeding 3. As 1 and 7 are both 1 modulo 3, if our number is prime, it must be either 1 or 2 modulo 3. Thus, if our number is 1 modulo 3, we can append at most one 1 or one 7, at which point our new number is now 2 modulo 3, and from this point onward, we can only use the digits 3 and 9.

While the fraction of numbers at most x that are prime is approximately $1/\log x$, which tends to zero, the fraction which are square-free tends to $1/\zeta(2) = 6/\pi^2$, or about 60.79%. Thus, there are tremendously more square-free numbers available than primes. In particular, once our number is large, it is unlikely that *any* digit can be appended to create another prime. Thus it should be impossible to walk to infinity among the primes by appending just one digit on the right. However, for square-free numbers, we expect to have several digits that we can append and stay square-free, leading to exponential growth in the number of paths.

Explicitly, we consider the following random processes. Given a sequence whose last term is x, we want to assign an appropriate probability of being able to append an additional digit to the right. We assume each term is independent of the previous, and the probability that a digit can be appended to x is $p(x)$. Thus the probability will decrease as x increases for primes but is essentially constant for square-free numbers. Furthermore, for prime walks, we present two different models: the first one randomly selects a digit among 1, 3, 7, 9 and appends it to the number, while the second (refined) random model first checks what digits yield a prime number in the next step and then randomly selects one from the set. We assume all numbers with the same number of digits are equally likely to be in the sequence for simplicity. For the primes base 10, we cannot append a digit that is even or a 5, whereas, for square-free numbers, we cannot append a digit such that the sum of the digits is 9 or the last two digits are a multiple of 4. One could consider more involved models taking these into account.

We approximate that if a number has k digits, the number of primes of k digits in base b is

$$\frac{b^k}{\log b^k} - \frac{b^{k-1}}{\log b^{k-1}} = \frac{b^{k-1}}{\log b} \cdot \left(\frac{b}{k} - \frac{1}{k-1}\right) = \frac{b^{k-1}((k-1)b-k)}{k(k-1)\log b} \approx \frac{(b-1)\cdot b^{k-1}}{k\log b}.$$

As there are $(b-1)b^{k-1}$ such numbers, we assume the probability that a $k-$digit number is prime is $1/(k\log b)$, and assume that the events of two distinct numbers being prime are independent.

Our main focus is the expected value and distribution of lengths of walks among these random primes and random square-free numbers. Such probabilistic models have had remarkable success in modeling other problems, such as the $3x+1$ map and its generalizations; see [KL]. As remarked, this toy model has several issues. In particular, we assume the numbers formed by appending the digits under consideration are all independent in our desired sequence. However, this yields a simple model with easily computed results on how long we expect to be able to walk in the various sequences from different starting points.

1.3. **Results.** We compare the random model with observations of the actual sequences. We present the two random models for prime walks and show that the refined one is very close to the actual sequence. In particular, when considering prime walks with starting point less than a million, the difference of expected lengths of the walks between our refined greedy model and the real primes is 0.14, less than 7% of the expected length of the real prime walks of 2.07.

Furthermore, we note that the model becomes more precise as the starting point increases, and the prime numbers become more sparse. As the starting point increases, the number of primes from which we randomly choose to continue decreases. Then, we also look at the frequency of the digits added at each step and see that the refined model approximates the real world extremely well. Lastly, while we discuss infinite prime walks, we extend our predictions for the case when we are allowed to insert a digit anywhere, rather than only to the right, and verify them using the Miller-Rabin probabilistic test.

On the other hand, when investigating square-free walks, we consider the results of its random models, remark on the discrepancies in the frequencies of added digits, and give the number-theoretic reasons.

Although we use stochastic models for prime and square-free walks, there are some sequences and restrictive scenarios for which we can prove several results regarding walks to infinity, for example, prime walks in base 2 and 5, perfect squares, and Fibonacci numbers. In parallel, we also examine the behaviour of prime walks in different number fields and plan to discuss it in a subsequent article.

The main results of the current study are as follows:

Prime walks

- The expected values of our models for prime walks in base 10 are given by (2.6), (2.7), (2.9) and (2.10);
- Comparison of the two prime walk models and the actual primes can be found in Tables 2, 3, 4, and 5;
- The expected values for prime walks obtained by inserting a digit anywhere are presented in Tables 7 and 8;
- A proof that it is impossible to walk to infinity on primes in base 2 by appending no more than 2 digits is given in Theorem 2.5, while Lemmas 2.6 and 2.7 show that it is impossible to walk to infinity on primes in base 4 and 5 by appending one digit to the right.

Square-free walks

- The expected values of square-free walks given by our models are presented in Tables 9 and 11, while Theorem 3.2 shows that almost surely there exists an infinite random square-free walk from any starting point;
- Table 10 presents the frequencies of the digits of square-free walks, and Remark 3.13 explains them;
- Theorems 3.7 and 3.8 give a tight bound on the expected value of the length of square-free walks in base 2 and 10 respectively, while Theorem 3.17 does the same for fourth-power-free walks.

Walks on other sequences

- Lemma 4.1 shows that we can't walk to infinity on perfect squares by appending one digit to the right, while Lemma 4.3 gives a condition on the terms of a walk to infinity on squares;
- Theorem 4.7 shows that walks on terms of the Fibonacci sequence constructed by appending one digit to the right have length at most 2, while Theorems 4.9 and 4.11 yield similar results when we append exactly N, respectively at most N digits at a time.

2. Modeling Prime Walks

2.1. **Models.** We now estimate the length of these random walks in base b, so there are b digits we can append. If our number has k digits, then from §1.2, the probability a digit yields a successful appending is approximately $1/(k \log b)$, as we are assuming all possible numbers are equally likely to be prime. For example, if $b = 10$, we are not removing even numbers or 5 or numbers that make the sum a multiple of 3. Thus, the probability that at least one of the b digits works is 1 minus the probability they all fail, or

$$1 - \left(1 - \frac{1}{k \log b}\right)^b. \tag{2.1}$$

The first stochastic model for primes is as follows. Each possible appended number is independently declared to be a random prime with probability as described above. Choose one digit uniformly at random and check if the obtained number is prime; if it is not, stop and record the length; otherwise, continue the process. This algorithm can be imagined as a greedy prime walk, as we are not looking further down the line to see which of many possible random primes would be best to choose to get the longest walk possible. We call this the greedy model. Furthermore, note that we may easily improve the model in base 10 by appending from $\{1, 3, 7, 9\}$. We discuss this improvement later and compare it to the initial greedy model.

In order to compute the expected length of such a walk, starting at a one digit random prime in base b, we count the probabilities in two different ways; note that the expected length is just the infinite sum of the probabilities that we stop at the n^{th} step times n. For brevity, let A_n be *the event that the walk has length at least* n, and B_n *the event that the walk has length exactly* n. It is obvious that B_i, B_j are pairwise independent and that $A_n = \cup_{i=n}^{\infty} B_i$. Since the B_i's are pairwise independent, we have that

$$\mathbb{P}[A_n] \ = \ \mathbb{P}[\cup_{i=n}^{\infty} B_i] \ = \ \sum_{i=n}^{\infty} \mathbb{P}[B_i].$$

Therefore, we have the following system of equations:

$$\begin{aligned} \mathbb{P}[A_1] \ &= \ \mathbb{P}[B_1] \ + \ \mathbb{P}[B_2] + \mathbb{P}[B_3] \ + \ \cdots \\ \mathbb{P}[A_2] \ &= \ \mathbb{P}[B_2] + \mathbb{P}[B_3] \ + \ \cdots \\ \mathbb{P}[A_3] \ &= \mathbb{P}[B_3] \ + \ \cdots \\ &\vdots \end{aligned} \tag{2.2}$$

Summing the above over all n we obtain

$$\sum_{i=1}^{n} \mathbb{P}[A_i] \ = \ \sum_{i=1}^{n} i\mathbb{P}[B_i], \tag{2.3}$$

which implies

$$\sum_{n=1}^{\infty} \mathbb{P}[\text{walk has length at least n}] \;=\; \sum_{n=1}^{\infty} n\mathbb{P}[\text{walk has length exactly n}]. \quad (2.4)$$

Note that the sum on the right is the expected value of the walk in our greedy model, while the sum on the left equals

$$\sum_{n=0}^{\infty}\prod_{k=1}^{n-1}\left(1-\left(1-\frac{1}{k\log b}\right)^{b}\right), \quad (2.5)$$

where each term in the sum represents the probability that there is a random prime with which we can extend the walk for the first $n-1$ steps, without considering the $n-$th step. In particular, the expected length in base 10 when starting with a single digit is 4.690852. Furthermore, by multiplying by the approximate number of primes with exactly r digits and dividing by the expected number of primes with at most s digits, we get that the expected length of a walk with a starting point at most s digits is about

$$\begin{aligned} &\frac{1}{\frac{b^s}{s\log b}}\left(\sum_{r=1}^{s}\frac{(b-1)b^{r-1}}{r\log b}\left(\sum_{n=0}^{\infty}\prod_{k=r}^{n-1}\left(1-\left(1-\frac{1}{k\log b}\right)^{b}\right)\right)\right)\\ =\;&\frac{s(b-1)}{b^s}\left(\sum_{r=1}^{s}\frac{b^{r-1}}{r}\left(\sum_{n=0}^{\infty}\prod_{k=r}^{n-1}\left(1-\left(1-\frac{1}{k\log b)}\right)^{b}\right)\right)\right). \quad (2.6)\end{aligned}$$

We present in Table 1 the expected lengths as we vary the starting point and base. As remarked earlier, one can view this model as a greedy random prime walk because we always take another step if possible, with no regard to how many steps we may be able to take afterward; thus, all decisions are local.

Note that the expected length of the walk in base 10 starting with a one-digit number, 4.22, is different than the one we computed earlier, 4.69. This is because we multiplied 4.69 by the approximation $(b-1)/b$; i.e., 0.9. More importantly, note that in base 10 we can only append $\{1,3,7,9\}$ and hope to stay prime since primes greater than 5 are odd and not divisible by 5.

This suggests a simple improvement to the model base 10: *we only allow the four digits 1, 3, 7 and 9 to be appended on the right.* Henceforth, we will only use this improved version. To do this, we have to make a couple of changes in the formula (2.6): replace the numerator of $1/(k\log b)$ with $10/4$ and, instead of raising $1-10/(4k\log b)$ to the b (in this case 10), we raise it to the fourth power as now there are only four options of digits to add. We shall call this the *refined greedy model*, whose expected value is presented in Table 1 as 10', and we note that this is relatively close to the real expected length of the greedy model presented in Table 2. Modifying our earlier analysis, we see that the formula for the expected length of the greedy model in base 10 is

$$\frac{s(b-1)}{b^s}\left(\sum_{r=1}^{s}\frac{b^{r-1}}{r}\left(\sum_{n=0}^{\infty}\prod_{k=r}^{n-1}\left(1-\left(1-\frac{10}{4k\log b}\right)^{4}\right)\right)\right). \quad (2.7)$$

Number of digits of starting point

Base	1	2	3	4	5	6	7
2	5.20	9.90	11.62	11.45	10.40	9.08	7.79
3	5.05	7.75	7.60	6.53	5.40	4.49	3.80
4	4.87	6.55	5.86	4.79	3.92	3.29	2.85
5	4.71	5.79	4.92	3.96	3.25	2.78	2.45
6	4.57	5.27	4.34	3.48	2.89	2.49	2.22
7	4.46	4.89	3.95	3.17	2.65	2.31	2.08
8	4.37	4.59	3.67	2.95	2.49	2.19	1.98
9	4.29	4.36	3.45	2.79	2.37	2.09	1.91
10	4.22	4.17	3.28	2.66	2.28	2.20	1.85
10'	4.54	4.55	3.55	2.83	2.38	2.09	1.90

TABLE 1. Expected length of prime walks given by our formula, 10' is the refined greedy model.

While the greedy model is useful for computations, we turn our attention to an improvement which better approximates the reality. The refined greedy model can only append $1, 3, 7, 9$ in base 10. We will show that this model approximates the real world prime walks more accurately than the greedy algorithm in terms of both the expected value of the walks ($< 10\%$ difference) and the frequency of digits added.

2.2. **Results and comparison of models.** According to the random probabilistic model of prime walks in §2.1, the expected length of a greedy prime walk, starting with a single digit prime in base 10, is 4.69. We compare this heuristic estimate with the primes.

We present the results of the greedy and refined greedy models in the following tables, which show the results of our computer simulations. The refined greedy model is rather close to the real data whereas the greedy one still predicts some behaviors of the walks. The data for the actual primes is computed by the program that exhaustively searches for the longest prime walk given a starting point. First, let us observe how the number of digits of the starting point affects the expected length of the models in Table 2.

We consider different starting points to eliminate small number bias. As a result, the above table shows that the expected length of the walks decreases as the starting point increases in both our random model and in the real world.

Furthermore, we analyze the frequency of digits added in the prime walks, both for the actual primes and in our models. In particular, we remark that it was

Start has x digits	1	2	3	4	5	6
greedy model	1.89	1.60	1.41	1.30	1.25	1.20
refined greedy model	4.33	3.37	2.76	2.37	2.08	1.90
primes	8.00	4.71	3.48	2.71	2.28	2.03

Table 2. Comparing the expected value of the walk lengths. The refined greedy model is significantly closer to the actual value compared to the greedy one.

Number of appended	1's	3's	7's	9's
random model	15.6%	33.0%	19.9%	31.3%
refined greedy model	11.8%	36.7%	14.2%	37.1%
primes	12.1%	40.2%	11.1%	36.5%

Table 3. Frequency of added digits in prime walks with starting point less than 100,000.

Number of appended	1's	3's	7's	9's
random model	15.4%	32.7%	18.5%	33.2%
refined greedy model	12.5%	35.9%	14.7%	36.8%
primes	13.1%	38.8%	12.2%	35.6%

Table 4. Frequency of added digits in prime walks with starting point less than 1,000,000.

expected that 3 and 9 appear more often than 1 and 7. This is because 1 and 7 can never be appended if we start with a prime that is 2 modulo 3, and at most one number in our prime walk can be 2 modulo 3. We present the frequency of digits in Table 4 when the starting point is less than 1,000,000. As expected, in both our models and the real prime walks, the number of appended 3's is very close to the number of appended 9's while the number of appended 1's is very close to the number of appended 7's. One surprising result is that there appear significantly more 7's in the random models than in the real prime walks. We observe how the starting point affects the frequency of the digits added in Table 3 and Table 4.

As mentioned above, we observe that the number of appended $3'$s and $9'$s is larger than the number of appended $1'$s and $7'$s. This is due to the fact that by modulo 3 considerations, we can only append 3 or 9 to a number 2 (mod 3). In particular, this means that when starting with a prime 1 (mod 3), we can only

Number of appended	1's	3's	7's	9's
random model	16.3%	32.3%	18.5%	32.8%
refined greedy model	12.7%	35.8%	14.8%	36.4%
primes	13.3%	38.6%	12.4%	35.5%

TABLE 5. Frequency of added digits in prime walks with starting point greater than 100,000 but less than 1,000,000.

Start has x digits	1	2	3	4	5	6
greedy model	2.83	1.94	1.64	1.45	1.34	1.28
refined greedy model	3.49	3.22	2.43	2.04	1.77	1.62
primes	8.00	3.81	2.64	2.12	1.81	1.64

TABLE 6. Expected value of the walks with starting point 2 modulo 3.

append 1 or 7 at most once in our walk, whereas there are no such constrains for $3, 9$. We present our models when starting with $3, 9$ in the following section.

2.3. **Starting with 2 mod 3.**

In this subsection, we compare our models with the primes when our starting number is $2 \mod 3$. This experiment is motivated because we can only append 3 or 9 to such a prime while hoping to remain prime; any other digit would lead to a composite number divisible by 2, 3, or 5. Therefore, we refine our model only to append 3 or 9. In this case, the walks are shorter, but the model predictions are closer to the primes. Note that the longest prime walk with starting point 2 modulo 3 less than 1,000,000 has length 10, and is

$$\{809243, 8092439, 80924399, 809243993, 8092439939, 80924399393, 809243993933,$$

$$8092439939333, 80924399393333, 809243993933339\}.$$

Since there are now only two possible digits to append, instead of four, the expected length of the walk is given by

$$\frac{s(b-1)}{b^s}\left(\sum_{r=1}^{s}\frac{b^{r-1}}{r}\left(\sum_{n=0}^{\infty}\prod_{k=r}^{n-1}\left(1-\left(1-\frac{10}{2k\log b}\right)^2\right)\right)\right). \tag{2.8}$$

We compare our model (2.8) to the primes in Table 6. The refined greedy model approximates the real world extraordinary well, especially as the initial number increases. This is due to primes' sparseness, as usually at most one of $\{1, 3, 7, 9\}$ can be appended as the number increases.

2.4. **Other walks along primes.** Since the probability of a given number being prime decreases the larger it gets, it seems quite reasonable, albeit extremely difficult to prove, that walking to infinity along the primes by appending or prepending digits is impossible. We might consider other sequences. For example, is it possible to walk to infinity along the primes by *inserting digits anywhere*? For some quick intuition, we can estimate the probability that a number n is prime at $1/\log n$. In base b, there are about $\log_b n$ digits in n, so there are about $b\log_b n$ possible numbers reachable by adding digits anywhere to n. This means that the expected number of primes reachable from n is about $b\log_b n/\log n = b/\log b$. This is quite interesting because it does not depend on n, strongly suggesting that walks to infinity are possible.

Now that we have an intuition on what may happen if we are allowed to append a digit at a time anywhere, we follow the idea of §2.1 to create a model to find the expected walk length of prime walks in this case.

Similar to (2.5), the expected value for the walk length for an $m-$digit number can be written as

$$\sum_{n=1}^{\infty}\prod_{k=m}^{n-1}\left(1-\left(1-\frac{1}{k\log b}\right)^{b(k+1)-1-k}\right). \tag{2.9}$$

Notice that the only difference between (2.5) and (2.9) is the exponent of $1-1/(k\log b)$. The number $b(k+1)-1-k$ is obtained by considering how many distinct numbers we could produce by adding a digit anywhere in the k-digit number in base b. For example, adding another 4 right before of after the existing 4 in 3141 yields the same number, so we do not count that twice.

By computing this expression, we have the expected walk length for the basic case at $b=10, k=1$, which is 61.57. This number is tremendously larger than an expected value of 4.69 for appending on the right only. More interestingly, as the base increases the expected value obtained by this expression diverges. Results for different bases, b, and different starting lengths, k, are presented in Table 7.

Again, by multiplying by the approximate number of primes with exactly r digits and dividing by the expected number of primes with at most s digits, we get that the expected length of a walk with starting point at most s digits is about

$$\frac{s(b-1)}{b^s}\left(\sum_{r=1}^{s}\frac{b^{r-1}}{r}\left(\sum_{n=1}^{\infty}\prod_{k=m}^{n-1}\left(1-\left(1-\frac{1}{k\log b}\right)^{b(k+1)-1-k}\right)\right)\right). \tag{2.10}$$

The expected values for starting length up to some s for different bases b is shown in Table 8. These values also respect the pattern seen in Table 7 that higher bases have larger expected values. One example that we can confirm its primality is the following walk in base 10 of length 17.

{ 7, 17, 137, 1637, 18637, 198637, 1986037, 19986037, 199860337, 1998660337, 19998660337, 199098660337, 1949098660337, 19490986560337, 194909865603317, 1949098656033817, 19490983656033817}.

Although the length 17 is not very large and does not suggest we can walk to infinity, it is the longest example our code can find. However, to show the

Base	Starting length: 1	2	3	4	5	6	7	8	9	10
2	6.22	6.74	5.89	5.35	4.99	4.73	4.54	4.40	4.28	4.18
3	10.01	9.01	8.25	7.74	7.37	7.09	6.88	6.71	6.58	6.46
4	13.32	12.33	11.55	10.99	10.58	10.26	10.01	9.80	9.63	9.49
5	17.56	16.57	15.76	15.16	14.69	14.33	14.03	13.79	13.58	13.40
6	22.90	21.90	21.07	20.42	19.90	19.49	19.15	18.87	18.63	18.42
7	29.59	28.59	27.73	27.04	26.48	26.03	25.65	25.32	25.05	24.81
8	37.96	36.97	36.08	35.36	34.76	34.26	33.85	33.49	33.17	32.90
9	48.45	47.45	46.55	45.79	45.16	44.63	44.17	43.78	43.43	43.13
10	61.57	60.57	59.65	58.87	58.21	57.64	57.15	56.72	56.35	56.01

TABLE 7. Expected value for small starting lengths evaluated up to $n = 1000$.

Base	s: 1	10	100
2	3.11	5.08	3.37
3	6.67	6.97	5.36
4	9.99	9.94	7.97
5	14.05	13.86	11.39
6	19.08	18.90	15.84
7	25.36	25.33	21.59
8	33.22	33.49	29.00
9	43.07	43.79	38.50
10	55.41	56.76	50.61

TABLE 8. Expected value for starting lengths up to s.

existence of a longer walk, we consider the Miller-Rabin probabilistic test, which is extremely fast, even for very large numbers, but does not completely prove that a number is prime. We then find many sequences of length 60, and much longer, but cannot completely guarantee that they are valid. Nevertheless, we can be quite confident, since the error is estimated at most 4^{-40}.

2.5. **Some proofs related to prime walks.** As mentioned in the introduction, it is possible to walk to infinity on primes by appending an unbounded number of digits to the left at each step. Dirichlet's theorem on arithmetic progressions states that given two positive coprime integers a and d, there are infinitely many primes congruent to $a \pmod d$. Thus given an initial prime p_0 other than 2 or 5, we can take n such that $10^{n_0} > p_0$, and find a prime p_1 such that $p_1 \equiv p_0 \pmod{10^{n_0}}$. The congruence mod 10^{n_0} corresponds to the left-appending of some number of digits. This gives us the next step in our walk; we can now seek a prime of the form $p_1 \pmod{10^{n_1}}$, and the process continues indefinitely unto infinity.

We now show that this statement's counterpart is also true, namely that it is possible to walk to infinity on primes by appending an unbounded number of digits to the right.

Theorem 2.1. *Let p_0 be a prime. Then there exists a sequence of infinitely many primes $p_0, p_1, ...$ such that for all $i \geq 1$, p_i is equal to $10^{n_i} \cdot p_{i-1} + k_i$, for positive integers n_i and k_i with $k_i < 10^{n_i}$.*

Proof. We can restate our goal as follows: given an arbitrary but fixed prime p, we must show that there exists an n such that there is a prime between $10^n \cdot p$ and $10^n \cdot p + 10^n - 1 = 10^n(p+1) - 1$; i.e., in the interval $[p10^n, (p+1)10^n)$.

To do so, we note that for a given p, and for a fixed but arbitrary real r such that $0 < r < 1$, there exists an n such that

$$p \;<\; 10^{\frac{1-r}{r}n} - 1. \tag{2.11}$$

Moreover, given such an n, then it is possible to find a real, positive x such that

$$p10^n \;=\; x - x^r. \tag{2.12}$$

Then, using first (2.11) and then (2.12), we have that

$$\begin{aligned} p10^n &< 10^{\frac{n}{r}} - 10^n \\ x - x^r &< 10^{\frac{n}{r}} - 10^n. \end{aligned}$$

This second inequality implies that $x^r < 10^n$, for when $x^r = 10^n$, then $x - x^r = 10^{\frac{n}{r}} - 10^n$, and moreover, $x - x^r$ is strictly increasing (once it is positive).

Given that $x^r < 10^n$, then $x - x^r > x - 10^n$. This means that $p10^n > x - 10^n$, and so

$$x \;<\; (p+1)10^n. \tag{2.13}$$

All that remains is finding an r such that there is always a prime in the interval $[x - x^r, x]$. Results of this nature are plentiful; most recently, Baker, Harman, and Pintz show that a value of $r = 0.525$ suffices for x greater than some lower bound x_0. Because there exists a prime in the interval $[x - x^{0.525}, x]$ for $x > x_0$, then using our definitions above there must be a prime contained in the interval $[p10^n, (p+1)10^n)$. Note that in order to guarantee $x > x_0$, it is necessary to choose an n such that $n > \log_{10}((x_0 - x_0^r)/p)$ (and such that (2.11) holds as well).

That there is a prime in $[p10^n, (p+1)10^n)$ implies that there are n and k such that $p10^n + k$ is prime, with $k < 10^n$. This gives the next prime in our sequence, which thus goes on infinitely. □

Definition 2.2. *An "extended Cunningham chain" is the infinite sequence $e_1, e_2, ...$, generated by an initial prime p and the relation $e_k = 2e_{k-1} + 1$ (for $k \geq 1$ and*

$e_0 = p$). In other words, we have that

$$\begin{aligned} e_0 &= p, \\ e_1 &= 2p+1, \\ e_2 &= 4p+3, \\ &\vdots \\ e_i &= 2^i p + 2^i - 1, \\ &\vdots \end{aligned}$$

We show that such extended Cunningham chains, no matter their initial prime p, contain a sequence of consecutive composite e_i's of arbitrarily length. To do so, we begin with the following lemma.

Lemma 2.3. *Given $k \geq \lceil \log_2(p+1) \rceil + 2$, then $2^k - (p+1)$ is not a power of 2.*

Proof. Suppose that there exist k and n such that $2^k - (p+1) = 2^n$. Then it is the case that $2^k - 2^n = p+1$. Moreover, we have that $2^k - 2^n \geq 2^k - 2^{k-1} = 2^{k-1}$. We can thus find a solution for n only if $k < \lceil \log_2(p+1) \rceil + 2$, for if we take $k \geq \lceil \log_2(p+1) \rceil + 2$, then $2^{k-1} \geq 2^{\lceil \log_2(p+1) \rceil + 1} > p+1$. We thus have that $p+1 = 2^k - 2^n \geq 2^{k-1} > p+1$, which is a contradiction. □

With this result in hand, we move on to the main theorem.

Theorem 2.4. *In every such extended Cunningham chain, given any $n \in \mathbb{Z}_+$, there is $i \in \mathbb{Z}_+$ such that $e_i, e_{i+1}, \ldots, e_{i+n-1}$ are composite.*

Proof. Set $k = \lceil \log_2(p+1) \rceil + 2$, and let us consider $i = c \cdot \phi(2^k - p - 1) \cdot \phi(2^{k+1} - p - 1) \cdots \phi(2^{k+n-1} - p - 1)$, where $c \in \mathbb{Z}_+$ is arbitrary. Moreover, for each of $2^{k+j} - p - 1$ (with $0 \leq j \leq n-1$), take an odd positive divisor $d_j \mid 2^{k+j} - p - 1$ that is greater than 1. We can find such d_j because we have chosen k via Lemma 2.3 such that none of $2^{k+j} - p - 1$ are powers of 2. Because $p + 1 \equiv 2^{k+j} \pmod{2^{k+j} - p - 1}$, it is also the case that $p + 1 \equiv 2^{k+j} \pmod{d_j}$. Thus we have that

$$e_{i-(k+j)} = 2^{i-(k+j)}(p+1) - 1 \equiv 2^{i-(k+j)} 2^{k+j} - 1 \equiv 2^i - 1 \pmod{d_j}. \tag{2.14}$$

However, as 2 is coprime with d_j, Euler's theorem gives $2^{\phi(d_j)} \equiv 1 \pmod{d_j}$. Moreover, it is the case that $\phi(d_j) \mid \phi(2^{k+j} - p - 1)$, since $d_j \mid (2^{k+j} - p - 1)$. Hence we have that

$$\begin{aligned} e_{i-(k+j)} &\equiv 2^i - 1 \equiv 2^{c \cdot \phi(2^k - p - 1) \cdot \phi(2^{k+1} - p - 1) \cdots \phi(2^{k+n-1} - p - 1)} - 1 \\ &\equiv (2^{\phi(d_j)})^{K_j} - 1 \equiv 0 \pmod{d_j}, \end{aligned} \tag{2.15}$$

such that K_j is an integer ($K_j = c \cdot [\phi(2^k - p - 1) \cdots \phi(2^{k+n-1} - p - 1)] / \phi(d_j)$).

We have thus show that $e_{i-k}, e_{i-k-1}, \ldots, e_{i-(k+n-1)}$ are composite. Notice that with c sufficiently large, i can be made greater than $k + n - 1$, allowing us to find a subsequence of n composite elements for any n. Renaming the indices gives the desired result. □

From here we can obtain the following corollary, which relates to the broader problem of walking to infinity on primes.

Theorem 2.5. *It is impossible to walk to infinity on primes in base* 2 *by appending no more than* 2 *digits at a time to the right.*

Proof. Given one or two digits, we can append either 0, 1, 00, 01, 10, or 11. For parity reasons, one cannot append any of 0, 00, or 10.

Appending 01_2 to a prime p gives $4p+1$. If $p \equiv 2 \pmod 3$, then $4p+1$ is divisible by 3 and thus not prime. Moreover, given $p \equiv 2 \pmod 3$, then $2p+1$ and $4p+3$ are equivalent to $2 \pmod 3$ as well. Thus if $p \equiv 2 \pmod 3$, we can walk to infinity from that point onward only by appending 1_2 or 11_2.

If $p \equiv 1 \pmod 3$, then $4p+1 \equiv 2 \pmod 3$. This brings us to the above case, now applied to $4p+1$. No matter the value of our initial prime p, we can therefore append 01 at most once in our walk to infinity. It is thus sufficient to consider the point at which we append only 1 or 11 to eternity. We can then apply Theorem 2.4 with $n = 2$. Namely, continuously appending 1 to a prime in a base 2 creates a generalized Cunningham chain, which we know contains prime gaps of size 2; hence there will be some point in the prime sequence for which $2p+1$ and $4p+3$ are both composite, and we can walk no further. □

Applying the ideas of the above results allows us to make observations in bases 4 and 5.

Lemma 2.6. *It is impossible to walk to infinity on primes in base* 4 *by appending a single digit at a time to the right.*

Proof. We confine ourselves to considering only odd digits. Because $4 \equiv 1 \pmod 3$, appending 1 to a prime $p \equiv 2 \pmod 3$ gives $4p+1 \equiv 0 \pmod 3$, a composite. One can thus append 1 at most a single time in walking to infinity, and so it suffices to consider the infinite subsequence over which only 3's are appended. Call the elements of this subsequence $p_0, p_1, \dots$. Then, in similar fashion to the extended Cunningham chains, these elements take the form

$$\begin{aligned} p_1 &= p_1, \\ p_2 &= 4p_1 + 3, \\ p_3 &= 16p_1 + 15, \\ &\vdots \\ p_i &= 4^{i-1}p_1 + 4^{i-1} - 1, \\ &\vdots \end{aligned}$$

But then

$$p_{p_1} \equiv 4^{p_1-1}p_1 + 4^{p_1-1} - 1 \equiv 0 \pmod{p_1}, \tag{2.16}$$

by Fermat's little theorem. Hence p_{p_1} is composite, and it is impossible to walk to infinity by appending just one digit at a time. □

Finally, we apply a similar argument to base 5.

Lemma 2.7. *It is impossible to walk to infinity on primes in base* 5 *by appending a single digit at a time to the right.*

Proof. In base 5, parity mandates that we append either 2 or 4 at each step. If we have a prime $p \equiv 1 \pmod 3$, then $5p + 4 \equiv 0 \pmod 3$, and so we must append a 2. Moreover, if $p \equiv 1 \pmod 3$ then $5p + 2 \equiv 1 \pmod 3$ as well, so we must append *another* 2, and so on until infinity.

If $p_1 \equiv 1 \pmod 3$, then we have that

$$p_i \;=\; 5^{i-1}p_1 + \frac{5^{i-1}-1}{2}. \tag{2.17}$$

Then it is the case that $2p_i \equiv 5^{i-1} - 1 \pmod{p_1}$, and so $2p_{p_1}$ is divisible by p_1 according to Fermat's little theorem. Therefore p_{p_1} is composite.

On the other hand, if we have a $p \equiv 2 \pmod 3$, then $5p+2 \equiv 0 \pmod 3$, so we must append a 4. But $5p + 4 \equiv 2 \pmod 3$ when $p \equiv 2 \pmod 3$, thus requiring that we append 4's unto infinity.

Given $p_1 \equiv 2 \pmod 3$, we find that

$$p_i \;=\; 5^{i-1}p_1 + 5^{i-1} - 1. \tag{2.18}$$

Fermat's little theorem, therefore, allows us to conclude that p_{p_1} is composite. The exception is when $p_1 = 5$, in which case we write $p_i = 5^{i-1}p_1 + 5^{i-1} - 1 = 5^{i-2}(5p_1 + 4) + 5^{i-2} - 1 = 5^{i-2}p_2 + 5^{i-2} - 1$, and observe that p_{p_2+1} is divisible by p_2.

Hence, no matter our initial prime, there must be a composite element in the sequence produced by appending one digit to the right. □

In conclusion, the use of stochastic models suggests that there is no infinite prime walk given adding one digit at a time to the right, but that it is likely to have ones if we can add a digit anywhere. On the other hand, for other less restrictive problems, namely, if we can walk to infinity given appending an unbounded number of digits to the right and if appending a digit at a time in base $2, 4$, and 5, we show the former case is possible, but the latter is not.

3. Modeling Square-free Walks

3.1. **Model.** We now turn our attention to square-free walks whose density is positive, contrasting to a zero density sequence like primes. We thus expect that it is possible to walk to infinity using square-free numbers. However, to verify our conjecture, one must append the digits carefully. For example, 231546210170694222 is a square-free number such that successively removing the rightmost digit always yields a square-free number, but appending any digit to the right yields a non-square-free one. We now present our model for estimating the length of square-free walks.

Definition 3.1. *A square-free integer is an integer that is not divisible by any perfect square other than* 1*.*

If $Q(x)$ denotes the number of square-free positive integers less than or equal to x, it is well-known (see for example [MT-B]) that

$$Q(x) \;\approx\; x\prod_{\text{p prime}}\left(1-\frac{1}{p^2}\right) \;=\; x\prod_{\text{p prime}}\frac{1}{1+\frac{1}{p^2}+\frac{1}{p^4}+\cdots} \;=\; \frac{x}{\zeta(2)} \;=\; \frac{6x}{\pi^2}. \tag{3.1}$$

Our first stochastic model for square-free numbers is defined as follows. Each possible appended number is independently declared to be a square-free number with probability $p = 6/\pi^2$. Choose one digit uniformly at random and append it: if the obtained number is not square-free, stop and record the length; otherwise, continue the process. One can view this algorithm as a greedy square-free walk because we are not looking further down the line to see which of many possible random square-free would be best to choose to get the longest walk possible. Henceforth, we will call this the greedy model. Still, comparing the greedy model for square-free walks with that for prime walks, the square-free walks are longer than prime walks. We present the comparison in Table 9.

Start has x digits	1	2	3	4	5	6
greedy square-free walk	2.81	2.76	2.72	2.71	2.71	2.70
greedy prime walk	2.83	1.94	1.64	1.45	1.34	1.28

TABLE 9. Comparing the expected value of the walks in base 10.

To gain more understanding in the behavior of square-free walks, we first find the probability that the square-free walk is of length exactly k. Let X denote the number of steps in our random square-free walk. Then

$$\Pr[X = k] \ = \ p^k(1-p) \ = \ \frac{6^k(\pi^2-6)}{\pi^{2k+2}}. \tag{3.2}$$

Since

$$\sum_{k=1}^{\infty} \Pr[X = k] \ = \ \sum_{k=0}^{\infty} \left(p^k - p^{k+1}\right) \ = \ 1 \tag{3.3}$$

and $\Pr[X = k] \geq 0$ for all k, this is a probability space; furthermore, X is a geometric random variable. Using the fact that

$$\sum_{k=0}^{\infty} p^k \ = \ \frac{1}{1-p} \tag{3.4}$$

and differentiating, we obtain

$$\sum_{k=0}^{\infty} kp^{k-1} \ = \ \frac{1}{(1-p)^2} \ \Rightarrow \ \sum_{k=0}^{\infty} kp^k(1-p) \ = \ \frac{p}{1-p}. \tag{3.5}$$

Therefore, we find the expected walk length of square-free walks as

$$\mathbb{E}[X] \ = \ \sum_{k=0}^{\infty} k \Pr[X = k] \ = \ \sum_{k=0}^{\infty} kp^k(1-p) \ = \ \frac{p}{1-p} \ = \ \frac{6}{\pi^2-6} \ \approx \ 1.55. \tag{3.6}$$

Similarly, we have that

$$\sum_{k=0}^{\infty} kp^k \ = \ \frac{p}{(1-p)^2} \ \Rightarrow \ \sum_{k=0}^{\infty} k^2p^k(1-p) \ = \ \frac{p(p+1)}{(1-p)^2}. \tag{3.7}$$

Therefore,

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{p(p+1)}{(1-p)^2} - \frac{p^2}{(1-p)^2} \\ &= \frac{p}{(1-p)^2} = \frac{6\pi^2}{(\pi^2-6)^2} \approx 3.95. \end{aligned} \tag{3.8}$$

Given a square-free number, let us now compute the probability that the longest walk starting with it is at most k. Let P_i be the probability that the longest square-free walk has length at most i. In particular, P_1 is the probability that the longest square-free walk has a length of exactly one, i.e., the walk is the starting point. In other words, appending any digit yields a non-square-free number, so

$$P_1 = (1-p)^{10} = \frac{(\pi^2-6)^{10}}{\pi^{20}} \approx 8.5835 \times 10^{-5}. \tag{3.9}$$

Now, consider the probability that the longest square-free walk has length at most 2. Indeed, there are 10 possible cases where exactly i digits work in the first appending, i.e., $0 \le i \le 9$. Then, by using (3.9) we have that

$$\begin{aligned} P_2 &= P_1 + \binom{10}{1}(1-p)^9 pP_1 + \binom{10}{2}(1-p)^8(pP_1)^2 + \cdots + \binom{10}{10}p^{10}P_1^{10} \\ &= \binom{10}{0}(1-p)^{10} + \binom{10}{1}(1-p)^9 pP_1 + \cdots + \binom{10}{10}p^{10}P_1^{10} \\ &= \sum_{i=0}^{10}\binom{10}{i}(1-p)^{10-i}(pP_1)^i = (1-p+pP_1)^{10} \approx 8.5950 \times 10^{-5}. \end{aligned} \tag{3.10}$$

To compute P_3, let i denote the number of digits that we can append in the first step while staying square-free. Then, there are $10i$ possible numbers after the second appending. Like P_2, we consider cases when there are exactly $0 \le k \le 10i$ numbers work. Note that, when $i = 0$ or $j = 0$, we have a walk of length $1, 2$ respectively, so such cases are included in P_2. Therefore, by (3.9) and (3.10), we have that

$$\begin{aligned} P_3 &= P_2 + \sum_{i=1}^{10}\binom{10}{i}p^i(1-p)^{10-i}\Big(\sum_{k=1}^{10i}\binom{10i}{k}(1-p)^{10i-k}(pP_1)^k\Big) \\ &= P_2 + \sum_{i=1}^{10}p^i(1-p)^{10-i}\big(((1-p)+pP_1)^{10i} - (1-p)^{10i}\big) \\ &= P_2 + \sum_{i=1}^{10}p^i(1-p)^{10-i}\big(P_2^i - P_1^i\big) \\ &= P_2 + \sum_{i=1}^{10}p^i(1-p)^{10-i}P_2^i - \sum_{i=1}^{10}p^i(1-p)^{10-i}P_1^i \\ &= P_2 + ((1-p+pP_2)^{10} - (1-p)^{10}) - ((1-p+pP_1)^{10} - (1-p)^{10}) \\ &= (1-p+pP_2)^{10} \approx 8.5950 \times 10^{-5}. \end{aligned} \tag{3.11}$$

The next step is to compute P_k which can be done by induction. Suppose that, for $2 \le m \le k, P_k = (1-p+pP_{k-1})^{10}$. Similar to the idea of computing P_2 and P_3, we have that

$$\begin{aligned}
& P_{k+1} \\
= \;& P_k + \sum_{a_1=1}^{10} \binom{10}{a_1} p^{a_1}(1-p)^{10-a_1} \sum_{a_2=1}^{10a_1} \binom{10a_1}{a_2} p^{a_2}(1-p)^{10a_1-a_2} \cdots \\
& \sum_{a_{k-1}=1}^{10a_{k-2}} \binom{10a_{k-2}}{a_{k-1}} p^{a_{k-1}}(1-p)^{10a_{k-2}-a_{k-1}} \sum_{a_k=1}^{10a_{k-1}} \binom{10a_{k-1}}{a_k} p^{a_k}(1-p)^{10a_{k-1}-a_k} p_1^{a_k} \\
= \;& P_k + \sum_{a_1=1}^{10} \binom{10}{a_1} p^{a_1}(1-p)^{10-a_1} \sum_{a_2=1}^{10a_1} \binom{10a_1}{a_2} p^{a_2}(1-p)^{10a_1-a_2} \cdots \\
& \sum_{a_{k-1}}^{10a_{k-2}} \binom{10a_{k-2}}{a_{k-1}} p^{a_{k-1}}(1-p)^{10a_{k-2}-a_{k-1}} \left((1-p+pP_1)^{10a_{k-1}} - (1-p)^{10a_{k-1}}\right) \\
= \;& P_k + \sum_{a_1=1}^{10} \binom{10}{a_1} p^{a_1}(1-p)^{10-a_1} \sum_{a_2=1}^{10a_1} \binom{10a_1}{a_2} p^{a_2}(1-p)^{10a_1-a_2} \cdots \\
& \sum_{a_{k-1}}^{10a_{k-2}} \binom{10a_{k-2}}{a_{k-1}} p^{a_{k-1}}(1-p)^{10a_{k-2}-a_{k-1}} \left(P_2^{a_{k-1}} - P_1^{a_{k-1}}\right). \qquad (3.12)
\end{aligned}$$

By repeating the same procedure as in calculating P_3, we are able to reduce the above expression to

$$\begin{aligned}
P_{k+1} &= P_k + \sum_{a_1=1}^{10} \binom{10}{a_1} p^{a_1}(1-p)^{10-a_1} \left(P_k^{a_1} - P_{k-1}^{a_1}\right) \\
&= P_k + (1-p+pP_k)^{10} - (1-p+pP_{k-1})^{10} \\
&= (1-p+pP_k)^{10}, \qquad (3.13)
\end{aligned}$$

which holds true for any positive integer $k \ge 1$.

We now prove that P_k approaches some constant as $k \to \infty$. Using (3.13), we have that

$$P_k \;=\; (1-p+pP_{k-1})^{10} \;\ge\; 0.$$

Furthermore, if $P_{k-1} \;\le\; 1/2$, then

$$P_k \;\le\; \left(1-p+\frac{p}{2}\right)^{10} \;=\; \left(1-\frac{3}{\pi^2}\right)^{10} \;<\; 0.7^{10} \;<\; \frac{1}{2}.$$

Then, by induction, when the base case is $P_1 \approx 8.5835 \times 10^{-5}$ (from (3.9)), we have that $P_k \le 1/2$ for any $k \ge 1$. Lastly, note that $P_2 \;>\; P_1$, and using strong induction and (3.13), we get that

$$P_{k+1} \;=\; (1-p+pP_k)^{10} \;\ge\; (1-p+pP_{k-1})^{10} \;=\; P_k.$$

In other words, $\big(P_k\big)_{k\geq 1}$ is an increasing sequence. By the monotone convergence theorem, we get that there exists $l \in [0, 1/2]$ such that

$$\lim_{k\to\infty} P_k \;=\; l.$$

Sending $k \to \infty$ in (3.13), we get that

$$l \;=\; (1-p+pl)^{10}.$$

Using Mathematica, we see that the only rational root in the range $[0, 1/2]$ is

$$l \;\approx\; 8.5950 \times 10^5. \tag{3.14}$$

The limit of P_k, or l, stands for the probability that, starting at some fixed number x, there is a bounded limit N, which can be very large, that no square-free walk can exceed length N. That is, if the limit of P_k is as small as 8.5950×10^5, it implies the following theorem.

Theorem 3.2. *Given we append one digit at a time, the probability that there is an infinite random square-free walk from any starting point is as least* $1-l \approx 0.99991$*. In other words, almost always there is such a walk from any starting point.*

Remark 3.3. *While this implies with high probability that we can always walk to infinity on square-free numbers, there exist square-free numbers such that you cannot. For example,* 231546210170694222 *is a square-free number, but if we append any digit to the right we get a non-square-free number. In particular, if we delete any number of digits to the right we get a square-free number as well, so this proves we can reach a stopping point when starting with* 2 *and append digits to the right randomly. Furthermore, our example implies that the walk is not constructive, i.e., if we start with a square-free walk and append a digit at random that yields a new square-free number, we might reach a point where we couldn't move forward.*

3.2. **Results.** From §3.1, according to the greedy model of square-free walks, the expected length of square-free walks is $6/(\pi^2-6)$ in *any* base. In reality, however, this is not always the case.

Dropping the probabilistic assumption about the square-free numbers, a random square-free walk is to start with the empty string, randomly append digits to it, and stop when the number is not square-free. We let E_b denote the expected length of such a walk in base b, and SF the set of square-free numbers.

We first introduce some notations.

Definition 3.4 (Right Truncatable Square Free)**.** *We set* RTSF_b *to be the set of square-free numbers base* b *such that if we successively remove the rightmost digit, each resulting number is still square-free. Equivalently, let* $b^{k-1} \leq x < b^k$*. Then,* $x \in \mathrm{RTSF}_b$ *if and only if for all* $\ell \in \{0, 1, \ldots, k-1\}$ *we have* $\lfloor x/b^\ell \rfloor$ *is square-free.*

Definition 3.5. *Define*

$$L_{b,k} \;:=\; \left|\mathrm{RTSF}_b \cap [b^{k-1}, b^k)\right|. \tag{3.15}$$

Thus $L_{b,k}$ counts the number of right-truncatable square-free numbers with exactly k digits in base b.

Lemma 3.6. *We have*

$$E_b \;=\; \sum_{k=1}^{\infty} \frac{L_{b,k}}{b^k}. \tag{3.16}$$

Proof. The proof follows from identical reasoning as that in proving the equivalent formula, (2.4), for the primes. □

Theorem 3.7. *We have E_2 satisfies the following bounds:*

$$2.31435013 \;<\; \frac{636163720502}{2^{38}} \;\le\; E_2 \;\le\; \frac{636163930777}{2^{38}} \;<\; 2.31435090. \tag{3.17}$$

Proof. A straightforward calculation yields $\{L_{2,n}\}_{1\le n\le 40} = (1, 2, 3, 5, 7, ..., 168220)$.
Let

$$S_1 \;:=\; \sum_{i=1}^{40} \frac{L_{2,i}}{2^i} \;=\; \frac{318081860251}{2^{37}}, \tag{3.18}$$

$$S_2 \;:=\; \sum_{i=41}^{\infty} \frac{L_{2,i}}{2^i}$$

and note that

$$S_1 + S_2 \;=\; E_2. \tag{3.19}$$

Moreover, let $L^O_{2,k} \;=\; \left|\mathrm{RTSF}_2 \cap [2^{k-1}, 2^k) \cap (2\mathbb{Z}+1)\right|$ be the number of odd right truncatable square-free binary numbers of length-k binary numbers, and similarly $L^E_{2,k} \;=\; \left|\mathrm{RTSF}_2 \cap [2^{k-1}, 2^k) \cap 2\mathbb{Z}\right|$ the even ones. By modulo 4 considerations, we have that $L^O_{2,k+1} \;\le\; L^O_{2,k} + L^E_{2,k}$ and $L^E_{2,k+1} \;\le\; L^O_{2,k}$.

$$\begin{aligned}
S_2 \;&=\; \frac{L^O_{2,41}+L^E_{2,41}}{2^{41}} + \sum_{i=41}^{\infty} \frac{L^O_{2,i+1}+L^E_{2,i+1}}{2^{i+1}} \;\le\; \frac{L^O_{2,41}+L^E_{2,41}}{2^{41}} + \sum_{i=41}^{\infty} \frac{2L^O_{2,i}+L^E_{2,i}}{2^{i+1}} \\
&=\; \frac{L^O_{2,41}+L^E_{2,41}}{2^{41}} + \frac{S_2}{2} + \frac{L^O_{2,41}}{2^{42}} + \sum_{i=41}^{\infty} \frac{L^O_{2,i+1}}{2^{i+2}} \\
&\le\; \frac{3L^O_{2,41}+2L^E_{2,41}}{2^{42}} + \frac{S_2}{2} + \sum_{i=41}^{\infty} \frac{L^O_{2,i}+L^E_{2,i}}{2^{i+2}} \\
&\le\; \frac{5L^O_{2,40}+3L^E_{2,40}}{2^{42}} + \frac{3S_2}{4}.
\end{aligned}$$

Thus, we have

$$S_2 \;\le\; \frac{5L^O_{2,40}+3L^E_{2,40}}{2^{40}} \;\le\; \frac{5L_{2,40}}{2^{40}} \;=\; \frac{210275}{2^{38}}. \tag{3.20}$$

As clearly $S_2 \;\ge\; 0$, $\sum_{i=1}^{40} \frac{L_{2,i}}{2^i} \;\le\; E_2 \;=\; \sum_{i=1}^{40} \frac{L_{2,i}}{2^i} + S_2$. Substituting the numerical results from (3.18) yields the bound. □

Although we do not use the base $b = 2$ model for square-free walks anywhere else, it is listed here since the proof of this theorem can be adapted to other bases.

Theorem 3.8. *$2.63297479 \le E_{10} \le 2.720303756$.*

Proof. The proof is similar to that of Theorem 3.7. We use

$$\{L_{10,n}\}_{1\le n\le 8} = (6, 39, 251, 1601, 10143, 64166, 405938, 2568499), \tag{3.21}$$

and the inequalities

$$L^O_{10,k+1} \le 5L^O_{10,k} + 5L^E_{10,k}$$

and

$$L^E_{10,k+1} \le 3L^O_{10,k} + 2L^E_{10,k}.$$

□

As stated earlier, the same proof can be adapted for any base $b = 3, \dots, 9$. This method fails for larger bases due to computational reasons.

3.3. **Discussion of the behaviors of square-free walks.** We first introduce some notation. Given a number x and a digit i in base b, $\overline{xi} = b \cdot x + i$; in other words, we append i to the right of x. The following are some remarks relating to some behaviors of square-free walks.

Remark 3.9. *The fact that $E_{10} > 6/(\pi^2 - 6)$ was expected, since we know that $\overline{xi}$ is more likely to be square-free if x is square-free. This is due to the fact that if x is square-free, then $x \not\equiv 0 \bmod p^2$ for every prime p. In particular, this implies that $[\overline{x0}, \overline{x9}]$ can be any segment of $\mathbb{Z}/p^2\mathbb{Z}$ except $[0, 9]$, hence the chance that $\overline{xi} \not\equiv 0 \pmod{p^2}, \forall i \in [0, 9]$ is slightly bigger. Notice that this behavior is consistent for any base b.*

Remark 3.10. *A computer program yields that when $x \in \{1, 2, \dots, 1{,}000{,}000\}$ is square-free, the probability of $\overline{xi}$ is also square-free is around 0.5944, and when $x \in \{1, 2, \dots, 1{,}000{,}000\}$ is not square-free, the probability of $\overline{xi}$ being square-free is around 0.5669. Note that both these values are larger than $6/\pi^2$. This is because small numbers have a larger chance of being square-free. Furthermore, when x is smaller, i.e. $x \in \{1, 2, \dots, 10^n\}, n < 6$, these probabilities are even larger. As x increases, we expect the two probabilities to decrease, but they still have a small difference.*

Remark 3.11. *We also explore how the starting point affects the length of the walk. As in the prime walks, the expected value of the walk's length decreases as the starting point increases since small numbers have a bigger chance of being square-free. This is shown in Table 9. Note that the expected length of around 2.71 (when the starting point increases) is inside the interval given by Theorem 3.8.*

Remark 3.12. *We also consider the frequency of the digits added in our square-free walk and how this changes when we vary the walk's starting point. The result is shown in Table 10.*

Remark 3.13. *We make the following observations based on the frequency of the digits in base 10.*

- *Odd digits appear more often than even digits. This is because if x is square-free, then it cannot be a multiple of 4, hence even digits appear less.*

- *The frequencies of* 2 *and* 6 *are less than* $0, 4,$ *and* 8*. This is because if* x *and* $\overline{xi}$ *are square-free and* i *is even, then if* x *is odd, by modulo* 4 *considerations* i *is* $0, 4,$ *or* 8*, and if* x *is even, then* i *is* 2 *or* 6*. However,* x *is almost twice more likely to be odd; hence the frequency of* $0, 4, 8$ *is bigger than* 2 *and* 6*.*
- 5 *appears less than any other odd digit. Similar to the above,* $\overline{x5}$ *is not square-free if* x *ends with* 2 *or* 7*.*
- 9 *appears more often than any other digit. This is because if* x *is square-free, then* $x \not\equiv 0 \pmod 9$*, hence* $\overline{x9} \not\equiv 0 \pmod 9$*.*
- *As the starting point increases, the frequencies stabilize.*

Remark 3.14. *By looking at the last digit, we can make informed decisions on what digit to append at each step to increase the chance the number is square-free using the Remark 3.13.*

Digit added \ Number of digits of starting point	1	2	3	4	5	6
0	10.1%	7.4%	7.6%	7.5%	7.5%	7.5%
1	14.0%	13.6%	13.2%	13.4%	13.4%	13.4%
2	8.4%	5.5%	5.3%	5.3%	5.3%	5.3%
3	13.5%	13.5%	13.4%	13.4%	13.4%	13.3%
4	5.1%	8.1%	8.0%	8.0%	8.0%	8.0%
5	12.1%	10.8%	10.9%	10.8%	10.8%	10.8%
6	8.3%	5.5%	5.4%	5.3%	5.3%	5.3%
7	13.4%	13.5%	13.2%	13.3%	13.3%	13.3%
8	4.9%	7.4%	8.0%	8.0%	8.0%	8.0%
9	9.7%	14.2%	14.5%	14.6%	14.6%	14.6%

TABLE 10. Comparing the frequency of the digits of square-free walks in base 10.

3.4. **Refined greedy model.** Lastly, we present an alternative to the greedy square-free walk. As stated in Remark 3.13, odd digits appear most frequently. Using this, we created a different model: if we start with an odd square-free number not divisible by 5, we can always append 0 to get a square-free number, since the initial number is not divisible by 2 or 5. Then randomly append one of $1, 3, 7, 9$. If the number is square-free, repeat the process, otherwise stop and record the length. Using (3.1), we get that the probability that a random odd integer, non-divisible by 5, is square-free is

$$p = \prod_{p \text{ prime } \neq 2,5} \left(1 - \frac{1}{p^2}\right) \;=\; \frac{1}{\zeta(2)} \cdot \frac{1}{1-\frac{1}{4}} \cdot \frac{1}{1-\frac{1}{25}} = \frac{25}{3\pi^2}.$$

Let X denote the number of steps in our refined square-free walk. Note that the number of steps can only be even, since appending 0 to an odd, non-divisible by 5 number yields a square-free number. Therefore, we have that

$$\mathbb{P}[X = 2k] \;=\; p^k(1-p) \;=\; \frac{25^k}{3^k\pi^{2k}} \cdot \frac{3\pi^2 - 25}{3\pi^2} \;=\; \frac{25^k(3\pi^2-25)}{3^{k+1}\pi^{2k+2}}.$$

Analogously to (3.6), we have that

$$\mathbb{E}[X] \;=\; \frac{2p}{1-p} \;=\; \frac{50}{3\pi^2 - 25} \;\approx\; 10.84,$$

which is a lot larger than the expected value in the normal model computed in (3.6). We present the comparison in Table 11.

Start has x digits	1	2	3	4	5	6
greedy square-free walk	2.81	2.76	2.72	2.71	2.71	2.70
refined greedy square-free walk	11.44	9.92	9.79	9.48	9.14	8.80

TABLE 11. Comparing the expected value of greedy square-free models the walks in base 10.

3.5. **Higher-power-free walks.** In general, by following the same procedure as in (3.1), $Q_n(x)$, the number of n^{th}-power-free numbers less than or equal to x, is approximately $x/\zeta(n)$. Thus, the probability any number x is an n-power-free is $1/\zeta(n)$.

It is known that the special values $\zeta(2n)$ can be computed as follows:

$$\zeta(2n) \;=\; (-1)^{n+1}\frac{B_{2n}(2\pi)^{2n}}{2(2n)!}.$$

However, there is no known closed formula for $\zeta(2n+1)$, but, according to [WK], the approximation of $\zeta(3)$ is $1.20205\ldots$. Hence, we obtain the following probabilities.

$$0.831905\ldots \;\leq\; P[x = \text{cube-free}] \;\leq\; 0.831912\ldots,$$
$$P[x = \text{fourth-power-free}] \;=\; 90/\pi^4 \;=\; 0.92393\ldots\;.$$

Since there is no significant change in dealing with cube-free compared with fourth-power-free, we focus only the fourth-power-free sequence due to its precise representation $90/\pi^2$. Similar to 3.1, letting X_4 denote the number of steps in our random fourth-power-free walks, we obtain the expected length and its variance.

$$\mathbb{E}[X_4] \;=\; \frac{90}{\pi^4 - 90} \;=\; 12.14723\ldots,$$

$$\text{Var}(X_4) \;=\; \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \;=\; \frac{p}{(1-p)^2} \;=\; \frac{90\pi^4}{(\pi^4-90)^2} \;=\; 159.7026.$$

Then, we observe if there is any subtle difference in the expected length of random fourth-power-free walks in difference bases: base 2 and base 10.

Definition 3.15. *$RT4F_b$ to be the set of fourth-power-free numbers base b such that if we successively remove the rightmost digit, each resulting number is still square-free.*

Definition 3.16. *$L_{b,k}$ as in definition 3.4, $L_{b,k} \;:=\; |RT4F_b \cap [b^{k-1}, b^k)|$.*

Theorem 3.17. *Let E_b denote the expected length of fourth-free walks in base b.*

$$10.002745850302745 \;\leq\; E_2 \;\leq\; 13.0679694,$$

and

$$5.081603865 \;\leq\; E_{10} \;\leq\; 15.02018838.$$

Proof. Following Theorem 3.7, and obtain $E_b = \sum_{k=1}^{\infty} \frac{L_{b,k}}{b^k}$. Given $S_{1,b} = \sum_{k=1}^{m} \frac{L_{b,k}}{b^k}$ and $S_{2,b} = \sum_{k=m+1}^{\infty} \frac{L_{b,k}}{b^k}$, we use $E_b = S_{1,b} + S_{2,b}$ to estimate E_b. The accuracy of E_b depends on the number of terms, m, of the sum $S_{1,b}$ we are able to compute.

Considering walks in base 2 and given E denotes even and O denotes odd, we find the following inequalities.

(1) $L^E_{2,k+4} \;\leq\; L^O_{2,k+3} + L^O_{2,k+2} + L^O_{2,k+1}$,
(2) $L^O_{2,k+1} \;\leq\; L^O_{2,k} + L^E_{2,k}$.

Given that $A_i = L_{2,m+i}/2^{m+i}$, the relationships above give us a lower and an upper bound of E_2,

$$S_{1,2} \;\leq\; E_2 \;\leq\; S_{1,2} + A_0 + 4A_1 + 8A_2 + 16A_3. \tag{3.22}$$

Similarly, for base 10, the following relationships can be found.

(1) $L^O_{10,k+1} \;\leq\; 5L^O_{10,k} + 5L^E_{10,k}$
(2) $L^E_{10,k+3} \;\leq\; 3L^O_{10,k+2} + 2L^E_{10,k+2} + 13L^O_{10,k+1} + 12L^E_{10,k+1} + 63L^O_{10,k} + 62L^E_{10,k}$.

Also, given that $B_i = L_{10,m+i}/10^{m+i}$, we find a lower and an upper bound of E_{10},

$$S_{1,10} \;\leq\; E_{10} \;\leq\; S_{1,10} + \frac{B_0}{125} + \frac{52B_1}{125} + \frac{24B_2}{5} + 16B_3. \tag{3.23}$$

Substituting the numerical values $S_{1,2} = \sum_{k=1}^{36} \frac{L_{2,k}}{2^k} \approx 10.002745850302745$ in (3.22) and $S_{1,10} = \sum_{k=1}^{7} \frac{L_{10,k}}{10^k} \approx 5.081603865$ in (3.23) yields the bounds. □

Nevertheless, Theorem 3.17 does not help us distinguish E_2 and E_{10} for fourth-power-free walks, but this is a computational issue that can be fixed. One way to do is to compute $S_{1,2}$ and $S_{1,10}$ up to a larger m, more precisely, when $m \approx 110$ in base 2 and $m \approx 50$ in base 10. Another way is to observe the difference between the limits of $L_{2,k+1}/L_{2,k}$ and $L_{10,k+1}/L_{10,k}$, as k approaches ∞, and go from there. For whichever approach, we expect $E_2 < E_{10}$ as we have for square-free walks. Note that, as the exponent n gets larger, the probability that a number is n^{th}-power-free, $1/\zeta(n)$, approaches 1. Hence, the existence of a walk to infinity exists in n^{th}-power-free is even more likely as n gets larger. In such a case, there is nothing much to investigate, so we turn our attention to other sequences whose asymptotic density is 0 just like primes.

4. Walks on Other Sequences

4.1. **Perfect squares.** Similar to how we model prime and square-free walks, the density of perfect squares is approximately $1/\sqrt{n}$ because the number of perfect squares less than n is about $\sqrt{n}$. Then, the asymptotic density of perfect squares is 0, just like primes, because as n approaches ∞, $1/\sqrt{n}$ goes to 0. Although, for sufficiently large n, $1/\sqrt{n} < 1/\log n < 6/\pi^2$ implies that perfect squares are sparser than primes and square-frees, this sequence is still a good choice to study because it has an explicit pattern, $(n^2)_{n\in\mathbb{N}}$. Hence, we seek to investigate if perfect squares have any walk to infinity in this section.

Lemma 4.1. *It is impossible to walk to infinity on perfect squares by appending a bounded, odd number of digits to the right.*

Proof. Suppose some infinite sequence of squares $s_1^2, s_2^2, \ldots$ exists, subject to the following relationship:

$$10^{2n_i-1}s_i^2 + k_i \;=\; s_{i+1}^2, \tag{4.1}$$

such that each n_i is a positive integer less than or equal to some upper bound N, and each k_i is a nonnegative integer such that $k_i < 10^{2n_i-1}$. Then, it is also the case that

$$10^{2n_{i+1}-1}s_{i+1}^2 + k_{i+1} \;=\; s_{i+2}^2, \tag{4.2}$$

with $k_{i+1} < 10^{2n_{i+1}-1}$. Using (4.1) to substitute for s_{i+1}, we have that

$$\begin{aligned} 10^{2n_{i+1}+2n_i-2}s_i^2 + 10^{2n_{i+1}-1}k_i + k_{i+1} &= s_{i+2}^2 \\ 10^{2n_{i+1}-1}k_i + k_{i+1} &= (s_{i+2} + 10^{n_{i+1}+n_i-1}s_i)(s_{i+2} - 10^{n_{i+1}+n_1-1}s_i). \end{aligned}$$

Neither k_i nor k_{i+1} can be equal to 0, for otherwise (4.1) would have an integer of the form $10^{2n_i-1}s_i^2$ be a perfect square, which is impossible (and similarly for (4.2)). Therefore $s_{i+2} - 10^{n_{i+1}+n_i-1}s_i$ is non-zero and so $s_{i+2} > 10^{n_{i+1}+n_i-1}s_i$, implying that $s_{i+2} + 10^{n_{i+1}+n_i-1}s_i > 2\cdot 10^{n_{i+1}+n_i-1}s_i$. Thus

$$\begin{aligned} 10^{2n_{i+1}-1}k_i + k_{i+1} &= (s_{i+2} + 10^{n_{i+1}+n_i-1}s_i)(s_{i+2} - 10^{n_{i+1}+n_i-1}s_i) \\ &> (2\cdot 10^{n_{i+1}+n_i-1}s_i)(1) = 2\cdot 10^{n_{i+1}+n_i-1}s_i. \end{aligned} \tag{4.3}$$

But because $k_i < 10^{2n_i-1}$ and $k_{i+1} < 10^{2n_{i+1}-1}$, it is also the case that

$$10^{2n_{i+1}-1}k_i + k_{i+1} \;<\; 10^{2n_{i+1}+2n_i-2} + 10^{2n_{i+1}-1}, \tag{4.4}$$

and so combining (4.3) with (4.4) yields

$$2\cdot 10^{n_{i+1}+n_i-1}s_i \;<\; 10^{2n_{i+1}+2n_i-2} + 10^{2n_{i+1}-1}. \tag{4.5}$$

Since our sequences of squares is infinite, one can choose an arbitrarily large s_i, while the values of n_i and n_{i+1} are bounded. Therefore it is possible to find an

s_i such that this inequality does not hold. Hence, such an infinite sequence of squares does not exist. □

We can apply a similar argument to the case of appending a bounded, even number of digits. First, however, we observe that it *is* possible to walk to infinity in such a manner.

Lemma 4.2. *It is possible to walk to infinity on perfect squares by appending a bounded, even number of digits to the right.*

Proof. Let s^2 be a perfect square. Then 10^2s^2, 10^4s^2, 10^6s^2, ... are all perfect squares. Thus, appending 00 to the right at each step allows us to walk to infinity on squares. □

This existence proof can be generalized to show that after a certain point, only 0's can be appended to obtain the next square in the sequence.

Lemma 4.3. *Let $s_1^2, s_2^2, ...$ be an infinite sequence of squares such that s_i^2 is generated by appending an even, bounded number of digits to s_{i-1}^2. Let the number of digits appended at each step be no greater than $2N$. Then for $s_i^2 \geq 10^{2N}/4$, it is the case that $s_{i+1}^2/s_i^2 = 10^{2n}$, for some positive integer $n \leq N$.*

Proof. Given, $s_1^2, s_2^2, ...$, we have that

$$10^{2n_i}s_i^2 + k_i \;=\; s_{i+1}^2 \tag{4.6}$$

such that each n_i is a positive integer less than or equal to some upper bound N, and each k_i is a nonnegative integer such that $k_i < 10^{2n_i}$. Then

$$k_i \;=\; (s_{i+1} + 10^{n_i}s_i)(s_{i+1} - 10^{n_i}s_i). \tag{4.7}$$

For a given i, either $k_i = 0$ or it does not. If it does, then we have $10^{2n_i}s_i^2 = s_{i+1}^2$, which corresponds to appending $2n_i$ zeros to the right. Otherwise, if $k_i \neq 0$, then $s_{i+1} > 10^{n_i}s_i$, so $s_{i+1} + 10^{n_i}s_i > 2 \cdot 10^{n_i}s_i$. This means that

$$\begin{aligned} k_i &= (s_{i+1} + 10^{n_i}s_i)(s_{i+1} - 10^{n_i}s_i) \\ &> (2 \cdot 10^{n_i}s_i)(1) = 2 \cdot 10^{n_i}s_i. \end{aligned}$$

But at the same time, $k_i < 10^{2n_i}$, so it must be the case that

$$10^{2n_i} \;>\; 2 \cdot 10^{n_i}s_i. \tag{4.8}$$

For large enough s_i, this is false, because n_i is bounded. In particular, we arrive at a contradiction once $s_i \geq 10^{n_i}/2$. Because $n_i \leq N$, this means that k_i cannot be nonzero once

$$s_i^2 \;\geq\; \frac{10^{2N}}{4}. \tag{4.9}$$

Once this condition is attained, we must have $k_i = 0$ and $s_{i+1}^2/s_i^2 = 10^{2n_i}$, as desired. □

The natural next question is to consider the general case: appending any number of bounded digits, even or odd.

Suppose we have squares $s_i^2, s_{i+1}^2, ..., s_{i+j+2}^2$, for some positive integer j, subject to the following relationships:

$$\begin{aligned} 10^{2n_1-1}s_i^2 + k_1 &= s_{i+1}^2, \\ 10^{2m_1}s_{i+1}^2 &= s_{i+2}^2, \\ 10^{2m_2}s_{i+2}^2 &= s_{i+3}^2, \\ &\vdots \\ 10^{2m_j}s_{i+j}^2 &= s_{i+j+1}^2, \\ 10^{2n_2-1}s_{i+j+1}^2 + k_2 &= s_{i+j+2}^2, \end{aligned}$$

such that $n_1, n_2, m_1, m_2, ..., m_j$ are positive integers less than or equal to some upper bound N, and k_1 and k_2 are nonnegative integers with $k_1 < 10^{2n_1-1}$, $k_2 < 10^{2n_2-1}$. In other words, we append an odd number of digits to s_i^2, then repeatedly append an even number of 0's for j steps, and then append another odd number of digits.

Substituting for s_{i+j+1}^2 in terms of s_{i+j}^2, then s_{i+j}^2 in terms of s_{i+j-1}^2, etc., gives

$$10^{2(n_1+n_2+m_1+\cdots+m_k-1)}s_i^2 + 10^{2(n_2+m_1+\cdots+m_j)-1}k_1 + k_2 \;=\; s_{i+j+2}^2. \tag{4.10}$$

Using difference of squares and the same bounding techniques as in Lemmas 4.1 and 4.3, we find that

$$2 \cdot 10^{n_1+n_2+m_1+\cdots+m_j-1}s_i \;<\; 10^{2(n_2+m_1+\cdots+m_j)-1}k_1 + k_2, \tag{4.11}$$

and using the fact that $k_1 < 10^{2n_1-1}$ and $k_2 < 10^{2n_2-1}$, we have that

$$2 \cdot 10^{n_1+n_2+m_1+\cdots+m_j-1}s_i \;<\; 10^{2(n_1+n_2+m_1+\cdots+m_j-1)} + 10^{2n_2-1}. \tag{4.12}$$

While it is still the case that we can choose arbitrarily large s_i, we can also guarantee that the right side of the inequality remains larger; while each m_i is bounded, j is not, and so given an s_i, the value of the sum $m_1 + \cdots + m_j$ can be chosen to satisfy the inequality.

This observation shows that we cannot rule out the possibility that there are always steps on our walk to infinity in which we append an odd number of digits. On the other hand, neither does it prove that there exists a walk containing infinitely many such odd steps. We can conclude that these odd steps must grow increasingly sparse as the squares grow larger, but the question of whether there could be infinitely many remains open.

Although the answer to the most general case is not complete, determining the existence of infinite walks on perfect squares turns out to be simpler than primes and square-frees due to its discernible pattern. With this observation, we turn our attention to another well-known sequence with zero density and evident pattern, Fibonacci numbers.

4.2. **Fibonacci Numbers.** The Fibonacci sequence is one of the most famous number theory sequences with a great number of remarkable properties and applications (see [Kos]) while its definition is not complicated.

Definition 4.4 (Fibonacci numbers)**.** *Let F_n be the n-th Fibonacci number and $F_0 = 0, F_1 = 1$. For $n \geq 2$, $F_n = F_{n-1} + F_{n-2}$.*

Based on Definition 4.4, it is known that the ratio of any consecutive Fibonacci numbers is approximately the golden ratio, about 1.618, so the number of Fibonacci numbers less than n is about $\log_{1.618} n$. Thus, the density of Fibonacci numbers is around $\log_{1.618} n/n$, which is even less than $1/\sqrt{n}$, the density of perfect squares, and hence less than the density of primes. With this low density and the recurrence relation that depends on the two previous terms, we speculate no walks to infinity in the Fibonacci sequence and present our proof in this section.

First, we establish some relations between any two Fibonacci with k order apart, i.e., F_m and F_{m+k}, to help prove our claim.

Lemma 4.5. *For all* $m, k \in \mathbb{N}$, $F_{k+1}F_m \leq F_{m+k} \leq F_{k+2}F_m$.

Proof. Let m be any positive integer. We show that the statement is true for all $k \in \mathbb{N}$ by strong induction.

For the base cases $k = 1$ and $k = 2$, by Definition 4.4 and the fact that the Fibonacci sequence is increasing, $F_m \leq F_{m+1} \leq 2F_m$ holds, and $2F_m \leq F_{m+2} \leq 3F_m$ follows by adding F_m throughout the prior inequality.

For the inductive step, suppose that for all k with $2 \leq k \leq r \in \mathbb{N}$, $F_{k+1}F_m \leq F_{m+k} \leq F_kF_m$. We have $F_rF_m \leq F_{m+r-1} \leq F_{r+1}F_m$ and $F_{r+1}F_m \leq F_{m+r} \leq F_{r+2}F_m$ from our supposition. Then combining both inequalities above gives us

$$F_{r-1}F_m + F_rF_m \ \leq\ F_{m+r-1} + F_{m+r} \ \leq\ F_rF_m + F_{r+1}F_m,$$

which is, by Definition 4.4, equivalent to

$$F_{r+1}F_m \ \leq\ F_{m+r+1} \ \leq\ F_{r+2}F_m.$$

□

Lemma 4.6. *For all* $m > k \in \mathbb{N}, k > 2,\ F_{m+k} = (F_{k+2} - F_{k-2})F_m + (-1)^{k+1}F_{m-k}$.

Proof. Let $\varphi = \frac{1+\sqrt{5}}{2}$ be the golden ratio. Computing the two sides, we have that

$$\varphi^4 - 1 \ =\ \sqrt{5}\varphi^2. \tag{4.13}$$

Furthermore, Binet's formula tells us that

$$F_n \ =\ \frac{\varphi^n + (-\varphi)^{-n}}{\sqrt{5}}. \tag{4.14}$$

By (4.14), we have that $F_{m+k} = \frac{\varphi^{m+k}+(-\varphi)^{-m-k}}{\sqrt{5}}$, and

$$\begin{aligned} &(F_{k+2} - F_{k-2})F_m + (-1)^{k+1}F_{m-k} \\ =\ &\frac{(\varphi^4 - 1)(\varphi^{k-2} + (-\varphi)^{-k-2})(\varphi^m + (-\varphi)^{-m})}{5} + (-1)^{k+1}\frac{(\varphi^{m-k} + (-\varphi)^{k-m})}{\sqrt{5}}. \end{aligned}$$

Now, by (4.13), we get that the above expression is equal to

$$\frac{\varphi^2(\varphi^{k-2}+(-\varphi)^{-k-2})(\varphi^m+(-\varphi)^{-m})}{\sqrt{5}}+(-1)^{k+1}\frac{(\varphi^{m-k}+(-\varphi)^{k-m})}{\sqrt{5}}$$
$$=\frac{\varphi^{m+k}+(-\varphi)^{-m-k}+\varphi^{m-k}(-1)^{-k}+\varphi^{k-m}(-1)^{-m}}{\sqrt{5}}-\frac{(-1)^k\varphi^{m-k}+(-1)^m\varphi^{k-m}}{\sqrt{5}}$$
$$=\frac{\varphi^{m+k}+(-\varphi)^{-m-k}}{\sqrt{5}} \;=\; F_{m+k}.$$

□

Lemmas 4.5 and 4.6 imply the following theorem.

Theorem 4.7. *It is impossible to construct an infinite walk on the Fibonacci sequence by appending exactly one digit at a time. In particular, all such walks have length at most 2.*

Proof. Starting with some Fibonacci number $F_m \geq 1$, if we append $d \in \{0,1,2,\ldots,9\}$, the newly appended number is $10F_m + d$. From Lemmas 4.5 and 4.6, we have the following relations:

$$5F_m \;\leq\; F_{m+4} \;\leq\; 8F_m \;\leq\; F_{m+5} \;\leq\; 13F_m \;\leq\; F_{m+6} \;\leq\; 21F_m, \qquad (4.15)$$

and

$$\text{For all, } m > 5 \in \mathbb{N},\; F_{m+5} \;=\; 11F_m + F_{m-5}. \qquad (4.16)$$

From (4.15), since $10F_m+d$ must be a Fibonacci number and $8F_m \leq 10F_m \leq 13F_m$, $10F_m + d$ is either F_{m+5}, or F_{m+6} in some odd cases.

- If $10F_m + d = F_{m+6}$, we have that $10F_m + d > 13F_m$ and thus $d > 3F_m$. Since d is a single-digit number, the possible values of F_m are 1 and 2.
- If $10F_m+d = F_{m+5}$, (4.16) tells us that if $m > 5$, $10F_m+d = 11F_m+F_{m-5}$, so $d = F_m + F_{m-5}$. Again, since $0 \leq d \leq 9$, $F_m \leq 9$. As $m > 5$, the only possible value is $F_m = 8$. For other cases when $m \leq 5$, we have to manually check it.

From both cases, we conclude that any walks must start from some $F_m = 1, 2, 3, 5,$ or 8. This fact gives us that there are only 5 possible walks of length 2, namely, $1 \to 13$, $2 \to 21$, $3 \to 34$, $5 \to 55$, and $8 \to 89$. □

Now, let us apply the same technique to a more general case when we are allowed to appended exactly N digits at a time. By appending exactly N digits, we also include appending as 0's leading numbers, for example, 001 or 0000002123.

Lemma 4.8. *There is no k such that $F_{k+2} - F_{k-2} = 10^N$, for all natural numbers N and k, where $k \geq 2$.*

Proof. Based on [SU] where the first 300 Fibonacci numbers are listed, $F_{61} \equiv F_1 \pmod{10}$ and $F_{62} \equiv F_2 \pmod{10}$. Then, by the Fibonacci definition, we prove by induction that for any positive integer n, $F_{60+n} \equiv F_n \pmod{10}$.

The periodic property tells us that if we cannot find a pair of Fibonacci, F_{k+2}, F_{k-2}, where $2 \le k \le 62$, such that $F_{k+2} - F_{k-2} \equiv 0 \pmod{10}$, then there exists no k such that $F_{k+2} - F_{k-2} \equiv 0 \pmod{10}$; as a result, it is impossible to have $F_{k+2} - F_{k-2} = 10^N$. This is because if there is no such a pair when $2 \le k \le 62$, neither does when $2 + 60m \le k \le 62 + 60m$, where $m \in \mathbb{Z}^+$. Hence there are no possible pairs for any integer $k > 2$.

Going through the list on [SU], there exists no such a pair in the first 62 Fibonacci numbers which completes our proof. □

Lemma 4.8 then serves as a tool to draw a conclusion for some cases in the following theorem.

Theorem 4.9. *It is impossible to construct an infinite walk on the Fibonacci sequence by appending exactly N digits at a time, where N is a fixed positive integer. In particular, any appendable step in the walk must be of length at most* $8/7 \cdot (10^N - 1)$.

(Note: In this case, an appendable step refers to a step in a walk such that we can append some N-digit number and still get a Fibonacci. When a step is not appendable, the walk terminates.)

Proof. In any walk, let N be a fixed positive integer and F_m be the starting number in the walk. Similar to Theorem 4.7 the next step in the walk can be written as

$$10^N F_m + d, \text{ where } 0 \le d \le 10^N - 1.$$

By Lemma 4.5, we know there exists k such that

$$F_{k+1}F_m \ \le\ F_{m+k} \ \le\ F_{k+2}F_m, \tag{4.17}$$

when $F_{k+1} \le 10^N$ and $F_{k+2} > 10^N$. Again, while the most likely case is when $10^N F_m + d = F_{m+k}$, there are two unlikely cases: $10^N F_m + d < F_{k+1}F_m$ and $10^N F_m + d > F_{k+2}F_m$. We start with the last two atypical cases.

Case 1: $10^N F_m + d < F_{k+1}F_m$. (4.17) fails because $F_{k+1} \le 10^N$ and d is positive.

Case 2: $10^N F_m + d > F_{k+2}F_m$. Since $2 \cdot 10^N \ge F_{k+2} > 10^N$ and $0 \le d \le 10^N - 1$, we have that

$$10^N F_m + d \ \ge\ (10^N + 1)F_m,$$

meaning that any appendable F_m in the walk must be $\le d \le 10^N - 1$.

Now we consider the most common case.

Case 3: $10^N F_m + d = F_{m+k}$. From Lemma 4.6, we have

$$\begin{aligned} 10^N F_m + d &= (F_{k+2} - F_{k-2})F_m + (-1)^{k+1}F_{m-k} \qquad (4.18)\\ d &= (F_{k+2} - F_{k-2} - 10^N)F_m + (-1)^{k+1}F_{m-k} \\ 10^N - 1 &\ge (F_{k+2} - F_{k-2} - 10^N)F_m + (-1)^{k+1}F_{m-k} \ \ge\ 0. \end{aligned}$$

Thus, $F_{k+2} - F_{k-2} \ge 10^N$; otherwise, (4.18) would not hold. If $F_{k+2} - F_{k-2} \ge 10^N + 2$, we obtain that

$$10^N - 1 \ \ge\ 2F_m + (-1)^{k+1}F_{m-k} \ \ge\ F_m.$$

Hence, any appendable F_m in the walk must be $\leq 10^N - 1$. However, it is more complicated when $F_{k+2} - F_{k-2}$ is 10^N or $10^N + 1$.

If $F_{k+2} - F_{k-2} = 10^N$, from Lemma 4.8, we know that this case is not possible.

On the other hand, if $F_{k+2} - F_{k-2} = 10^N + 1$, from (4.18) and $\leq d \leq 10^N - 1$, we have

$$\begin{aligned} 10^N F_m + d &= (10^N + 1)F_m + (-1)^{k+1} F_{m-k} \\ 10^N - 1 &\geq F_m + (-1)^{k+1} F_{m-k} \geq 0. \end{aligned}$$

If k is odd, it is clear that F_m has to be $\leq 10^N - 1$ just like the result we have had so far.

If k is even, we have $10^N - 1 \geq F_m - F_{m-k} \geq 0$. Therefore, $F_m \leq 10^N - 1 + F_{m-k}$, so if we can approximate an upper bound of F_{m-k} in terms of F_m, we can find a bound for F_m. We know that $k \geq 5$ because when $N = 1$, $k = 5$ by Theorem 4.7. Then, $F_{m-k} \leq F_{m-5} \leq F_m/8$ by using the bounding technique in Lemma 4.5. Therefore, we have that $F_{m-k} \leq F_m/8$, so $10^N - 1 \geq 7F_m/8$, or $F_m \leq 8/7 \cdot (10^N - 1)$.

Thus, since the bound $8/7 \cdot (10^N - 1)$ is greater than $10^N + 1$, we conclude that any appendable step in the walk must be less than $8/7 \cdot (10^N - 1)$. This implies no walks to infinity on the Fibonacci sequence, given we append exactly N digits each time. □

Corollary 4.10. *The implication of Theorem 4.9 is that any appendable step in a walk must contain at most* $\log(8/7 \cdot (10^N - 1)) \approx 0.058 + N = N + 1$ *digits, given we append exactly* N *digits each time. Since any number at most* $8/7 \cdot (10^N - 1)$ *will contain at least* $N + 1$ *digits after appended by* N *digits one time, the walk must be at most of length 3.*

Theorem 4.11. *Given we can append at most* N *digits each time and the starting number contains* $N_1 \geq 2$ *digits, the length of the longest walk is at most* $\lfloor \log_2 \frac{N}{N_1 - 1} \rfloor + 2$. *If* $N_1 = 1$*, the length of the longest walk is at most* $\lfloor \log_2 \frac{N}{N_1} \rfloor + 2$.

Proof. Given that we start with a Fibonacci number A_1 that has N_1 digits. Clearly, $10^{N_1 - 1} \leq A_1 \leq 10^{N_1} - 1$. From Theorem 4.9, $8/7 \cdot (10^{N_1 - 2} - 1) \leq 10^{N_1 - 1} \leq A_1$ tells us that we cannot append $0, 1, \ldots, N_1 - 2$ digits to A_1. Thus, we can only append $N_1 - 1$ digits or above in the first appending. Notice that after the first appending, the newly appended number, A_2, now contain at least $N_1 + N_1 - 1 = 2N_1 - 1$. Then, by the same analysis, $10^{2N_1 - 1} \leq A_1$ tells us that we can only append $2N_1 - 2$ or above number of digits in the second appending. Repeating the process above, we are required to append at least $2^{M-1}(N_1 - 1)$ digits at the M-th step. Hence, we can determine the largest M as follows.

$$\begin{aligned} 2^{M-1}(N_1 - 1) &\leq N \\ M &\leq \log_2 \frac{N}{N_1 - 1} + 1. \end{aligned}$$

Therefore, the length of the longest walk is at most $\lfloor \log_2 \frac{N}{N_1 - 1} \rfloor + 2$ when including the starting number. Notice that this formula does not work for $N_1 = 1$ since we do not want to append $N_0 - 1 = 0$ digit in the first time. Still, by similar analysis, we obtain that $\lfloor \log_2 \frac{N}{N_1} \rfloor + 2$ is the length of the longest walk starting with a single-digit number. □

By exploiting several special relations among Fibonacci numbers, we conclude that there is no walk to infinity on the Fibonacci sequence, given we append at most N digits at a time to the right. In addition, the length $\lfloor \log_2 \frac{N}{N_1 - 1} \rfloor + 2$ in Theorem 4.11 suggests us that the length of any walk on Fibonacci is relatively small compared to N, which is a fixed positive integer.

5. Conclusion

In the exploration to find a walk to infinity along some number theoretical sequences given we append a bounded number of digits, we establish several results for different sequences. Our study choices depend on their density: from the least to the highest one, we study Fibonacci, perfect squares, primes, and N^{th}-power-free. Where we couldn't prove anything concrete, we used stochastic models that approximate the real world fairly well in the ranges we study.

Utilizing some stochastic models leads us to a conjecture that there are no walks to infinity for primes, a sequence of zero density with no discernible pattern in its occurrence, but that it is very likely to have one for square-free numbers whose density is a positive constant. Additionally, we provide the answers to this question for other sequences, namely perfect squares, Fibonacci numbers, and primes in smaller bases. We show that it is impossible to walk to infinity on primes in base 2, 4, or 5 if appending 1 or 2 digits at a time. Similarly, we obtain that the answer is no for Fibonacci with any bounded step size. Lastly, we find a way to append an even bounded number of digits indefinitely for perfect squares.

Stochastic models give us a strong inclination to determine whether we can walk to infinity along some number theory sequence. The results presented in this paper suggest simple speculation that small density leads to the absence of the walks to infinity. However, as we mainly observe sequences based on their density, it remains to be determined how much other factors, such as the sequence's pattern or structure, may contribute as well.

References

[KL] A. V. Kontorovich and J. C. Lagarias, *Stochastic models for the $3x + 1$ and $5x + 1$ problems and related problems*, in The Ultimate Challenge: The 3x+1 Problem (Jeffrey C. Lagarias, Editor), American Math Society 2010.

[Kos] T. Koshy, *Fibonacci and Lucas Numbers with Applications*, Wiley-Interscience, New York, 2001.

[MT-B] S. J. Miller and R. Takloo-Bighash, *An Invitation to Modern Number Theory*, Princeton University Press, 2006.

[MS] H. Montgomery and K. Soundararajan, *Beyond pair correlation*, Paul Erdős and his mathematics, Vol. I (Budapest, 1999), Math. Studies 11, Bolyai Society (Budapest), 2002, pp. 507–514.

[NN] *Normal Number*, Retrieved from http://mathworld.wolfram.com/NormalNumber.html.

[SU] *Fibonacci list*, Retrieved from http://www.maths.surrey.ac.uk/hosted-sites/R.Knott/Fibonacci/fibtable.html

[TP] *Trucnatable primes*, Retrieved from https://primes.utm.edu/glossary/page.php?sort=RightTruncatablePrime.

[WK] $\zeta(3)$ *approximation*, Retrieved from https://en.wikipedia.org/wiki/Riemann_zeta_function

(Steven J. Miller) Department of Mathematics and Statistics, Williams College, Williamstown, MA 01267
Email address: sjm1@williams.edu

(Fei Peng) Carnegie Mellon University, Pittsburgh, PA 15213
Email address: fpeng1@andrew.cmu.edu

(Tudor Popescu) Carnegie Mellon University, Pittsburgh, PA 15213
Email address: tpopescu@andrew.cmu.edu

(Joshua M. Siktar) University of Tennessee-Knoxville, Knoxville, TN 37916
Email address: jsiktar@vols.utk.edu

(Nawapan Wattanawanichkul) Bowdoin College, Brunswick, ME 04011
Email address: nwattana@bowdoin.edu