On Bernoulli Convolutions and Related Topics

Mason Rogers, Arianna Serafini, Anav Sood

August 2017

Abstract

A Bernoulli convolution is the distribution ν_λ associated with the random series $\sum_{i=0}^{\infty} \pm\lambda^i$. We construct and study T_r, a graphical structure referred to as a r-relational binary tree, that is closely related to Bernoulli convolutions for certain $\lambda \in (0,1)$. In particular, we develop methods for counting the number of nodes of T_r given restrictions on r, and relate this to determining the number of distinct values particular finite Bernoulli convolutions can take. We also develop methods for studying the asymptotics of number of overlapping paths in T_r through analogizing to m-Bonacci sequence representations of integers and careful treatment of generating functions.

Additionally, treating the Bernoulli convolution as a geometrically scaled random walk along the real line, we generalize the Bernoulli convolution to n dimensions. We show that the generalized n-dimensional Bernoulli convolution is absolutely continuous with respect to Lebesgue measure in $\mathbb{R}^n$ for all $\lambda \in (0,1)$ and $n \geq 2$.

1 Introduction and Statement of Results

1.1 Historical Background

A Bernoulli convolution ν_λ is the probability measure associated to the random variable

$$Y_\lambda = \sum_{i=0}^{\infty} \lambda^i X_i,$$

where $\lambda \in (0,1)$ and the X_i are i.i.d. random variables which take values ± 1 each with probability $\frac{1}{2}$. Because $\lambda < 1$, the series converges absolutely, and the associated measure is compactly supported on the set $[\frac{-1}{1-\lambda}, \frac{1}{1-\lambda}]$. It is known that the Bernoulli convolution measure ν_λ is of pure type depending on λ, meaning the measure associated to any choice of λ is either totally singular or absolutely continuous with respect to Lebesgue measure. Which type ν_λ is depends on the size of λ as well as its algebraic properties. Study of Bernoulli convolutions dates back to the 1930s beginning with a series of papers published in 1935 by Aurel Wintner and collaborators [4] [10] [11], since which mathematicians have approached this question with a wide array of techniques. Still, the measures are only understood for a few select classes of λ.

It has long been understood that for $\lambda \in (0, \frac{1}{2})$, Y_λ exists on a set with Lebesgue measure 0, so ν_λ must be totally singular for such λ. In 1939, Paul Erdős [5] showed that when λ is the reciprocal of a Pisot-Vijayaraghavan (PV) number, the resulting law ν_λ is totally singular. A PV number is an algebraic integer greater than 1, all of whose Galois conjugates have modulus less

than 1; the golden ratio is an example. The argument shows that because powers of PV numbers converge exponentially to the set of integers, the characteristic function $\widehat{\nu_\lambda}(\xi)$ does not tend to zero as ξ grows to infinity. The Riemann-Lebesgue lemma then ensures that ν_λ cannot be absolutely continuous, and as it is of pure type, it must be singular. This is the only class of numbers where λ is known to be singular for $\lambda \in (\frac{1}{2}, 1)$. Some notable results include:

(a) The set of $\lambda \in (\frac{1}{2}, 1)$ for which ν_λ is singular has Hausdorff dimension 0 [8].

(b) In 1962, Garsia [6] showed for λ such that λ^{-1} is an algebraic integer with Mahler measure 2, then ν_λ absolutely continuous. This includes all integer roots of $\frac{1}{2}$.

The random variable Y_λ can also be thought of as a geometrically scaled random walk on the real line. In this way we can represent the set of outcomes of Y_λ as instantiations of the random walk, where on the k-th step a walker takes a step of length λ^{k-1} either to the left or to the right. After finitely many steps, the distribution governing a walker's location is discrete, and it makes sense to talk about the information-theoretic entropy H of the distribution.

When λ is an algebraic integer, the entropy can be intimately connected to its algebraic properties. In 1963, Garsia [7] extended the notion of entropy to the limit of the infinite random walks in an attempt to quantify the singularity of such measures. Letting N_n denote the number of possible locations of a walker after n steps, indexing these locations by i, and letting π_i denote the probability of being at the i-th location after n steps, he defined the entropy for the random walk as

$$H_\lambda = \lim_{n\to\infty} \frac{-\sum_{i=1}^{N_n} \pi_i \ln(\pi_i)}{n \ln(1/\lambda)}.$$

Note that both the numerator and denominator are positive as written. Garsia proved that $H_\lambda < 1$ is sufficient to conclude ν_λ is singular. He then showed that when λ is the reciprocal of a PV number, one has $H_\lambda < 1$; however, he required that ν_λ was singular to prove this.

Motivated by Garsia's work, we construct and study graphical structures called r-relational binary trees which encapsulate entropic information about Bernoulli convolutions in particular situations where λ is an algebraic integer. The r-relational binary trees relate to the random walk perspective of Bernoulli convolutions and are interesting objects themselves, and we discuss both of these vantage points. Overall, this discussion reveals some interesting results that arise from the algebraic properties of particular λ values. In addition, we extend the random walk motivated by Bernoulli convolutions into the n-dimensional case, and explore its associated generalized Bernoulli convolution.

1.2 Main Theorems

Theorem 3.13. For an r-relational binary tree T_r such that r has length ℓ and no gaps, we have $N_n = \sum_{i=0}^{n+1} F_i^{\ell-1}$, where $\{F_i^m\}_{i\in\mathbb{Z}}$ is the sequence of m-bonacci numbers.

Theorem 3.20 and **Corollary 3.21.** Let ν_λ be a Bernoulli convolution with graph G_λ such that λ has minimal polynomial $m_\lambda(x) = \sum_{i=0}^{\ell-1} r_{i+1}x^i$ over $\mathbb{Q}$, and T_r be the r-relational binary tree where $r = (r_1, \ldots, r_\ell)$. The number of nodes on the n-th level of G_λ is bounded above by the number of nodes on the n-th level of T_r, namely $\mathcal{N}_n \leq N_n$. In the case that $r_j \neq 0$ for $j \in \{1, \ldots, \ell\}$, we have $\mathcal{N}_n \leq \sum_{i=0}^{n+1} F_i^{\ell-1}$.

Theorem 3.25 and **Corollary 3.26.** Suppose $\lambda_d \in (0,1)$ is the root of the polynomial $1 - x - \cdots - x^d$, then $G_{\lambda_d} \cong T_r$ where $r = (1, \underbrace{-1, \ldots, -1}_{d \text{ times}})$ and thus we have that $\mathcal{N}_n = \sum_{i=0}^{n+1} F_i^d$ for G_{λ_d}.

Theorem 4.12. There is a two-to-one correspondence between integers that can be written in n distinct ways as a sum of Fibonacci numbers and a certain set of size-n equivalence classes of outcomes for truncated Bernoulli random walks.

Theorem 4.14. We can compute the number of such equivalence classes, which yields a formula for the number of values a truncated Bernoulli random walk can achieve by a specified number of paths.

Proposition 4.17. We compute that the number of integers with two Tribonacci representations of length n (G_n) obeys (for sufficiently large n) a recursion $G_n = G_{n-1} + G_{n-2} + 4F_{n-3}$, where F_n is the Fibonacci sequence. This contrasts a result for integers with a unique Tribonacci representation by Kocábová et al [9].

Theorem 5.3. For integers $n \geq 2$, the n-dimensional Bernoulli convolution $\sum_{i=0}^{\infty} X_i \lambda^i$, where the X_i's are i.i.d. random variables chosen uniformly from the set $\{x \in \mathbb{R}^n \ : \ |x| = 1\}$, is absolutely continuous for all $\lambda \in (0,1)$.

2 Theoretical Background

Bernoulli convolutions—here defined as the law of a random variable—are natural to study in the language of probability theory. Thus a (very) basic knowledge of probability and measures will be necessary. We refer the reader to Durrett's textbook [3], and here we will only attempt to cover the essential definitions and facts, glossing over some technicalities when appropriate.

Some of the fundamental features of probability are observed in everyday scenarios. For instance, if we role two dice and compute their sum, we have some sense that 7 is the number that appears most often, and 12 and 2 are the numbers that appear least. We also know that if you take a room full of people and line them up, their heights will tend to clump toward the average. As a final example, it is apparent that a weighted coin showing heads $\frac{1}{4}$ of the time must show tails for about $\frac{3}{4}$ of its tosses. In the first example above, the dice summing to 7 and the dice summing to 12 are two different "events", i.e. possible outcomes when you roll two dice, and they each have a certain likelihood of occurring. Our intuition tells us that two rolls summing to 7 is more likely than two rolls summing to 12, and this should be reflected in their respective probabilities. In the second example, the average height serves as an "expected value" toward which all the heights in the room tend. In the third example, a coin showing tails is the "complement" of it showing heads, so their two probabilities must sum to 1, representing the certainty that when a coin is flipped, it will necessarily either come up heads or tails.

To give ourselves a setting in which to discuss these concepts mathematically, we introduce the notion of a probability space.

Definition 2.1. A *probability space* is a triple $(\Omega, \mathcal{F}, P)$, where Ω is a set of outcomes, $\mathcal{F}$ is a *σ-algebra*: a collection of subsets of Ω satisfying 1) if $A \in \mathcal{F}$ then $A^C \in \mathcal{F}$ and 2) if $\{A_i\} \in \mathcal{F}$ is a countable sequence of sets then $\bigcup A_i \in \mathcal{F}$. P is a function from $\mathcal{F}$ to $[0,1]$ which takes every element in $\mathcal{F}$ and assigns it a probability. This function, called a probability measure, also must

satisfy several rules: $P(\Omega) = 1$, $P(A) \geq P(\emptyset) = 0$ for every $A \in \mathcal{F}$, and if $\{A_i\}$ is a countable sequence of disjoint events in $\mathcal{F}$, then $P(\bigcup_i A_i) = \sum_i P(A_i)$.

To interpret this definition, one can consider elements of Ω to be the simplest building blocks of "things that could happen," and $\mathcal{F}$ to be a set of events, i.e. the combinations of these outcomes that we care about investigating. P provides us with a formal definition of likelihood for the elements in this set. The rules that $\mathcal{F}$ and P must obey then become common sense: for example, if we are able to talk about the probability that any event occurs, we must be able to talk about the probability that the "opposite" of that event occurs, motivating our requirement that $A \in \mathcal{F} \Rightarrow A^C \in \mathcal{F}$.

However, this definition is still rather general, so we will specify what it means on the real line (that is, when $\Omega = \mathbb{R}$). In this case, most of the time $\mathcal{F}$ is taken to be the Borel σ-algebra, which is the smallest σ-algebra which contains all open sets of the real line. There are many different measures than can be placed on the real line, and even more if we drop the condition that $P(\Omega) = 1$; by doing so we would allow all measures, not just probability measures. The most common measure on the real line is *Lebesgue measure*, which corresponds to the usual notion of length. For example, the Lebesgue measure of $(a, b]$ is simply $b - a$, the Lebesgue measure of a single point is 0, and the Lebesgue measure of an unbounded interval is ∞. While rigorously defining Lebesgue measure is cumbersome, understanding it intuitively as length will be sufficient for our purposes.

To talk about other measures, specifically probability measures, on the real line, it is helpful to introduce the concept of a random variable and its distribution.

Definition 2.2. Given a probability space as defined above, a *random variable* is a function $X : \Omega \to \mathbb{R}$ that is well-behaved in that for every Borel set in $\mathbb{R}$, we have $X^{-1}(B) \in \mathcal{F}$. Predictably, a *random vector* is defined identically except that its codomain is $\mathbb{R}^n$.

A random variable naturally defines a probability measure μ from the Borel sets to $[0, 1]$ called its *distribution* or *law*, defined by

$$\mu(A) = P(X \in A) = P(X^{-1}(A)).$$

In other words, we have $\mu = P \circ X^{-1}$, meaning that μ is equivalent to pulling back any well-behaved $A \subset \mathbb{R}$ and then looking at its probability under the original measure. We say that X takes value within A *almost surely* if $\mu(A) = 1$. A random variable formalizes such things as dice rolls, height measurements, and coin flips, if for instance we let landing on heads $= 1$ and landing on tails $= -1$. It can be thought of as a repeatable experiment which, although not giving the same outcome every time it is sampled, will behave in the expected way as the experiment is conducted over and over again. By convention, random variables are denoted by capital letters, most often X or Y, while their distributions will be typically labelled μ or ν.

To completely describe the law of a given random variable X, we can consider its *distribution function* (which is also called its *cumulative distribution function* or CDF) defined by

$$F(x) := P(X \leq x).$$

This function's derivative (or lack thereof) is also important.

Definition 2.3. The measure of a random variable X is said to be *absolutely continuous* if there exists some function f such that

$$F(x) = \int_{-\infty}^{x} f(y)\, dy,$$

where $F(x)$ is the measure's CDF as defined above. We call this f (if it exists) the *density function* of X.

Equivalently, a measure μ associated to a random variable X is absolutely continuous if there exists some function f such that

$$\mu(D) = \int_D f(y)\ dy$$

for every $D \in \mathcal{F}$ (this is the exact same function f as above).

This density function can almost be thought of as being equivalent to $P(X = x)$, even though $P(X = x) = 0$ for any x due to absolute continuity. However, it "weights" every x in relation to other points. For example, if we define a random variable such that $\frac{1}{3}$ of the time it takes on value uniformly within $[0, 1]$ and $\frac{2}{3}$ of the time it takes on value uniformly within $(1, 2]$, then $f(\frac{1}{2}) = \frac{1}{3} = \frac{1}{2}f(\frac{3}{2})$. It is important to note that f must be nonnegative because $F(x)$ is increasing, and that because $\lim_{x\to\infty} F(x) = 1$, $\int_{-\infty}^{\infty} f(y)\ dy = 1$.

Additionally, the second definition for above can easily be extended into higher dimensions: the measure μ of a random *vector* X is *absolutely continuous* if there exists some function f such that

$$\mu(A) = \int_A \cdots \int f(\vec{y})\ dV$$

for every A in the higher-dimensional Borel sets. The first definition is most easily generalized when there is enough symmetry that we can look at a CDF in one variable.

The opposite of absolute continuity is singularity:

Definition 2.4. The measure μ of a random variable X is said to be *singular* (with respect to Lebesgue measure) if there exists a set $A \subset \mathbb{R}$ such that $\mu(A) = 1$ but the Lebesgue measure of A is 0. In other words, singularity means that we can find a zero-length subset of the reals such that X will almost surely take on a value within this set.

While it is not necessarily a given that every random variable has either absolutely continuous or singular measure, this is indeed true of Bernoulli convolutions. While a direct result of the self-similar nature of these convolutions (we have that $\nu_\lambda = \lambda\nu_\lambda \pm 1$), this result is actually a specific case of a more general theorem.

Theorem 2.5. (Law of Pure Type, Jessen and Wintner, 1935 [4])
Any convergent infinite convolution of discrete measures is of pure type, i.e. it is either absolutely continuous or singular with respect to Lebesgue measure.

We have covered almost everything we need to properly discuss Bernoulli convolutions; let us address two questions left outstanding. First, how can we be sure that the weighed sum of random variables is itself a random variable, and second, why are their laws called convolutions?

Fortunately and by design, random variables are closed under standard manipulations: sums, products, quotients, scalar multiplication, roots, powers, etc. Specific to our case, for any constant c_i and random variable X_i, c_iX_i is also a random variable. Additionally, the sum of a finite number of random variables is also a random variable, as is the limit of a sequence of random variables. Taken together, these facts tell us that $Y = \sum_{i=0}^{\infty} \lambda^i X_i$ is a random variable and thus has an associated probability measure. Furthermore, this law of the sum is the convolution of the laws of the summands.

Definition 2.6. The *convolution* of μ and ν, denoted $\mu * \nu$, is the function defined by

$$(\mu * \nu)(x) = \int_{-\infty}^{\infty} \mu(x-t)\nu(t)dt.$$

Convolving two functions creates a third function via local averaging, which has the effect of smoothing the original functions. For instance, convolving an absolutely continuous measure with any other measure results in another absolutely continuous distribution, no matter "how singular" the second measure may be.

It is now clear why Bernoulli convolutions carry the name that they do. They are the probability measure associated to the infinite sum of random variables, each of which is scaled version of the classical ± 1 Bernoulli random variable, and so they are equal to the infinite convolution of the measures corresponding to those variables. With this information in hand, we are ready to state and prove results, starting with an investigation into graph-theoretical counting methods.

3 Graph-Based Counting

As discussed above, one way to interpret a Bernoulli convolution $Y_\lambda = \sum_{i=0}^{\infty} \lambda^i X_i$ for $\lambda \in (0,1)$ is as an infinite geometrically scaled random walk on the real line. One can draw the possible realizations of the random walk in a tree-like graphical structure. When distinct random walk paths intersect (such as when λ is the inverse of a PV number), this graph can be a good source of intuition regarding the convolution's behavior. In the following section we formalize and study the graph of a general Bernoulli convolution. In further discussion we adopt the convention that the natural numbers start at 0.

3.1 Relational Binary Trees

To further understand the nature of certain Bernoulli graphs, we will construct and study a more general graphical structure. Before delving into formalism, let us set forth some terminology. A binary tree with n levels is a graph that can be constructed following a recursive algorithm. First, start with a single node; we will say that this node is on the 0-th level. Add a "left edge" and a "right edge" extending below from said node, making it the parent node of a new "left child" and "right child" node. If the parent node is at level k, then we define its children to be at level $k+1$. Repeat this process for every node on the level below until one would construct children past level n. It is clear that the number of nodes at the k-th level is 2^k, and that there is a unique downward path from our first constructed node to any given node in the tree.

Suppose we introduced further constraints on our tree, such that if one was to go right, left, and then left, the resulting node would be the same as if one made mirror decisions: left, right, and then right. Such a graph certainly would not have unique paths to every node, and would have strictly fewer than 2^k nodes at the k-th level for $k > 2$. We give a few pictorial examples. In the below pictures, Figures 1 and 3 are binary tree up to the 4-th level, where we have quotiented out paths by a particular relation; the relation in Figure 1 being that right, right, left is equivalent to left, left, right, and the relation in Figure 3 being that right, right, right is equivalent to left, left, left. We color nodes that are led to by equivalent paths the same non-black color. The corresponding graphs below said figures, Figures 2 and 3, are possible graph representations of the r-relational binary trees corresponding to the relation imposed on the graph above them.

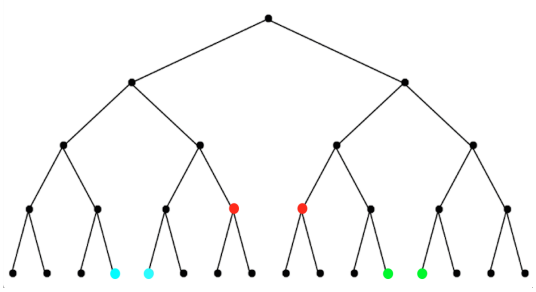

Figure 1: Nodes of the same color represent equivalent paths

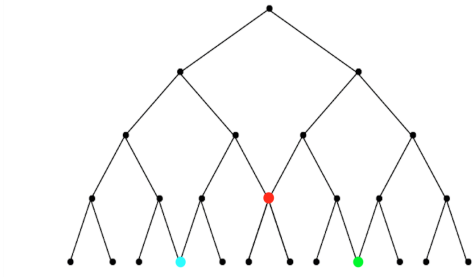

Figure 2: r-relational binary tree where right, left, left is equivalent to left, left, right

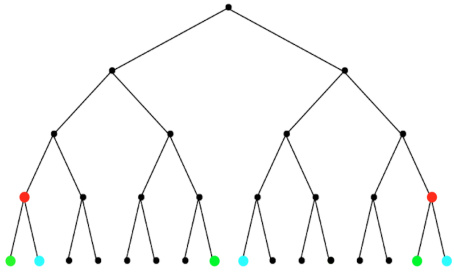

Figure 3: Nodes of the same color represent equivalent paths

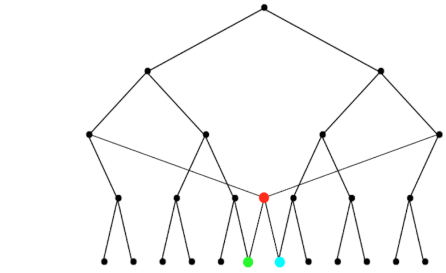

Figure 4: r-relational binary tree where right, right, right is equivalent to left, left, left

Note from Figure 2 we can see that it is fairly easy to physically construct the r-relational binary tree where right, left, left is equivalent to left, right, right. We see from Figure 4, however, that certain r-relation binary trees lose the physical intuition of right and left fairly quickly. This can be explained by the fact that travelling left some finite distance and then right some finite distance can certainly be equivalent to traveling right some finite distance and the left some finite distance. Travelling only right some finite distance, however, is never physically equivalent to traveling only left some finite distance. Thus, although providing illustrated examples of such graphs can be helpful, when one is not careful such examples can be misleading. To study these structures further, we make the following definitions.

Definition 3.1. A *path* of length n is an element of the set $\{-1, 1\}^n$. An *infinite path* is an element of $\{-1, 1\}^\infty$.

In essence, a path is a sequence of right and left steps down the tree, where 1 represents stepping right, and -1 represents stepping left. For simplicity, we will denote the set of all paths with length n as $\mathcal{P}_n$, and the set of all paths up to length n as $\mathcal{P}^n = \bigcup_{i=0}^{n} \mathcal{P}_n$. We define the set of all paths to be $\mathcal{P} = \lim_{n\to\infty} \mathcal{P}^n$. Occasionally, we will write $|p|$ to indicate the size of $p \in P$. In the case that p is infinite, $|p| = \infty$. In general, if we have some path $p \in P$, then p_i represents its i-th component. If the path itself is indexed as p_i, the we will write its j-th component as $(p_i)_j$. Suppose $p = (p_1, \ldots, p_n)$ and $q = (q_1, \ldots, q_m)$. We define the *negation* of path p as $-p = (-p_1, \ldots, -p_n)$. We will write $(p_1, \ldots, p_n, q_1, \ldots, q_n)$, the concatenation of paths p and q, as pq. In the case that we

concatenate p with the negation of some path $-q$, we write the concatenation $p(-q)$. In the special case we wish to concatenate p with ± 1, will write $p^+ = (p_1, \ldots, p_n, 1)$ and $p^- = (p_1, \ldots, p_n, -1)$.

Definition 3.2. A *relation* of length ℓ is an ℓ-tuple $(r_1, \ldots, r_\ell)$ such that $r_1, r_\ell \in \{-1, 1\}$ and $r_j \in \{-1, 0, 1\}$ for all $j \in \{2, \ldots, \ell - 1\}$. The number of *gaps* in a relation is the number of 0-valued components it has.

For $r = (r_1, \ldots, r_\ell)$, we extend our definition of negation, so that $-r = (-r_1, \ldots, -r_\ell)$. Using a relation $r = (r_1, \ldots, r_\ell)$, we can define a notion of equivalence between paths. Take $p \in \mathcal{P}$ for which there exists a $k \in \mathbb{N}$ such that we have either

- $r_i = p_i$ for all $i \in \{k - \ell + 1, \ldots, k\}$ such that $r_i \neq 0$; or
- $r_i = -p_i$ for all $i \in \{k - \ell + 1, \ldots, k\}$ such that $r_i \neq 0$.

Let the set of paths that meet this criterion be denoted $R_{r,k}$. Note that $R_{-r,k} = R_{r,k}$. We define a map $a_{r,k} : R_{r,k} \to R_{r,k}$ such that $a_{r,k}(p) = p'$ where

$$p'_i = \begin{cases} p_i & i \notin \{k - \ell + 1 \ldots, k\} \text{ or } r_i = 0 \\ -p_i & i \in \{k - \ell + 1 \ldots, k\} \text{ and } r_i = 1, -1. \end{cases}$$

We can refer to this action as applying the relation r to the path p at index k. Note that $a_{r,k}$ is its own inverse, and $a_{r,k} = a_{-r,k}$. For the sake of simplicity, if the relation has been specified, we will write $a_{r,k}$ as a_k and $R_{r,k}$ as R_k. If $p \in R_k$, then we say that a_k is a *well-defined* operation on p. Colloquially, we say $a_k(p)$ is well defined. Suppose we have $(a_{i_k} \circ \cdots \circ a_{i_1})(p)$ such that the action of each a_{i_j} is well defined. Then if $A = a_{i_k} \circ \cdots \circ a_{i_1}$, we say that $A(p)$ is well defined.

Definition 3.3. Consider a relation r of length ℓ. The *line of action* of g_k is the set $\{k - \ell + i : i \in \{1, \ldots, \ell\} \text{ such that } r_i \neq 0\}$

Proposition 3.4. Suppose for some $p \in \mathcal{P}$ that $a_{k_1}(p)$ and $a_{k_2}(p)$ are both well defined for some $k_1 \neq k_2$. Then $(a_{k_1} \circ a_{k_2})(p)$ and $(a_{k_2} \circ a_{k_1})(p)$ are both well defined if and only if the intersection of the lines of action of a_{k_1} and a_{k_2} is trivial.

Proof. Let the line of action of a_{k_1} be $\{\alpha_1, \ldots, \alpha_m\}$ and the line of action of a_{k_2} be $\{\beta_1, \ldots, \beta_m\}$. Let $\{\gamma_1, \ldots \gamma_m\}$ be the ordered set of indices such that $r_{\gamma_i} \neq 0$. Note that $a_{k_1}(p)$ being well defined is exactly equivalent to having

- $p_{\alpha_i} = r_{\gamma_i}$ for $i \in \{1, \ldots, m\}$; or
- $p_{\alpha_i} = -r_{\gamma_i}$ for $i \in \{1, \ldots, m\}$

This is similarly true for $a_{k_2}(p)$ being well defined. Suppose that the intersection of the line of actions of a_{k_1} and a_{k_2} is trivial. Then we see that $(a_{k_1}(p))_{\beta_i} = p_{\beta_i}$ and $(a_{k_2}(p))_{\alpha_i} = p_{\alpha_i}$ for $i \in \{1, \ldots, m\}$. Namely, after applying a_{k_1}, the value of the components in the line of action of a_{k_2} are not altered and vice-versa. Because we have that $a_{k_1}(p)$ and $a_{k_2}(p)$ are initially well defined actions, from the equivalent definition above we must have that $(a_{k_1} \circ a_{k_2})(p)$ and $(a_{k_2} \circ a_{k_1})(p)$ are both well defined, proving the reverse direction.

To prove the forward direction, we show the contrapositive. Suppose there is some shared element s_1, of both lines of action. Without loss of generality, assume that $(a_{k_1} \circ a_{k_2})(p)$ is well

defined. Because $k_1 \neq k_2$, we must have some s_2 unique to the line of action of a_{k_1}. We thus see that $(a_{k_2}(p))_{s_1} = -p_{s_1}$ and $(a_{k_2}(p))_{s_2} = p_{s_2}$. Again by the above definition, because both s_1 and s_2 are in the line of action of a_{k_1}, $(a_{k_1} \circ a_{k_2})(p)$ being well defined contradicts that $a_{k_1}(p)$ is well defined, completing the proof.

□

Obviously, if for some $p \in \mathcal{P}$, $(a_\alpha \circ a_\beta)(p)$ and $(a_\beta \circ a_\alpha)(p)$ are both well defined then $(a_\alpha \circ a_\beta)(p) = (a_\beta \circ a_\alpha)(p)$. Thus if such is the case, we say that a_β and a_α commute.

Definition 3.5. We say that $p, q \in \mathcal{P}$ are *equivalent paths* by the relation r if there exists a finite set $\{i_1, \ldots, i_k\}$ such that $(a_{i_k} \circ \cdots \circ a_{i_1})(p) = q$. We denote this equivalency $p \sim_r q$ or simply $p \sim q$ if the relation in question has already been specified.

Thus two paths are equivalent by a relation if and only if there is an ordered list of indices at which one can sequentially apply the relation to obtain one path from the other. From this definition, it follows that $p \sim_r q$ if and only if $p \sim_{-r} q$.

Example 3.6. Consider the path $p = (1, -1, -1, -1, -1)$ and relation $r = (1, -1, -1)$. We see that $a_3(p) = (-1, 1, 1, -1, -1)$ and $(a_5 \circ a_3)(p) = (-1, 1, -1, 1, 1)$. Thus, when considering the above relation, we see that the paths $(1, -1, -1, -1, -1)$, $(-1, 1, 1, -1, -1)$, and $(-1, 1, -1, 1, 1)$ are all equivalent.

Suppose we have a set of paths A. We see that r partitions A into sets of equivalent paths. We define $A/\sim_r$ to be the collection of these partitioned sets. Again, if the relation in use has been specified, we will write $A/\sim_r$ as $A/\sim$. In subsequent discussion it will suffice to simply represent an element of $A/\sim$ by any of its equivalent paths. Using these definitions, we can describe how to construct a relational binary tree.

Definition 3.7. The r-relational binary tree T_r is a graph whose nodes are in bijection with elements of $\mathcal{P}/\sim$, and has edges between nodes η_1 and η_2 if and only if their corresponding elements of $\mathcal{P}/\sim$ are path represented by paths p_1 and p_2 with length m and $m+1$ respectively, and either $p_1^+ \sim p_2$ or $p_1^- \sim p_2$.

From this definition it immediately follows that the identity map is a graph isopmorphism between T_r and T_{-r}.

As the set of nodes of a r-relational binary tree are in bijection with $\mathcal{P}/\sim$, it will be useful to refer to a node by its corresponding element in $\mathcal{P}/\sim$. Since we can refer to an element of $\mathcal{P}/\sim$ by any of its representative paths, we adopt the convention of representing a node by any of these equivalent paths. We say a path *leads to* a node if it can represent that node.

Definition 3.8. Suppose η is a node in the r-relational binary tree T_r. The *level* of η is the length of any path that leads to it, and the *frequency* of η is the number of such paths.

If a node η has level n, we say it is on the n-th level. In examining relational graphs, we are mainly interested in two characterizing features. The first is the number of nodes on the n-th level, a quantity we denote N_n; the second is the number of those nodes on the n-th level with frequency f, a quantity we denote N_n^f. To study the properties of a certain family of relational binary trees, we require the following lemmas.

Lemma 3.9. Let r be a relation of length ℓ with no gaps. If $p, q \in \mathcal{P}_n$ are such that $p \not\sim q$ and $s \in \mathcal{P}_m$, then $ps \not\sim qs$

Proof. We prove the following: Let r be a relation of length ℓ with no gaps, and take $p, q \in \mathcal{P}_n$. If $p \not\sim q$ then $p^+ \not\sim q^+$ and $p^- \not\sim q^-$. Our desired result follows by inducting on this claim.

We show that $p \not\sim q \implies p^+ \not\sim q^+$; the proof for $p \not\sim q \implies p^- \not\sim q^-$ is identical. For the sake of contradiction, assume there is some p, q such that $p \not\sim q$ but $p^+ \sim q^+$. We consider all finite sets $\{i_1, \ldots, i_k\}$ such that $(a_{i_k} \circ \cdots \circ a_{i_1})(p^+) = q^+$. For such sets, k represents the number of times the relation must be applied to p^+ to get to q^+, so naturally, $k = 0$ implies that $p^+ = q^+$. we will show that $k = 0$, thus contradicting our condition that $p \not\sim q$.

We choose $\{i_1, \ldots, i_k\}$ such that k is minimized and let $A = a_{i_k} \circ \cdots \circ a_{i_1}$. Assume again for the sake of contradiction that $k > 0$. From the fact that $p \not\sim q$, we see that we must be able to write A as

$$(a_{j_{k'}} \circ \cdots \underbrace{\circ\, a_{n+1} \circ \cdots \circ a_{n+1} \circ \cdots \circ a_{n+1} \circ \cdots \circ a_{n+1}}_{\text{even number of } a_{n+1}} \circ \cdots \circ a_{j_1})$$

Note we must have an even number of a_{n+1} as we must maintain that $(A(p^+))_{n+1} = 1$, and we must have at least one a_{n+1} as otherwise we would have $p \sim q$ immediately. We consider the innermost pair of a_{n+1}. Consider the relations applied between them to be B_1, so we can write A as

$$a_{j_{k'}} \circ \cdots \circ a_{n+1} \circ B_1 \circ a_{n+1} \circ \cdots \circ a_{j_1}$$

where every relation in B_1 is applied at an index less than $n+1$. Suppose that there is no a_b in the composition B_1 such that the line of action of a_b intersects that of a_{n+1}. Then we could commute B_1 out of its encasing a_{n+1}'s, and write $\cdots \circ a_{n+1} \circ B_1 \circ a_{n+1} \circ \cdots$ as $\cdots \circ B_1 \circ \cdots$, implying that k not minimal. If there is some a_b such that the line of the line of action of a_b intersects that of a_{n+1}, we let β_1 be the largest index the relation is applied in B_1. For $a_{n+1} \circ B_1 \circ a_{n+1}$ to ever be well defined, we require that a_{β_1} appear an even number of times in B_1. Thus in B_1, we will find a similar structure $\cdots \circ a_{\beta_1} \circ B_2 \circ a_{\beta_1} \circ \ldots$, where relation in B_2 is applied at an index less than β_1. In the case that B_2 commutes out, we again have that k not minimal, else we can repeat the above process. After some finite m repetitions, we must eventually find that the largest and smallest index in B_m are the same, again contradicting that k is minimal. This means that $k \not> 0$, so in fact $k = 0$ as desired.

□

Remark 3.10. Lemma 3.9 does not apply for relations with gaps. For example, take $r = (1, 0, 1, 1, 1)$ and consider the paths

$$p^+ = (-1, -1, -1, -1, 1, -1, 1, 1, 1) \quad \text{and} \quad q^+ = (-1, 1, -1, -1, 1, 1, 1, 1, 1).$$

It is easy to verify that $p^+ \sim q^+$ as $(a_9 \circ a_5 \circ a_6 \circ a_9)(p^+) = q^+$. However we also see that $p \not\sim q$ as there is no index such that the relation can be applied to p.

Now we can prove our main results.

Definition 3.11. For any non-negative integer m, the *m-bonacci numbers* are a sequence $\{F_i^m\}_{i \in \mathbb{Z}}$ defined as follows. When $m = 0$, we set $F_1^0 = 1$ and $F_i^0 = 0$, $i \in \mathbb{Z}/\{1\}$. For $m > 0$ we define the sequence by the recursion

$$F_i^m = \begin{cases} 0 & i < 1 \\ 1 & i = 1 \\ \sum_{j=i-m}^{i-1} F_j^m & i > 1. \end{cases}$$

Lemma 3.12. Consider an r-relational binary tree T_r such that r has length ℓ and no gaps. The number of nodes at level n is given by the recursion

$$N_n = \begin{cases} 2^n & n < l \\ 2N_{n-1} - N_{n-\ell} & n \geq l \end{cases}$$

Proof. When $\ell = 1$ it is obvious that $N_n = 1$ for all $n \in \mathbb{N}$ which is what the recursion yields. We now consider the case when $\ell > 1$.

First consider n such that $n < \ell$. Any path with length less than ℓ cannot be equivalent to another path, as there is no index the relation can be applied at. Thus each node at a level less than ℓ can only be represented by a unique path, obviously implying that $N_n = 2^n$ for $n < \ell$.

Now we examine $n > \ell$. We will denote the set of nodes at the $n-1$st level as $\{\eta_1, \dots \eta_{N_{n-1}}\}$, and we pick a path representative p_i for each node η_i. We show that any node η of level n has a path representative in the set $\{p_1^+, p_1^-, \dots, p_{N_{n-1}}^+, p_{N_{n-1}}^-\}$. Take the some path representative $o^\pm = (o_1, \dots, o_{n-1}, \pm 1)$ of η. Without loss of generality, assume our path representative is of the form o^+. We must have that $o \sim p_i$ for some $i \in \{1, \dots, N_n\}$. Correspondingly, we see that $o^+ \sim p_i^+$, meaning that p_i^+ is a path representative for η. Thus we know every node of level n is represented in the set $\{p_1^+, p_1^-, \dots, p_{N_{n-1}}^+, p_{N_{n-1}}^-\}$. We do not know, however, if it is represented uniquely.

Since $p_i \not\sim p_j$ for $i \neq j$, we know that $p_i^+ \not\sim p_j^+$ and $p_i^- \not\sim p_j^-$ for $i \neq j$ from Lemma 3.9. It is fairly easy to show that $p_i^+ \not\sim p_i^-$ for all $i \in \{1, \dots, N_{n-1}\}$. Assume for the sake of contradiction that there exists some i such that $p_i^+ \sim p_i^-$. Given our relation $r = (r_1, \dots r_\ell)$ take a root $\lambda \in \mathbb{C}$ of the polynomial $p(x) = \sum_{k=0}^{\ell} r_{k+1} x^k$ such that $\lambda \neq 0$. Note we can always do this because $p(x)$ has a non-zero constant term. From $p_i^+ \sim p_i^-$, it is easy to show that $\sum_{k=0}^{n} (p_i^+)_{k+1} x^k = \sum_{k=0}^{n} (p_i^-)_{k+1} x^k$. This would exactly imply that $-\lambda^n = \lambda^n$, contradicting that $\lambda \neq 0$.

From the fact that $p_i^+ \not\sim p_i^-, p_i^+ \not\sim p_j^+$, and $p_i^- \not\sim p_j^-$ for any i, j such that $i \neq j$, we see that, of our constructed path representatives, the ones which do not represent a unique node come in pairs. We can define a set of such pairs

$$E = \{\{p_i^+, p_j^-\} : i, j \in \{1, \dots, N_{n-1}\}, i \neq j \text{ such that } p_i^+ \sim p_j^-\}$$

and seek to determine the size of this set.

We denote the set of nodes at the $(n-\ell)$-th level as $\{\zeta_1, \dots \zeta_{N_{n-\ell}}\}$, and pick a path representative q_i for each node ζ_i. From this we can define a set

$$E' = \{q_i r : i \in \{1, \dots, N_{n-\ell}\}\}$$

and show that there is a bijection between E and E'

To show our desired bijection we construct a map $\phi : E \to E'$ defined as follows. Take an element $\{p_i^+, p_j^-\} \in E$, so have that $p_i^+ \sim p_j^-$. By definition we must have the existence of $\{\kappa_1, \dots \kappa_s\}$ such that $(a_{\kappa_s} \circ \dots \circ a_{\kappa_1})(p_i^+) = p_j^-$. Because $(p_i^+)_n \neq (p_j^-)_n$, there must be some $\kappa_t = n$. Namely, both $p_i^+ \sim p_j^-$ are equivalent to at least one path where the relation can be applied at the index n. We pick a particular such path by the following algorithm. If w a fintie path with at least m components of value 1, define c_m to be the smallest index such that $\{w_1, \dots, w_m\}$ has m 1's. We attempt to pick our above mentioned equivalent path so that c_1 is minimized. If multiple such

paths meet this requirement we attempt to pick the path such that c_2 is minimized, so on and so forth. We call the path chosen by the algorithm h.

Say $h = (h_1, \ldots, h_{n-\ell}, h_{n-\ell+1}, \ldots, h_n)$. Since $(h_1, \ldots, h_{n-\ell}) \in P_{n-\ell}$, we know that it is equivalent to exactly one q_ι. We define our map $\phi : E \to E'$ such that $\phi(\{p_i^+, p_j^-\}) = q_\iota r$. Note, since h is such that the relation can be applied at index n we have that $(h_{n-\ell+1}, \ldots, h_n) = r$ or $-r$, which makes it apparent that $h \sim q_\iota r$, and thus $p_i^+ \sim q_\iota r$ and $p_j^- \sim q_\iota r$. Therefore, the map must be injective, as each pair of paths in E are explicitly not equivalent to any other pair.

We can also show the map is surjective. For some $q_\iota r \in E'$, we consider $\{q_\iota r, q_\iota(-r)\}$, which itself a pair of equivalent paths with level n. We see that $(q_\iota r)_n = -(q_\iota(-r))_n$ and $q_\iota r \sim q_\iota(-r)$. Thus by how E has been constructed, we must certainly have $q_\iota r \sim a$ and $q_\iota(-r) \sim b$ such that $\{a, b\} \in E$. Then we must have $\phi(\{a, b\}) = q_\iota r$, as if this was not this case, then we would have some q_τ such that $q_\iota \not\sim q_\tau$ but $q_\iota r \sim q_\tau r$, which violates Lemma 3.9. Thus have established our bijection and see that $|E| = |E'| = N_{n-\ell}$ is exactly the number of our constructed paths that represent the same node as another. From this we see $N_n = 2N_{n-1} - N_{n-\ell}$ as desired.

□

Theorem 3.13. For an r-relational binary tree T_r such that r has length ℓ and no gaps, we have $N_n = \sum_{i=0}^{n+1} F_i^{\ell-1}$.

Proof. This follows directly from the recursion shown in Lemma 3.12. The statement is obvious when $\ell = 1$ so we only consider $\ell > 1$. First we note that for $0 \leq n < \ell$, $N_n = 2^n$ and also

$$\begin{aligned}\sum_{i=0}^{n+1} F_i^{\ell-1} &= 1 + \sum_{i=0}^{n-1} 2^i \\ &= 1 + \frac{1-2^n}{1-2} \\ &= 2^n\end{aligned}$$

Using this we can prove our desired result through strong induction. Assume as our strong inductive hypothesis that for some $k \geq \ell - 1$, $N_q = \sum_{i=0}^{q+1} F_i^{\ell-1}$ for all $q \leq k$. From above we have $N_{k+1} = 2N_k - N_{k-\ell+1}$. Applying the hypothesis we see

$$\begin{aligned}N_{k+1} &= 2\sum_{i=0}^{k+1} F_i^{\ell-1} - \sum_{i=0}^{k-\ell+2} F_i^{\ell-1} \\ &= \sum_{i=0}^{k+1} F_i^{\ell-1} + \sum_{i=k-\ell+3}^{k+1} F_i^{\ell-1} \\ &= (\sum_{i=0}^{k+1} F_i^{\ell}) + F_{k+2}^{\ell-1} \\ &= \sum_{i=0}^{k+2} F_i^{\ell-1}\end{aligned}$$

which confirms our desired result.

□

Remark 3.14. In the case $\ell = 3$, the expression above can be simplified to $N_n = F_{n+3}^2 - 1$.

Proposition 3.15. Consider the relations $r = (r_1, \ldots, r_\ell)$ and $s = (s_1, ..., s_\ell)$, where $s_i = r_i$ if i is odd and $s_i = -r_i$ if i is even. Then $T_r \cong T_s$.

Proof. We consider a map $f : \mathcal{P} \to \mathcal{P}$ which maps $p \mapsto f(p)$, where $(f(p))_i = p_i$ if i is odd and $(f(p))_i = -p_i$ if i is even. Note that f is obviously invertible as it is its own inverse. It is clear from construction that $a_{r,k}(p)$ is well defined if and only if $a_{s,k}(f(p))$ is well defined. Additionally, in the case $a_{s,k}(f(p))$ is well defined, we have that $a_{s,k}(f(p)) = f(a_{r,k}(p))$.

We define our graph isomoprhism ϕ as follows. Suppose η is a node in T_r path represented by p, then ϕ is such that $\phi : \eta \mapsto \phi(\eta)$, where $\phi(\eta)$ is the node in T_s path represented by $f(p)$. By above we see that this map is well defined. If two nodes η_1 and η_2 are connected by an edge e, then they can be path represented by some p and $p^{\pm}$ respectively. We see that $f(p^{\pm})$ can be written in the form $(f(p))^{\pm}$, implying that $\phi(\eta_1)$ and $\phi(\eta_2)$ are connected some edge. We define ϕ such that $\phi : e \mapsto \phi(e)$, where if e is an edge connecting η_1 and η_2, $\phi(e)$ the edge connecting $\phi(\eta_1)$ and $\phi(\eta_2)$. □

Proposition 3.16. Consider the "reversed relations" $r = (r_1, \ldots, r_\ell)$ and $s = (r_\ell, \ldots, r_1)$. Then for the graphs T_r and T_s, N_n^f is the same for all $n, f \in \mathbb{N}$.

Proof. First we define a permutation σ on the set $\{1, ..., n\}$ such that $\sigma(i) = n+1-i$ for $i \in \{1, ..., n\}$. We define a function $f : \mathcal{P}_n \to \mathcal{P}_n$, where if $p = (p_1, \ldots, p_n)$, then $f(p) = (p_{\sigma(1)}, \ldots, p_{\sigma(n)})$. From this definition one can see that $a_{r,k}(p)$ being well defined is equivalent to $a_{s,\sigma(k-\ell+1)}(f(p))$ being well defined. We also see that $a_{s,\sigma(k-\ell+1)}(f(p)) = f(a_{r,k}(p))$. This guarantees our desired result. □

Remark 3.17. An interesting fact to note is that although all graphs T_r such that r has length ℓ and no gaps seem to share a great deal of structure, not all such graphs are isomorphic. As it turns out, although N_n is the same for all such graphs, one can easily find examples where the values of N_n^f do not match; this is a clear indication that such graphs are not isomorphic. As an example, if we consider T_r where $r = (1, 1, 1)$, we can easily check that $N_4^3 = 2$. However if $s = (1, -1, -1)$ one can check that $N_4^3 = 0$ for T_s. This is shown in Figures 2 and 4 at the beginning of the section.

3.2 Results on Bernoulli Graphs

One means of studying the Bernoulli convolution ν_λ associated with the random variable $Y_\lambda = \sum_{i=0}^{\infty} X_i \lambda^i$ is to examine the n-finite convolution $\nu_{\lambda,n}$ associated with $Y_{\lambda,n} = \sum_{i=0}^{n} X_i \lambda^i$. If λ satisfies a suitable polynomial, then the number of distinct values that $Y_{\lambda,n}$ can assume may be less than 2^{n+1}. In this case, it is natural to ask how many possible values n-finite variable $Y_{\lambda,n}$ can realize. It is obvious that the number of "outcomes" the variable $Y_{\lambda,n}$ can result in is 2^{n+1}, where an outcome corresponds to a particular string of pluses and minuses. These outcomes, however, do not necessarily represent unique values. We define the function $\mathrm{Val} : P \to \mathbb{R}$ by $\mathrm{Val}(p) = \sum_{i=0}^{|p|} p_{i+1} \lambda^i$ if p is a path with nonzero length, and $\mathrm{Val}(p) = 0$ for the trivial path $p = ()$. We call $\mathrm{Val}(p)$ the *value* of p (with respect to λ). We can now rigorously define the graph of the convolution ν_λ, which we will often denote G_λ.

Definition 3.18. Let ν_λ be a Bernoulli convolution. The *graph of* ν_λ is a graph with nodes represented by the ordered pair $(\mathrm{Val}(p), n)$ for some $p \in \mathcal{P}_n, n \in \mathbb{N}$. Two nodes are connected by an edge if and only if they can be written as (c, n) and $(c \pm \lambda^n, n + 1)$.

Definition 3.19. Suppose (c, n) represents a node of G_λ. We say that such a node is on the n-th level and has frequency f, where f is the number of $p \in \mathcal{P}_n$ with $\text{Val}(p) = c$.

For Bernoulli convolution graphs we adopt the convention that $\mathcal{N}_n$ is the number of nodes of level n, and $\mathcal{N}_n^f$ is the number of nodes with level n and frequency f as to avoid ambiguity.

Theorem 3.20. Let ν_λ be a Bernoulli convolution with graph G_λ such that λ has minimal polynomial $m_\lambda(x) = \sum_{i=0}^{\ell-1} r_{i+1}x^i$ over $\mathbb{Q}$, and T_r be the r-relational binary tree where $r = (r_1, \ldots, r_\ell)$. The number of nodes on the n-th level of G_λ is bounded above by the number of nodes on the n-th level of T_r, namely $\mathcal{N}_n \leq N_n$.

Proof. We construct an injective map $\phi : \Lambda_{G_\lambda}^n \Rightarrow \Lambda_{T_r}^n$, where $\Lambda_{G_\lambda}^n$ is the set nodes in G_λ with level n, and $\Lambda_{T_r}^n$ is the set of nodes in T_r with level n. Suppose $\eta \in \Lambda_{G_\lambda}^n$, then η is represented by the tuple (c, n) and there is some path $p \in \mathcal{P}_n$ such that $\text{Val}(p) = c$. We construct our map so that ϕ sends $\eta \mapsto \phi(\eta)$, where $\phi(\eta)$ is the node in T_r path represented by p. Since $p \sim q \implies \text{Val}(p) = \text{Val}(q)$, we must have ϕ is injective. Take two nodes $\eta_1, \eta_2 \in \Lambda_{G_\lambda}^n$ represented by $(c_1, n), (c_2, n)$, where there are $p_1, p_2 \in \mathcal{P}_n$ such that $\text{Val}(p_1) = c_1$ and $\text{Val}(p_2) = c_2$. If η_1 and η_2 mapped to the same node in T_r, then we would have $p_1 \sim p_2$, so then $c_1 = c_2$ and thus $\eta_1 = \eta_2$, guaranteeing injectivity. □

From Theorem 3.13 and Theorem 3.20 we have the following Corollary:

Corollary 3.21. Let ν_λ be a Bernoulli convolution with graph G_λ such that λ has minimal polynomial $m_\lambda(x) = \sum_{i=0}^{\ell-1} r_{i+1}x^i$, $r_j \neq 0$ for $j \in \{1, \ldots, \ell\}$ over $\mathbb{Q}$. Then $\mathcal{N}_n \leq \sum_{i=1}^{n+1} F_i^{\ell-1}$.

Lemma 3.22. Let ν_λ be a Bernoulli convolution with graph G_λ such that λ has minimal polynomial $m_\lambda(x) = \sum_{i=0}^{\ell-1} r_{i+1}x^i$ over $\mathbb{Q}$, and for every $p, q \in \mathcal{P}_n$, we have $p \sim q$ if and only if $\text{Val}(p) = \text{Val}(q)$. Then the r-relational binary tree T_r where $r = (r_1, \ldots, r_\ell)$ is isomorphic to G_λ.

Proof. First we note that we always have $p \sim q \implies \text{Val}(p) = \text{Val}(q)$, so if we have a node η in T_r we can define $\text{Val}(\eta)$ to be $\text{Val}(h)$ where h is a path representative of η. We define a graph isomorphism ϕ that sends $\eta \mapsto (\text{Val}(\eta), |\eta|)$. The condition that $\text{Val}(p) = \text{Val}(q) \implies p \sim q$ guarantees that the value of each node in T_r at a particular level is unique, implying injectivity. The map is also surjective as if (c, n) is some node in G_λ, then there is some path $\overline{h}$ such that $\text{Val}(\overline{h}) = c$ and $|\overline{h}| = n$, and $\overline{h}$ represents a node in T_r that maps to (c, n).

Suppose we consider η_1 and η_2 in T_r such that there is an edge e connecting them. We can thus path represent η_1 with p_1 of length k and η_2 with $p_1^{\pm}$. This guarantees that there is an edge connecting $\phi(\eta_1)$ and $\phi(\eta_2)$, as $\eta_1 \mapsto (\text{Val}(p_1), k)$ and $\eta_2 \mapsto (\text{Val}(p_1) \pm \lambda^k, k+1)$. We define ϕ such that $e \mapsto \phi(e)$, where $\phi(e)$ is the edge in G_λ that connects $\phi(\eta_1)$ and $\phi(\eta_2)$, completely establishing the isomorphism. □

Remark 3.23. There are certainly $\lambda \in (0, 1)$ for which Lemma 3.22 does not apply. Take $\lambda \in (\frac{1}{2}, 1)$ such that it has minimal polynomial $1 - \lambda - \lambda^2 - \lambda^3 + \lambda^4 = 0$. We see that thus

$$0 = (1 + \lambda^3)(1 - \lambda - \lambda^2 - \lambda^3 + \lambda^4) = 1 - \lambda - \lambda^2 - \lambda^5 - \lambda^6 + \lambda^7$$

If we consider paths $p = (1, -1, -1, 1, 1, -1, -1, 1)$ and $q = (-1, 1, 1, 1, 1, 1, 1, -1)$, then $\text{Val}(p) = \text{Val}(q)$, however it is obvious that $p \not\sim_r q$ where $r = (1, -1, -1, -1, 1)$, as there is no index where we can application of r to p is well defined.

Lemma 3.24. Suppose $\lambda \in (0,1)$ is the positive root of the polynomial $1 - x - \cdots - x^d$. Given a strictly increasing indexing set $\{k_1, k_2, \ldots, k_i\}$ where $k_1 \geq -1$ and having no d elements consecutive,

$$\sum_{j=1}^{i} \lambda^{k_j} < \lambda^{k_1 - 1}$$

.

Proof. First, note that given any indexing set, $\{k_1, k_2, \ldots, k_i\}$, proving $\sum_{j=1}^{i} \lambda^{k_j} < \lambda^{k_1-1}$ is equivalent to proving that

$$\sum_{j=1}^{i} \frac{\lambda^{k_j}}{\lambda^{k_1-1}} = \sum_{j=1}^{i} \lambda^{k_j - k_1 + 1} < 1,$$

which amounts to shifting our indexing set by a constant to fix $k_1 = 1$. Therefore, without loss of generality we can assume that $k_1 = 1$ and thus our goal becomes proving that

$$\sum_{j=1}^{i} \lambda^{k_j} < 1.$$

To proceed, we will use proof by strong induction on the size of k_i. Our base cases are when $k_i \leq d$. Because we can't have d consecutive terms, and our largest index is at most d, we have that

$$\sum_{j=1}^{i} \lambda^{k_j} \leq \lambda + \lambda^2 + \cdots + \lambda^{d-1} + \lambda^d < 1,$$

so our claim holds in these cases.

For our inductive step, assume that our claim holds for all $k_i < n$.

Given $k_i = n$, let k_ℓ be the lowest non-consecutive index (that is, k_ℓ is the first index such that $k_\ell - k_{\ell-1} > 1$). Because $\{k_1, k_2, \ldots, k_i\}$ does not contain any d consecutive terms, we have that $k_\ell \leq d + 1$. Then, from our inductive hypothesis, we have that

$$\begin{aligned}
\sum_{j=1}^{i} \lambda^{k_j} &< \lambda + \lambda^{k_2} + \cdots + \lambda^{k_{\ell-1}} + \lambda^{k_\ell - 1} \\
&< \lambda + \lambda^2 + \lambda^3 + \cdots + \lambda^d \\
&\leq 1.
\end{aligned}$$

Therefore our claim also holds when $k_i = n$, so by strong induction it is true for all sizes of k_i and thus all indexing sets of the kind allowed above.

□

Theorem 3.25. Suppose $\lambda_d \in (0,1)$ is the root of the polynomial $1 - x - \cdots - x^d$. Then $G_{\lambda_d} \cong T_r$ where $r = (1, \underbrace{-1, \ldots, -1}_{d \text{ times}})$.

Proof. We always have that, given $p, q \in \mathcal{P}_n$, $p \sim q \implies \text{Val}(p) = \text{Val}(q)$. To prove that $G_{\lambda_d} \cong T_r$ with $r = (1, \underbrace{-1, \ldots, -1}_{d \text{ times}})$, we must show that $\text{Val}(p) = \text{Val}(q) \implies p \sim q$, and then our result follows from Lemma 3.22.

This proof is closely related to our work in Section 4, and as such uses the convention that the X_i's are chosen from $\{0, 1\}$ as opposed to $\{-1, 1\}$. This change amounts merely to a shift and scale of the associated Bernoulli convolution, and therefore does not influence its graphical representation G_λ up to isomorphism. Similarly, this change simply constitutes a relabeling from -1 to 0 for every path representative of a node in T_r.

Let $p, q \in \mathcal{P}_n$ such that $\text{Val}(p) = \text{Val}(q)$. To show that $p \sim q$, we first develop an algorithm to put both p and q into "canonical form." Given any path p, we search from left to right until we find d consecutive 1's. Formally, let n be the smallest index such that $p_{n-d+1}, p_{n-d+1}, \ldots, p_n = 1$. Then if p_{n-d} exists it must $= 0$, otherwise $n - d + 1 = 1$. If so, create an additional 0-th slot in our path, which is filled by a 0.

Then, apply the relation at n, i.e. consider $a_n(p)$. Our new path is relationally equivalent to our original one, but has fewer 1's because we've just replaced $(\ldots, 0, \underbrace{1, \ldots, 1}_{d \text{ times}}, \ldots)$ with $(\ldots, 1, \underbrace{0, \ldots, 0}_{d \text{ times}}, \ldots)$.

Repeat this process until there are no more blocks of d consecutive 1's, noting that this algorithm will always terminate because each step decreases the number of 1's found in our path.

Let these "canonical forms" of p and q be denoted p' and q', and extend our definition of $\text{Val} : \mathcal{P} \to \mathbb{R}$ to $\text{Val}(p) = \sum_{i=-1}^{|p|} p_{i+1}\lambda^i$, where $p_0 = 0$ unless flipped to a 1 through the algorithm above.

If these "canonical forms" are identical, then clearly $p \sim q$. Therefore assume that they are not.

For every integer $i \in [-1, n]$ such that $p'_i = q'_i = 1$, replace both 1's by 0's, calling our new paths p'' and q'' respectively. We have, again, that $\text{Val}(p'') = \text{Val}(q'')$. This allows us to consider only the places in which p' and q' differ.

Without loss of generality, if p'' was only composed of 0's, then because $\text{Val}(p'') = \text{Val}(q'')$ and all terms in the sums are positive, then q'' would also only contain 0's. Therefore neither p'' nor q'' are empty.

Let n be the smallest index such that $p''_n = 1$, and let m be the smallest index such that $q''_m = 1$. By construction, $n \neq m$, so again without loss of generality let $n > m$.

By Lemma 3.24, we have that

$$\text{Val}(p'') < \lambda^{n-1} < \lambda^m < \text{Val}(q''),$$

with this last inequality holding because all summands in $\text{Val}(q'')$ are positive, and we know that $q''_m = 1$. However, by assumption, $\text{Val}(p'') = \text{Val}(q'')$.

Therefore p and q must have the same canonical form, i.e. $p \sim p' = q' \sim q$, and thus $p \sim q$. Thus $\text{Val}(p) = \text{Val}(q) \implies p \sim q$, so $G_{\lambda_d} \cong T_r$.

□

From Theorem 3.13 and Theorem 3.25 we have the following Corollary:

Corollary 3.26. Suppose $\lambda_d \in (0, 1)$ is the root of the polynomial $1 - x - \cdots - x^d$. Then the n-finite Bernoulli convolution Y_{n,λ_d} retains exactly $\sum_{i=1}^{n+1} F_i^d$ distinct values.

3.3 Additional Algebraic Results

As discussed above, we are interested in when two paths of the same length correspond to the same value (i.e. node) of the graph G_λ. Such overlap, if it occurs, is related to algebraic properties of λ.

Definition 3.27. A polynomial is said to be *supermonic* if all its coefficients are -1, 1, or 0.

We will write $\mathcal{S}$ for the set of supermonic polynomials, and define $\mathcal{S}_\lambda = \{p(x) \in \mathcal{S} : p(\lambda) = 0\}$. Note that $\mathcal{S}_\lambda$ is necessarily empty if λ is not an algebraic integer, for instance when λ is rational.

Proposition 3.28. Let G_λ be a graph of the Bernoulli convolution Y_λ, and $m_\lambda(x)$ be the minimal polynomial of λ in $\mathbb{Q}[x]$. Then $N_n = 2^n$ for all $n \in \mathbb{N}$ if and only if $\mathcal{S}_\lambda \cap (m_\lambda(x)) = \emptyset$.

Proof. First we prove the forward direction. Suppose that there is some $s(x) \in S_\lambda \cap (m_\lambda(x))$. It is easy to check that it must have degree $k-1, k > 2$, so one can write $s(x) \in (m_\lambda(x))$, where $s(x) = \sum_{j=0}^{k-1} a_{j+1}x^i$. Suppose $a_{i_i}, ..a_{i_m} \neq 0$. We define $p = (\alpha_1, \ldots, \alpha_k)$ where $\alpha_j = a_j$ if $j \in (i_i, ..i_m)$, and 1 otherwise. Now we define p' such that $p'_j = -p_j$ if $j \in (i_i, ..i_m)$, and 1 otherwise. Then we see that $\mathrm{Val}(p) = \mathrm{Val}(p')$, meaning certainly that $N_k \neq 2^k$. The contrapositive of this result proves the forward direction.

To prove the reverse direction, we note that $N_k \neq 2^k$ for some $k \in \mathbb{N}, k > 2$. if and only if there are some paths $t, s \in \mathcal{P}_k$ such that $\mathrm{Val}(t) = \mathrm{Val}(s)$, namely $\sum_{i=0}^{k-1} t_{i+1}\lambda^i = \sum_{i=0}^{k-1} s_{i+1}\lambda^i$. If we define $s(x) = \frac{1}{2}\sum_{i=0}^{k-1}(t_{i+1} - s_{i+1})\lambda^i$, we have that $s(x) \in S_\lambda$ as certainly $s(x) \in S$ and $s(\lambda) = 0$. Since $\mathcal{Q}[x]$ is a Euclidean domain, $s(\lambda) = 0$, and $m_\lambda(x)$ is the minimal polynomial of λ in Q_x, we have that $s(x) = d(x)m_\lambda(x)$ for some $d(x) \in \mathbb{Q}[x]$, namely $s(x) \in (m_\lambda(x))$ and thus $s(x) \in S_\lambda \cap (m_\lambda)$. The contrapositive of this result exactly proves the reverse direction. □

Remark 3.29. It is not true that $\mathcal{S}_\lambda$ non-empty $\implies m_\lambda(x) \in \mathcal{S}_\lambda$. Take for example λ such that $m_\lambda(x) = 1 - x + 2x^2 - x^3$. We see that $m_\lambda(x) \notin \mathcal{S}_\lambda$, however $1 + x^2 + x^3 - x^4 \in \mathcal{S}_\lambda$ as $1 + x^2 + x^3 - x^4 = m_\lambda(x) \cdot (x + 1)$.

Corollary 3.30. Suppose $\lambda \in (0, \frac{1}{2})$ has minimal polynomial $m_\lambda(x)$ over $\mathbb{Q}$. Then $\mathcal{S}_\lambda \cap (m_\lambda(x)) = \emptyset$.

Proof. Suppose we have two distinct paths $p, q \in \mathcal{P}_n$. Let k be the index the paths first differ. We see that

$$\left|\sum_{i=0}^{k-1} p_{i+1}x^i - \sum_{i=0}^{k-1} q_{i+1}x^i\right| = 2\lambda^{k-1}$$

however we can also see that

$$\left|\sum_{i=k}^{n-1} p_{i+1}x^i - \sum_{i=k}^{n-1} q_{i+1}x^i\right| \leq 2\sum_{i=k}^{\infty} \lambda^i = 2\frac{\lambda^k}{1-\lambda} < 2\lambda^{k-1}$$

which exactly shows that we cannot have $\mathrm{Val}(p) = \mathrm{Val}(q)$. Thus the number of values the n-finite convolution takes on must be 2^n, and our desired result follows from Proposition 3.28. □

We conclude this section with the following observation regarding supermonic polynomials.

Proposition 3.31. Let $p, q \in \mathbb{Z}_{\neq 0}$ be coprime integers, $p \neq \pm 1$. Then no supermonic polynomials divide $p + qx$. In other words, $\big((p + qx)\big) \cap \mathcal{S} = \emptyset$.

Proof. Suppose for the purpose of contradiction that there exists some polynomial $a(x)$ such that $(p+qx)a(x) = s(x) = \sum_{i=0}^{m} s_i x^i \in \mathcal{S}$. Then we can express $a(x)$ as a sum $a(x) = \sum_{i=0}^{\infty} a_i x^i$ such that for some $N \in \mathbb{N}$, $a_n = 0$ for all $n > N$. We aim to show that such a finite N cannot exist.

If $s_0 = 0$, then clearly $a_0 = 0$, and the problem is simply shifted up one index. Additionally, flipping the sign of $s(x)$ can be achieved by flipping the sign of $a(x)$. Without loss of generality, then, let $s_0 = 1$, so $a_0 = \frac{1}{p}$.

We now proceed inductively. Assume that for all $i \leq n$, $a_i = \frac{c_i}{p^{i+1}}$, where the fraction is expressed in lowest terms. By the Cauchy product formula, we then have $s_{n+1} = a_{n+1}p + a_n q$. Rearranging,

$$\begin{aligned} a_{n+1} &= \frac{a_n q - s_{n+1}}{p} \\ &= \frac{qc_n - p^n s_{n+1}}{p^{n+1}}. \end{aligned}$$

We have assumed $qc \neq 0$, and accordingly the fraction is irreducible. We define $c_{n+1} = qc_n - p^n s_{n+1} \not\equiv 0 \mod p$, which is of the desired form. Thus, there must be infinitely many nonzero a_i for the product $(p+qx)a(x)$ to be supermonic, and our claim follows. □

4 Fibonacci Representation-Based Counting

Given the importance of supermonic polynomials satisfied by λ in the graph-based counting approach of the previous section, we would like a more complete understanding of the set $\mathcal{S}_\lambda$. Unfortunately, $\mathcal{S}_\lambda$ is difficult to study even in specific cases, and basic properties such as the degrees of its elements do not obey any obvious patterns. For *certain* λ, instead of broaching $\mathcal{S}_\lambda$ directly we can study an intimately related recursive sequence to compute $\mathcal{N}_n^f$. All of the ideas presented in Section 3 have a natural translation to this new setting, and we thus gain a new set of tools that do not immediately arise from considering polynomials. Integers take the role of paths, while the number of representations of these integers in a certain base capture the path frequencies. We will begin with the concrete example of $\lambda = \phi^{-1}$, the inverse of the golden ratio and the positive root of the supermonic polynomial $1 - \lambda - \lambda^2 = 0$. In this case, the associated sequence is the Fibonacci sequence $F_n = F_{n-1} + F_{n-1}, F_0 = 0, F_1 = 1$. After developing tools in that setting, we will generalize some of these results to the m-Bonacci case, in which the supermonic polynomial of interest is $1 - \lambda - \ldots - \lambda^m = 0$ and the associated sequence is $A_n = \sum_{i=1}^{m} A_{n-i}, A_1 = 1, A_{i \leq 0} = 0$.

Before starting, we make a slight change in convention. We will still represent a Bernoulli convolution as a random variable $Y_i = \sum \lambda^i X_i$; however, instead of choosing the X_i from $\{\pm 1\}$, we will choose them from $\{0, 1\}$. It is clear that the associated measure ν_λ under this definition is a simple shift and scaling of the previously associated measure. Moreover, all of the results on supermonic polynomials as relations still apply. The only difference is cosmetic; instead of representing a path as a word in the alphabet $\{-, +\}$, we use binary strings. This simply brings the role of sequences and sequence representations to the foreground.

4.1 Fibonacci Case

For this section, fix $\lambda = \phi^{-1}$, which arises as a root of the supermonic polynomial $p(x) = 1 - x - x^2$. Given λ, $p(x)$ tells us that in the random walk perspective, one step to the right (left) is exactly undone by the next two shorter steps to the left (right). If we want to represent step lengths

with terms of an integer sequence, then choosing a sequence F_n which satisfies $F_n = F_{n+1} + F_{n+2}$ generates the same combinatorial structure. This is recognizable as the Fibonacci sequence in reverse. Accordingly, for a walk of n steps, we suppose that the first (and largest) step is of length F_{n+1}, the second of length F_n, and so on, until the penultimate step has length $F_3 = 2$, and the last has length $F_2 = 1$. As before, we can interchange the subpath 011 with 100 at any position, but can make no substitutions that do not result from applying this basic one repeatedly.

The value of any length-n path is simply the sums of its step lengths, which is an integer between 0 and $\sum_{i=2}^{n+1} F_i = F_{n+3} - 1$. The frequency of an outcome $k \in \{0, 1, \ldots, F_{n+3} - 1\}$ is then simply the number of ways that k can be written as a sum of distinct Fibonacci numbers; see [12].

Definition 4.1. A *sequence representation* of an integer k in the sequence A_n is a finite, increasing set of indices $\{i_1 < \cdots < i_n\}$ such that $\sum_{j=1}^{n} A_{i_j} = k$.

Definition 4.2. The *representation count* of an integer k is the number of distinct sequence representations of k.

There are many methods for computing representation counts of increasing sequences. In principle, this gives us a procedure for computing $\mathcal{N}_n^f$ in the Fibonacci case: Simply write out the integers from 0 to $F_{n+3} - 1$, compute their representation counts, and count the quantity with a representation count of f. As studied in [9], the number $\mathcal{N}_n^f$ eventually increases by a constant amount each level. By exploring the representation count strategy a bit more carefully, we can find out exactly what this constant is, and at which level we can be assured that $\mathcal{N}_n^f$ will increase linearly with n.

As an example for the remainder of the section, let the path $w = 110100100101$. It can be computed that w represents the integer 449, which has 10 Fibonacci representations, of which 6 have length 12 and 4 have length 13. The frequency of the associated length-12 path is thus 6. We will use this example to make sense of some of the concepts defined abstractly in the section.

We begin with a definition of the fundamental building blocks for the paths we seek to study. Understanding these blocks and how to embed them as substrings in longer paths will yield the asymptotic behavior for $\mathcal{N}_n^f$ in the Fibonacci case.

Definition 4.3. A (*frequency-f*) *block* is a path p with frequency f meeting the following criteria:

1. It is possible to change all digits of p via (perhaps multiple) applications of the relation.
2. p is not a concatenation of smaller blocks.

Ultimately, we will understand how to count the number of frequency-f blocks in terms of integers with representation count f, and then embed blocks as subpaths of longer paths. We need a few preliminary results on blocks before proceeding.

Theorem 4.4. (Zeckendorf) Every non-negative integer has a unique Fibonacci representation with no consecutive 1's.

For a proof of Zeckendorf's theorem, see [1]. Zeckendorf's theorem gives us a canonical representation of each non-negative integer k. In the Fibonacci case, it is known that any representation of k can be transformed to the Zeckendorf representation by iterative application of the relation $011 \mapsto 100$. As a caveat, it is possible that the Zeckendorf representation of a word is one digit longer than an arbitrary one. Even so, it is still possible to apply the relation to a representation of k without changing the length until the lexicographically greatest representative of the given length

is achieved. This representative will have no consecutive 1's with the expection of a possible leading string of 1's. For example, our path $w = 110100100101$ has length 12 and is the lexicographically greatest representation of 449 with length 12. We will call such a string 'lexicographically maximal'. The Zeckendorf representation of 449 is given by 1000100100101 but has length 13; we will call such a string a 'Zeckendorf' string. In general, we will only be interested in representations of the desired path length. However, our computational methods will be untroubled by this distinction, as we will ultimately only consider Zeckendorf-form substrings. We now move on to an important result about blocks.

Lemma 4.5. If a path p that is a Fibonacci representation of an integer k is a block, then so are all paths p of the same length also representing k.

Proof. Given two paths p and q which represent k, it is possible to transform one into the other by repeatedly applying the relation [12]. This transformation clearly does not affect whether property 1 of a block is satisfied, so we move to property 2, which we will check inductively. Suppose p is a concatenation of smaller blocks b_i. To be able to toggle each digit of each b_i, each b_i must begin with 10 or 01 and end in 00 or 11. No application of the relation within a block can change this fact, for that would violate property 1. Thus, the two digits on either side of a concatenation point of blocks can be any of four possibilities: 0001, 0010, 1101, or 1110. In all of these cases, it is impossible to apply the relation anywhere; accordingly, individual applications of the relation can only affect individual blocks. We can now proceed by induction. Assume the lemma holds for paths of length smaller than p. If p is a concatenation of blocks b_i with lengths ℓ_i, then by the above, so too must be q. Reversing the roles of p and q, it is clear that if p is not a concatenation of shorter blocks b_i, than q cannot be either. We are thus left to check that the lemma holds on the smallest blocks, which are clearly 100 and 011. Since these are the exactly the set of Fibonacci representations of the number 3, we can conclude the lemma. □

We also proved the following useful corollary:

Corollary 4.6. Single applications of the relation can only affect individual blocks.

From Zeckendorf's theorem and the above lemma, we can study paths that represent k simply by looking at the the lexicographically greatest representation of k with the desired length. We now want to know which such strings are blocks. Fortunately this question has a concrete answer; to explain it, we use an exponential shorthand for concatenating binary strings. For example, we can rewrite $w = 110100100101$ as $110(10^2)^2101$.

Proposition 4.7. The lexicographically maximal strings of a given length that correspond to blocks are of form $\left(\prod_{i=1}^{m} 10^{r_i}\right) 0$, where the r_i are odd positive integers.

Proof. Let p be a binary string of the specified form. That property 1 of blocks is satisfied can be checked by induction on m and the sizes of the r_i. Property 2 follows from another induction on the size of the block and the fact that 10^{r_m+1} is a block, which is also straightforward to check.

Now suppose p is a lexicographically maximal string (i.e. no consecutive 1's after the leading string) of another form. If property 1 is to be satisfied without changing the length of the string, then p must begin with exactly one 1 and end in an even number of 0's, so we move to the case suggested in the lemma except for choosing some r_i for $i < m$ to be even. In this case, p is a concatenation of at least 2 blocks by the above, and accordingly not a block itself. □

We now know that blocks look like a specific type of Zeckendorf string, and we are ready to discuss how to assemble them. In the case of w, we see two subblocks: the substring 10100 beginning with the second digit, and the substring 100 that immediately follows. It is apparent that w is a concatenation of two blocks with a few digits to the left and right. We will proceed with a few lemmas which show that this is indeed how all lexicographically maximal strings behave. First, we will show that the frequency of a concatenation of blocks is the product of the frequencies of the blocks. Second, we will show that all lexicographically maximal strings of a given length are simply concatenations of blocks with possible frills attached. Finally, we will demonstrate how to embed concatenations of blocks as substrings of longer paths with the same overall frequency.

Lemma 4.8. If $\{b_i\}_{i=1}^n$ is a finite collection of blocks with frequencies f_i, then the concatenation $\prod_{i=1}^m b_i$ is a path with frequency $\prod_{i=1}^m f_i$.

Proof. In the proof of lemma 4.5, we remarked how any individual application of a relation cannot affect the blocks on both sides of a concatenation point. Accordingly, each block b_i can exist in all of its f_i representations independently of the other b_i. Thus, the total number of representations of the concatenation $\prod_{i=1}^m b_i$ is simply the product of the numbers of representations of each block, i.e. $\prod_{i=1}^m f_i$. □

Definition 4.9. A Zeckendorf string with a positive even number of trailing zeros and without any leading zeros is a *unit*.

Lemma 4.10. All units are concatenations of blocks.

Proof. Let p be a unit. Recall that the general form of a such a string p is $\prod_{i=1}^m 10_i^r$, where the r_i are positive integers and r_m is even. Applying Proposition 4.7, we see that we can demarcate the blocks as ending wherever r_i is even. Thus, p is a concatenation of blocks. □

Lemma 4.11. If p is a unit and $\ell_1, \ell_2 > 0$, then there are exactly two strings h of length ℓ_1 which can be added to the beginning of p and exactly two strings t of length ℓ_2 which can be added to the end of p such that the concatenation hpt has the same frequency as p.

Proof. The first digit of p can, perhaps after applying the relation several times, be chosen as either 0 or 1. Suppose h ends in 1. If $\ell_1 > 1$, then the preceding digit in h must also be 1; otherwise, an instance of 100 could appear in hp after applying the relation to p, which would increase the frequency of p. Inducting, we must have either $h = 0^{\ell_1}$ or $h = 1^{\ell_1}$. It is clear that both of these cases do not increase the overall frequency of hp.

Likewise, the last digit of p can be toggled to either 0 or 1. Suppose that t begins in 0 and $\ell_2 > 1$. Then the second digit of p must be 1, or else an instance of 100 would appear for any representation of p ending in 1. Inducting again, we observe that t must be an alternating string of 1's and 0's. Like in the first case, it is possible for this string to start with 1 or 0, and we have two possibilities for t such that the frequency of hpt is the same as that of p. □

Returning to our example w, we see that we can split w into a concatenation of three substrings: 1, 10100100, and 101. The first and third are respectively of the forms of h and t in Lemma 4.11. The middle string is a unit, which is a concatenation of the two blocks 10100 and 100. The first block has frequency 3; the second has frequency 2. Accordingly, the path w has frequency $2 \cdot 3 = 6$, corresponding to the six length-12 representations of 449.

We now have all of the necessary structural tools in place to compute the asymptotic behavior of $\mathcal{N}_n^f$ in the Fibonacci case. We know that given a node k at level n of the graph, we can choose the path corresponding to the lexicographically greatest representation q of k as a sum of n Fibonacci numbers. Invoking Lemma 4.11, we can reduce q to the form hpt, where p has the same frequency as q, no extraneous leading zeros or ones, and an even number of zeros at the end. We know that p must be a unit by the choice of q as the lexicographically greatest representation of p.

In principle, this allows us to proceed in either of two ways. We can compute the frequency of p by looking at the frequencies of each of its constituent blocks, in which case we reduce the problem of solving for $\mathcal{N}_n^f$ to a problem of classifying blocks. In the general m-Bonacci case, this is what we must do; this will be elaborated in the next section. However, in the Fibonacci case specifically, we can alternatively determine frequencies of full units by invoking a slight change in what constitutes a Fibonacci representation of an integer k. Then we must classify the number of units with frequency f, which we achieve in the following theorem:

Theorem 4.12. There is a two-to-one correspondence between integers with Fibonacci representation count f (allowing two 1's, one 2, etc.) and units of frequency f.

Remark 4.13. This employs a different representation counting convention than mentioned above. To compute the frequency of a node k, we may only permit representations of k with one 1. Allowing two 1's would permit a transposition of the last two binary digits, which introduces additional frequency. The counting convention used in this theorem is not for computing the frequency of a path, but rather the frequency of a unit which may occupy any spot within a longer path of the same frequency.

Proof. Let p be a unit of the required form. Let the least significant bit of p represent 0, the next least significant bits then representing $1, 1, 2, 3$, etc. b then represents an integer k, and we stipulate that all other representations of k have matching last two bits. Thus, if the two 1s are swapped, the 0 must also be toggled. In this way, we still only allow manipulations of p which are interchange 100 and 011, and the frequency of p is thus the same as the representation count of k in the new scheme. The new scheme precisely allows representations to have two 1s, one 2, etc. Note that it does not allow for the 0 to be included independently, as the 0 bit must always be set to the same position as the last 1 bit.

The Zeckendorf representation of every positive integer k (including the second one bit and the zero bit both switched to 0) has either an even number of trailing zeros or an odd number of trailing zeros. There is a one-to-one correspondence between these types of integers given by bitshifting the associated Zeckendorf representations: If a Zeckendorf representation has an even number of zeros, identify it with the Zeckendorf representation with one additional zero. Clearly the two associated Zeckendorf representations have the same frequency as strings.

Identifying each pair of integers with the block corresponding to the Zeckendorf representation of the one with an even number of trailing zeros, we have arrived at a correspondence between pairs of integers and units. Injectivity follows from the uniqueness part of Zeckendorf's theorem, and surjectivity from the existence part. Thus, the correspondence is bijective, and there is precisely one unit of frequency f for each pair of integers with Fibonacci representation count f, permitting two 1's. □

The previous theorem tells us that there is a two-to-one correspondence between integers with representation count f (under the two-1's-allowed counting scheme) and units of frequency f.

Serendipitously, the number of integers with representation count f under these rules is finite and easily computed [13]. We are now ready to present the main result of this section.

Theorem 4.14. Let m_f be the number of integers with Fibonacci representation count f (allowing two 1's), and let n be large enough such that the length of the longest unit of frequency f is strictly smaller than n (specifically $n > 2f$ [13]). Then $\mathcal{N}_n^f = 2m_f n - c_f$, where c_f is a constant depending on f.

Proof. For now, assume $f > 1$. Let $\{p_i\}_{i=1}^{m_f/2}$ be the set of non-equivalent units with frequency f, and let ℓ_i be the length of p_i. That there are $m_f/2$ of these follows from the previous theorem. We will embed the p_i as subpaths of a longer path q. We can position the first index of p_i anywhere between zero and $n - \ell_i$, in correspondence with $n - \ell_i - 2$ positions in the interior of the q and 2 positions on the exterior of q.

For each interior case, invoking lemma 4.11, there are four choices of q containing each p_i, which can start at position $j \in \{1, 2, \ldots, n - \ell_i - 1\}$. This accounts for $4(n - \ell_i - 1)$ choices of q for each p_i. For the exterior cases, again via Lemma 4.11, there are only two choices of q for each p_i. (In the language of the lemma, either the length of h or t has been shrunk to zero.) In total, there are $4(n - \ell_i)$ ways to construct q containing p_i and maintaining frequency f.

Combining Corollary 4.6 and Lemma 4.10, we know that positioning each p_i differently results in a nonequivalent path q. Thus, we can conclude

$$\mathcal{N}_n^f \geq \sum_{i=1}^{m_f/2} 4(n - \ell_i) = 2m_f n - 4 \sum_{i=1}^{m_f/2} \ell_i \sim 2m_f n - c_f.$$

For the reverse inequality, note that any path must contain all of its frequency in the form of the unit by Lemma 4.11. We can always strip a path q of frequency f down to a frequency f unit, so we know that $\mathcal{N}_n^f$ is less than or equal to the number of different ways to embed frequency-f units in a path of length n.

The frequency-1 case is somewhat analogous to the higher frequency case, except the only unit is the empty unit, representing zero. We still get to choose between leading 1's or 0's as well as the first position in which the opposite digit appears, yielding $\mathcal{N}_n^f = 2n$. This corresponds to $m_1 = 1$, as there is only one integer with representation count 1 instead of a pair.

In conclusion, we have shown that $\mathcal{N}_n^f = 2m_f n - c_f$ for all $n > 2f$, where c_f depends only on the lengths of the p_i and not n. □

As a final example for the Fibonacci case, we can apply this counting method to our string $w = 110100100101$. The unit within w is given by 10100100, which is the Zeckendorf string corresponding to the pair of integers $\{19, 31\}$, each of which has frequency 6 under the two-1's-allowed rules. We have placed the unit starting at the second digit of w, after the leading 1; we could have chosen the first, third, fourth, or fifth as well, which would yield a total of $4(n - 8)$ ways to encase w in a length-12 string of frequency 6. We also could have chosen any of five other pairs of integers with six representations and encased them in longer strings. Multiplying all of the possibilities together, it is clear that $\mathcal{N}_n^6$ should scale as $24n - c_6$, where c_6 accounts for the differences in the lengths of the frequency-6 units.

4.2 Tribonacci and m-Bonacci Generalizations

Instead of choosing λ to be the positive root of $1 - x - x^2$, we can choose λ to be the root of $1 - x - \cdots - x^m$, and many of the above results generalize moderately well. In particular, the Fibonacci recursion $F_n = F_{n-1} + F_{n-2}$ must be replaced with the m-Bonacci recursion $F_n = \sum_{i=1}^{m} F_{n-i}$. We no longer have Zeckendorf's theorem, but we still can choose lexicographically greatest representations of paths without more than $m-1$ consecutive 1's (apart from a leading string) [13]. Making this choice, the proofs of Corollary 4.6 and Lemma 4.8 carry through nicely, albeit with a bit more casework as blocks begin with form 10^{m-1} or 01^{m-1} and end with form 0^m or 1^m. In Proposition 4.7, we must change the restriction that the r_i are odd to that the $r_i \equiv -1 \mod m$.

From this point, the strategy remains much the same, though we must be careful with our argument. There are two main differences. The biggest difference is that it is possible for units to consist of blocks separated by 'filler strings' which do not affect the overall frequency of the unit. Consequently, the generalization of 4.11 requires more careful counting. The second main difference is that there is no choice of representation scheme in which the number of integers with representation count f becomes finite, so classifying blocks is a harder problem. Still, if we know what the blocks of frequency f' are for all $f' \mid f$, and if we know all of the ways we can string the blocks together into paths of frequency f, we can still produce meaningful counting results. The answers to the first question have not been thoroughly pursued at this time, so we will restrict to easy cases. The answers to the second question, however, fit together quite nicely in the language of generating functions. We outline the general approach to computing the number of frequency-f nodes at level n of the m-Bonacci graph below, letting p and q denote a blocks of frequency f and f':

1. Construct a generating function $A^m = \sum_{n=0}^{\infty} A_n^m x^n$ where A_n is the number of length-n prefixes h such that hp has frequency f. By symmetry, A_n^m will also count the length-n suffixes t such that pt has frequency f.

2. Construct a generating function $B^m = \sum_{n=0}^{\infty} B_n^m x^n$ where B_n^m is the number of length-n 'filler strings' m such that pmq has frequency (ff').

3. Using the Cauchy product formula and a list of possible block structures of a path of particular frequency k, construct a generating function $G^{m,k}(x) = \sum_{n=0}^{\infty} G_n^{m,k} x^n$ such that $G_n^{m,k}$ is the desired quantity of frequency-k nodes at level n.

As mentioned above, the list of possible block structures in step 3 is not immediately attainable. However, we will demonstrate the above computation procedure in the Tribonacci case for frequencies $k = 2, 3, 4$. The case $k = 1$ has been computed in the general m-Bonacci case by Kocábová et al [9], who observe that $\mathcal{N}_n^1$ follows an $(m-1)$-Bonacci-like recursion. In the case $m = 3$, they show that $\mathcal{N}_n^1 = \mathcal{N}_{n-1}^1 + \mathcal{N}_{n-2}^1 + 2$. We will show that such is not always true in the Tribonacci case for arbitrary frequencies. For the rest of this section, we will assume $m = 3$ and omit superscripts.

Computing $A(x)$

Given a block p, we seek to count the number of strings h of length-n such that hp has the same frequency as p. Clearly a first requirement is that h itself have frequency 1, so $A_n \leq \mathcal{N}_n^1$. We know that any path equivalent to p begins either 100 or 011, so to avoid introducing additional frequency,

we demand that h cannot end in 011 or 100. The numbers of length $1, 2$, and 3 paths for which satisfy this requirement are $2, 4$, and 6, respectively. For longer paths of length n, we can simply take the number of frequency-1 paths and subtract off the ones that end in these suffixes.

Proposition 4.15. For $n \geq 4$, the number of length-n, frequency-1 paths which end in 011 or 100 is $N_{n-3} + A_{n-3}$

Proof. Let q be such a path. Clearly the first $n-3$ digits of q, which we will call r, is a string of frequency 1. We can thus form all such strings q by appending the suffixes 011 and 100 to frequency-1, length-$(n-3)$ strings, as long as the appended strings still have frequency 1.

If r does not end in 011 or 100, then either suffix can be appended without introducing any additional frequency. This generates $2A_{n-3}$ paths q.

If r does end in either suffix, then only the same suffix can be appended without introducing additional frequency. This generates $N_{n-3} - A_{n-3}$ paths q, and exhausts all possibilities for r.

In sum, there are $2A_{n-3} + N_{n-3} - A_{n-3} = N_{n-3} + A_{n-3}$ such paths q. □

We thus have an order-3 recursion given by $A_n = N_n - N_{n-3} - A_{n-3}$ and three initial values for A. Additionally imposing that $A_0 = 1$, we can construct the generating function

$$A(x) = \frac{1 + x + x^2}{1 - x - x^2}.$$

The counting argument for suffixes t is the same, only replacing strings 011 and 100 with 000 and 111 and swapping the words 'suffix' and 'prefix'. The same generating function therefore also works.

Computing $B(x)$

Given blocks p and q with frequencies f and f', we want to count the length-n strings m such that pmq has frequency (ff'). We know that m must be frequency-1, and because q can be represented as beginning with a 0 or a 1, we also know that m may not end in the suffixes 011 or 100. Accordingly, $B_n \leq A_n$. Additionally, we cannot allow m to begin in 000 or 111 since p can be represented as ending in either 0 or 1. We proceed similarly to before, computing how much we must subtract from A_n to compute B_n. This time, we aim to subtract off the number of frequency-1 strings that start with 000 or 111 from the set of frequency-1 strings that do not end in 100 or 011.

Proposition 4.16. For $n \geq 4$, the number of frequency-1 strings that do not end in 011 or 100 but begin with 000 or 111 is $A_{n-3} + B_{n-3}$

Proof. Let q be such a string. Then the last $n-3$ digits of q comprise a frequency-1 string that does not end in 011 or 100, which we will call r. We can therefore construct all such strings q by prefixing strings of the form of r with 000 or 111 as long as we do not affect the frequency of r or the fact that r does not end in 100 or 011. The latter condition is vacuous, as adding either 000 or 111 as a prefix can never force the suffix of q to look like 011 or 100. We thus must only worry about frequency.

If r does not begin with 000 or 111, then either 000 or 111 can be attached as a prefix, and the resulting string will still have frequency 1. This generates $2B_{n-3}$ possibilities for q.

If r does begin with 000 or 111, only the same prefix can be attached without increasing frequency. This generates $A_{n-3} - B_{n-3}$ possibilities for q and exhausts all possibilities for r.

Adding together, we have a total of $2B_{n-3} + A_{n-3} - B_{n-3} = A_{n-3} + B_{n-3}$ possibilities for q. □

Again, we have an order-3 recursion for B_n given by $B_n = A_n - A_{n-3} - B_{n-3}$. The constraints on the prefix and suffix are sufficient whenever $n \geq 3$; it can be seen that the first few values of B_n are $(B_1, B_2, B_3) = (2, 2, 4)$, corresponding to choices $m \in \{0, 1, 01, 10, 001, 010, 101, 110\}$. Enforcing $B_0 = 1$, this again yields a generating function

$$B(x) = \frac{-1 + x^2 - 2x^3 - x^4}{-1 + 2x - x^2 + x^4}.$$

Computing $G^k(x)$

The Cauchy product formula tells us that given generating functions $f = \sum_i f_i x^i$ and $g = \sum_j g_j x^i$, the x^k coefficient of the product fg will take the form $\sum_{i+j=n} f_i g_j$. In our case, n represents the total number of digits we add to a block or combination of blocks. For instance, the x^n coefficient of $A(x)^2$ represents how many ways one can choose strings h and t of combined length n such that for any block p, the string hpt has the same frequency as p. In general, different products of $A(x)$ and $B(x)$ can be chosen to reflect the number of possible block constituents of a path q.

We begin with the frequency $k = 2$, there is only one equivalence class of blocks under the relation, namely 1000. As a generating function for $\mathcal{N}_n^2$, then, all we must choose is a way to allocate $n-4$ digits to a prefix and suffix. In consideration of the above paragraph, $G^2(x) = x^4 A(x)^2$ suffices such that $\mathcal{N}_n^2 = G_n^2$.

In the case $k = 3$, there are two equivalence classes of blocks, namely 1000000 and 1001000. They are both length-7, so we want to allocate $n - 7$ digits to a prefix and suffix. To allow for the two choices of block, we see that $G^3(x) = 2x^7 A(x)^2$ enforces $\mathcal{N}_n^3 = G_n^3$.

The case $k = 4$ requires a bit more care. Again, there are only two frequency-4 blocks: 1000000000 and 1001001000. They are both length-10 and result in a generating function $2x^{10}A(x)^2$. However, since 4 is a composite number, there is also the possibility of a frequency-4 path containing two frequency-2 blocks. In this case, we need to allocate $n - 8$ digits to choosing one prefix h, one suffix t, and one filler string m. Again because of the Cauchy product rule and the fact that the blocks are identical, the appropriate generating function is $x^8 A(x)^2 B(x)$. Adding together the two terms, we see that $G^4(x) = x^8 A(x)^2 (B(x) + 2x^2)$ ensures that $G_n^4 = \mathcal{N}_n^4$.

In this way, given a classification of all ways to decompose a frequency-k path into blocks, we can straightforwardly compute $\mathcal{N}_n^f$ by constructing an appropriate generating function. More work is necessary to understand these block decompositions and ultimately construct a general formula for $\mathcal{N}_n^f$ in the general Tribonacci or even m-Bonacci case. However, this framework proves powerful in computations of $\mathcal{N}_n^f$ for small f.

As a final remark in this section, we observe that the frequency-2 nodes in the Tribonacci case do not follow the same pattern as frequency-1 nodes. Indeed, the frequency-1 nodes obey a recursion $\mathcal{N}_n^1 = N_{n-1}^1 + N_{n-2}^1 + 2$ for sufficiently large n [9]. One may hope that this formula would generalize easily, replacing the '+2' with a constant depending on the number of frequency-f blocks. However, there is no sensible 'adjustment of the rules' like allowing additional 1's to encourage a finite number of integers with representation count f in the Tribonacci setting as there is in the Fibonacci setting. It is thus less clear to quantify blocks of frequency f, and we rely on our generating function methods.

Proposition 4.17. For $n \geq 7$, $G_n^2 = G_{n-1}^2 + G_{n-2}^2 + 4F_{n-3}$, where F_n is the Fibonacci sequence.

Proof. We begin with $x^{-4}G^2(x) = A(x)^2$; we keep out the x^4 term since it just shifts indices, which we will handle later. We have the following partial fractions decompositions:

$$\begin{aligned}
x^{-4}G^2(x) &= \frac{4}{(x^2+x-1)^2} + \frac{4}{x^2+x-1} + 1 \\
&= \left(\frac{4}{\phi+\phi^{-1}} - \frac{8}{(\phi+\phi^{-1})^3}\right)\left(\frac{-1}{x+\phi} + \frac{1}{x-\phi}\right) \\
&\quad + \frac{4}{(\phi+\phi^{-1})^2}\left(\frac{1}{(x+\phi)^2} + \frac{1}{(x-\phi^{-1})^2}\right) + 1 \\
&= \left(\frac{8}{(\phi+\phi^{-1})^3} - \frac{4}{\phi+\phi^{-1}}\right)\left(\frac{\phi}{1-\phi x} + \frac{(-\phi)^{-1}}{1-(-\phi^{-1}x)}\right) \\
&\quad + \frac{4}{(\phi+\phi^{-1})^2}\left(\frac{\phi^2}{(1-\phi x)^2} + \frac{\phi^{-2}}{(1-(-\phi^{-1}x))^2}\right) + 1.
\end{aligned}$$

In this form, we can easily write the x^n coefficient of $x^{-4}G^2(x)$. Let

$$\alpha = \left(\frac{8}{(\phi+\phi^{-1})^3} - \frac{4}{\phi+\phi^{-1}}\right), \beta = \frac{4}{(\phi+\phi^{-1})^2}.$$

Then

$$G^2_{n+4} = \alpha(\phi^{n+1} - (-\phi)^{-(n+1)}) + \beta(n+1)(\phi^{n+2} + (-\phi)^{-(n+2)}); \quad n \geq 1.$$

Assuming $n \geq 7$, we can subtract G^2_{n-1} and G^2_{n-2} from G^2_n using this formula to see if the difference is constant like in the $f = 1$ case. We have

$$\begin{aligned}
G^2_n - G^2_{n-1} &= \alpha(\phi^{n-5} - (-\phi)^{-(n-5)}) + \beta(n-4)(\phi^{n-4} + (-\phi)^{-(n-4)}) + \beta(\phi^{n-2} + (-\phi)^{-(n-2)}) \\
&= G^2_{n-2} + \beta(\phi^{n-2} + \phi^{n-4} + (-\phi)^{-(n-2)} + (-\phi)^{-(n-4)}).
\end{aligned}$$

Immediately, it is apparent that the remainder depends on n and is thus not constant. A little more manipulation reveals that the remainder is actually a Fibonacci number. Let the remainder $R = \beta(\phi^{n-2} + \phi^{n-4} + (-\phi)^{-(n-2)} + (-\phi)^{-(n-4)})$. Factoring,

$$\begin{aligned}
R &= \frac{4}{\phi+\phi^{-1}}(\phi^{n-3} - (-\phi^{-1})^{-(n-3)}) \\
&= 4\sum_{i=0}^{n-4} \phi^{n-i}(-\phi^{-1})^i.
\end{aligned}$$

Recognizing the Cauchy product formula at work, R is simply the n-th coefficient of the generating function

$$\frac{1}{1-\phi x}\frac{1}{1+\phi^{-1}x} = \frac{1}{1-x-x^2}.$$

This is a generating function for the Fibonacci sequence, and under our convention we have that $R = 4F_{n-3}$. Thus, not only is the remainder nonconstant in n; it grows as ϕ^n. As a fraction of the whole, however, still tends to zero as n^{-1}, for G^2_n itself grows as $n\phi^n$. □

5 Generalizations to Higher Dimension

While the Bernoulli convolutions considered thus far live strictly on the real line, generalizing Bernoulli convolutions to higher dimensions is a natural area of further investigation. Much of the literature on this topic works in the complex plane, selecting λ to be a complex number with magnitude less than 1 but otherwise defining the convolution identically. For more on this kind of generalization, see Shmerkin and Solomyak's paper [2]. To examine another possible generalization, we follow a different approach.

Definition 5.1. An *n-dimensional Bernoulli convolution* is the measure ν_λ of the infinite sum

$$\sum_{i=0}^{\infty} X_i \lambda^i,$$

where the X_i's are i.i.d. random variables chosen uniformly from the set $\{x \in \mathbb{R}^n \ : \ |x| = 1\}$, and $\lambda \in (0, 1)$.

In other words, our generalized convolution is the distribution of a geometrically scaled random walk in n dimensions, which reduces to our traditional Bernoulli convolution when $n = 1$.

We will first consider the two-dimensional case, where again we are interested in determining singularity versus continuity.

Theorem 5.2. Every 2-dimensional Bernoulli convolution is absolutely continuous.

Proof. Let $\lambda \in (0, 1)$, and let ν_λ be its associated 2-dimensional Bernoulli convolution. Additionally, define μ_λ to be the distribution of $X_0 + \lambda X_1$, i.e. the first two "steps" of our walk. Because

$$\sum_{i=0}^{\infty} X_i \lambda^i = X_0 + \lambda X_1 + \sum_{i=2}^{\infty} X_i \lambda^i,$$

we have that

$$\nu_\lambda = \mu_\lambda * \text{the distribution measure of all other terms in the sum.}$$

An absolutely continuous measure convolved with any other measure will remain absolutely continuous, so to show continuity, it suffices to prove that μ_λ is absolutely continuous.

Note that because the direction for each "step" is chosen uniformly, μ_λ is a radially symmetric distribution (as is ν_λ), and it is supported on the annulus $\{x \in \mathbb{R}^2 \mid 1 - \lambda \leq |x| \leq 1 + \lambda\}$. Thus we can determine the CDF of $X_0 + \lambda X_1$ by conditioning on a single radius. Specifically, assume that $X_0 + \lambda X_1$ lands somewhere on the positive x-axis (between $x = 1 - \lambda$ and $x = 1 + \lambda$).

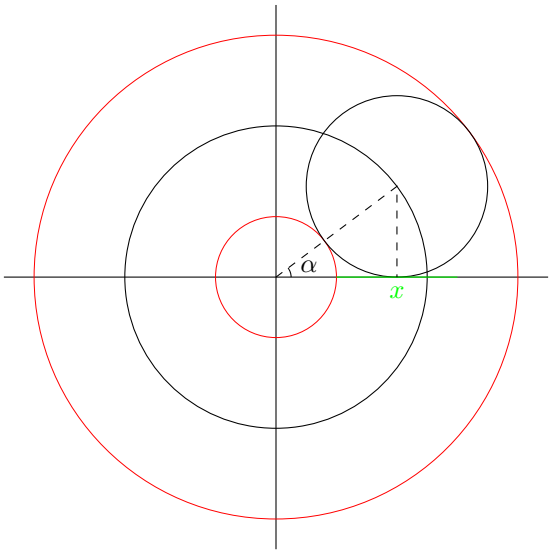

Figure 5: 2-Dimensional Case

Now, consider the interval on the x-axis $[1-\lambda, 1-\lambda+x]$, with x varying from 0 to 2λ, as highlighted in green in Figure 5. To derive the CDF of our random variable, we'll determine the probability that $X_0 + \lambda X_1$ falls within this line segment as a function of x, its length.

First observe that for our second step – which has length λ – to land back on the positive x-axis, our first step – which has length 1 – cannot get too far away from the positive x-axis. Concretely, define θ as the angle of our first step as measured from the positive x-axis. Assuming that our two-step walk lands in $(1-\lambda, 1+\lambda)$ implies that θ has magnitude less than α, where $\alpha = \arcsin\lambda$ is the angle of the first step at which the circle representing all possible second steps is tangent to the positive x-axis. If we were to choose an angle greater than α for our first step, our second step wouldn't be long enough to return back to the positive x-axis.

The point of tangency, $(\sqrt{1-\lambda^2}, 0)$, is also important. For any given θ such that $|\theta| < \alpha$, the circle of all possible outcomes after the second step will intersect the positive x-axis in two places: $(\cos\theta + \sqrt{\lambda^2 - \sin^2\theta}, 0)$ and $(\cos\theta - \sqrt{\lambda^2 - \sin^2\theta}, 0)$. This is because our two intersection points lie at a distance of $\sqrt{\lambda^2 - \sin^2\theta}$ away from the projection of the smaller circle's center onto the x-axis. See the Figure 6 for more clarity.

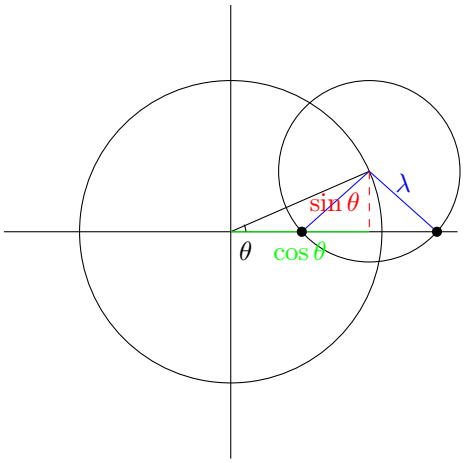

Figure 6: 2-Dimensional Case

Because both $\cos\theta + \sqrt{\lambda^2 - \sin^2\theta}$ and $\cos\theta - \sqrt{\lambda^2 - \sin^2\theta}$ are monotonic as functions of θ, and the two intersection points coincide at $(\sqrt{1-\lambda^2}, 0)$ when $\theta = \alpha$, we know that if two intersection points exist, one will lie to the left of $(\sqrt{1-\lambda^2}, 0)$ and one will lie to the right.

Thus if $x < \sqrt{1-\lambda^2} - (1-\lambda)$ – i.e. the green interval in Figure 5 does not cross $(\sqrt{1-\lambda^2}, 0)$ – for every angle θ which generates an opportunity for the second step to land within the green interval, there is only one way to do so because there is only one intersection point. However, if $x > \sqrt{1-\lambda^2} - (1-\lambda)$ – i.e. the green interval does cross the point of tangency – small angles will only give one way to land within the interval, but larger angles will give two. Therefore, we must split our function of x into two separate cases.

Case 1: $x < \sqrt{1-\lambda^2} - (1-\lambda)$

For θ of small magnitude, we know that there will be exactly one point of intersection until

$$\begin{aligned}
&1 - \lambda + x = \cos\theta_A - \sqrt{\lambda^2 - \sin^2\theta_A} \\
\Rightarrow\ &(1 - \lambda + x - \cos\theta)^2 = \lambda^2 - \sin^2\theta_A \\
\Rightarrow\ &2 + x^2 - 2\lambda + 2x - 2x\lambda - 2(1 - \lambda + x)\cos\theta_A = 0 \\
\Rightarrow\ &\theta_A = \arccos\left(1 + \frac{\frac{1}{2}x^2 - x\lambda}{1 - \lambda + x}\right),
\end{aligned}$$

at then θ of greater magnitude will produce no intersections. Thus our cumulative distribution function, conditioning on landing on the positive x-axis, will count once all the angles from $-\theta_A$ to θ_A, but will divide by *twice* the angles from $-\alpha$ to α, because each of those angles generates two intersections. Our function is therefore

$$f(x) = \frac{2\theta_A}{2 \cdot 2\alpha} = \frac{\arccos\left(1 + \frac{\frac{1}{2}x^2 - x\lambda}{1 - \lambda + x}\right)}{2\arcsin\lambda}.$$

Case 2: $x > \sqrt{1-\lambda^2} - (1-\lambda)$

For θ of small magnitude, we again see that there will be one point of intersection close to $(1-\lambda, 0)$, as the other x-axis intersection starts at $(1+\lambda, 0)$ and moves inward, so it will initially be too far to the right to intersect with our desired interval. However, once we find θ_B satisfying

$$\begin{aligned} 1-\lambda+x &= \cos\theta_B + \sqrt{\lambda^2 - \sin^2\theta_B} \\ \Rightarrow (1-\lambda+x-\cos\theta)^2 &= \lambda^2 - \sin^2\theta_B \\ \Rightarrow \theta_B &= \arccos\left(1 + \frac{\frac{1}{2}x^2 - x\lambda}{1-\lambda+x}\right), \end{aligned}$$

all θ with greater magnitude will intersect the green interval in two places, so we have to count those angles twice. We will continue to find two intersections until $|\theta| = \alpha$, at which point we will have exactly one angle of tangency, and then no intersections for $|\theta| > \alpha$. Thus our cumulative distribution function, again with the same conditions, will have the same denominator as above, but a numerator that counts twice the angles with magnitude between θ_B and α but counts once the angles from 0 to θ_B. Explicitly, we have

$$f(x) = \frac{2(2(\alpha-\theta_B)+\theta_B)}{2\cdot 2\alpha} = \frac{2\alpha-\theta_B}{2\alpha} = 1 - \frac{\arccos\left(1+\frac{\frac{1}{2}x^2-x\lambda}{1-\lambda+x}\right)}{2\arcsin\lambda}.$$

Note that θ_A and θ_B are equivalent angles, but labeled differently because they represent different things based on the size of x.

Putting both parts together gives us our CDF $f : (0, 2\lambda) \to [0, 1]$:

$$f(x) = \begin{cases} \frac{\arccos\left(1+\frac{\frac{1}{2}x^2-x\lambda}{1-\lambda+x}\right)}{2\arcsin\lambda} & \text{for } 0 < x < \sqrt{1-\lambda^2} - (1-\lambda) \\ 1 - \frac{\arccos\left(1+\frac{\frac{1}{2}x^2-x\lambda}{1-\lambda+x}\right)}{2\arcsin\lambda} & \text{for } \sqrt{1-\lambda^2} - (1-\lambda) \le x < 2\lambda \end{cases}.$$

To prove that our measure μ is absolutely continuous, we must show that our function f is differentiable (and continuous), because radial symmetry dictates that the probability density function at any given point on the annulus will be the derivative of f with respect to x divided by 2π. Both parts of f are clearly continuous and differentiable on their own ranges, so the only potential problem spot is where we piece the functions together, i.e. when $x = \sqrt{1-\lambda^2} - (1-\lambda)$. We must show for this value of x both that

$$\frac{\theta_A}{2\alpha} = 1 - \frac{\theta_B}{2\alpha} \text{ and that } \frac{d}{dx}\frac{\theta_A}{2\alpha} = \frac{d}{dx}\left(1 - \frac{\theta_B}{2\alpha}\right).$$

To prove the former, note that θ_A is the solution to the equation $1-\lambda+x = \cos\theta_A - \sqrt{\lambda^2 - \sin^2\theta_A}$. When $x = \sqrt{1-\lambda^2} - (1-\lambda)$ becomes

$$\sqrt{1-\lambda^2} = \cos\theta_A - \sqrt{\lambda^2 - \sin^2\theta_A},$$

which is satisfied by $\theta_A = \alpha = \arcsin\lambda$.

Similarly, $\theta_B = \alpha$ satisfies $1-\lambda+x = \cos\theta_B + \sqrt{\lambda^2 - \sin^2\theta_B}$ for this specific x. (Geometrically, this makes sense, because this value of x corresponds exactly to the point of tangency.) Thus

$$\frac{\theta_A}{2\alpha} = \frac{\alpha}{2\alpha} = \frac{1}{2} = 1 - \frac{1}{2} = 1 - \frac{\alpha}{2\alpha} = 1 - \frac{\theta_B}{2\alpha},$$

i.e. f is continuous.

For differentiability, we must show that

$$\frac{d}{dx}\frac{\arccos\left(1+\frac{\frac{1}{2}x^2-x\lambda}{1-\lambda+x}\right)}{2\arcsin\lambda} = \frac{d}{dx}\left(1 - \frac{\arccos\left(1+\frac{\frac{1}{2}x^2-x\lambda}{1-\lambda+x}\right)}{2\arcsin\lambda}\right),$$

i.e. that $\frac{d}{dx}\left(\frac{\arccos\left(1+\frac{\frac{1}{2}x^2-x\lambda}{1-\lambda+x}\right)}{2\arcsin\lambda}\right) = 0$, for $x = \sqrt{1-\lambda^2} - (1-\lambda)$. Fortunately, this derivative is computable, and evaluates to 0 for this specific value of x.

Therefore f is differentiable, so, as discussed above, μ is absolutely continuous, and finally ν_λ is also absolutely continuous. □

Intuitively, this simple geometric method should generalize to higher dimensions, and we find that it indeed does. In fact, the proof below is almost identical to the one above, and draws heavily from the work we've already done. Note that while the following proof applies for all Bernoulli convolutions of 2 or higher dimensions, the 3-dimensional case is by far the easiest to visualize, and as a result all graphics below represent 3-dimensional Bernoulli convolutions.

Theorem 5.3. For any integer $n \geq 2$, every n-dimensional Bernoulli convolution is absolutely continuous.

Proof. Given an integer $n \geq 2$, let $\lambda \in (0,1)$, and let ν_λ be its associated n-dimensional Bernoulli convolution. Again, define μ_λ to be the distribution our first two steps, $X_0 + \lambda X_1$. As before,

$$\nu_\lambda = \mu_\lambda * \text{the distribution of all other summands},$$

so to prove absolute continuity of ν_λ it suffices to prove absolute continuity of μ_λ.

Each X_i is chosen uniformly on S^{n-1} with radius 1, i.e. the unit $(n-1)$-sphere, so both μ_λ and ν_λ are radially symmetric, with μ_λ supported on the hyperspherical shell $\{x \in \mathbb{R}^n \mid 1-\lambda \leq |x| \leq 1+\lambda\}$. Therefore even though we have a higher-dimensional space, we can still determine the CDF of $X_0 + \lambda X_1$ by assuming this random variable lands along the positive x_1-axis. To do so, we will calculate the probability that, given this assumption, $X_0 + \lambda X_1$ has $x_1 \in [1-\lambda, 1-\lambda+x]$ as a function of x, which varies from 0 to 2λ. Our first "step" places us on the $(n-1)$-sphere of unit radius, so to examine our likelihood of returning to our target line segment, we consider λ-radius $(n-1)$-spheres with centers on the unit hypersphere.

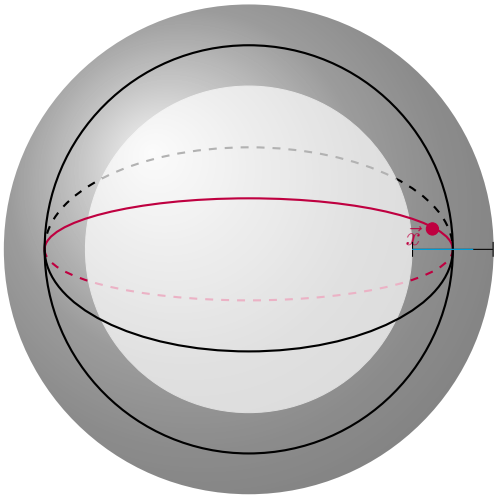

Figure 7: n-Dimensional Case

If we have a center with $x_3 = x_4 = \cdots = x_n = 0$ (that is, our center lies on the x_1x_2-plane), then our task of counting intersection points reduces immediately to the 2-dimensional case, which we've already studied in depth.

One may initially assume that a center not on the x_1x_2-plane, such as $\vec{x}$ in purple in Figure 7, might not be as simple. However, if we consider the plane containing $\vec{x}$, $(-1, 0, \ldots, 0)$, $(1, 0, \ldots, 0)$ (and therefore also our target line segment), it will intersect with our unit S^{n-1} in the great circle outlined in purple in Figure 7. The sphere of radius λ centered at $\vec{x}$ will intersect our target line segment only on its great circle that also lies in this plane. Therefore, we can reduce any $\vec{x}$ to our 2-dimensional case by considering this plane. In other words, it is only this plane which matters for determining intersections, and our angle of interest is the one made between the ray from the origin to $\vec{x}$ and the ray from the origin to $(1, 0, \ldots, 0)$.

Our calculations for θ_A, θ_B, and α therefore remain identical to those we performed in the previous proof. Once we find these angles, instead of directly comparing them as before, where we implicitly needed to calculate arc length because we were working on a circle, here we must calculate the surface area of hyperspherical caps with corresponding latitude angles. In the Figure 8, one such hyperspherical cap corresponding to a hypothetical α is outlined in red.

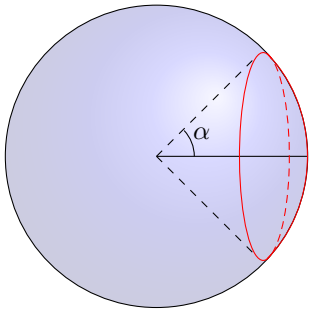

Figure 8: Hyperspherical Cap

Given the radius of the hypersphere (which in our case is always 1), explicit formulas exist for calculating these desired surface areas as functions of the latitudinal angle: denoting this area function as A_n as we have that

$$A_n(\theta) = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} I_{\sin^2\theta}\left(\frac{n}{2}, \frac{1}{2}\right),$$

where Γ is the gamma function and I is the regularized incomplete beta function [14]. This function is differentiable with respect to θ, and for our purposes this is the only information necessary.

We can now compile our piecewise function: for $x < \sqrt{1-\lambda^2} - (1-\lambda)$, we count $A_n(\theta_A)$ once and divide it by twice $A(\alpha)$, and for $x > \sqrt{1-\lambda^2} - (1-\lambda)$, we count $A_n(\theta_B)$ once, $A_n(\alpha) - A_n(\theta_B)$ twice, and divide again by twice $A_n(\alpha)$. Writing these angles as functions of x where appropriate, we have

$$f_n(x) = \begin{cases} \frac{A_n(\theta_A(x))}{2A_n(\alpha)} & \text{for } 0 < x < \sqrt{1-\lambda^2} - (1-\lambda) \\ \frac{A_n(\theta_B(x)) + 2(A_n(\alpha) - A_n(\theta_B(x)))}{2A_n(\alpha)} = 1 - \frac{A_n(\theta_B(x))}{2A_n(\alpha)} & \text{for } \sqrt{1-\lambda^2} - (1-\lambda) \leq x < 2\lambda \end{cases},$$

where

$$\arccos\left(1 + \frac{\frac{1}{2}x^2 - x\lambda}{1 - \lambda + x}\right) = \theta_A \text{ for } x \text{ small}$$
$$\text{and} = \theta_B \text{ for } x \text{ large.}$$

Composition and multiplication preserves differentiability, so we have that $f_n(x)$ is differentiable everywhere on both of its "pieces," and again all that remains to check is continuity and differentiability when $x = \sqrt{1-\lambda^2} - (1-\lambda)$. However, we have already shown that for this specific x,

$$\theta_A(x) = \alpha = \theta_B(x),$$

and

$$\frac{d}{dx}\theta_A = \frac{d}{dx}\theta_B = 0.$$

Thus we immediately have continuity, because

$$\frac{A_n(\theta_A(x))}{2A_n(\alpha)} = \frac{A_n(\alpha)}{2A_n(\alpha)} = 1 - \frac{A_n(\alpha)}{2A_n(\alpha)} = 1 - \frac{A_n(\theta_B(x))}{2A_n(\alpha)}.$$

Using the chain rule,

$$\frac{d}{dx}\frac{A_n(\theta_A(x))}{2A_n(\alpha)} = A_n'(\theta_A(x))\theta_A'(x) = 0 = -A_n'(\theta_B(x))\theta_B'(x),$$

so

$$\frac{A_n(\theta_A(x))}{2A_n(\alpha)} = 0 = 1 - \frac{A_n(\theta_B(x))}{2A_n(\alpha)},$$

which establishes differentiability.

Therefore our CDF has a density function, which implies that μ_λ and therefore ν_λ is absolutely continuous.

□

A Code for Section 3

Here we document two python scripts that were used to test results and develop intuition regarding the graphical structures discussed in section 3.

A.1 Drawing Bernoulli Graphs

The following script requests a value of λ from the user and a level n, and then plots corresponding the n-finite Bernoulli Convolution graph.

```
import matplotlib.pyplot as plt
from matplotlib import pyplot

# Prompts user for value of lambda and number of levels of the graph

root  = input("Enter a value for lambda: ")
treeSize = input("Enter the number of levels: ")

# Recursively draws a Bernoulli graph with the appropriate levels

def recDrawTree(parent, currHeight, totalHeight, root, heightUnit):
    addedLength = root**currHeight
    rightChild = [parent[0] + addedLength, parent[1] - heightUnit]
    leftChild = [parent[0] - addedLength, parent[1] - heightUnit]
    plt.plot(leftChild[0], leftChild[1], marker = 'o')
    plt.plot(rightChild[0], rightChild[1], marker = 'o')
    pyplot.plot([parent[0], leftChild[0]], [parent[1], leftChild[1]])
    pyplot.plot([parent[0], rightChild[0]], [parent[1], rightChild[1]])
    if currHeight != totalHeight - 1:
        recDrawTree(rightChild, currHeight + 1, totalHeight, root, heightUnit)
        recDrawTree(leftChild, currHeight + 1, totalHeight, root, heightUnit)

# Sets axes and plots graph

rightBound = 1/(1-root)
leftBound = -1 * rightBound
pyplot.axis([leftBound, rightBound, 0, treeSize + 1])
plt.plot(0, treeSize, marker = 'o')
recDrawTree([0, treeSize], 0, treeSize, root, 1)
plt.show()
plt.clf()
```

A.2 Counting in r-Relational Binary Trees

Given a relation r and level n, the following script will return to the user both N_k and N_k^f for all $k \in \{0, \ldots, n\}$.

```
#  Given some list of paths, this function a list of paths a level down

def createNewPaths(paths):
    newPaths = []
    for path in paths:
        leftPath = list(path)
        leftPath.append(-1)
        rightPath = list(path)
        rightPath.append(1)
        newPaths.append(rightPath)
        newPaths.append(leftPath)
    return newPaths


#  Given a set of paths returns the set of paths n levels down

def createPaths(paths, n):
    if n == 0:
        return []
    i = 1
    while i < n:
        paths = createNewPaths(paths)
        i += 1
    return paths

#  Given a relation r returns the negated relation -r

def negateRelation(relation):
    negatedRelation = []
    for i in range(0, len(relation)):
        negatedRelation.append(-relation[i])
    return negatedRelation

#  Given a path p, relation r, negated relation -r, and index k,
#  the following function determines whether it is well defined to
#  apply the relation to said path at the index k + l - 1,
#  where l is the length of the relation

def isAppliable(path, relation, negatedRelation,  k):
    if len(path) - k < len(relation):
        return False
```

```
    subPath = []
    for i in range(0, len(relation)):
        if relation[i] == 0:
            subPath.append(0)
        else:
            subPath.append(path[k + i])
    if subPath == relation or subPath == negatedRelation:
        return True
    else:
        return False

#  Given a path p, relation, and index k, the following function
#  returns the path one gets by applying the relation to p at
#  index k + l - 1, where l is the length of the relation

def applyRelation(path, relation, k):
    if isAppliable(path, relation, negateRelation(relation), k):
        newPath = list(path)
        for i in range(0, len(relation)):
            if relation[i] != 0:
                newPath[k + i] = -newPath[k + i]
        return newPath
    return path

#  Given a path p and relation r, the next to functions recursively
#  find and then return a list of all paths equivalent to p by the
#  relation r

def findEquivPaths(path, relation):
    equivalentPaths = []
    equivalentPaths.append(path)
    recFindEquivPaths(path, relation, equivalentPaths)
    return equivalentPaths

def recFindEquivPaths(path, relation, equivalentPaths):
    for i in range(0, len(path) - len(relation) + 1):
        equivalentPath = applyRelation(path, relation, i)
        if equivalentPath not in equivalentPaths:
            equivalentPaths.append(equivalentPath)
            recFindEquivPaths(equivalentPath, relation, equivalentPaths)

#  Given a set of paths and a relation r, the following function
#  returns a dictionary where each key corresponds to a particular
#  frequency, and the corresponding value of the key corresponds
#  to the number of paths with said frequency
```

```
def countFrequencies(allPaths, relation):
    frequencies = {}
    for path in allPaths:
        frequency = len(findEquivPaths(path, relation))
        frequencies.setdefault(frequency, 0)
        frequencies[frequency] += 1
    for key in frequencies:
        frequencies[key] = (frequencies[key])/key
    return frequencies

#  Prompts user for a relation and level, and prints all information
#  regarding the number of particular frequencies per level and total number
#  of nodes per level up to the entered level

def countGraph():
    relation = input("Enter relation: ")
    level = input("Enter final level: ")
    print {1:1}
    print 1
    for n in range(1, level + 1):
        totalNodes = 0
        frequencies = countFrequencies(createPaths([[1],[-1]], n), relation)
        print frequencies
        for frequency in frequencies:
            totalNodes += frequencies[frequency]
        print totalNodes

countGraph()
```


Acknowledgements

We'd like to thank Patrick Revilla for *many* helpful insights, particularly in help with proving Lemma 3.9; George Schaeffer for being an enthusiastic and supportive program director; and Erik Bates for his infinite patience and incredible mentorship.

References

[1] Brown, J. L. Jr. "Zeckendorf's Theorem and Some Applications." *Fib. Quart.* **2**, 163-168, 1964.

[2] Pablo Shmerkin and Boris Solomyak, On the Absolute Continuity of Complex Bernoulli Convolutions. arXiv:1504.00631.

[3] Rick Durrett, Probability: Theory and Examples. Cambridge University Press, 2010.

[4] Børge Jessen and Aurel Wintner, Distribution Functions and the Riemann Zeta Function. *Transactions of the American Mathematical Society.* **38** (1935), no. 1, 48-88.

[5] Paul Erdős, On a Family of Symmetric Bernoulli Convolutions. *American Journal of Mathematics.* **61** (1939), 974–975.

[6] Adriano Mario Garsia, Arithmetic properties of Bernoulli convolutions. *Transactions of the American Mathematical Society.* **102** (1962), 409–432.

[7] Adriano Mario Garsia, Entropy and singularity of infinite convolutions. *Pacific Journal of Mathematics.* **13** (1963), 1159–1169.

[8] Michael Hochman, On self-similar sets with overlaps and inverse theorems for entropy. *Annals of Mathematics.* **2** 180 (2014), no. 2, 773–822.

[9] Petra Kocábová, Zuzana Masáková, Edita Pelantová. Ambiguity in the m-bonacci numeration system. Discrete Mathematics and Theoretical Computer Science, DMTCS, 2007, 9 (2), pp.109–123.

[10] Richard Kershner and Aurel Wintner, On symmetric Bernoulli convolutions. *American Journal of Mathematics.* **57** (1935), 541–548.

[11] Aurel Wintner, On convergent Poisson convolutions. *American Journal of Mathematics.* **57** (1935), 827–838.

[12] J.C. Alexander and D.B. Zagier, The entropy of certain infinitely convolved Bernoulli measures. *L. London Math. Soc.* **44** (1991), pp. 121-134

[13] Marcia Edson and Luca Q. Zamboni, On representations of positive integers in the fibonacci base. *Theoretical Computer Science.* **326** (2004), 241-260.

[14] Shengqiao Li, Concise Formulas for the Area and Volume of a Hyperspherical Cap. *Asian Journal of Mathematics and Statistics.* **4** (1), 66-70.