Tossing a Coin

Olivia Lee, Katherine Worden, Alice Zhu

February 9th, 2024

Abstract

This paper explores cumulative distribution functions for infinite coin tosses, parameterized by the probability p of flipping heads. We graph the outcomes of simulated coin flips and study properties of the cumulative distribution function, analyzing its pathological behavior in terms of continuity, differentiability, and arc length.

1 Introduction

Coin flips are a classic probability problem, and some probability distributions that commonly result from these processes are the geometric and binomial distributions. In this problem, we consider an interesting twist on the classic coin problem with an infinite number of coin tosses. To begin, the result of n coin tosses can be represented by a binary number in the interval $[0, 1]$ with n digits, where the k-th digit is 0 if the kth toss comes up tails and 1 if it comes up heads. For example, in binary expansion for one toss of heads, $0.1 = \frac{1}{2}$, and tails, heads, heads is $0.011 = \frac{3}{8}$. In section 2, we will prove any number in the interval [0, 1] can be represented in a binary expansion.

Now let y be the outcome of an infinite toss, and consider the function on $[0, 1]$: $P(x) =$ probability that $y \leq x$, which is the cumulative distribution function. We let the probability of heads be p, so tails is $q = 1 - p$, where the coin is not necessarily fair. We will begin in Section 2 by deriving expressions for $P\left(\frac{x}{2}\right)$ and $P\left(\frac{1+x}{2}\right)$ in terms of $P(x)$. After this, in Sections 3 and 4, we generate an approximation of the cumulative distribution function for $P(x)$ for various values of p. In Section 5, we will remark on the different properties of the cumulative distribution functions. In Sections 6 and 7, we make connections to some other concepts in probability, such as the Galton board, and explore areas for future work.

2 Expressing $P(\frac{x}{2})$ and $P(\frac{1+x}{2})$ in terms of $P(x)$

Before we begin expressing $P(\frac{x}{2})$ and $P(\frac{1+x}{2})$ in terms of $P(x)$, it is first important to note the lemma any number in the interval $[0, 1]$ can be represented in a binary expansion.

Lemma 1. Any number in the interval $[0, 1]$ can be represented in a binary expansion.
Proof: Consider the interval $[0, 1]$. For any number m, we can divide the interval into 2^m many intervals, each of length $\frac{1}{2^m}$, where each interval goes from $[\frac{k}{2^m}, \frac{k+1}{2^m}]$, with $0 \leq k \leq 2^m - 1$. For example, when $m = 2$, we have 4 intervals, each of length $\frac{1}{4}$. Then we have the intervals $[0, \frac{1}{4}]$, $[\frac{1}{4}, \frac{2}{4}]$, $[\frac{2}{4}, \frac{3}{4}]$, $[\frac{3}{4}, \frac{4}{4}]$. At higher values of m, we have more and more intervals, each of smaller lengths. Now consider any number y in the interval $[0, 1]$. We can think of identifying this y by which side of an interval it falls on for every level m. First, for $m = 1$, is the number between $[0, \frac{1}{2}]$ or $[\frac{1}{2}, 1]$? If the former, the first digit is a 0, and if the latter, a 1. Let's assume the first digit is a 1. Next, for $m = 2$, we consider whether, knowing that $y \geq \frac{1}{2}$, is y between $[\frac{1}{2}, \frac{3}{4}]$, or $[\frac{3}{4}, \frac{4}{4}]$? Again, if y is in the left partition of this interval, add a 0; else add a 1. By following this process for higher values of m, we can describe y as a sum of these 1's or 0's for each decimal place, based on which side of the interval y falls in at each level. Then we can express y as $y = \sum_{m=1}^{N} \frac{x_m}{2^m}$, as the summation of small intervals between 0 and 1, where $x_m \in 0, 1$.
Now let's take $N = 1$ and consider the difference between some number y and the interval $\frac{x_n}{2^n}$: $y - \frac{x}{2}$. Since we can place y to the precision of either left or right of $\frac{1}{2}$, $y - \frac{x_n}{2}$ is bounded by the length of the smallest interval we've placed y in: $y - \frac{x_n}{2} < \frac{1}{2}$. Then, in the generalized case, as $N \to \infty$, $y - \sum_{m=1}^{N} \frac{x_m}{2^m} < \frac{1}{2^N} \to 0$, as the distance between the intervals becomes infinitely small. Now assume that $y - \sum_{m=1}^{N} \frac{x_m}{2^m} < \frac{1}{2^N}$, and consider $N = N + 1$: $y - \sum_{m=1}^{N+1} \frac{x_m}{2^m} < \frac{1}{2^{N+1}}$. If we expand out the final term, we recover $y - \sum_{m=1}^{N} \frac{x_m}{2^m} - \frac{x_m+1}{2^{m+1}} < \frac{1}{2^N}$, which demonstrates that y is always within $\frac{1}{2^N}$ from the closest

edge of the interval, where the whole interval is size $\frac{1}{2^N}$. $\square$

Note this binary expansion is not necessarily unique, e.g. $\frac{1}{2}$ could be represented as either 0.10000 or 0.01111. However, we argue that this will not pose a problem for finding $P(x)$. While the infinite sequences 0.10000... and 0.01111... represent very different coin flip sequences, they are essentially equivalent in terms of binary expansion: using the intervals interpretation, both numbers (in the limit) are essentially in the same position on the number line. However, since $P(x)$ represents the probability that the infinite coin toss $y \leq x$, this single example will not affect the value of $P(x)$. This is because the probability of any single outcome is 0 as the probability of a single outcome with say k heads and $(n-k)$ tails is $p^k(1-p)^{n-k} \to 0$ since $p, 1-p < 1$. In other words, the function represented by $P(x)$ is not dependent on the representation of y, because two different infinite binary expansions (or coin toss sequences) result in the same number y, but since the probability of each outcome is 0, this does not affect the cumulative distribution function.

Now, we will first express $P(\frac{x}{2})$ in terms of $P(x)$. We want to show that for any x representable by a binary expansion, $\frac{x}{2}$ can be represented by shifting one digit to the right (e.g. inserting a 0 in the front). By the lemma above, we have shown that any number in the interval [0, 1] has a binary expansion. So any y $\in [0,1]$ can be expressed as $y = \sum_{n=0}^{\infty} \frac{x_n}{2^n}$, where $x \in 0, 1$. Now consider $\frac{y}{2}$:

$$\frac{y}{2} = \frac{1}{2}\sum_{n=0}^{\infty} \frac{x_n}{2^n} = \sum_{n=0}^{\infty} \frac{x_n}{2} \cdot \frac{1}{2^n} = \sum_{n=0}^{\infty} \frac{x}{2^{n+1}}$$

The primary intuition behind deriving the expression for $P(\frac{x}{2})$ from $P(x)$ is that we shift all the digits for the outcomes satisfying $P(x)$ 1 digit to the right and prepend a 0 to the binary string. Let us consider the case where there is a digit 1 in the k^{th} position where $k \in \mathbb{Z}$, which adds $\frac{1}{2^k}$ to the binary expansion. Shifting this to the right by 1, there is now a 1 in the $(k+1)^{th}$ position, which adds $\frac{1}{2^{k+1}} = \frac{1}{2}(\frac{1}{2^k})$ to the binary expansion. If we do this for every digit in the binary expansion, the overall sum is halved.

For y, the outcome of our infinite toss, observe that $P(x)$ refers to the probability that $y \leq x$. When we consider $P(\frac{x}{2})$, we are considering the probability that $y \leq \frac{x}{2}$. Mathematically, for $y \in [0, 1]$, $y = \sum_{k=0}^{\infty} \frac{x_k}{2^k}$ where $x_k \in 0, 1$. Then, $\frac{y}{2} = \sum_{k=0}^{\infty} \frac{x_k}{2^{k+1}} = \sum_{k=0}^{\infty} \frac{1}{2}\frac{x_k}{2^k}$, as desired. This justifies the transition function of shifting all the digits for the outcomes satisfying $P(x)$ 1 digit to the right, which is equivalent to dividing the binary expansion by 2, and prepend a 0. Hence, dividing x by 2 is equivalent a right shift, so $P(\frac{x}{2})$ refers to x with a 0 inserted in the first position, followed by the original digits of x. Note that a leading zero refers to flipping tail. Hence $P(\frac{x}{2})$ refers to the event where we flip a tails first, and all subsequent flips are the same as the event x.

Since all of the coin flips are independent, we can therefore multiply $P(x)$ by the probability of the first coin flip resulting in tails, to get the probability of flipping a tails (or 0) first followed by all of the flips corresponding to binary sequences generating an infinite coin toss $y \leq x$. In general, $P(\frac{x}{2}) = (1-p) \cdot P(x)$, since the probability of the first coin flip being 0 or tails is $(1-p)$.

Therefore, we can express $P(\frac{x}{2})$ as

$$P\left(\frac{x}{2}\right) = (1-p) \cdot P(x) \tag{1}$$

Similarly, the intuition behind deriving the expression for $P\left(\frac{1+x}{2}\right)$ from $P(x)$ is that we shift all the digits for the outcomes satisfying $P(x)$ to the right by 1 and prepend a 1 to the binary string. The reasoning for this intuition is similar to the above, but adding a 1 digit in the first position adds a value of $\frac{1}{2}$ to the overall binary expansion, since we want to obtain $P\left(\frac{1}{2} + \frac{x}{2}\right)$ from $P(x)$. To derive this, we need to consider two cases:

- Case 1 - Outcomes in the interval $[0, \frac{1}{2}]$: In this case, we consider all the outcomes of infinite tosses where the first flip is 0, or tails. This occurs with probability $(1-p)$, since the probability of the first coin flip being tails is $(1-p)$.
- Case 2 - Outcomes in the interval $[\frac{1}{2}, \frac{1+x}{2}]$: In this case, we use the reasoning similar to the reasoning for $P\left(\frac{x}{2}\right)$, except the first digit prepended to the binary expansion is 1 instead of 0. We can therefore similarly multiply $P(x)$ by the probability of the first coin flip resulting in heads (or 1), which is $p \cdot P(x)$.

Considering these cases jointly, we can then express $P(\frac{1+x}{2})$ as follows:

$$P\left(\frac{1+x}{2}\right) = (1-p) + p \cdot P(x) \tag{2}$$

To restate our findings:

$$P\left(\frac{x}{2}\right) = (1-p) \cdot P(x)$$
$$P\left(\frac{1+x}{2}\right) = (1-p) + p \cdot P(x)$$

We can use a change of variables to represent $x = 2a$ (i.e. $a = \frac{x}{2}$) to obtain:

$$P(a) = (1-p) \cdot P(2a)$$

We can use a different change of variables to represent $x = 2b-1$ (i.e. $b = \frac{1+x}{2}$) to obtain:

$$P(b) = (1-p) + p \cdot P(2b-1)$$

We can combine these two cases to get a combined function for $P(z)$:

$$P(z) = \begin{cases} (1-p) \cdot P(2z) & 0 \leq z \leq \frac{1}{2} \\ (1-p) + p \cdot P(2z-1) & \frac{1}{2} \leq z \leq 1 \end{cases} \tag{3}$$

This function in Equation 3 determines $P(x)$ at *dyadic rationals*, or numbers that can be expressed as a fraction whose denominator is a power of two. Specifically, dyadic rationals have binary expansions with finitely many 1 digits. Since we know for a fact that $P(0) = 0$ and $P(1) = 1$, given an arbitrary dyadic rational x, we can determine $P(x)$ using finitely many applications of the two equations in the combined function. As an example, consider $x = \frac{3}{8} = 0.011$. We see that

$$\begin{aligned}
\frac{3}{8} = \frac{1}{2}\left(\frac{3}{4}\right) &= \frac{1}{2}\left(\frac{1}{2} + \frac{1}{4}\right) = \frac{1}{2}\left(\frac{1}{2} + \frac{1}{2}\left(\frac{1}{2}\right)\right) = \frac{1}{2}\left(\frac{1+\frac{1}{2}}{2}\right) \\
\Rightarrow P\left(\frac{3}{8}\right) &= P\left(\frac{1}{2}\left(\frac{3}{4}\right)\right) \\
&= (1-p) \cdot P\left(\frac{3}{4}\right) \text{ (Equation 3a)} \\
&= (1-p) \cdot P\left(\frac{1+\frac{1}{2}}{2}\right) \\
&= (1-p) \cdot \left((1-p) + p \cdot P\left(\frac{1}{2}\right)\right) \text{ (Equation 3b)} \\
&= (1-p) \cdot ((1-p) + p \cdot ((1-p) \cdot P(1))) \text{ (Equation 3a)} \\
&= (1-p)^2 + p(1-p)^2 \text{ (since } P(1) = 1) \\
&= (1-p)^2(1+p)
\end{aligned}$$

However, for the infinite coin toss, deriving a closed form solution for the cumulative distribution function is likely to be challenging (or impossible). That said, these values can be used to create a plot of the function from which we can observe interesting properties. From Equation 3 alone, we hypothesize that a graph of $P(x)$ is likely continuous since x varies continuously in the interval $[0, 1]$ and for every value of x, $P(x)$ can be determined using the corresponding equation. However, $P(x)$ will likely not be differentiable since $P(x)$ is modeled using different functions for $0 \leq z \leq \frac{1}{2}$ and $\frac{1}{2} \leq z \leq 1$, so there will be a sharp transition at the value $z = \frac{1}{2}$. We show this explicitly in the next section by plotting the cumulative distribution functions for different values of p.

3 Continuous random variables and probability distributions

From the problem statement, let y be the outcome of an infinite toss, and consider the function on the interval $[0, 1]$ defined as $P(x) =$ probability that $y \leq x$. At first glance, this seems to resemble a cumulative distribution function for a continuous random variable in statistics, and in this section we push this idea further.

From above, we defined $P(x) = p(y \leq x)$, and we can think of y as the outcome of a series of random events taking on any value in the range $[0, 1]$, meaning y is a continuous random variable. We are ascertaining whether the value y lies within an interval $[0, x]$ (i.e., $p(y \leq x) = p(y \in [0, x])$. This means $P(x) = P(y \leq x)$ expresses the cumulative distribution function, which is parameterized by different values of p, meaning the value of p will determine the underlying distribution of y[1]. We prove the theorem below:

Theorem 1. The probability of y being any single outcome is zero, i.e., $p(y = x) = 0$.
Proof: We can prove this for the infinite coin toss: as the number of digits in the binary expansion (and analogously, the number of partitions in the interval $[0, 1]$) $n \to \infty$, the probability of a single outcome with say k heads and $(n - k)$ tails is $p^k(1 - p)^{n-k} \to 0$ since $p, 1 - p < 1$. Therefore, y is a continuous random variable, and we represent $P(x)$ as a cumulative distribution function since it is a continuous random variable, rather than typical (finite) coin tosses which are discrete random variables. Therefore, our cumulative distribution function is continuous. □

To explore this notion, consider first the simplest case of a fair coin, meaning the probability of any toss coming up heads (or 1) is equal to that of tails (or 0), and they are both $\frac{1}{2}$. As described above all outcomes of an infinite coin toss have the same probability; specifically, 0. Further, because heads and tails occur with the same probability, y is equally likely to be inside of any given equally sized interval between $[0, 1]$.

Hence we can model the probability density of the infinite tosses with the uniform distribution. The underlying probability density function for y would therefore be the uniform distribution, and $P(x)$ will be the cumulative distribution function of a uniform random variable. We can see the probability density function (Figure 1a) and cumulative distribution function (Figure 1b) of the fair coin in the figure below, where the interval is $[0, 1]$.

(a) Probability density function of a uniform random variable.

(b) Cumulative distribution function of a uniform random variable.

Figure 1: Probability distributions for the fair coin.

Now, we seek to generalize to the unfair coin, meaning the probability of any toss coming up heads p and probability of any toss coming up tails q are not equal, and $p + q = 1$. This is far less straightforward and our primary next steps involve exploring this "pathological" function as stated in the problem statement.

4 Graphing the Pathological Function

Attempting to derive a closed form solution for the cumulative distribution function for the unfair coin is challenging, or potentially impossible. Rather than deriving a formula from first principles, we can graph the probabilistic outcomes of large number of coin flips (to approximate approaching an infinite number of coin flips). We can then analyze various properties about the "pathological" function, such as its continuity and differentiability. We can also get a better understanding of the $P(x)$ for various values of x and see how this corresponds with the function we derived above in Equation 3. The code we used to generate the cumulative distribution functions for unfair coins is shown below:

```
def calculate(p, n, iters):
    values = [1/2 + 1/(2**(k+1)) for k in range(0, n + 1)] + [1/(2**k) for k
    in range(1, n + 1)]
    print(values)
    probabilities = [(1-p) + p*(1-p)**k for k in range(0, n + 1)] + [(1-p)**k
     for k in range(1, n + 1)]
    print(probabilities)
    for i in range(iters):
```

[1]Another remark, from our derivations above, is that since $p(y \leq x) = p(y \in [0, x])$, then $p(y \leq \frac{x}{2}) = p(y \in [0, T(x)])$, where T represents some transition function. In the case of $T(x) = \frac{x}{2}$, T represents a shift to the right by 1 digit and prepending a 0 to the binary expansion.

```
        values = [(value/2) + 0.5 for value in values] + [value/2 for value in values]
        probabilities = [p * prob + (1-p) for prob in probabilities] + [prob*(1-p) for prob in probabilities]
        print(i)
    return values, probabilities


if __name__ == '__main__':
    p, n, iters = 0.9, 30, 17
    calculated_values, calculated_probabilities = calculate(p, n, iters)
    plt.scatter(calculated_values, calculated_probabilities, s = 1)
    ... # formatting
    plt.show()
```

We plot $(\frac{1}{2}^i,\ P(\frac{1}{2}^i))$ and $(\frac{1}{2}+\frac{1}{2}^i, P(\frac{1}{2}+\frac{1}{2}^i))$ values for $1 \leq i \leq n$ using our given equations. Then, we iteratively divide these x values by $\frac{1}{2}$ and extended the array to all of these x values to x and $x+\frac{1}{2}$, creating new y values for corresponding x along the way. By iteratively doing this, we are able to obtain more x values closer to $\frac{1}{2}$, until we are able to fill in many many points between 0 and 1. Below are the plots the cumulative distribution functions for the uniform and "pathological" cases:

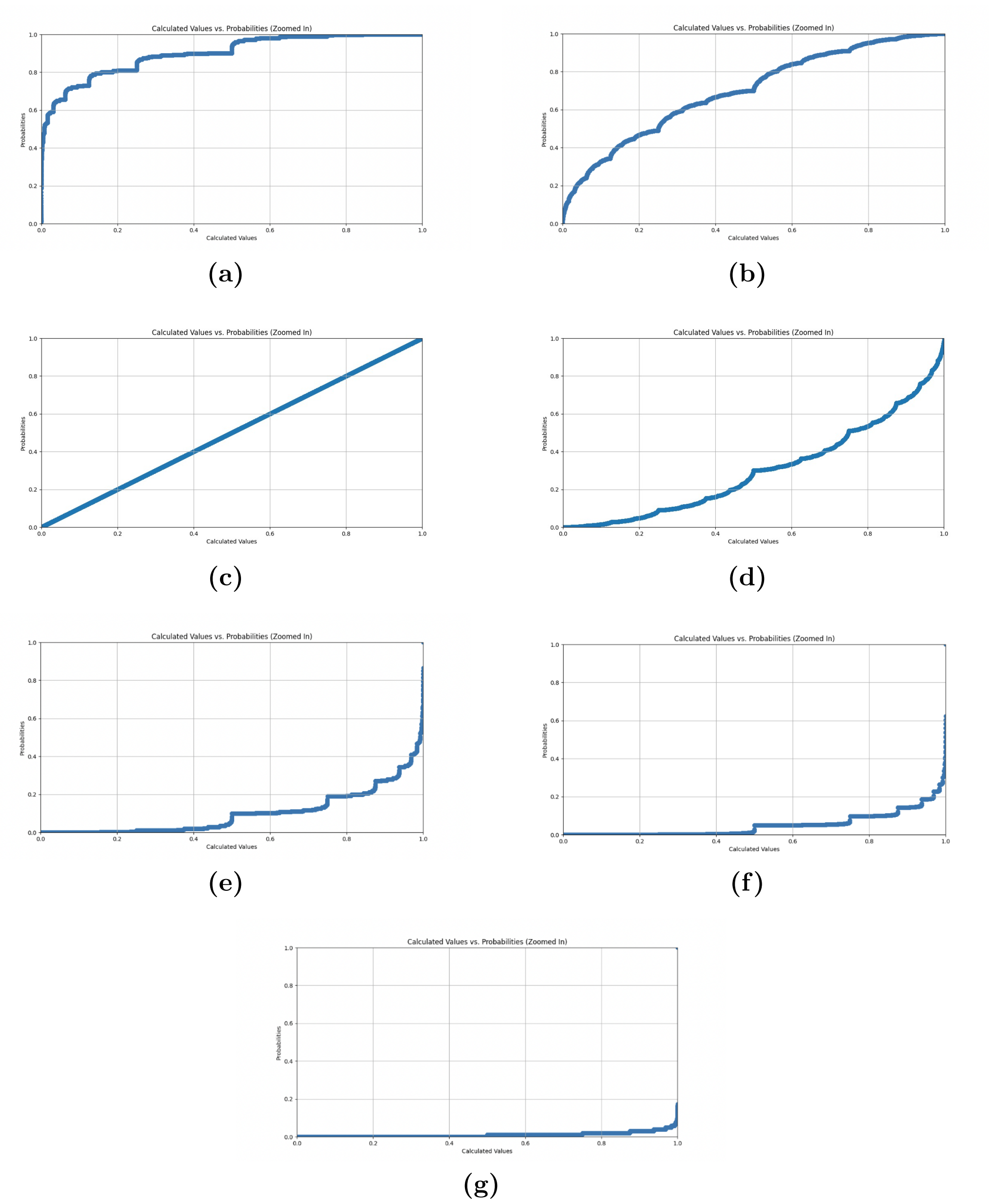

Figure 2: Cumulative distribution functions for various values of p (the probability of flipping heads), where (a) p=0.1 (b) p=0.3 (c) p=0.5 (d) p=0.7 (e) p=0.9 (f) p=0.95 (g) p=0.99.

Observe that at smaller values of p, when the coin is biased in favor of tails, the graph is concave down, with a steep curve around the lower calculated values. In contrast, at higher p values, the graph has a slow take-off with a steep curve closer to one. In the former case, since the coin favors tails, we expect it to take longer until we see a 1 appear (e.g. if p = 0.01, we expect to see one 1 in 100 trials). This means that runs of 0's are more common, so we see more clusters around numbers with more leading zeros. Hence these outcomes are clustered around smaller numbers (before $\frac{1}{2}$. Similarly, for higher p values, we expect it to take longer until zeroes appear, so outcomes are clustered around the end of the interval, where ones appear earlier in the binary expansion. Another observation pertains to the symmetry of the graphs. For the graph of $P(x)$ generated with $p = a$, the graph of $P(x)$ generated with $p = 1 - a$ is symmetric about the point $(\frac{1}{2}, \frac{1}{2})$, specifically a 180-degree rotation in the same plane. In Figure 2, this can be seen by comparing Figures 2(a) and 2(e), and Figures 2(b) and 2(d).

In the next section, we will analyze the "pathological" behaviors and characteristics of these functions.

5 Analyzing the Cumulative Distribution Function

5.1 Continuity

Definition 1. A cumulative distribution function is continuous if and only if $P(X = x) = 0$, i.e., the probability that a random variable X takes on any specific value x is 0.

With the above definition, we prove that the function $P(x)$ is continuous.

Theorem 2. $P(x)$ is continuous.
Proof: We want to show that for our infinite coin toss y, $p(y = x) = 0$, which would imply that our cumulative distribution function $P(x)$ is continuous by Definition 1. We know that the probability of the binary sequence outcome x is $p^h(1-p)^t$, where h is the number of heads and t is the number of tails flipped. We know that $p, 1-p < 1$. There are an infinite number of coin flips so there are either an infinite number of heads or tails flipped. Then, since $\lim_{h\to\infty} p^h(1-p)^t = 0$ and $\lim_{t\to\infty} p^h(1-p)^t = 0$, $p^h(1-p)^t \to 0$. Then, $p(y = x) = 0$, so by Definition 1, our cumulative distribution function $P(x)$ is continuous.

5.2 Differentiablity

From the graphs in Section 4, we observe that the most prominent kinks frequently occur at the values $x = \frac{1}{4}, x = \frac{1}{2}, x = \frac{3}{4}...$, i.e., at locations corresponding to the dyadic numbers. Specifically, these locations correspond to places where the binary expansion consists of a finite number of ones, e.g. where the outcome of a single coin toss is added to the binary expansion. Most notably, at $x = \frac{1}{2}$, there is an obvious kink, due to the process of generating the points for the graph corresponding to $(\frac{1}{2}^i,\ P(\frac{1}{2}^i))$ and $(\frac{1}{2} + \frac{1}{2}^i, P(\frac{1}{2} + \frac{1}{2}^i))$. We use this observation to prove that $P(x)$ is not differentiable at $x = \frac{1}{2}$:

Theorem 3. $P(x)$ is not differentiable at $x = \frac{1}{2}$ for values of $p \neq \frac{1}{2}$.
Proof: Notice that the numbers between $\frac{1}{2} = 0.1$ and $\frac{1}{2} + \frac{1}{2^n} = 0.10...01$ (where there are $n-2$ 0s before the second 1) are all in the form 0.10...0.... The number must have a 1 as its first binary digit as it needs to be greater than $\frac{1}{2}$, and since it needs to be less than $\frac{1}{2} + \frac{1}{2^n}$, the number needs to be followed with $n-1$ 0s, and the numbers after the first n digits will not matter. (The cases where the number is equal to $\frac{1}{2}$ and $\frac{1}{2} + \frac{1}{2^n}$ is not significant to our calculations as the probability of that occuring is 0.) Thus, $P(\frac{1}{2} + \frac{1}{2^n}) - P(\frac{1}{2})$ is the probability a number occurs in the form 0.10...0..., with 1 followed by $n-1$ 0s.
The derivative at $\frac{1}{2}$ approaching from the right side is

$$\lim_{n\to\infty} \frac{P(\frac{1}{2} + \frac{1}{2^n}) - P(\frac{1}{2})}{\frac{1}{2^n}} = 2^n[p(1-p)^{n-1}] = \frac{p}{1-p} * (2-2p)^n$$

Notice that when $p < \frac{1}{2}$, $(2-2p)^n \to \infty$ and when $p > \frac{1}{2}$, $(2-2p)^n \to 0$, and when $p = \frac{1}{2}$, $(2-2p)^n \to 1$.
Now we examine what numbers are between $\frac{1}{2} = 0.1$ and $\frac{1}{2} - \frac{1}{2^n} = 0.01...1$ where there are $n-1$ 1s after the first 0. The number must have a 0 as its first binary digit as it needs to be less than $\frac{1}{2}$ and the next $n-1$ digits must be 1, for the number to be greater than $\frac{1}{2} - \frac{1}{2^n}$. The digits after the first n digits will not affect whether the number is between $\frac{1}{2}$ and $\frac{1}{2} - \frac{1}{2^n}$ or not. (The cases where the number is equal to $\frac{1}{2}$ and $\frac{1}{2} - \frac{1}{2^n}$ is not significant to our calculations as the probability of that occuring is 0.) Thus, $P(\frac{1}{2}) - P(\frac{1}{2} - \frac{1}{2^n})$ is the probability a number occurs in the form 0.01...1..., with 0 followed by $n-1$ 1s.
Now, examine the derivative at $\frac{1}{2}$ approaching from the left side:

$$\lim_{n\to\infty} \frac{P(\frac{1}{2}) - P(\frac{1}{2} - \frac{1}{2^n})}{\frac{1}{2^n}} = 2^n[(1-p)p^{n-1}] = \frac{1-p}{p} \cdot (2p)^n$$

Notice that when $p > \frac{1}{2}$, $(2p)^n \to \infty$, and when $p = \frac{1}{2}$, $(2p)^n \to 1$, and when $p < \frac{1}{2}$, $(2p)^n \to 0$. Our derivative is only defined at $p = \frac{1}{2}$, where the derivative is equal to 1 since $2 - 2p = 2 - 2(\frac{1}{2}) = 1$. $\square$

From Theorem 3, we extend this reasoning to all the powers of two, i.e. $\frac{1}{2^i}$ for $i \in \mathbb{N}$:

Theorem 4. $P(x)$ is not differentiable at $x = \frac{1}{2^i}$ for values of $p \neq \frac{1}{2}$.
Proof: Notice that the numbers between $\frac{1}{2^i} = 0.0...01$ (there are $i-1$ 0s before the 1) and $\frac{1}{2^i} + \frac{1}{2^n} = 0.0...010...01$ (there are $n-1$ digits before the second 1) are all in the form 0.0...010...0. The number must have $i-1$ 0s followed by a 1 as its first i digits as it needs to be greater than $\frac{1}{2^i}$, and since it needs to be less than $\frac{1}{2^i} + \frac{1}{2^n}$, there must be $n-i$ 0s after the first i digits. The numbers after the first n digits will not matter. (The cases where the number is equal to $\frac{1}{2^i}$ and $\frac{1}{2^i} + \frac{1}{2^n}$ is not significant to our calculations as the probability of that occuring is 0.) Thus, $P(\frac{1}{2^i} + \frac{1}{2^n}) - P(\frac{1}{2^i})$ is the probability a number occurs in the form 0.0...010...0..., with $i-1$ 0s followed by a 1 followed by $n-i$ 0s.
The derivative at $\frac{1}{2^i}$ approaching from the right side is

$$\lim_{n\to\infty} \frac{P(\frac{1}{2^i} + \frac{1}{2^n}) - P(\frac{1}{2^i})}{\frac{1}{2^n}} = 2^n[p(1-p)^{n-1}] = \frac{p}{1-p} \cdot (2-2p)^n$$

Notice that when $p < \frac{1}{2}$, $(2-2p)^n \to \infty$ and when $p > \frac{1}{2}$, $(2-2p)^n \to 0$, and when $p = \frac{1}{2}$, $(2-2p)^n \to 1$.
Now examine what numbers are between $\frac{1}{2^i} = 0.0...01$ and $\frac{1}{2} - \frac{1}{2^n} = 0.0...01...1$ (which has i 0s followed by $n-i$ 1s). The number must have i 0s as its first i digits as it needs to be less than $\frac{1}{2^i}$ and the next $n-i$ digits must be 1, for the number to be greater than $\frac{1}{2^i} - \frac{1}{2^n}$. The digits after the first n digits will not affect whether the number is between $\frac{1}{2^i}$ and $\frac{1}{2^i} - \frac{1}{2^n}$ or not. (The cases where the number is equal to $\frac{1}{2^i}$ and $\frac{1}{2^i} - \frac{1}{2^n}$ is not significant to our calculations as the probability of that occurring is 0.) Thus, $P(\frac{1}{2^i}) - P(\frac{1}{2^i} - \frac{1}{2^n})$ is the probability a number occurs in the form 0.0...01...1..., with i 0s followed by $n-i$ 1s.
Now, examine the derivative at $\frac{1}{2^i}$ approaching from the left side:

$$\lim_{n\to\infty} \frac{P(\frac{1}{2^i}) - P(\frac{1}{2^i} - \frac{1}{2^n})}{\frac{1}{2^n}} = 2^n[(1-p)^i p^{n-i}] = (\frac{1-p}{p})^i (2p)^n$$

Notice that when $p > \frac{1}{2}$, $(2p)^n \to \infty$, and when $p = \frac{1}{2}$, $(2p)^n \to 1$, and when $p < \frac{1}{2}$, $(2p)^n \to 0$. Our derivative is only defined at $p = \frac{1}{2}$. $\square$

Having proved that $P(x)$ is not differentiable at the powers of 2, it is intuitive that the graph will not be differentiable at dyadic rationals. We prove this formally below:

Theorem 5. $P(x)$ is not differentiable at $x = \frac{k}{2^i}$ for values of $p \neq \frac{1}{2}$.
Proof: Let $\frac{k}{2^i}$ be in simplest form (for simplicity of calculations below). Notice that $\frac{k}{2^i}$ has a 1 as its i-th digit in its binary expansion, and $\frac{k}{2^i} + \frac{1}{2^n}$ appends $n-i-1$ 0s to $\frac{k}{2^i}$ followed by a 1. Any number between $\frac{k}{2^i}$ and $\frac{k}{2^i} + \frac{1}{2^n}$ must be larger than $\frac{k}{2^i}$ so its first digits will be those of $\frac{k}{2^i}$. It must be less than $\frac{k}{2^i} + \frac{1}{2^n}$, so its remaining digits must be $n-i$ 0s. Note that the digits after the first n digits do not affect whether the number satisfies our conditions. Below, we are going to refer to h as the number of 1s and t as the number of 0s in the binary expansion of $\frac{k}{2^i}$.
The derivative at $\frac{k}{2^i}$ approaching from the right side is

$$\lim_{n\to\infty} \frac{P(\frac{k}{2^i} + \frac{1}{2^n}) - P(\frac{k}{2^i})}{\frac{1}{2^n}} = 2^n[p^h(1-p)^{n-h}] = (\frac{p}{1-p})^h * (2-2p)^n$$

Notice that when $p < \frac{1}{2}$, $(2-2p)^n \to \infty$ and when $p > \frac{1}{2}$, $(2-2p)^n \to 0$, and when $p = \frac{1}{2}$, $(2-2p)^n \to 1$.
Now, note that $\frac{k}{2^i} - \frac{1}{2^n}$ has the same first $i-1$ digits as $\frac{k}{2^i}$ followed by a 0 and $n-i$ 1s. Numbers between $\frac{k}{2^i} - \frac{1}{2^n}$ and $\frac{k}{2^i}$ must have the same first $i-1$ digits as $\frac{k}{2^i}$ in order to be between $\frac{k}{2^i} - \frac{1}{2^n}$ and $\frac{k}{2^i}$. Further, the digits following the first $i-1$ digits must be a 0 followed by $n-i$ 1s to be bigger than $\frac{k}{2^i} - \frac{1}{2^n}$. The digits after the first n digits do not affect whether the number is between $\frac{k}{2^i} - \frac{1}{2^n}$ and $\frac{k}{2^i}$ or not.

Now, examine the derivative at $\frac{k}{2^i}$ approaching from the left side:

$$\lim_{n\to\infty} \frac{P(k/2^i) - P(k/2^i - \frac{1}{2^n})}{\frac{1}{2^n}} = 2^n[(1-p)^{t+1}p^{n-t-1}] = (\frac{1-p}{p})^{t+1}(2p)^n$$

Notice that when $p > \frac{1}{2}$, $(2p)^n \to \infty$, and when $p = \frac{1}{2}$, $(2p)^n \to 1$, and when $p < \frac{1}{2}$, $(2p)^n \to 0$. Our derivative is only defined at $p = \frac{1}{2}$. We have shown that there are infinitely many dyadic rationals $\frac{k}{2^m}$ in the interval $[0, 1]$ where $P(x)$ is not differentiable, so these numbers are dense on $[0, 1]$. Furthermore, because the cumulative distribution function is not differentiable, there is no function that we can integrate over to obtain the function. This is because the cumulative distribution function is algebraically found by integrating over the probability density function, this means that $P(x)$ has no closed-form probability density function. □

5.3 Arc Length

Estimating the arc length of a curve can be done by partitioning the curve into segments, each of which are approximated by a straight line, and adding up the lengths of the straight lines, like in the diagram below.

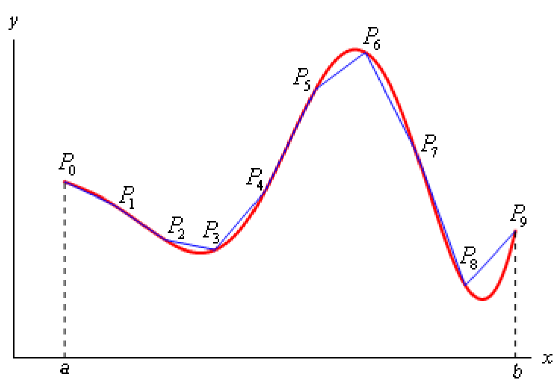

Figure 3: Example of calculating arc length of the graph of a function.

For a single straight line segment, a simple application of Pythagoras' theorem will yield the length of the line segment, ℓ:

$$\ell = \sqrt{(\Delta x)^2 + (\Delta y)^2} = \sqrt{(x_2 - x_1)^2 + (f(x_2) - f(x_1))^2}$$

The arc length of our pathological function will be difficult to estimate, but the intuition is that with a small enough Δx, i.e., an infinitesimally small interval, the graph at this point can be estimated with a straight line segment, and adding up these infinitesimally small line segments will yield an estimate for the arc length.

Theorem 6. The arc length of $P(x)$ is approximated as $\frac{1}{2^m}\sum_{\lambda=0}^{m}\binom{m}{\lambda}\sqrt{1 + (2^m p^\lambda (1-p)^{m-\lambda})^2}$, where m is the number of coin flips so far (analogously, the number of partitions of the interval $[0, 1]$), λ is the number of heads flipped.

Proof: Consider the interval at the m-th level of binary expansion (where $m \in \mathbb{N}$ is very large), meaning the interval size is $\frac{1}{2^m}$. As $m \to \infty$, meaning the number of coin flips or analogously, the number of partitions of the interval $[0, 1]$, becomes very large, the approximation of the arc length improves. Now consider the segment of the graph between the k-th and $k+1$-th interval, i.e. $0 \leq \frac{k}{2^m} < \frac{k+1}{2^m} \leq 1$. Given the estimation of arc length using Pythagoras' theorem above, we can estimate the arc length in the interval $[\frac{k}{2^m}, \frac{k+1}{2^m}]$ as follows:

$$\begin{aligned}
\ell_{k,k+1} &= \sqrt{\left(\frac{k+1}{2^m} - \frac{k}{2^m}\right)^2 + \left(P\left(\frac{k+1}{2^m}\right) - P\left(\frac{k}{2^m}\right)\right)^2} \\
&= \sqrt{\left(\frac{1}{2^m}\right)^2 + \left(P\left(\frac{k+1}{2^m}\right) - P\left(\frac{k}{2^m}\right)\right)^2} \\
&= \frac{1}{2^m}\sqrt{1 + 2^{2m}\cdot\left(P\left(\frac{k+1}{2^m}\right) - P\left(\frac{k}{2^m}\right)\right)^2}
\end{aligned}$$

Summing this up over all the intervals, we get

$$\ell = \frac{1}{2^m}\sum_{k=0}^{2^m-1}\sqrt{1 + 2^{2m}\cdot\left(P\left(\frac{k+1}{2^m}\right) - P\left(\frac{k}{2^m}\right)\right)^2}$$

Now, one might ask how we determine $P\left(\frac{k+1}{2^m}\right) - P\left(\frac{k}{2^m}\right)$, seeing as we do not have a closed form for the cumulative distribution function due to its pathological nature. Consider an

alternative interpretation of the intervals where binary expansions representing the outcomes of the m coin flips lie in these intervals. The difference $P\left(\frac{k+1}{2^m}\right)-P\left(\frac{k}{2^m}\right)$ gives the probability of being in the binary interval $[\frac{k}{2^m}, \frac{k+1}{2^m}]$, which determines the binary expansion up to the m-th digit. The binary expansions that lie within the $[\frac{k}{2^m}, \frac{k+1}{2^m}]$ interval with λ heads and $m-\lambda$ tails have probability $p^\lambda(1-p)^{m-\lambda}$. Furthermore, the choices of intervals for different values of k exhaust all possible binary sequences of length m. Therefore, the binomial coefficient $\binom{m}{\lambda}$ determines the number of binary expansions with λ many heads (or 1s), each corresponding to probability $p^\lambda(1-p)^{m-\lambda}$.
We can therefore rewrite this expression using a binomial coefficient as follows:

$$\begin{aligned}\ell &= \frac{1}{2^m}\sum_{k=0}^{2^m-1}\sqrt{1+2^{2m}\cdot\left(P\left(\frac{k+1}{2^m}\right)-P\left(\frac{k}{2^m}\right)\right)^2}\\ &\equiv \frac{1}{2^m}\sum_{\lambda=0}^{m}\binom{m}{\lambda}\sqrt{1+(2^m p^\lambda(1-p)^{m-\lambda})^2}\end{aligned}$$

To verify this with the case of the fair coin (i.e., $p=\frac{1}{2}$), which is the only case with a non-pathological cumulative distribution function, we saw previously that the cumulative distribution function is a straight line with gradient 1 since the probability distribution function is a uniform distribution function. In this case, the arc length is $\sqrt{1^2+1^2}=\sqrt{2}$. Indeed, substituting in the value $p=\frac{1}{2}$ into the formula we derived above, we get

$$\begin{aligned}\ell &= \frac{1}{2^m}\sum_{\lambda=0}^{m}\binom{m}{\lambda}\sqrt{1+\left(2^m\cdot\frac{1}{2}^\lambda\cdot\left(1-\frac{1}{2}\right)^{m-\lambda}\right)^2}\\ &= \sqrt{1+\left(2^m\cdot\frac{1}{2}^\lambda\cdot\left(1-\frac{1}{2}\right)^{m-\lambda}\right)^2} \quad (\text{since } \sum_{\lambda=0}^{m}\binom{m}{\lambda}=2^m)\\ &= \sqrt{1+\left(2^m\cdot\frac{1}{2^m}\right)^2}\\ &= \sqrt{1+1}\\ &= \sqrt{2}\end{aligned}$$

Hence, this verifies the formula we derived for arc length in the one non-pathological case of the fair coin. □

6 Connection to Galton Board visualizations

We draw from the intuition of infinitesimally small binary intervals introduced in Section 2, the notion that any number of interest (representing a probability) lies somewhere on a number line with binary intervals $\frac{1}{2^n}$, and as the $n\to\infty$, the position of the number on the number line becomes increasingly specific. An alternative visualization is to think of each set binary intervals of size $\frac{1}{2^k}$ for various level numbers $k\geq 0$ to be a bucket, and a ball is dropped into the very first interval of size 1, representing an infinite coin flip.

As the ball descends through various levels of intervals, at each level $k+1$ with binary interval size $\frac{1}{2^{k+1}}$, the ball in the previous level k has a certain probability of being sorted into the left interval or right interval. Being sorted into the right interval is equivalent to flipping a heads (which has probability p) and appending a 1 to the growing binary expansion, and conversely being sorted into the left interval is equivalent to flipping a tails (which has probability $q=1-p$) and appending a 0 to the growing binary expansion. At some large level number, say level n, which approximates the infinite coin toss, the ball being dropped ends up in one of the binary intervals sized $\frac{1}{2^n}$.

Dropping a very large number of balls will result in a distribution of balls across the different buckets or intervals at level n, and this represents the probability distribution for a specific value of p. This is analogous to a Galton Board, illustrated below in Figure 5.

One important distinction is that for the case with "infinite" levels (or where the number of coin flips is infinitely many), the ball dropping experiment on the Galton board demonstrates the law of large numbers, specifically that

$$P(\text{ball ends up in interval } i_k)=\frac{\text{number of ball that ended up in interval } i_k}{\text{total number of balls dropped}}$$

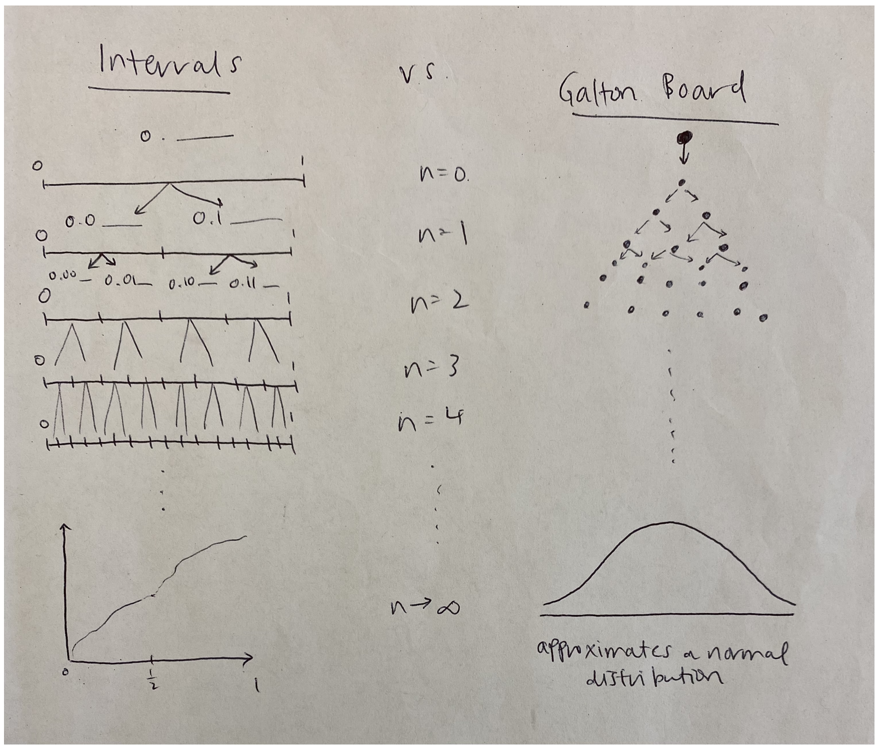

Figure 4: Comparing intervals and the Galton Board

Figure 5: A Galton Board, a device invented to demonstrate the law of large numbers, in particular that with sufficient sample size the binomial distribution approximates a normal distribution.

However, for our intervals visualization, we demonstrated in Section 2 that the probability distribution will be uniform. In a (physical) Galton board, the probability of a dropped ball falling to the left or right at each level is going to be equal (i.e., $p = \frac{1}{2}$), and it demonstrates that in the limit, such ball dropping experiments reveal a normal distribution, by the law of large numbers (when the number of rows in the Galton Board as well as the number of balls dropped are very large). One can also imagine a hypothetical Galton board for $p \neq \frac{1}{2}$, and determining how many balls fell in each interval as a proportion of the total number of balls dropped will give an approximation of the probability of falling in each interval, as well as a visualization of the probability distribution overall.

7 Future Work

From our observations about the properties of the graphs of cumulative distribution functions for infinite tosses of the unfair coin, we noted that there were kinks in the graph at the dyadic rationals (i.e., binary expansions with with finite number of 1 digits), which suggests that the cumulative distribution functions are not differentiable at the binary numbers. We give a proof for non-differentiability at the dyadic rationals in Section 5.2, but we leave a rigorous proof of why the function is non-differentiable at numbers other than the dyadic numbers for future work.

One potential application of this problem that can be explored in the future is its applicability to betting games, since, like a coin flip, a person could have a certain probability p of winning a bet, and would like to decide betting amounts such that (in expectation) one could maximize earnings. Relating back to Equation 3, a possible way to frame this function is in decision making for betting money: let one's initial wealth be represented by x, which lies in the interval $[0, 1]$, and the probability of losing any arbitrary bet be p (and correspondingly, the probability of winning any arbitrary bet is $1 - p$). If one wins a bet, their current wealth doubles. The aim of the betting game is to reach a wealth value of 1, where the probability of winning given wealth x is given by $P(x)$ and represented by the same function as in Equation 3.

If one's initial wealth $x \leq \frac{1}{2}$ (the first expression in Equation 3), then one must win the first time, otherwise one will go broke in the limit of infinite bets. Therefore, one wins the first time with probability $(1 - p)$ and then multiply that with the probability of winning starting from one's new wealth, which is now $2x$, and this probability if $P(2x)$. So we get $(1 - p) \cdot P(2x)$. In the second case, if one's initial wealth $x > \frac{1}{2}$, then one only needs $1 - x$ to reach 1, and that should be the amount that one bets. If one wins the first time then they are done, which has probability $(1 - p)$. One might lose the first time with probability p and can still win overall - losing the first time results in a wealth of $x - (1 - x) = 2x - 1$, and then one wins starting from the new wealth of $2x - 1$, which is $p \cdot P(2x - 1)$. So this second case can be represented by summing the two probabilities $(1 - p) + p \cdot P(2x - 1)$. Depending on whether $2x - 1$ is more or less than $\frac{1}{2}$, the player will be able to determine their probability of winning with either the first or second equation of the function.

Future work can investigate the applicability of this infinite coin toss problem to a variety of betting games with different rules - the betting game above specifically states that winning a bet doubles ones earnings, but other games with different rules could result in different optimal betting amounts depending on one's wealth. This game also has a fixed probability of winning throughout, whereas more dynamic games might have changing probabilities. More broadly, future work can explore the applicability of the infinite coin toss problem to modeling long sequences of decision making, where each decision has a certain probability, and used to optimize decisions made at each step of the process.