Choose between "digit" and "bit," and stick to it.

TOSSING A COIN

J.M. NÁTER, P. WEAR, M. COHEN

Authors are always listed alphabetically by 1st letter of last name

ABSTRACT. We study the properties of a function that takes $x \in [0, 1]$ as input and determines the probabilty that the number obtained by writing a decimal point and then tossing a coin infinitely many times, writing a 1 after the point when the outcome is heads and a 0 when the outcome is tails, is less than or equal to x.

No. See comments

1. INTRODUCTION

The result of n tosses of a two-headed coin can be represented by an n-digit binary number in the interval [0,1]. The kth digit is 0 if the kth toss comes up tails and 1 if it comes up heads. These representations correspond to rational numbers with denominators of the form 2^k for some k, a.k.a. dyadic rationals. Similarly, an infinite series of tosses gives us a binary representation of any real number in the interval [0,1]. Now let y be the outcome of an infinite toss. For any given real number $x \in [0, 1]$ we would like to determine the probability that $y \leq x$ and we denote this probabilty by $f_p(x)$ where $p \in (0, 1)$ is the probability that a coin toss comes up heads.

Not what you mean. Better: ... binary expansion that can represent any real number.

Indent ¶

For an idea of how to go about compute let us compute $f_p(\frac{1}{3})$. The binary expansion for $\frac{1}{3}$ is $.\overline{01}$. Now we consider the possible outcomes of an infinite sequence y of coin tosses. For $.\overline{01} \leq x$ the first must necessarily come up tails, which contributes $1 - p$ to the probability. If the second toss comes up tails the inequality is still satisfied, however if it comes up heads, for the rest of the inequality to be satisfied the remaining tosses must represent a number less than or equal to the remaining digits of $\frac{1}{3}$, which also have the form $.\overline{01}$. So then

"In order for y to be less than $.\overline{01}$, ..."

comma splice.

$$f_p\Big(\frac{1}{3}\Big) = (1 - p)(1 - p + pf_p\Big(\frac{1}{3}\Big).$$

Solving that equation we get

$$f_p\Big(\frac{1}{3}\Big) = \frac{(1 - p)^2}{p^2 - p + 1}.$$

Date: March 1, 2013.

my artifact

The follow images should provide some intuition about the behavior of f_p.

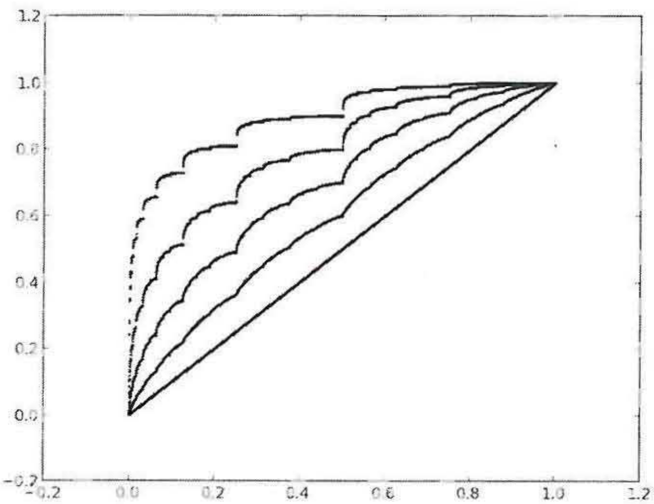
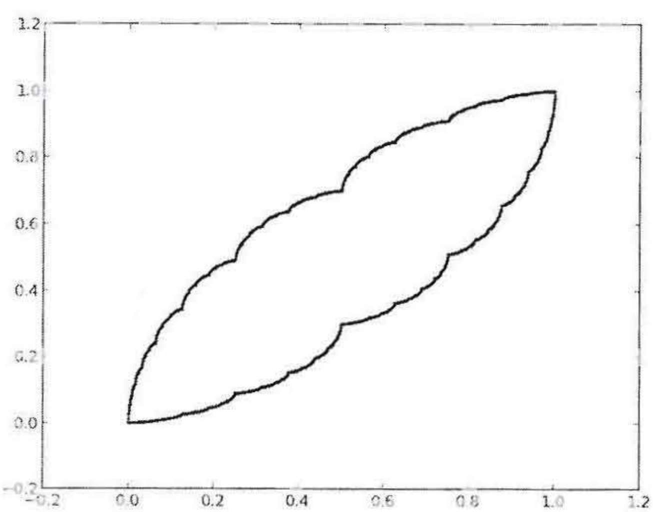


FIGURE 1. Graphs of f_p for $p = .1, .2, .3, .4, .5$



What is that relationship?

FIGURE 2. Graphs of f_p for $p = .3, .7$ to illustrate the relation between f_p and f_{1-p}.

For most values of p, the function f_p is pathological, but it has many interesting properties. In the following sections we prove continuity of f_p for $p \in (0,1)$, show that $f_p(x)$ is not nowhere-differentiable and give a definition of arc length for f_p.
Sections 1, 3 by J.M. Náter
Sections 2, 6 by P. Wear
Sections 4, 5 by M. Cohen

2. Continuity

Given a binary representation of some number $x \in [0,1]$, the mapping $x \mapsto \frac{x}{2}$ corresponds to inserting a 0 between the decimal point and the first digit of x. Similarly, $x \mapsto \frac{x}{2} + \frac{1}{2}$ corresponds to inserting a 1 between the decimal point and the first digit of x. We now introduce two functional equations that give us a method for evaluating f_p on any dyadic number. Given a dyadic x, for an infinite flip sequence to be less than $\frac{x}{2}$ the outcome of the first toss must be tails and the rest of the tosses must represent a number less than x. The probability of the first toss being tails is $(1-p)$ and the probability of the rest of the flips being smaller than x is $f_p(x)$, so we have

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

which immediately generalizes to $f_p(\frac{x}{2^k}) = (1-p)^k f_p(x)$. For the infinite toss sequence to give a number smaller than $\frac{x}{2} + \frac{1}{2}$ the first toss can come out either heads or tails. If it is tails the sequence will necessarily be smaller. If it is heads, then the rest of the sequence must give a number smaller than x, and so we have the second equation:

$$f_p\left(\frac{x}{2} + \frac{1}{2}\right) = 1 - p + pf_p(x). \tag{2}$$

These two functional equations allow us to calculate f_p for any dyadic number, since every such number can be represented by a finite binary sequence (preceded by a decimal point of course) ending in a 1, and so we can start with $f_p(.1) = (1-p)$ and keep iterating (1) and (2) depending on the bits until we reach the desired dyadic.

Now we are ready to prove continuity. We will use the two equations and monotonicity, which follows from the basic measure-theoretic argument that if $y > x$ the probability that a toss sequence is less than y cannot be less than the probability that a toss sequence is less than x.

Because we have monotonicity it suffices to show that for any x and any $\epsilon > 0$ there are numbers $y < x$ and $y' > x$ such that $f_p(x) - f_p(y) < \epsilon$ and $f_p(y') - f_p(x) < \epsilon$. Without loss of generality assume $p \geq 1 - p$.

For any $x \in (0,1)$ and for any positive integer N there exists $n > N$ such that the nth digit of x is 0. If this were not the case then there would be some point after which all the digits were 1, in which we could use the substitution $.0\overline{1} = .1\overline{0}$ to obtain the desired form. Now let $y' = x + 2^{-n}$, where the nth digit of x is 0. The only toss sequences which correspond to number smaller than y' but greater than x are those for which the first $n-1$ tosses agree with the first $n-1$ digits of x, so because $p \geq 1-p$ we have $f_p(y') - f_p(x) \leq p^{n-1}$. As n approaches infinity $f_p(y') - f_p(y)$ will approach 0, so given any $\epsilon > 0$ we can always choose an appropriate y'.

We can find $y < x$ similarly, as there will be infinitely many 1s in the binary expansion of x and in this case we want to choose a 1 arbitrarily far down the binary expansion and flip it to a 0. Continuity follows immediately.

3. Differentiability at $x = \frac{1}{3}$

Although a thorough characterization of the sets on which f_p is differentiable is not available yet, we at least know f_p is not nowhere-differentiable. We prove this by showing differentiabilty at $x = \frac{1}{3}$. First notice that the binary representation of $\frac{1}{3}$ is $.\overline{01}$, so that the probability that the outcome of $2n$ coin tosses matches the first $2n$ digits of $.\overline{01}$ is $p^n(1-p)^n$. Now denote the derivative limit $\lim_{h\to 0} \frac{f_p(x+h) - f_p(x)}{h}$ by $f_p'(x)$. As we did for continuity, we can choose a 0 arbitrarily far down the binary representation of $\frac{1}{3}$. Now let the $(2k+1)$th digit be 0, so that setting $h = \frac{1}{2^{2k+1}}$ and adding h to x will flip that digit to a 1. Then we can bound $f'(\frac{1}{3})$ by $2^{2k+1} \cdot (p(1-p))^k = 2 \cdot 4^k \cdot (p(1-p))^k$. Also notice by the inequality of arithmetic and geometric means we have

$$\frac{p + (1-p)}{2} \geq \sqrt{p(1-p)}$$

$$\frac{1}{2} \geq \sqrt{p(1-p)}$$

$$\frac{1}{4} \geq p(1-p)$$

Equality is achieved only for $p = \frac{1}{2}$ so assume $p \neq \frac{1}{2}$ and take the inequalities to be strict. So then $4 \cdot p(1-p) < 1$ and so

$$\lim_{k\to\infty} 2 \cdot (4p(1-p))^k = 0,$$

1) "Its total arc length" is not a question.
2) A function doesn't have an arclength

which, because as k approaches infinity h approaches 0, is equivalent to saying $f'(\frac{1}{3}) = 0$. In the case $p = \frac{1}{2}$ the function $f_{\frac{1}{2}}(x)$ is exactly the line $y = x$ which is also differentiable. — probably omit.

4. Defining arc length

An interesting question to ask about f_p is its total arc length. In order to rigorously investigate this, however, we will need an actual definition of arc length. The traditional definition of arc length, as seen in introductory calculus courses, is defined using the derivative of the function:

Definition 4.1. *Let f be a function defined and continuously differentiable on $[a, b]$. Then the arc length of f on $[a, b]$ is*

$$s = \int_a^b \sqrt{1 + f'(x)^2}\, dx \tag{3}$$

so don't give it

This definition clearly does not work for f_p, since f_p is undifferentiable on a dense set of points in its domain. However, there is a natural definition of arc length which applies to all functions (although it may be infinite). To introduce it, we must first define a partition:

Definition 4.2. *A partition P of the closed interval $[a, b]$ is a finite sequence of n points x_i satisfying $x_1 = a$, $x_n = b$, and $x_i \leq x_{i+1}$ for all i where both are defined. The fineness of P, $F(P)$, is defined as the largest value of $x_{i+1} - x_i$. $[a, b]$ is the set of all partitions of $[a, b]$.*

A partition can be viewed as a way to split $[a, b]$ into the subintervals $[x_i, x_{i+1}]$. Note that this notion of a partition is also used in the definition of Riemann integration. We define a notion of an approximate arc length using a partition:

Definition 4.3. *Let f be a function defined on $[a, b]$, and let P be a partition of $[a, b]$, consisting of x_i for $1 \leq i \leq n$. Then the P-length of f is:*

$$L_P(f) = \sum_{k=1}^{n-1} \sqrt{(x_{k+1} - x_k)^2 + (f(x_{k+1}) - f(x_k))^2} \tag{4}$$

The P-length essentially gives an approximate arc length, defined with the granularity given by the partition. It is the arc length that f would have if it consisted of a collection of line segments, each covering a segment from P, but with the correct value on the endpoints of each segment. We can now define the actual arc length:

This is a standard definition for the arc length of a curve, adapted to the case of the graph of a function. Nice development, not really part of this project; certainly needs a reference.

Definition 4.4. *Let f be a function defined on $[a,b]$. Then the arc length of f on $[a,b]$ is*

$$s = \sup_{P \in [a,b]} L_P(f) \tag{5}$$

The motivation for this definition is that the P-lengths define the lengths of arbitrarily fine approximations to f, but the P-lengths should always be at most the actual arc length (since lines are the shortest path between two points). In fact, this supremum is also a sort of limit:

Lemma 4.5. *Let f be a function defined on $[a,b]$, with finite arc length s defined according to 4.4. Then for any ϵ, there exists a δ such that for all partitions P with fineness at most δ, $|s - L_P| < \epsilon$.*

This lemma can be proved with a relatively simple bounding argument (essentially, given a P with arc length close to the supremum, all sufficiently fine partitions must have arc length almost that of P, while they are still bounded above by s). The detailed proof is omitted here, since it is not the focus of this paper. The lemma could be taken as giving an alternative, possibly more natural definition for the arc length of s; this definition is very similar to that of the Riemann integral.

Note that both of these definitions are equivalent to 4.1 for continuously differentiable functions. This can also be proved relatively simply (by showing that the value of $\sqrt{1+f'(x)^2}\Delta x$ is close to $\sqrt{(\Delta x)^2 + (\Delta y)^2}$ for sufficiently fine partitions). Again, the detailed proof is not given here.

Finally, consider that $\sqrt{(x_{k+1}-x_k)^2 + (f(x_{k+1}) - f(x_k))^2}$ is upper-bounded (by the triangle inequality) by $(x_{k+1}-x_k) + |f(x_{k+1}) - f(x_k)|$. In the special case when f is monotonically increasing, $f(x_{k+1}) - f(x_k)$ is always nonnegative, so we can drop the absolute value there: $\sqrt{(x_{k+1}-x_k)^2 + (f(x_{k+1}) - f(x_k))^2} \le (x_{k+1}-x_k) + (f(x_{k+1}) - f(x_k))$.

That can be used to bound $L_P(f)$ for any partition P of $[a, b]$:

$$
\begin{aligned}
L_P(f) &= \sum_{k=1}^{n-1} \sqrt{(x_{k+1} - x_k)^2 + (f(x_{k+1}) - f(x_k))^2} \\
&\leq \sum_{k=1}^{n-1} (x_{k+1} - x_k) + (f(x_{k+1}) - f(x_k)) \\
&= (\sum_{k=1}^{n-1} x_{k+1} - x_k) + (\sum_{k=1}^{n-1} f(x_{k+1}) - f(x_k)) \\
&= (x_n - x_1) + (f(x_n) - f(x_1)) \\
&= (b - a) + (f(b) - f(a))
\end{aligned} \tag{6}
$$

Since the arc length is the suprema of the L_P, that gives rise to the following lemma:

Lemma 4.6. *Let f be a monotonically increasing function defined on $[a, b]$. Then the arc length of f is at most $(b - a) + (f(b) - f(a))$, and in particular is finite.*

5. Arc length of f_p

We now have the machinery to investigate the arc length of the f_p on $[0, 1]$. For the special case of $p = \frac{1}{2}$, the arc length is clearly just $\sqrt{2}$, since it is a straight line. For other values of p, we still know that f_p is monotonically increasing, and that $f_p(0) = 0$ and $f_p(1) = 1$. Then by 4.6 the arc lengths must be at most 2.

In this section, we will show that that bound is in fact tight: the arc length of f_p is 2. This, on its face, is somewhat surprising. Despite the fact that f_p is continuous, its arc length is the same as it would be if it were a monotonic step function covering the same range.

In fact, the proof can be interpreted as showing that f_p is "almost a step function" in that it can be broken down into intervals which are mostly completely flat, but where the actual increase of f_p mostly happens over intervals that are very steep, almost vertical.

We will lower bound the P_n-lengths for particular partitions P_n, where P_n consists of the points $x_i = \frac{i-1}{2^n}$ for $1 \leq i \leq 2^n + 1$. These P_n have the property that $x_{i+1} - x_i$ is always $\frac{1}{2^n}$: they divide $[0, 1]$ into 2^n equal segments. To obtain bounds, we will estimate the distribution of $f(x_{i+1}) - f(x_i)$.

The x_i (for $1 \leq i \leq 2^n$) are precisely those numbers whose binary expansion is all zeroes after the first n places after the decimal point.

To examine $f(x_{i+1}) - f(x_i)$ we define the function

$$D(y,m) = f\left(y + \frac{1}{2^m}\right) - f(x) \tag{7}$$

so that $D(x_i, n) = f(x_{i+1}) - f(x_i)$. D satisfies the following:

Lemma 5.1. *For all nonnegative integers m, all y in $[0,1)$ such that $2^m y$ is an integer, $D(y,m)$ is $p^a(1-p)^b$, where a is the number of ones in the binary expansion of y (up to the mth place) and b is the number of zeroes.*

Proof. We will prove this by induction on m. If $m = 0$, it is trivial: y must be 0, and $D(0,0) = f_p(1) - f_p(0) = 1 = p^0(1-p)^0$, as expected.

For $m > 0$, we will use the functional equations (given in the introduction) that apply for all x in $[0,1]$: $f_p(\frac{x}{2}) = (1-p)f_p(x)$, and $f_p(\frac{1}{2} + \frac{x}{2}) = 1 - p + pf_p(x)$.

First, note that if y is in $[0, \frac{1}{2})$, $y + \frac{1}{2^m}$ is in $[0, \frac{1}{2}]$ (because both of them, when multiplied by 2^m, are integers and they differ by 1; they can't skip over the integer 2^{m-1}). Otherwise, both must be in $[\frac{1}{2}, 1]$. The former case corresponds precisely to the first bit after the decimal place being 0, and the latter corresponds to it being 1.

- In the former case, we can apply the first functional equation with $x = 2y$ and $x = 2\left(y + \frac{1}{2^m}\right)$ to get $f_p(y) = (1-p)f_p(2y)$ and $f_p\left(y + \frac{1}{2^m}\right) = (1-p)f_p\left(2y + \frac{1}{2^{m-1}}\right)$. $f_p\left(y + \frac{1}{2^m}\right) - f_p(y)$ then comes out to $(1-p)M(2y, m-1)$. Replacing y by $2y$ and m by $m-1$ is precisely stripping the leading 0 from the binary expansion, while otherwise keeping the numbers of zeroes and ones up to the mth place the same. The requirements for the lemma are preserved. Thus, if the lemma holds for $m-1$, $M(2y, m-1)$ will be $p^a(1-p)^{b-1}$, so $M(y,m)$ will be $p^a(1-p)^b$, satisfying the lemma.
- The latter case is similar. Here, we apply the second functional equation with $x = 2y - 1$ and $x = 2\left(y + \frac{1}{2^m}\right) - 1$, getting $f_p(y) = 1 - p + pf_p(2y-1)$ and $f_p\left(y + \frac{1}{2^m}\right) = 1 - p + pf_p(2y - 1 + \frac{1}{2^{m-1}})$. $f_p\left(y + \frac{1}{2^m}\right) - f_p(y)$ then comes out to $pM(2y-1, m-1)$. Replacing y by $2y-1$ and m by $m-1$ is stripping the leading 1 but otherwise keeping the bits the same, and the requirements for the lemma are again preserved. Thus, if the lemma holds for $m-1$, $M(2y-1, m-1)$ will be $p^{a-1}(1-p)^b$, so $M(y,m)$ will again be $p^1(1-p)^b$, again satisfying the lemma.

The lemma then holds for $m = 0$ and holds for m if it holds for $m-1$, so by induction it holds for all m. □

This lemma implies that $f(x_{i+1}) - f(x_i)$ is $p^a(1-p)^b$, where a is the number of ones and b the number of zeroes in the binary expansion of x_i, up to the nth place. If we define

$$d_k = \begin{cases} p & \text{if the } k\text{th bit in the binary expansion of } x_i \text{ is } 1 \\ 1-p & \text{if the } k\text{th bit in the binary expansion of } x_i \text{ is } 0 \end{cases} \tag{8}$$

then we can alternatively write

$$f(x_{i+1}) - f(x_i) = \prod_{k=1}^{n} d_k \tag{9}$$

We can then get

$$\log_2(f(x_{i+1}) - f(x_i)) = \sum_{k=1}^{n} \log_2 d_k \tag{10}$$

We will now look at x_i as a random variable, with i chosen uniformly out of the integers from 1 to 2^n. It is important to note that each digit in the binary expansion of x_i is independent of all the rest, so the d_k (and $\log_2 dk$) are independent random variables. Furthermore, each of d_k (and each of $\log_2 d_k$) has the same distribution (since the probability of each bit being 0 is always $\frac{1}{2}$). We let μ be the mean value of $\log_2 d_k$ and σ^2 be the variance. Note that the probability distribution of an individual d_k does not depend on n, so neither do μ or σ. Since the probability of picking each value is $\frac{1}{2}$,

$$\begin{aligned} \mu &= \frac{1}{2}(\log_2 p + \log_2(1-p)) \\ &= \log_2 \sqrt{p(1-p)} \\ &< \log_2 \frac{1}{2} \text{ (by AM-GM inequality)} \\ &= -1 \end{aligned} \tag{11}$$

Since $\mu < -1$, we can then pick some real number r such that $\mu < r < -1$. We will take any such r (again, not depending on n).

We need not calculate σ^2 explicitly; what is important is that it is constant over choice of n and that it is finite (since it applies to a discrete probability distribution).

Since $\log_2(f(x_{i+1}) - f(x_i))$ is the sum of n independent instances of the same probability distribution, it has mean $n\mu$ and variance $n\sigma^2$. Then we can apply Chebyshev's inequality to bound the probability that $\log_2(f(x_{i+1}) - f(x_i)) > nr$: Chebyshev's inequality says this probability is at most

$$\frac{n\sigma^2}{(nr - n\mu)^2} = \frac{1}{n} \cdot \frac{\sigma^2}{r - \mu} \tag{12}$$

Then for any $\epsilon > 0$, there exists an N such that if $n \geq N$, that probability will be at most $\frac{\epsilon}{2}$: we can simply set N to $\frac{2}{\epsilon} \cdot \frac{\sigma^2}{r-\mu}$.

Notably, exponentiating both sides shows that this is actually bounding the probability that $f(x_{i+1}) - f(x_i) > 2^{nr}$. Since $r < -1$, $\lim_{n \to -\infty} 2^{n(r+1)} = 0$. Applying the definition of a limit, this means that for any $\epsilon > 0$, there exists an N' such that if $n \geq N'$, $2^{n(r+1)} < \frac{\epsilon}{2}$.

Given any $\epsilon > 0$, we will then pick n as $\max(N, N')$. We divide the i (for i from 1 to 2^n) into "good" and "bad" values: "good" values satisfy $f(x_{i+1}) - f(x_i) \leq 2^{nr}$ while "bad" ones do not. For each "good" i,

$$\begin{aligned} f(x_{i+1}) - f(x_i) &\leq 2^{nr} \\ &= 2^{-n} \cdot 2^{n(r+1)} \\ &< \frac{\epsilon}{2} 2^{-n} \end{aligned} \tag{13}$$

Since there are only 2^n values of i, summing this over all good i gives less than $\frac{\epsilon}{2}$. On the other hand, summing $f(x_{i+1}) - f(x_i)$ over all i gives $f(x_{2^n+1}) - f(x_1) = 1$. Thus the sum of $f(x_{i+1}) - f(x_i)$ over all bad i gives $> 1 - \frac{\epsilon}{2}$. Furthermore, $\sqrt{(x_{i+1} - x_i)^2 + (f(x_{i+1}) - f(x_i))^2} \geq f(x_{i+1}) - f(x_i)$ by the triangle inequality, so the sum of $\sqrt{(x_{i+1} - x_i)^2 + (f(x_{i+1}) - f(x_i))^2}$ over all bad i is greater than $1 - \frac{\epsilon}{2}$.

Since all i were chosen with equal probability, the number of bad i is equal to 2^n times the probability than an i is bad, which is less than $\frac{\epsilon}{2}$, so this number is less than $2^n \frac{\epsilon}{2}$. Then the number of good i is greater than $2^n(1 - \frac{\epsilon}{2})$. Since $x_{i+1} - x_i = 2^{-n}$, $\sqrt{(x_{i+1} - x_i)^2 + (f(x_{i+1}) - f(x_i))^2}$ is always at least 2^{-n} for any i, so the sum of this over all x_i is at least $1 - \frac{\epsilon}{2}$. Then the sum of this over all i, good and bad, is at least $2 - \epsilon$.

This sum is precisely the L_P. Thus, for any $\epsilon > 0$, the arc length must be at least $2 - \epsilon$; thus the arc length must be at least 2. Since it cannot be > 2, it must equal 2.

Theorem 5.2. *The arc length of f_p, for any $p \neq \frac{1}{2}$, on $[0, 1]$, is 2.*

Better to avoid cutenesses like a.k.a.

6. Further Possibilities

A natural extension of this question is to consider n-sided coins a.k.a. dice. Many of the results from this paper can be generalized to dice with an arbitrary number of sides, but the graphs of the resulting functions become even more complex. One interesting case arises when we take a 3-sided coin such that the probabilities of two of the faces are $1/2$ each and the probability of the third face is 0. This gives the Cantor function a.k.a. the Devil's staircase, as we are essentially converting binary numbers to trinary.

Reference (

"ternary" is better

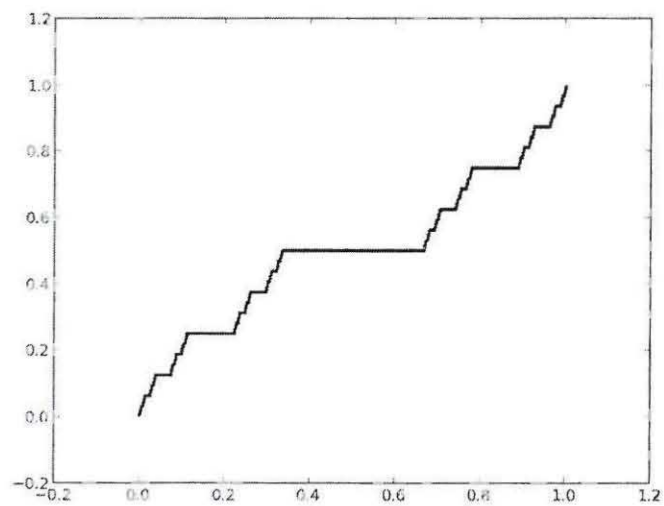


Figure 3. The Devil's staircase.