TOSSING A COIN

M. COHEN, J.M. NÁTER, P. WEAR

ABSTRACT. We can associate any sequence of tosses of an unfair coin with a binary number between 0 and 1. In this paper, we explore some properties of the cumulative probability distribution of this random variable.

1. INTRODUCTION

The result of n tosses of a two-headed coin can be represented by an n-bit binary number in the interval [0,1]. The kth bit 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, also known as the dyadic rationals. Similarly, an infinite sequence of tosses gives us a binary representation that can represent 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. The following image should provide some intuition about the behavior of f_p.
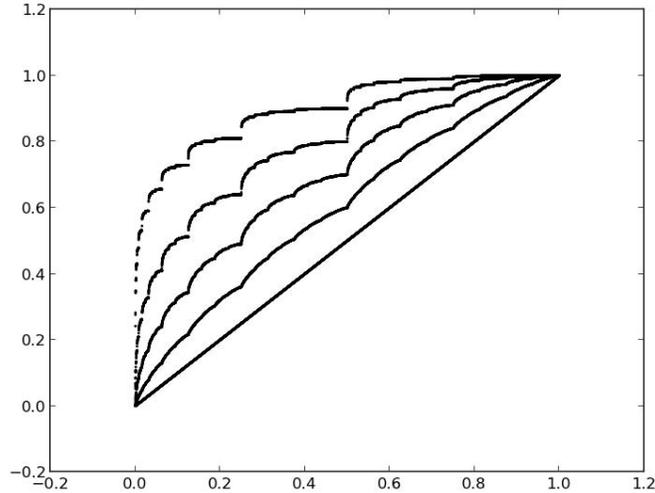
FIGURE 1. Graphs of f_p for $p = .1, .2, .3, .4, .5$

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 differentiable at the dyadics but not nowhere-differentiable and find the arc length of f_p.

Sections 1, 3 by J.M. Náter
Sections 2, 6 by P. Wear
Sections 4, 5 by M. Cohen

2. BASIC PROPERTIES

In this section, we go over some of the basic properties of f_p. We will show that f_p is monotonically increasing, give a method of computing rational values of f_p, define some useful functional equations describing f_p, and use these equations to find the integral of f_p and show continuity.

One special case to consider comes when $p = \frac{1}{2}$. We can see that $f_{\frac{1}{2}}(x) = x$ in this case. The coin is fair, so the probability that a set of flips is less than x is equal to the probability that a random number in $[0, 1]$ is less than x. This is just x, so we have the desired equality. In the rest of the paper, we will assume that $p \neq \frac{1}{2}$ unless we specifically state otherwise.

Proposition 2.1. *For all $p \in (0, 1)$, f_p is monotonically increasing.*

Proof. We claim that if $y > x$ then $f_p(y) > f_p(x)$. We can see that $f_p(y) - f_p(x)$ is equal to the probability that a set of flips corresponds to a number between y and x. This probability is positive and non-zero, so we have the desired property. $\square$

Proposition 2.2. *For any $p \in (0, 1)$, f_p is a 180 degree rotation of f_{1-p} around the point $(1/2, 1/2)$. To write this more mathematically, $f_p(x) = 1 - f_{1-p}(1 - x)$.*

Proof. $f_p(x)$ is the probability that a sequence of tosses of an unfair coin will correspond to a number less than x. The probability that a sequence of tosses will correspond to a number greater than x will then be $1 - f_p(x)$. If we switch from the probability of heads being p to being $1 - p$, then the probability of a sequence being less than x becomes the probability of a sequence being greater than $1 - x$. So $1 - f_{1-p}(1 - x)$ will be equal to $f_p(x)$ and we have the desired symmetry. $\square$
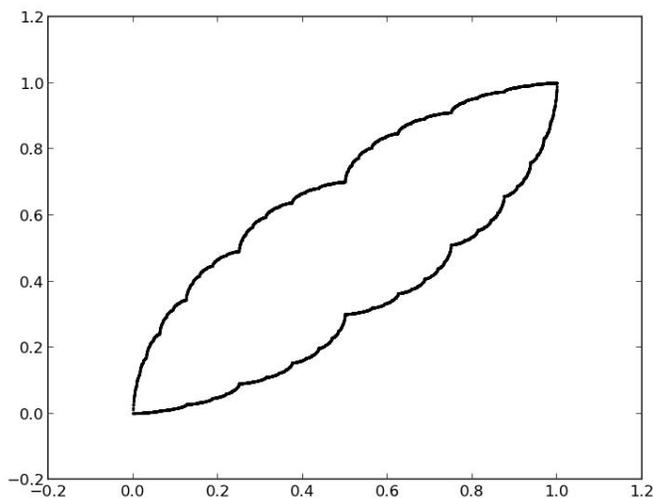
FIGURE 2. Graphs of f_p for $p = .3, .7$ to illustrate the relation between f_p and f_{1-p}.

We will give a computation of $f_p(\frac{1}{3})$. This method can be seen to extend to any non-dyadic rational, as they will all have repeating binary expansions. The binary expansion for $\frac{1}{3}$ is $.\overline{01}$. We consider the possible outcomes of an infinite sequence y of coin tosses. In order for y to be less than $.\overline{01}$, the first flip must come up tails. This will happen with probability $1 - p$. 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 bits of $\frac{1}{3}$, which also have the form $.\overline{01}$. So then

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

Solving this equation we get

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

We now introduce two functional equations that give us a method for evaluating f_p on any dyadic number. Given a dyadic x in $[0,1]$, 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

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

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:

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

Every dyadic 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.

These functional equations can also be used to find the integral of the functions f_p and prove that each function is continuous.

Proposition 2.3. *We have the equality* $\displaystyle\int_0^1 f_p(x)dx = 1-p.$

Proof. Using equation (1) on the entire interval $[0,1]$ will contract f_p onto the interval $[0,1/2]$ and using equation (2) will contract f_p onto $[1/2,1]$. So

$$\int_0^1 f_p(x)dx = \frac{1}{2}\left(\int_0^1 (1-p)f_p(x)dx + \int_0^1 1-p+pf_p(x)dx\right).$$

Combining these two integrals gives

$$\int_0^1 f_p(x)dx = \frac{1}{2}\int_0^1 1 - p + f_p(x)dx.$$

This leads to the equality $\frac{1}{2}\int_0^1 f_p(x)dx = \frac{1-p}{2}$, so $\int_0^1 f_p(x)dx = 1-p$ as desired. $\square$

We end this section with a proof of continuity.

Theorem 2.4. *For all $p \in (0,1)$, the function f_p is continuous on the interval $[0,1]$.*

Proof. 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 bit of x is 0. If this were not the case then there would be some point after which all the bits were 1, in which we could use the substitution $.0\bar{1} = .1\bar{0}$ to obtain the desired form.

Now let $y' = x + 2^{-n}$, where the nth bit of x is 0. Any toss sequences corresponding to a number smaller than y' but greater than x will agree with the first $n - 1$ bits 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. $\square$

3. Non-Differentiability at dyadic points

Here we prove is not differentiable at the dyadic points.

Proposition 3.1. *If $x \in (0,1)$ is a dyadic then the limit*

$$f_p'(x) = \lim_{h \to 0} \frac{f_p(x+h) - f_p(x)}{h}$$

does not exist.

Proof. We use the fact that any dyadic number can be represented by starting with $f_p(.1) = (1-p)$ and then iterating between (1) and (2). Let x be a dyadic whose last nonzero bit is in the kth place and choose $h = \frac{1}{2^n}$. Also, AWLOG $p > 1 - p$. Using the iteration process to compute $f_p(x+h)$, if we stop the process after having $n - k + 1$ steps

(where the first step is $f(.1) = 1 - p$) we have $q + p(1-p)^{n-k}$. Starting from the end, this covers the bits up to and including the last nonzero bit of x. The rest of the iterations will multiply this quantity by a factor of at least $(1-p)^k - 1$. The rest of the terms are shared by $f_p(x+h)$ and $f_p(x)$. So then we know the difference

$$f_p(x+h) - f_p(x) = (1-p)^{k-1}(1 - p + p(1-p)^{n-k}) - (1-p)^{k-1}$$

$$= p(1-p)^{k-1} + (1-p)^{n-1}.$$

Since the limit

$$\lim_{n \to \infty} 2^n \cdot (p(1-p)^{k-1} + (1-p)^{n-1})$$

does not converge $f_p'(x)$ does not exist.

$\square$

4. DIFFERENTIABILITY AT $x = \frac{1}{3}$

Despite f_p not being differentiable on a dense set of its domain we know it is not nowhere-differentiable. We show this by computing its derivative at the point $x = \frac{1}{3}$. 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 $2n$ bits of $.\overline{01}$ is $p^n(1-p)^n$. Now consider the definition of the derivative $f_p'(x) = \lim_{h \to 0} \dfrac{f_p(x+h) - f_p(x)}{h}$. As in the proof of continuity, we can choose a 0 arbitrarily far down the binary representation of $\frac{1}{3}$. In particular, we know the odd bits in the binary representation of $\frac{1}{3}$ are equal to 0. So then choosing $h = \frac{1}{(2k+1)}$ and adding it to $\frac{1}{3}$ will flip the $(2k+1)$th bit 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

$$(3) \qquad\qquad \frac{1}{4} \geq p(1-p)$$

Equality is achieved only for $p = \frac{1}{2}$, but recall we are not considering that case so we can take the inequalities to be strict. So then $4 \cdot p(1 - p) < 1$ and so

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

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.

5. DEFINING ARC LENGTH

An interesting question to ask about f_p is: what is the arc length of its graph? In order to rigorously investigate this, however, we will need an actual definition of arc length. To introduce one, we must first define a partition:

Definition 5.1. *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. $\mathrm{Part}[a, b]$ 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 5.2. *Let c be a (vector-valued) function (parametrizing a curve) 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 c is:*

$$(4) \qquad L_P(c) = \sum_{k=1}^{n-1} |c(x_{k+1}) - c(x_k)|$$

The P-length essentially gives an approximate arc length, defined with the granularity given by the partition. It is the arc length that c 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:

Definition 5.3. *Let c be a parametrization of a curve defined on $[a, b]$. Then the arc length of c on $[a, b]$ is*

$$(5) \qquad s = \sup_{P \in \mathrm{Part}[a,b]} L_P(c)$$

The motivation for this definition is that the P-lengths define the lengths of arbitrarily fine approximations to c, 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 5.4. *Let c be a parametrization of a curve defined on $[a, b]$, with finite arc length s defined according to 5.3. Then for any ϵ, there exists a δ such that for all partitions P with fineness at most δ, $|s - L_P| < \epsilon$.*

In other words, not only do there exist partitions with arc lengths arbitrarily close to s, but *all* sufficiently fine partitions have arc lengths arbitrarily close to s. The lemma can be proved with a relatively

simple bounding argument; 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. It will not be used for the remainder of this paper, but is mentioned because it justifies using the supremum definition which will be very convenient.

In this paper, we are specifically considering the arc lengths of the graphs of functions. The graph of a function f on $[a, b]$ can be parametrized as

$$(6) \qquad c(x) = \langle x, f(x) \rangle$$

The P-length for a partition consisting of x_i is then

$$(7) \qquad L_P(c) = \sum_{k=1}^{n-1} |\langle x_{k+1} - x_k, f(x_{k+1}) - f(x_k) \rangle|$$

We can use structural properties of a function to bound the arc length of its graph on an interval. Consider that $|\langle x_{k+1} - x_k, f(x_{k+1}) - f(x_k) \rangle|$ is bounded above (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: $|\langle x_{k+1} - x_k, f(x_{k+1}) - f(x_k) \rangle| \leq (x_{k+1} - x_k) + (f(x_{k+1}) - f(x_k))$. That can be used to bound $L_P(f)$ for any partition P of $[a, b]$:

$$
\begin{aligned}
L_P(f) &= \sum_{k=1}^{n-1} |\langle x_{k+1} - x_k, f(x_{k+1}) - f(x_k) \rangle| \\
&\leq \sum_{k=1}^{n-1} (x_{k+1} - x_k) + (f(x_{k+1}) - f(x_k)) \\
&= \left(\sum_{k=1}^{n-1} x_{k+1} - x_k \right) + \left(\sum_{k=1}^{n-1} f(x_{k+1}) - f(x_k) \right) \\
&= (x_n - x_1) + (f(x_n) - f(x_1)) \\
&= (b - a) + (f(b) - f(a))
\end{aligned}
$$

(8)

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

Lemma 5.5. *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.*

6. ARC LENGTH OF f_p

We now have the machinery to investigate the arc length of the f_p on $[0, 1]$. There are two quite distinct cases: $p = \frac{1}{2}$ and $p \neq \frac{1}{2}$. In the former case, the arc length is clearly just $\sqrt{2}$, since it is a straight line. The remainder of this section will assume that $p \neq \frac{1}{2}$.

For these other values of p, we don't have such an immediately obvious answer. However, we do know that f_p is monotonically increasing, and that $f_p(0) = 0$ and $f_p(1) = 1$. Then by 5.5 the arc lengths must be at most 2.

In this section, we will show that that bound is in fact tight.

Theorem 6.1. *The arc length of the graph of f_p, for any $p \neq \frac{1}{2}$, on $[0, 1]$, 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 particular, the graph of f_p can be broken down into segments where almost all of the increase in x is covered by segments that are nearly horizontal, but almost all of the increase of $f_p(x)$ happens over intervals that are very steep, almost vertical.

We will prove a lower bound on the P_n-lengths for particular partitions P_n. The partition P_n consists of the points $x_i = \frac{i-1}{2^n}$ for $1 \leq i \leq 2^n + 1$. These partitions 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 $f_p(x_{i+1}) - f_p(x_i)$: that is, the increase in f_p over each segment of the partition.

The x_i (for $1 \leq i \leq 2^n$) are precisely those numbers in $[0, 1]$ whose binary expansion is all zeroes after the first n places after the decimal point, because they are obtained by dividing integers by 2^n (which shifts the binary expansion by n places to the right).

It turns out that we can give a quite simple description of the value of $f_p(x_{i+1}) - f_p(x_i)$:

Lemma 6.2. *For all i satisfying $1 \leq i \leq 2^n$, let a be number of ones in the binary expansion of x_i (up to the nth place) and b the number of zeroes. Then $f_p(x_{i+1}) - f_p(x_i) = p^a(1-p)^b$.*

Proof. The proof will precede by induction. For notational convenience, we define a function

$$(9) \qquad D(y, m) = f\left(y + \frac{1}{2^m}\right) - f(x)$$

so that

$$f(x_{i+1}) - f(x_i) = D(x_i, n) \tag{10}$$

We will prove that for all nonnegative integers m, all y in $[0, 1)$ such that $2^m y$ is an integer, $D(y, m) = p^a (1-p)^b$ (with a and b defined analogously to the statement of the lemma, using y and m in place of x_i and n). This directly implies the truth of the lemma by 10.

The proof will proceed 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 1 and 2 that apply for all x in $[0, 1]$.

First, note that if y is in $[0, \frac{1}{2})$, $y + \frac{1}{2^m}$ is in $[0, \frac{1}{2}]$ since they are all multiples of $\frac{1}{2^m}$. 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 1 with $x = 2y$ and $x = 2\left(y + \frac{1}{2^m}\right)$ to get

$$f_p(y) = (1-p)f_p(2y)$$

$$f_p\left(y + \frac{1}{2^m}\right) = (1-p)f_p\left(2y + \frac{1}{2^{m-1}}\right)$$

This gives

$$f_p\left(y + \frac{1}{2^m}\right) - f_p(y) = (1-p)D(2y, m-1) \tag{11}$$

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, by the induction hypothesis, $D(2y, m-1) = p^a(1-p)^{b-1}$, so $D(y, m) = p^a(1-p)^b$.

- The latter case is similar. Here, we apply 2 with $x = 2y - 1$ and $x = 2\left(y + \frac{1}{2^m}\right) - 1$, getting

$$f_p(y) = 1 - p + pf_p(2y - 1)$$

$$f_p\left(y + \frac{1}{2^m}\right) = 1 - p + pf_p(2y - 1 + \frac{1}{2^{m-1}})$$

$$f_p\left(y + \frac{1}{2^m}\right) - f_p(y) = pD(2y - 1, m - 1)$$

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, by the induction hypothesis, $D(2y - 1, m - 1) = p^{a-1}(1 - p)^b$, so $D(y, m) = p^a(1 - p)^b$.

This completes the induction. $\square$

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

(12) $\quad b_k = \begin{cases} p & \text{if the } k\text{th bit in the binary expansion of } x_i \text{ is } 1 \\ 1 - p & \text{if the } k\text{th bit in the binary expansion of } x_i \text{ is } 0 \end{cases}$

then we can alternatively write

$$(13) \qquad\qquad f(x_{i+1}) - f(x_i) = \prod_{k=1}^{n} b_k$$

We will now look at x_i as a random variable, with i chosen uniformly out of the integers from 1 to 2^n. In order to apply standard probabilistic reasoning, it will be helpful to deal with a sum rather than a product. We thus write

$$(14) \qquad\qquad \log_2(f(x_{i+1}) - f(x_i)) = \sum_{k=1}^{n} \log_2 b_k$$

It is important to note that each bit in the binary expansion of x_i is independent of all the rest, so the b_k (and $\log_2 b_k$) are independent random variables. Furthermore, each of b_k (and each of $\log_2 b_k$) 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 $\frac{1}{2}$,

$$
\begin{aligned}
\mu &= \frac{1}{2}(\log_2 p + \log_2(1 - p)) \\
&= \log_2 \sqrt{p(1 - p)} \\
&< \log_2 \frac{1}{2} \text{ (by 3)} \\
&= -1
\end{aligned}
$$

(15)

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 q_n that $\log_2(f(x_{i+1}) - f(x_i)) > nr$: the inequality says this probability is at most

$$(16) \qquad \frac{n\sigma^2}{(nr - n\mu)^2} = \frac{1}{n} \cdot \frac{\sigma^2}{(r - \mu)^2}$$

This implies

Proposition 6.3. *For any $\epsilon > 0$, there exists an N such that if $n \geq N$, $q_n \leq \frac{\epsilon}{2}$.*

The truth of the proposition is immediate apparent: one can simply set N to $\frac{2}{\epsilon} \cdot \frac{\sigma^2}{(r-\mu)^2}$. Notably, exponentiating both sides shows that q_n is actually the probability that $f(x_{i+1}) - f(x_i) > 2^{nr}$. This effectively shows that for sufficiently large N, most of the intervals have relatively small increases of f.

Since $r < -1$, we also have the following fact about $2^{n(r-1)}$

Proposition 6.4. *For any $\epsilon > 0$, there exists an N' such that if $n \geq N'$, $2^{n(r+1)} < \frac{\epsilon}{2}$.*

This is just a statement that the limit of a decaying exponential is 0.

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,

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

Since there are only 2^n values of i, the sum of these differences over all good i is less than $\frac{\epsilon}{2}$. On the other hand, the sum over all i is $f(x_{2^n+1}) - f(x_1) = 1$. Thus the sum of $f(x_{i+1}) - f(x_i)$ over all bad i is greater than $1 - \frac{\epsilon}{2}$. Furthermore, $|\langle x_{i+1} - x_i, f(x_{i+1}) - f(x_i) \rangle| \geq f(x_{i+1}) - f(x_i)$, so the sum of $|\langle x_{i+1} - x_i, f(x_{i+1}) - f(x_i) \rangle|$ 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 $x_{i+1} - x_i = 2^{-n}$, $|\langle x_{i+1} - x_i, f(x_{i+1}) - f(x_i) \rangle|$ is always greater than 2^{-n} for any i, so the sum of this over all x_i is greater than $1 - \frac{\epsilon}{2}$. Then the sum of this over all i, good and bad, is greater than $2 - \epsilon$.

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

7. 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.
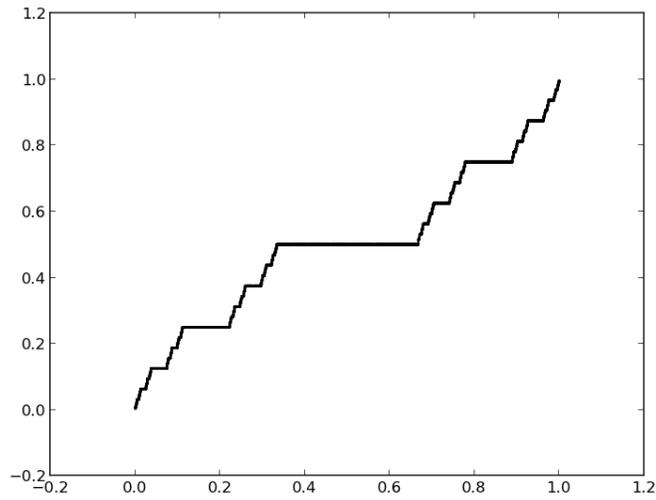
Figure 3. The Devil's staircase.

MIT OpenCourseWare

18.821 Project Laboratory in Mathematics
Spring 2013